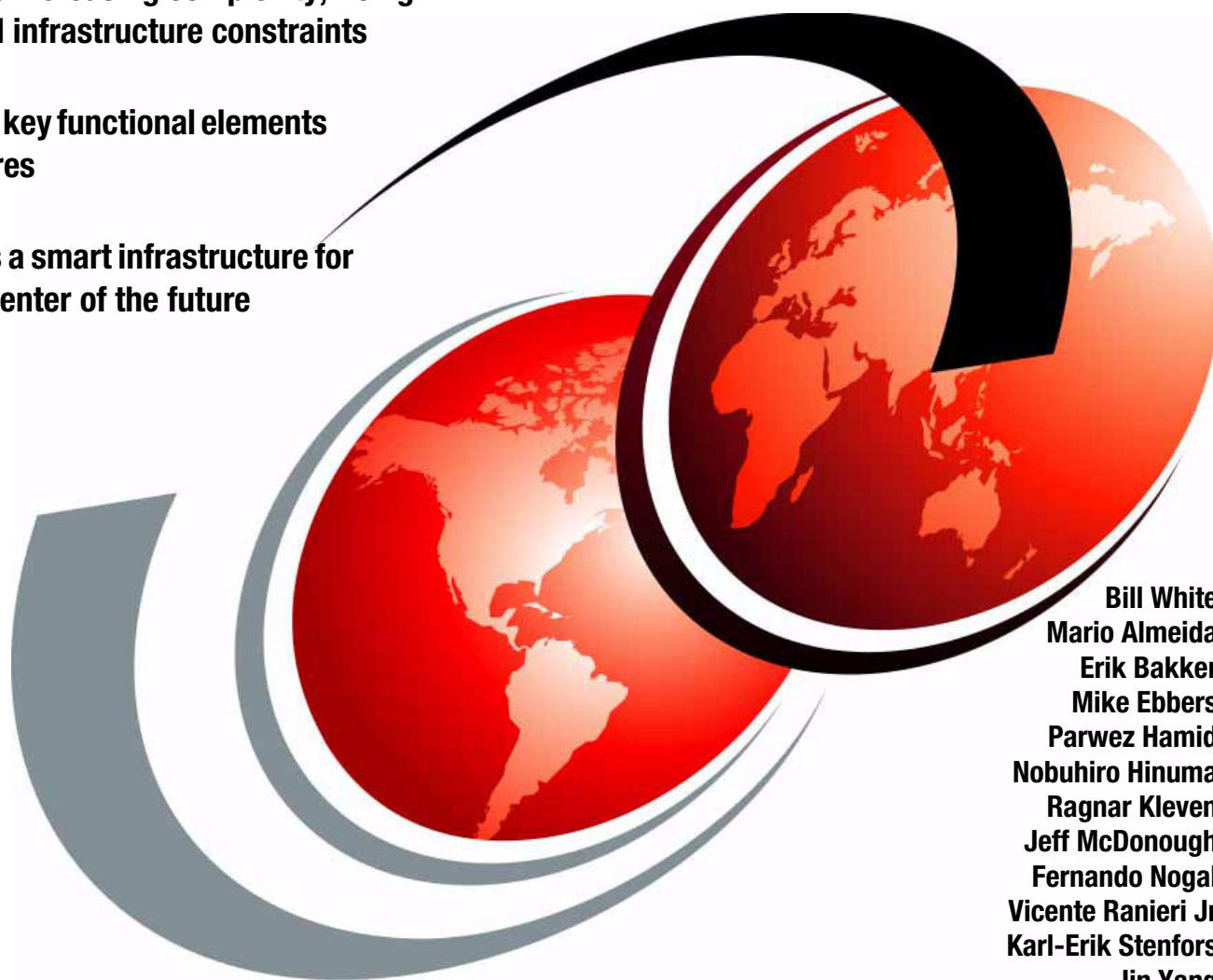


IBM zEnterprise System Technical Introduction

Addresses increasing complexity, rising costs, and infrastructure constraints

Describes key functional elements and features

Discusses a smart infrastructure for the data center of the future



Bill White
Mario Almeida
Erik Bakker
Mike Ebbers
Parwez Hamid
Nobuhiro Hinuma
Ragnar Kleven
Jeff McDonough
Fernando Nogal
Vicente Ranieri Jr
Karl-Erik Stenfors
Jin Yang

Redbooks



International Technical Support Organization

IBM zEnterprise System Technical Introduction

August 2010

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

First Edition (August 2010)

This edition applies to the IBM zEnterprise System.

This document created or updated on July 28, 2010.

Contents

Notices	ix
Trademarksx
Preface	xi
The team who wrote this book	xi
Now you can become a published author, too!	xiii
Comments welcome	xiv
Chapter 1. Proposing an infrastructure (r)evolution	1
1.1 z196 technical description	8
1.1.1 Central Processing Complex	10
1.1.2 I/O subsystem	11
1.1.3 I/O connectivity	13
1.1.4 zEnterprise BladeCenter Extension	15
1.1.5 Unified Resource Manager	16
1.2 Capacity On Demand	16
1.3 Software	16
Chapter 2. zEnterprise System hardware overview	19
2.1 z196 highlights	20
2.2 Models and model upgrades	20
2.3 The frames	22
2.3.1 Top exit I/O cabling	24
2.4 CPC cage and books	24
2.5 Multi-chip module (MCM)	25
2.6 z196 processor chip	26
2.7 Processor unit (PU)	27
2.8 Memory	28
2.9 I/O system structure	30
2.10 I/O cages, drawers, and features	31
2.10.1 ESCON channels	34
2.10.2 FICON Express8	35
2.10.3 FICON Express4	35
2.10.4 OSA-Express3	36
2.10.5 OSA-Express2	37
2.11 Cryptographic functions	38
2.11.1 CP Assist for Cryptographic Function	38
2.11.2 Crypto Express3 feature	39
2.11.3 TKE workstation	40
2.12 Coupling and clustering	40
2.12.1 ISC-3	40
2.12.2 Internal Coupling (IC)	41
2.12.3 Parallel Sysplex InfiniBand (PSIFB) coupling	41
2.12.4 System-Managed CF Structure Duplexing	41
2.12.5 Coupling Facility Control Code (CFCC) level 17	42
2.13 Time functions	42
2.13.1 External clock facility (ECF)	42
2.13.2 Server Time Protocol (STP)	42
2.13.3 Network Time Protocol (NTP) support	43

2.14	z196 HMC and SE	43
2.15	Power and cooling	44
2.15.1	Power consumption	44
2.15.2	Hybrid cooling system	45
2.15.3	Water cooling	45
2.15.4	High voltage DC power	45
2.15.5	Internal Battery Feature	45
2.15.6	IBM Systems Director Active Energy Manager	46
2.16	zEnterprise BladeCenter Extension	46
Chapter 3. Key functions and capabilities of the zEnterprise System		51
3.1	Virtualization	52
3.1.1	z196 hardware virtualization	52
3.1.2	z196 software virtualization	55
3.2	z196 technology improvements	56
3.2.1	Microprocessor	56
3.2.2	Large system images	58
3.2.3	Granular capacity and capacity settings	58
3.2.4	Memory	59
3.2.5	Connectivity	61
3.2.6	Cryptography	70
3.2.7	Hardware Management Console functionality	70
3.3	z196 common time functions	71
3.3.1	Server Time Protocol (STP)	72
3.4	z196 Capacity on Demand (CoD)	73
3.5	Throughput optimization with z196	75
3.6	zEnterprise BladeCenter Extension	76
3.6.1	IBM blades	76
3.6.2	IBM Smart Analytics Optimizer solution	77
3.7	z196 performance	78
3.8	Reliability, availability, and serviceability	80
3.8.1	RAS capability for zBX	81
3.9	High availability technology	82
Chapter 4. Achieving better infrastructure resource management		85
4.1	zEnterprise ensembles and virtualization	86
4.2	How can I tell if my business will benefit	87
4.2.1	Mainframe workloads	87
4.2.2	Heterogeneous platform deployments	87
4.3	Unified Resource Manager	90
4.3.1	Resource management suites	91
4.4	Physical resource management	92
4.4.1	Serviceability	93
4.5	Virtualization management	93
4.5.1	Network virtualization	93
4.5.2	Hypervisor management	95
4.5.3	Virtual server management	95
4.5.4	Storage virtualization	95
4.6	Performance management	96
4.7	Energy monitoring	97
4.8	Technical support services	97
Appendix A. Operating Systems support and considerations		99
	Software support summary	100

Support by operating system	103
z/OS	103
z/VM	105
z/VSE	107
Linux on System z	108
z/TPF	109
Software support for zBX	110
References	110
z/OS considerations	110
Coupling Facility and CFCC considerations	113
IOCP considerations	114
ICKDSF considerations	114
Appendix B. Frequently asked questions	115
Appendix C. Software licensing	131
Software licensing considerations	131
Workload License Charges (WLC)	132
System z New Application License Charges (zNALC)	133
Select Application License Charges (SALC)	133
Midrange Workload Licence Charges	133
System z International Program License Agreement (IPLA)	134
Appendix D. Channel options	135
Related publications	139
IBM Redbooks publications	139
Online resources	139
Other publications	139
How to get IBM Redbooks publications	140
Help from IBM	140
Index	141

Notices

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

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

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

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

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

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

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

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

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

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


COPYRIGHT LICENSE:

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

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>

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

1-2-3®	MVS™	System i®
AIX®	OMEGAMON®	System p®
BladeCenter®	Parallel Sysplex®	System Storage®
CICS®	Passport Advantage®	System x®
DB2 Connect™	Power Systems™	System z10®
DB2®	POWER7™	System z9®
Domino®	PowerVM™	System z®
DRDA®	POWER®	Tivoli®
DS8000®	PR/SM™	TotalStorage®
Dynamic Infrastructure®	Processor Resource/Systems Manager™	WebSphere®
ESCON®	RACF®	z/Architecture®
FICON®	Rational Rose®	zEnterprise™
HiperSockets™	Rational®	z/OS®
IBM Systems Director Active Energy Manager™	Redbooks®	z/VM®
IBM®	Redbooks (logo)  ®	z/VSE™
IMS™	Resource Link™	z10™
Lotus®	RMF™	z9®
MQSeries®	Sysplex Timer®	zSeries®

The following terms are trademarks of other companies:

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

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

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

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

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

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

Preface

Recently we have seen an explosion in applications, architectures, and platforms. With the generalized availability of the Internet and the appearance of commodity hardware and software, several patterns have emerged that have gained center stage. Workloads have changed. Many applications, including mission-critical ones, are deployed in heterogeneous infrastructures and the System z design has adapted to this change. IBM has a holistic approach to System z design, which includes hardware, software and procedures. It takes into account a wide range of factors, including compatibility and investment protection, thus ensuring a tighter fit with the IT requirements of the entire enterprise.

This IBM® Redbooks® publication introduces the revolutionary scalable IBM zEnterprise™ System, which consists of the IBM zEnterprise 196 (z196) and the IBM zEnterprise BladeCenter® Extension (zBX). IBM is taking a bold step by integrating heterogeneous platforms under the well-proven System z hardware management capabilities, while extending System z qualities of service to those platforms. The z196 is a general-purpose server that is equally at ease with compute-intensive workloads and with I/O-intensive workloads. The integration of heterogeneous platforms is based on IBM's BladeCenter® technology, allowing improvements in price and performance for key workloads, as well as enabling a new range of heterogeneous platform solutions. The z196 is at the core of the enhanced System z platform that is designed to deliver technologies that businesses need today along with a foundation to drive future business growth.

This book provides basic information about z196 and zBX capabilities, hardware functions and features, and its associated software support. It is intended for IT managers, architects, consultants, and anyone else who wants to understand the new elements of the zEnterprise System. For this introduction to the zEnterprise System, readers are not expected to be generally familiar with current IBM System z technology and terminology.

The team who wrote this book

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

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

Mario Almeida is an IBM Certified Consulting IT Specialist working as an STG technical consultant in Brazil. He has more than 30 years of experience working with IBM large systems. Mario has co-authored several IBM Redbooks publications. His areas of expertise include System z hardware, Parallel Sysplex, GDPS and capacity planning.

Erik Bakker is a Senior IT Specialist working for IBM Server and Technology Group in the Netherlands. During the past 24 years he has worked in various roles within IBM and with a large number of mainframe customers. For many years he worked for Global Technology Services as a systems programmer providing implementation and consultancy services at many customer sites. He currently provides pre-sales System z technical consultancy in support of large and small System z customers. His areas of expertise include Parallel Sysplex®, z/OS® and System z.

Mike Ebbers is a Consulting IT Specialist and Project Leader at the International Technical Support Organization, Poughkeepsie Center. He has worked with IBM mainframe hardware and software products since 1974 in the field, in education, and in the ITSO.

Parvez Hamid is an Executive IT Consultant working for IBM's Server and Technology Group. During the past 37 years he has worked in various IT roles within IBM. Since 1988 he has worked with a large number of IBM's mainframe customers and spent much of his time introducing new technology. Currently, he provides pre-sales technical support for IBM's System z product portfolio and is the lead System z technical specialist for UK and Ireland. Parvez co-authors a number of ITSO Redbooks and prepares technical material for the world-wide announcement of System z Servers. Parvez works closely with System z product development in Poughkeepsie and provides input and feedback for 'future' product plans. Additionally, Parvez is a member of IBM's IT Specialist profession certification board in the UK and is also a Technical Staff member of the IBM's UK Technical Council which is made of senior technical specialist representing all of IBM's Client, Consulting, Services and Product groups. Parvez teaches and presents at numerous IBM user group and IBM internal conferences.

Nobuhiro Hinuma is an IT Specialist at STG Systems Technical Sales in Japan. He has 15 years of experience in the System z technical sales support field including Parallel Sysplex customer support in Japan and AP countries, and z/OS Early Support Program. He is a member of zChampions team since 2006. His areas of expertise include Parallel Sysplex, z/OS and System z.

Ragnar Kleven is a Client IT Architect in Norway, supporting Finance Services Sector and Public Sector customers. He has 35 years of experience in IT which of 8 are with IBM, and has been working with mainframes in many different roles. He holds a Bachelor degree in Engineering. The areas of expertise include System z, z/OS and z/OS middleware software stack and transactional banking solutions in general.

Jeff McDonough is an IT Architect in the USA. He holds a degree in Computing Analysis from Missouri Southern State College, and has 33 years experience in IT. He has experience in the manufacturing, transportation, and retail industries. He has specialized in z/OS and parallel sysplex environments for 17 years and has previously written about Parallel Sysplex using InfiniBand, Server Time Protocol, and other z/OS topics.

Fernando Nogal is an IBM Certified Consulting IT Specialist working as an STG Technical Consultant for the Spain, Portugal, Greece, and Israel IMT. He specializes in on-demand infrastructures and architectures. In his 28 years with IBM, he has held a variety of technical positions, mainly providing support for mainframe customers. Previously, he was on assignment to the Europe Middle East and Africa (EMEA) zSeries® Technical Support group, working full time on complex solutions for e-business on zSeries. His job included, and still does, presenting and consulting in architectures and infrastructures, and providing strategic guidance to System z customers regarding the establishment and enablement of e-business technologies on System z, including the z/OS, z/VM®, and Linux® environments. He is a zChampion and a core member of the System z Business Leaders Council. An accomplished writer, he has authored and co-authored over 20 Redbooks and several technical papers. Other activities include chairing a Virtual Team from IBM interested in e-business on System z, and serving as a University Ambassador. He travels extensively on direct customer engagements and as a speaker at IBM and customer events, and trade shows.

Vicente Ranieri is an Executive IT Specialist at STG Advanced Technical Support (ATS) team supporting System z in Latin America. Ranieri has more than 30 years of experience working for IBM. Ranieri is a member of zChampions team, a worldwide IBM team to participate in the creation of System z technical roadmap and value proposition materials. Besides co-authoring several redbooks, Vicente has been an ITSO guest speaker since

2001, teaching the System z security update workshops worldwide. Ranieri also presents in several IBM internal and external conferences. His areas of expertise include System z security, Parallel Sysplex, System z hardware and z/OS. Vicente is a member of Technology Leadership Council – Brazil, an IBM Academy of Technology Affiliate.

Karl-Erik Stenfors is a Senior IT Specialist in the PSSC Customer Center in Montpellier, France. He has more than 40 years of working experience in the Mainframe environment, as a systems programmer, as a consultant with IBM's customers, and, since 1986 with IBM. His areas of expertise include IBM System z hardware and operating systems. He teaches at numerous IBM user group and IBM internal conferences, and he is a member of the zChampions work group. His current responsibility is to execute System z Early Support Programs in Europe and Asia.

Jin Yang is a Senior System Service Representative at the IBM Global Technical Services in Beijing, China. He joined IBM in 1999 to support and maintain System z products for clients throughout China. Jin has been working in the Technical Support Group (TSG) providing second level support to System z clients since 2009. His areas of expertise include System z hardware, Parallel Sysplex, and FICON connectivity.

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

Ivan Bailey, Connie Beuselinck, Patty Driever, Jeff Frey, Steve Fellenz, Michael Jordan, Gary King, Bill Kostenko, Jeff Kubala, Kelly Ryan, Lisa Schloemer, Jaya Srikrishnan, Peter Yocom, Martin Ziskind
IBM Poughkeepsie

Gwendolyn Dente, Harv Emery, Gregory Hutchison
IBM Advanced Technical Skills (ATS), North America

Friedemann Baitinger, Klaus Werner
IBM Germany

Brian Tolan, Brian Valentine, Eric Weinmann
IBM Endicott

Garry Sullivan
IBM Rochester

Jerry Stevens
IBM Raleigh

International Technical Support Organization:

Robert Haimowitz
IBM Raleigh

Ella Buslovich
IBM Poughkeepsie

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author - all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your

network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

ibm.com/redbooks

- ▶ Send your comments in an e-mail to:

redbooks@us.ibm.com

- ▶ Mail your comments to:

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



1

Proposing an infrastructure (r)evolution

The IBM zEnterprise System is the first of its kind. It was purposefully designed to help overcome fundamental problems of today's IT infrastructures and simultaneously provide a foundation for the future.

A great deal of experimentation has occurred in the marketplace during the last few decades, due to the explosion in applications, architectures, and platforms. Out of these experiments, and with the generalized availability of the Internet as well as the appearance of commodity hardware and software, several patterns have emerged that have taken center stage.

Exploitation of IT by enterprises continues to grow and the demands placed upon it are increasingly complex. The world is not stopping—in fact, the pace of business is accelerating. Internet pervasiveness fuels ever-increasing utilization modes by a growing number of users. And the most rapidly growing “type” of user is not people, but devices. All sorts of services are being offered and new business models are being implemented.

IT infrastructures today

Multitier workloads and their deployment on heterogeneous infrastructures are commonplace today. What is harder to find is the infrastructure setup needed to provide the high qualities of service required by mission-critical applications.

Creating and maintaining these high-level qualities of service from a large collection of distributed components demands significant knowledge and effort. It implies acquiring and installing extra equipment and software to ensure availability and security, monitoring and managing. Additional manpower and skills are required to configure, administer, troubleshoot, and tune such a complex set of separate and diverse environments. Due to platform functional differences, the resulting infrastructure will not be uniform regarding those qualities of service or serviceability.

Careful engineering of the workload's several server tiers is required to provide the robustness, scaling, consistent response, and other characteristics demanded by the users and lines of business.

Despite all efforts, these infrastructures do not scale well. What is a feasible setup with a few servers becomes difficult to handle with tens of servers—and a nightmare with hundreds of servers. And when it is possible, it is expensive. Often, by the end of the distributed equipment's life cycle, its residual value is nil, therefore requiring new acquisitions, new software licences, and recertification. Today's resource-constrained environments need a better way.

To complete this picture on the technology side, it is now clear that performance gains from increasing chips' frequency are diminishing. Thus, special-purpose compute acceleration will be required for greater levels of workload performance and scalability, which will result in additional system heterogeneity.

Future infrastructures

There is a growing awareness that the very foundation of IT infrastructures is not up to the job. Most existing infrastructures are too complex, too inefficient, and too inflexible. The demands placed on network and computing resources will reach a breaking point unless something changes. It is then necessary to define the target infrastructure and how to effect the change.

And, while they are changing, the need to improve service delivery, manage the escalating complexity, and maintain a secure enterprise continues to be felt. To compound it, there is a daily pressure to cost-effectively run a business, while supporting growth and innovation.

In the IBM vision of the future, transformation of the IT delivery model is strongly based on new levels of efficiency and service excellence for businesses, driven by and from the data center. This evolution will prepare systems to handle massive scale and integration, capture, store, manage, and retrieve vast amounts of data, *and* analyze and unlock the insights of the data.

Business service workloads will continue to be inherently diverse and will require dissimilar system structures on which to deploy them.

Aligning IT with the goals of the business and the speed of IT execution with the pace of business is an absolute top priority. This places further demands on the infrastructure which needs to be dynamic, automated, with policy-based resource provisioning, deployment, reallocation, and optimization. And the infrastructure needs to be managed in accordance with specified workload service level objectives.

A new technology is needed that can go to the next level, where smarter systems and smarter software work together to address the needs of the business. In a word, infrastructures and systems in them need to become smarter.

A smarter infrastructure will allow organizations to better position themselves, to adopt and integrate new technologies, such as Web 2.0 and cloud computing, and deliver dynamic and seamless access to IT services and resources.

It would seem then that the key to solving today's problems and unlocking the road to the future is based on a very smart IT infrastructure, composed of diverse systems, very flexible, highly virtualized, automated and tightly managed.

The zEnterprise System is following an evolutionary path that directly addresses those infrastructure problems. Over time, it will provide increasingly complete answers to the smart infrastructure requirements. The zEnterprise System, with its diverse platform management capabilities, already provides many of these answers, offering great value in a scalable solution that integrates and simplifies hardware and firmware management and support, as well as the definition and management of a network of virtualized servers, across multiple diverse platforms.

Introducing the zEnterprise System

The zEnterprise System brings about a revolution in the end-to-end management of diverse systems, while offering expanded and evolved traditional System z capabilities.

With zEnterprise, a *system of systems* can be created where the virtualized resources of both the zEnterprise 196 (z196) and selected IBM blade-based servers, housed in the zEnterprise BladeCenter Extension (zBX), are pooled together and jointly managed.

End-to-end solutions based on multi-platform workloads can be deployed across the zEnterprise System structure and benefit from System z's traditional qualities of service, including high availability, and simplified and improved management of the virtualized infrastructure.

Because many mission-critical workloads today have one or more components on System z, exploiting System z environments for database and other capabilities, the ability to co-locate all of the workload components under the same management platform and thereby benefit from uniformly high qualities of service should be quite appealing and provide tangible benefits and a rapid ROI.

Figure 1-1 shows the zEnterprise System with its management capabilities.

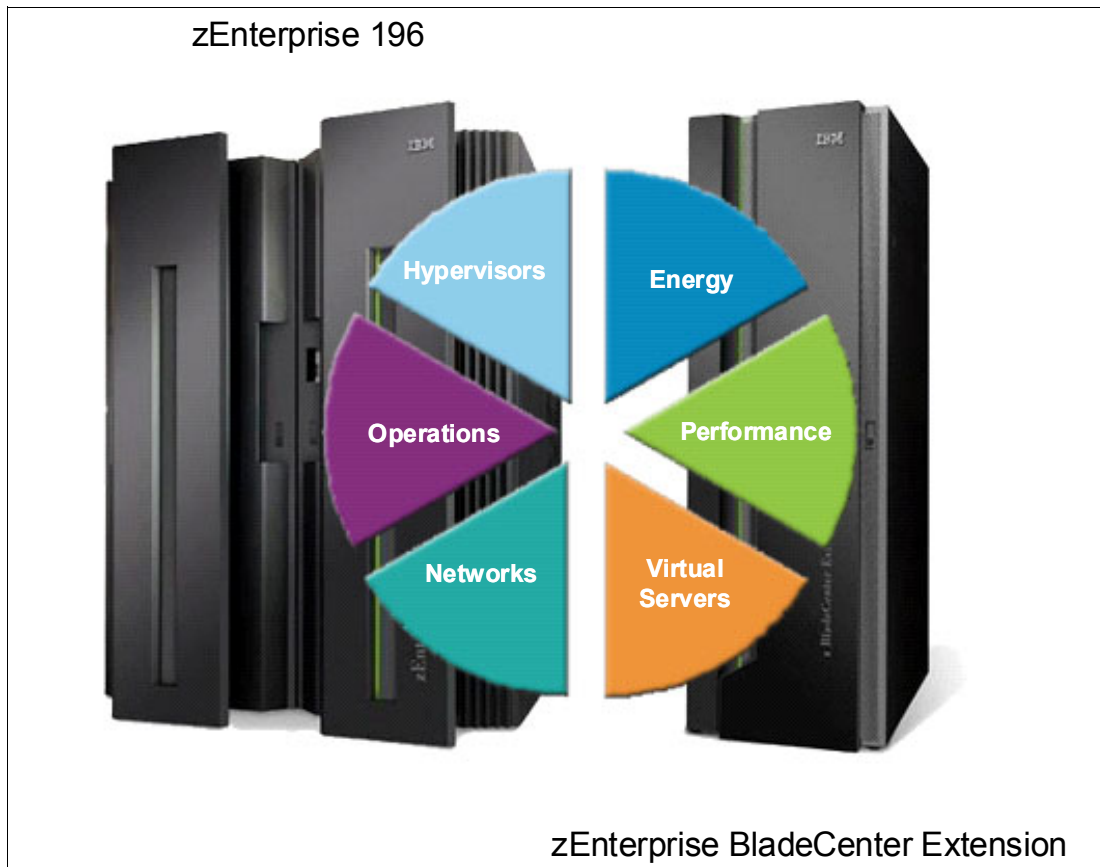


Figure 1-1 zEnterprise System and management capabilities

The z196 improves upon the capabilities of its predecessor, the System z10® Enterprise Class (z10 EC). The z196 has an upgraded 5.2 GHz system chip, which is the fastest quad-core processor in the industry. The z196 can be configured up to an 80-way, with up to 3.0 TB of memory, which doubles the z10 EC capacity, and has better reliability, availability and serviceability (RAS). The z196 continues to support the latest connectivity options, a

highlight of its open server characteristics. It delivers, in a single footprint, unprecedented performance and capacity growth with improved power efficiency, while drawing upon the rich heritage of previous z/Architecture servers. The z196 is a well-balanced, general-purpose server that is equally at ease on compute-intensive workloads as it is with I/O-intensive workloads.

z196 workload support capabilities are complemented by the zBX. Each zBX can have up to four frames housing up to eight BladeCenters, each with up to 14 selected IBM POWER-based blades, for a maximum of 112 blades.

The zBX supports two types of blades: POWER7™ blades, which can run a wide variety of applications; and special-purpose blades, which are dedicated to a specific task, as in the IBM Smart Analytics Optimizer solution.

High speed virtualized networks support data communication between applications running in the z196 and zBX, and management of the several hardware components.

IBM takes a holistic approach to System z design that includes hardware, software, and procedures, and considers a wide range of factors, including compatibility and investment protection, thus ensuring a tighter fit with the IT requirements of the entire enterprise with better total cost of ownership.

The System z server design makes no compromises to the traditional mainframe qualities of service and continues to follow the fundamental objective of simultaneous support of a large number of dissimilar workloads. The workloads in themselves have changed significantly and continue to evolve, so the design must adapt to those changes.

Fixing the IT infrastructure

For the first time it is possible to deploy an integrated hardware platform that brings mainframe and distributed technologies together - a system that can start to replace individual islands of computing and that can work to reduce complexity, improve security and bring applications closer to the data they need.

With the zEnterprise System a new concept in IT infrastructures is being introduced: zEnterprise *ensembles*. Ensembles, together with the virtualization, flexibility, security, and management capabilities provided by the zEnterprise System are key to solving the problems posed by today's IT infrastructure.

Virtualization is central to the effective exploitation of flexible infrastructures. System z has a long tradition in this area and the most advanced virtualization in the industry, while ensembles management further advance those capabilities.

zEnterprise ensembles management

IBM is widening the enterprise role and application domain of System z servers by integrating additional heterogeneous infrastructure components into System z and extending System z qualities of service to them.

System z has long been an integrated diverse platform, with specialized hardware and dedicated computing capabilities. Recall, for instance, in the mid-1980's, the IBM 3090 and its vector facility (occupying a separate frame). Or the Cryptographic processors and all the I/O cards, which are specialized dedicated hardware running non-System z code on non-System z processors, to offload processing tasks from the System z processor units (PUs). All of these specialized hardware components have been seamlessly integrated within the mainframe for over a decade.

Each z196 with its optional zBX make up a *node* of a zEnterprise ensemble. A zEnterprise ensemble is a collection of highly virtualized diverse systems that can be managed as a single logical entity where diverse workloads can be deployed.

A zEnterprise ensemble is composed of up to 8 members, with up to eight z196 servers and up to 896 blades housed in up to eight zBXs, dedicated integrated networks for management and data, and the Unified Resource Manager function. With the Unified Resource Manager, the z196 provides advanced end-to-end management capabilities for the diverse systems housed in the zBX.

The zBX components are configured, managed, and serviced the same way as the other components of the z196. Despite the fact that the zBX processors are not System z PUs and run specific software, including hypervisors, the software intrinsic to the zBX components does not require any additional administration effort or tuning by the user. In fact, it is handled as System z Licensed Internal Code. The zBX hardware features are part of the mainframe, not add-ons.

With zBX, IBM reaches a new level, creating “One Infrastructure” integration. Integration provides investment protection, reduction of complexity, improved resiliency, and lower cost of ownership

It is worth mentioning that the concept of “ensemble” has some similarity to that of “cloud.” Actually an ensemble would provide a perfect infrastructure to support a cloud, as the real purpose of an ensemble is to provide infrastructure resources in a way that ensures that the workloads running on it achieve their business requirement’s objectives. Those objectives are specified through policies, which the ensemble implements.

Diverse workloads span several platform infrastructures, so the ensemble owns the physical resources in those infrastructures and manages them in order to fulfill the workload policies. Ensemble resources can be shared by multiple workloads and optimized for each workload. Virtualization provides the most flexible and cost effective way to meet the policies’ requirements.

z196 virtualized environments

The z196 supports advanced server consolidation and offers the best virtualization in the industry. Up to 60 logical partitions (LPARs) can be deployed. Each one can run any of the supported operating systems:

- ▶ z/OS
- ▶ z/VM
- ▶ z/VSE™
- ▶ z/TPF
- ▶ Linux on System z

The z196 offers software virtualization through z/VM. z/VM’s virtualized z/Architecture servers, known as virtual machines, support all operating systems and other software supported in a logical partition. In fact, a z/VM virtual machine is the functional equivalent of a real server.

z/VM’s extreme virtualization capabilities, which have been perfected since its introduction in 1967, enable virtualization of thousands of distributed servers on a single z196 server.

In addition to server consolidation and image reduction by vertical growth under z/VM, z/OS provides a highly sophisticated environment for application integration and co-residence with data, especially for mission-critical applications.

In addition to the hardware-enabled resource sharing, other uses of virtualization include:

- ▶ Isolating production, test, training, and development environments
- ▶ Supporting back-level applications
- ▶ Enabling parallel migration to new system or application levels, and providing easy back-out capabilities

zBX virtualized environments

On the zBX, the IBM blades also have a virtualized environment that is similar to the one found in IBM Power Systems™ servers. Management of the zBX environment is done as a single logical virtualized environment by the Unified Resource Manager. The POWER7-based blades run the AIX® operating system.

Flexibility and security

The z196 expands the subcapacity settings offering to up to 15 central processors (CPs), delivering the scalability and granularity to meet the needs of medium-sized enterprises, while also satisfying the requirements of large enterprises having large-scale, mission-critical transaction and data-processing requirements. The first 15 processors can be configured at three different sub-capacity levels, giving a total of 125 distinct capacity settings in the system, and providing for a range of over 1:200 in processing power.

In the same footprint, the z196 80-way server can deliver up to 60% more capacity than the largest z10 EC (the largest z10 EC is a 64-way). The z196 continues to offer all the specialty engines available with System z10; see “PU characterization” on page 10.

Most hardware upgrades can be installed concurrently. As we describe later, the z196 reaches new availability levels by eliminating various pre-planning needs and other disruptive operations.

The z196 enhances the availability and flexibility of just-in-time deployment of additional server resources, known as Capacity on Demand (CoD). CoD provides flexibility, granularity, and responsiveness by allowing the user to dynamically change capacity when business requirements change. With the proper contracts, up to eight temporary capacity offerings can be installed on the server. Additional capacity resources can be dynamically activated, either fully or in part, by using granular activation controls directly from the management console, without the having to interact with IBM Support.

IBM has further enhanced and extended the z196 leadership with improved access to data and the network. The following list indicates several of many enhancements:

- ▶ Tighter security with a CP Assist for Cryptographic Function (CPACF) protected key and longer personal account numbers for stronger protection of data
- ▶ Enhancements for improved performance connecting to the network
- ▶ Increased flexibility in defining your options to handle backup requirements
- ▶ Enhanced time accuracy to an external time source

A fast-growing number of enterprises are reaching the limits of available physical space and electrical power at their data centers. The extreme virtualization capabilities of the z196 enable the creation of dense and simplified infrastructures that are highly secure and can lower operational costs.

Further simplification is possible by exploiting the z196 HiperSockets™¹ and z/VM virtual switch functions. These may be used, at no additional cost, to replace physical routers, switches, and their cables, while eliminating security exposures and simplifying configuration and administration tasks. In some actual simplification cases, cables have been reduced by 97%.

¹ For a description of HiperSockets, see “HiperSockets” on page 15. The z/VM virtual switch is a z/VM system function that uses memory to emulate switching hardware.

A recent paper² on Payment Card Industry compliance recognizes the inherent qualities of the mainframe and the simplification in the infrastructure it can provide.

It would seem that increased flexibility inevitably leads to increased complexity. However it does not have to be so. IT operational simplification greatly benefits from z196's intrinsic autonomic characteristics, the ability to consolidate and reduce the number of system images, and the management best practices and products which were developed and are available for the mainframe, in particular for the z/OS environment.

A cornerstone of a smart IT infrastructure

Summing up these characteristics leads to an interesting result:

- Capacity range and flexibility
- + A processor equally able to handle compute-intensive and I/O-intensive workloads
 - + Specialty engines for improved price/performance
 - + Extreme virtualization
 - + Additional platforms and unified resource management
-
- = A very wide range of workloads that can be seamlessly deployed and managed in an integrated heterogeneous environment

Many clients use their mainframe and application investments to support future business growth and to provide an important competitive advantage. Having chosen the mainframe as the platform to support their environment, these clients are demonstrating how to create a smart business.

An important point is that the System z *stack* consists of much more than just a server. This is because of the total systems view that guides System z development. The *z-stack* is built around services, systems management, software, and storage. It delivers a complete range of policy-driven functions, pioneered and most advanced in the z/OS environment, including:

- ▶ Access management to authenticate and authorize who can access specific business services and associated IT resources.
- ▶ Utilization management to drive maximum use of the system. Unlike other classes of servers, System z servers are designed to run at 100% utilization 100% of the time, based on the varied demands of its users.
- ▶ Just-in-time capacity to deliver additional processing power and capacity when needed.
- ▶ Virtualization security to enable clients to allocate resources on demand without fear of security risks.
- ▶ Enterprise-wide operational management and automation, leading to a more autonomic environment.

System z is the result of sustained and continuous investment and development policies. Commitment to IBM Systems design means that z196 brings all this innovation while helping customers leverage their current investment in the mainframe, as well as helping to improve the economics of IT.

The zEnterprise System can improve the integration of people, processes, and technology to help run the business more cost effectively while also supporting business growth and innovation. It is, thus, the most powerful tool available to reduce cost, energy, and complexity in enterprise data centers.

The z196 continues the evolution of the mainframe, building upon the z/Architecture definitions. IBM mainframes traditionally provide an advanced combination of reliability,

² Written by the atsec information security corporation. The paper can be found at:
http://www.atsec.com/downloads/white-papers/PCI_Compliance_for_LCS.pdf

availability, security, scalability, and virtualization. The z196 has been designed to extend these capabilities into heterogeneous servers and is optimized for today's business needs.

The zEnterprise System is a platform of choice for the integration of the new generations of applications with existing applications and data. The zEnterprise System truly is a cornerstone of a smart IT infrastructure.

1.1 z196 technical description

The z196 is a follow-on to the System z10 Enterprise Class (z10 EC). In this section we briefly review the most significant characteristics of the z196. Chapter 2, “zEnterprise System hardware overview” on page 19, provides further details.

The z196 employs leading-edge silicon-on-insulator (CMOS 12s-SOI) and other technologies, such as InfiniBand and Ethernet. The z196 provides benefits like very high frequency chips, additional granularity options, improved availability, and enhanced on demand options. In addition, it supports the latest offerings for data encryption.

Five models of the z196 are offered. These are named M15, M32, M49, M66, and M80. The names represent the maximum number of processors that can be configured in the model.

The z196 system architecture ensures continuity and upgradeability from the z10 EC and z9® EC designs.

z196 offers an air-cooled version, similar to z10 EC, and a water-cooled version. These characteristics are factory installed, and it is not possible to convert in the field from one to the other, so careful consideration should be given to current and future needs.

Figure 1-2 on page 9 provides a comparison of z196 with previous System z servers along four major attributes:

- ▶ Single engine processing capacity (based on Processor Capacity Index (PCI))
- ▶ Number of engines
- ▶ Memory (the z196 and z10 allow up to 1 TB per LPAR)
- ▶ I/O bandwidth (the z196 and z10 only exploit a subset of their designed I/O capability)

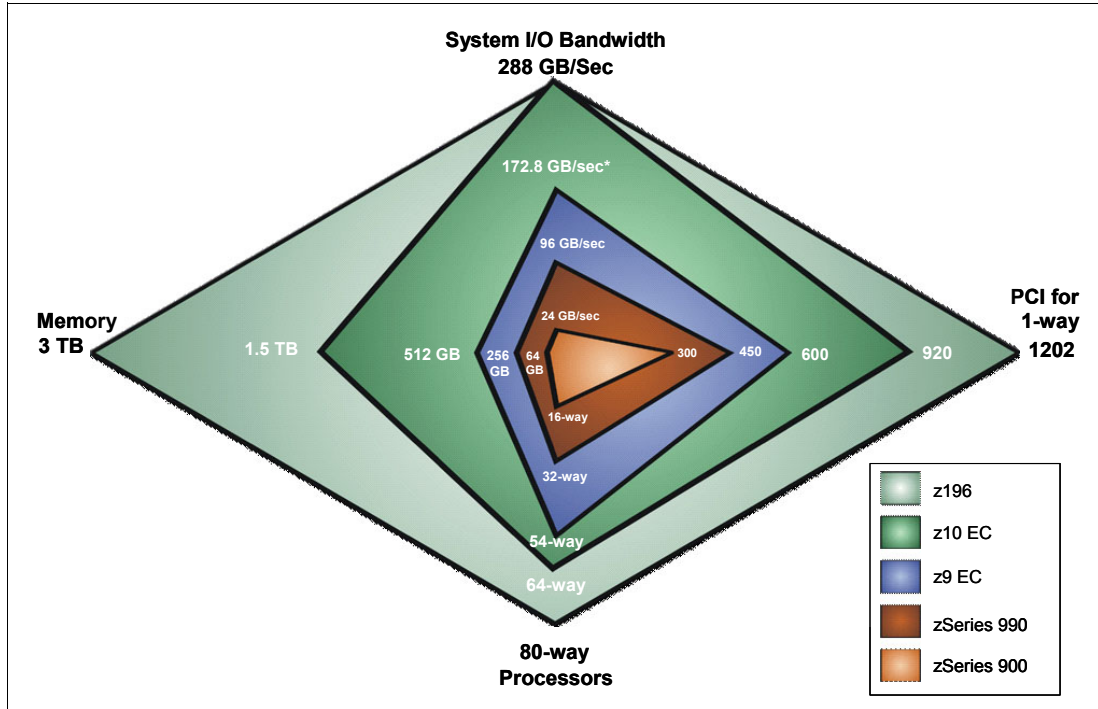


Figure 1-2 System z design comparison

The z196 server is a two-frame system. It has a machine type designation of 2817. The frames in the z196 are known as the A frame and the Z frame.

The A frame contains:

- ▶ The Central Processing Complex (CPC)
- ▶ Modular cooling units (different for water and air cooling)
- ▶ None or one I/O cage, or up to two I/O drawers
- ▶ Power supplies
- ▶ An optional internal battery feature (IBF)

The Z frame contains:

- ▶ Two system Support Elements (SEs)
- ▶ Up to four I/O drawers, or one additional I/O cage and up to two I/O drawers, or two additional I/O cages (two I/O cages in the Z frame requires an RPQ)
- ▶ Power supplies
- ▶ An optional IBF

The two redundant Support Elements (SEs) are used to configure and manage the z196 server.

If you are looking to build a green data center, water cooling and high voltage DC power allow a bold step into the future of cooler computing and increased energy efficiency without significantly changing the system physical footprint (the water cooling option adds 4 inches of depth to the back of both system frames).

1.1.1 Central Processing Complex

The Central Processing Complex (CPC) is housed in its own cage. The cage houses from one to four processor books that are fully interconnected. Each book contains a multi-chip module (MCM), memory and I/O cage connectors, and (optionally) coupling link connectors.

The z196 is built on the proven superscalar microprocessor architecture already deployed on the z10 EC. However, the new processor unit (PU) chip has several distinctive innovations, notably it is the first CMOS mainframe core with out-of-order instruction execution. Improvements have been made in error checking and correcting, namely in the memory design, and new specialized circuitry, for instance, to support out-of-order execution and decimal floating point operations. Additionally, it has a 5.2 GHz high-speed quad-core design.

Each book has one MCM that houses six PU chips and two storage control (SC) chips. Each PU chip has either three or four enabled cores. There are two cooling options for the MCM: modular refrigeration units (MRUs) with air-cooling backup, or chilled water.

In any model of the server, two cores are designated as spares, and each individual core can be transparently spared, as with the z10 EC. This contrasts with previous systems where the chip was the sparing unit.

Memory has been increased, as compared with the z10 EC. It is now implemented as a Redundant Array of Independent Memory (RAIM), for enhanced availability. In each book, up to 960 GB can be installed but part of it is redundant, so up to 768 GB of usable memory can be configured. In addition, 16 GB are part of the base and reserved for the hardware system area (HSA), making the maximum amount of purchasable memory 3056 GB, just short of 3 TB (with redundancy, a total of 3.75 GB are installed). Plan-ahead memory, a capability whereby memory can be installed but not enabled for use until needed, further enhances system availability for continuous operations.

PU characterization

At server initialization time, each purchased PU is *characterized* as one of a variety of types. It is also possible to dynamically characterize PUs. A PU that is not characterized cannot be used. A PU may be characterized as follows:

- CP** Central processor: the standard z196 processors. For use with any supported operating system and user applications.
- ICF** Internal Coupling Facility: used for z/OS clustering. ICFs are dedicated to this function and exclusively run the Coupling Facility Control Code (CFCC).
- IFL** Integrated Facility for Linux: exploited by Linux and for z/VM processing in support of Linux. z/VM is often used to host multiple Linux virtual machines (called guests). It is not possible to IPL operating systems other than z/VM or Linux on an IFL.
- SAP** System Assist Processor: offloads and manages I/O operations. Several are standard with the z196. More may be configured if additional I/O processing capacity is needed.
- zAAP³** z196 Application Assist Processor: exploited under z/OS for designated workloads, which include the IBM JVM and some XML System Services functions.
- zIIP³** z196 Integrated Information Processor: exploited under z/OS for designated workloads, which include various XML System Services, IPsec offload, certain parts of DB2® DRDA®, star schema, HiperSockets for large messages, and the IBM GBS Scalable Architecture for Financial Reporting.

³ z/VM V5 R4 and later support zIIP and zAAP processors for z/OS guest workloads.

Note: Work dispatched on zAAP and zIIP does not incur any IBM software charges. It is possible to run a zAAP-eligible workload on zIIPs^a if no zAAPs are installed on the server. This capability is offered to enable optimization and maximization of investment on zIIPs.

- a. This capability is available with z/OS V1.11 and later (and z/OS V1.9 and V1.10, with service) on all z9, z10 and z196 servers. Some additional restrictions apply.

Processor Resource/Systems Management

Processor Resource/Systems Management (PR/SM™) is responsible for hardware virtualization of the server. It is always active and has been enhanced to provide additional performance and platform management benefits. PR/SM technology on previous System z servers has received Common Criteria EAL5⁴ security certification. Each logical partition is as secure as an isolated server.

CP Assist for Cryptographic Function (CPACF)

The z196 continues to use the Cryptographic Assist Architecture, first implemented in 2003. Further enhancements have been made to the z196 CP Assist for Cryptographic Function (CPACF).

CPACF is physically implemented in the quad-core chip by the Compression and Cryptography Accelerator (CCA). Each of the two CCAs is shared by two cores. CPACF supported protocols include:

- ▶ Data Encryption Standard (DES)
- ▶ Triple Data Encryption Standard (TDES)
- ▶ Secure Hash Algorithm (SHA):
 - SHA-1: 160 bit
 - SHA-2: 224 bit, 256 bit, 384 bit, and 512 bit
- ▶ Advanced Encryption Standard (AES) for 128-bit, 192-bit, and 256-bit keys
- ▶ Protected key capabilities
- ▶ Pseudo Random Number Generation (PRNG)⁵
- ▶ Random number generation long (RNGL): 8 bytes to 8096 bytes
- ▶ Random number generation (RNG) with up to 4096-bit key RSA support
- ▶ Single-key and double-key message authentication code (MAC)
- ▶ The CPACF functions are supported by z/OS, z/VM, z/VSE, and Linux on System z

1.1.2 I/O subsystem

As with its predecessors, the z196 server has a dedicated subsystem to manage all input/output operations. Known as the channel subsystem, it is composed of:

SAP System Assist Processor (SAP) is a specialized processor that uses the installed PU cores⁶. Its role is to offload I/O operations and manage channels and the I/O operations queues. It relieves the other PUs of all I/O tasks, allowing them to be dedicated to application logic. An adequate number of SAP processors is automatically defined, depending on the number of installed books. These are part of the base configuration of the server.

HSA Hardware System Area (HSA) is a reserved part of the system memory containing the I/O configuration, and it is used by SAPs. On the z196 a

⁴ Evaluation Assurance Level with specific Target of Evaluation, Certificate for System z10 EC published October 29, 2008, pending for z196.

⁵ PRNG is also a standard function supported on the Crypto Express features.

⁶ Each z196 PU can be characterized as one of six different configurations. For more information see “PU characterization” on page 10.

fixed amount of 16 GB is reserved, which is not part of the customer purchased memory. This provides for greater configuration flexibility and higher availability by eliminating some planned and pre-planned outages.

Channels Channels are small processors that communicate with the I/O control units (CUs). They manage the data transfer between memory and the external devices. Channels are contained in the I/O card features.

Channel path Channel paths are the means by which the channel subsystem communicates with the I/O devices. Due to I/O virtualization, multiple independent channel paths can be established on a single channel, allowing sharing⁷ of the channel between multiple logical partitions, with each partition having a unique channel path.

Subchannels Subchannels appear to a program as a logical device and contain the information required to perform an I/O operation. One subchannel exists for each I/O device addressable by the channel subsystem. A third subchannel set is new with the z196.

The z/Architecture specifies an I/O subsystem to which all I/O processing is offloaded. This is a significant contributor to the performance and availability of the system, and it strongly contrasts with the architectures of other servers.

The z196 I/O subsystem direction is evolutionary, expanding on development from the z9 and z10 EC. It is based on I/O cages and I/O drawers, a new companion to I/O cages. I/O cages and I/O drawers house I/O cards, which are connected to the CPC through I/O buses. The I/O subsystem is supported by an I/O bus identical to the z10 EC one, and includes the InfiniBand infrastructure (replacing the self-timed interconnect features found in the prior System z servers). This infrastructure is designed to reduce overhead and latency and provide increased data throughput.

InfiniBand

InfiniBand is an industry-standard specification that defines a first-order interconnection technology, which is used to interconnect servers, communications infrastructure equipment, storage, and embedded systems. InfiniBand is a fabric architecture that leverages switched, point-to-point channels with data transfers of up to 120 Gbps, both in chassis backplane applications and through copper and optical fiber connections.

A single connection is capable of carrying several types of traffic, such as communications, management, clustering, and storage. Additional characteristics include low processing overhead, low latency, and high bandwidth. Thus, it can become quite pervasive.

InfiniBand is very scalable, as experience proves, from two-node interconnects to clusters of thousands of nodes, including high-performance computing clusters. It is a mature and field-proven technology, used in thousands of data centers.

InfiniBand is being exploited by the z196 server. Internally, in the server, the cables from the CPC cage to the I/O cages and I/O drawers carry the InfiniBand protocol. For external usage, Parallel Sysplex InfiniBand (PSIFB) links are available which can completely replace the ISC-3 and ICB-4 offerings available on previous servers. They are used to interconnect System z servers in a Parallel Sysplex.

⁷ The function that allows sharing I/O paths across logical partitions is known as the multiple image facility (MIF).

1.1.3 I/O connectivity

The z196 generation of the I/O platform, particularly through the exploitation of InfiniBand, OSA-Express3, FICON® Express8, and High Performance FICON for System z (zHPF), is intended to provide significant performance improvements over the previous I/O platform used for FICON Express4 and OSA-Express2. z196 also offers a new I/O infrastructure element, a companion to the I/O cage of previous systems, the I/O drawer.

I/O drawer

I/O drawers provide increased I/O granularity and capacity flexibility and can be concurrently added and removed in the field, an advantage over I/O cages, which also eases pre-planning. The z196 server can have up to four I/O drawers, two on the A frame and two on the Z frame. I/O drawers were first offered with the z10 BC and can accommodate up to eight I/O features, in any combination.

I/O cage

The z196 has a CPC cage and, optionally, one I/O cage in the A frame. The Z frame can accommodate an additional I/O cage. Each I/O cage can accommodate up to 28 I/O features, in any combination. Adding a third I/O cage requires an RPQ.

I/O features

The z196 supports the following I/O features, which can be installed in both the I/O drawers and I/O cages:

- ▶ ESCON®
- ▶ FICON Express8
- ▶ FICON Express4 (when carried forward on a migration)
- ▶ OSA-Express3
- ▶ OSA-Express2 (when carried forward on a migration, except OSA-Express2 10 GbE LR)
- ▶ Crypto Express3
- ▶ ISC-3 coupling links

ESCON channels

The Enterprise Systems Connection (ESCON) channels support connectivity to ESCON disks, tapes, and printer devices. Historically, they represent the first use of optical I/O technology on the mainframe. They are much slower than FICON channels. FICON Express8 is the preferred technology. The maximum number of supported ESCON features is 16 (up to 240 ports) on the z196. An RPQ will allow you to go beyond 16 features if necessary.

Statement of Direction: The z196 will be the last high-end server to offer ordering of ESCON channels. IBM intends not to offer ESCON channels on future servers.

FICON channels

Fibre Connection (FICON) channels follow the Fibre Channel (FC) standard and support data storage and access requirements as well as the latest FC technology in storage and access devices. FICON channels support the following protocols:

- ▶ Native FICON, Channel-to-Channel (CTC) connectivity, and zHPF traffic to FICON devices such as disks, tapes, and printers in z/OS, z/VM, z/VSE, z/TPF, and Linux on System z environments.
- ▶ Fibre Channel Protocol (FCP) in z/VM and Linux on System z environments support connectivity to disks and tapes through Fibre Channel switches and directors. z/VSE

supports FCP for SCSI disks only. The FCP channel can connect to FCP SAN fabrics and access FCP/SCSI devices.

It is possible to choose any combination of the FICON Express8 and FICON Express4 features. Depending on the feature, auto-negotiated link data rates of 1, 2, 4, or 8 Gbps are supported (1, 2, and 4 for FICON Express4; 2, 4, and 8 for FICON Express 8). FICON Express8 provides significant improvements in start I/Os and data throughput.

Statement of Direction: The z196 will be the last server to support FICON Express4 features.

Open Systems Adapter

The Open Systems Adapter (OSA) features provide local networking (LAN) connectivity and comply with IEEE standards. In addition, OSA features assume several functions of the TCP/IP stack that would normally be performed by the processor. This can provide significant performance benefits.

The z196 can have up to 24 OSA features (96 ports). It is possible to choose any combination of OSA-Express2 and OSA-Express3 features, with the exception of the OSA-Express2 10 GbE LR, which is not supported.

Cryptography

The Crypto Express3 feature provides for tamper-proof, high-performance cryptographic operations. Each feature has two PCI-X or PCI Express adapters. Each of the adapters can be configured as either a coprocessor or an accelerator:

- ▶ Crypto Express Coprocessor: for secure key-encrypted transactions (default)
 - Designed to support security-rich cryptographic functions, use of secure encrypted key values, and user-defined extensions (UDX)
 - Designed for Federal Information Processing Standard (FIPS) 140-2 Level 4 certification
- ▶ Crypto Express Accelerator: for Secure Sockets Layer (SSL) acceleration
 - Designed to support high-performance clear key RSA operations
 - Offloads compute-intensive RSA public-key and private-key cryptographic operations employed in the SSL protocol

Support for 13-digit through 19-digit personal account numbers is provided for stronger protection of data.

The tamper-resistant hardware security module, which is contained in the Crypto Express features, is designed to meet the FIPS 140-2 Level 4 security requirements for hardware security models.

The configurable Crypto Express3 feature is supported by z/OS, z/VM, and Linux on System z. z/VSE supports clear-key RSA operations only.

Coupling links

Coupling links are used in the Parallel Sysplex cluster configurations of System z servers. The links provide high-speed bidirectional communication between members of the sysplex. The z196 supports internal coupling links for memory-to-memory transfers, 12x InfiniBand for distances up to 150 meters (492 feet), and InterSystem Channel-3 (ISC-3) and 1x InfiniBand for unrepeated distances up to 10 km (6.2 miles).

Statement of Direction: The z196 will be the last server to offer ISC-3 features.

HiperSockets

The HiperSockets function is an integrated function of the z196 that provides users with attachments to up to 32 high-speed *virtual* local area networks with minimal system and network overhead.

HiperSockets is a function of the virtualization Licensed Internal Code (LIC) and performs memory-to-memory data transfers in a totally secure way. HiperSockets eliminates having to utilize I/O subsystem operations and having to traverse an external network connection to communicate between logical partitions in the same z196 server. Therefore, HiperSockets offers significant value in server consolidation by connecting virtual servers.

1.1.4 zEnterprise BladeCenter Extension

The zEnterprise BladeCenter Extension Model 002 (zBX) is available as an option with the z196 servers and consists of the following:

- ▶ Up to four IBM Enterprise racks
- ▶ Up to eight BladeCenter⁸ chassis with up to 14 blades⁹ each
- ▶ IBM blades, up to 112, based on POWER7 technology
- ▶ Two Top the Rack (TOR) 1000BASE-T switches for the intranode management network (INMN). The INMN provides connectivity for management purposes between the z196 support elements and zBX.
- ▶ Two Top the Rack (TOR) 10 GbE switches for the intraensemble data network (IEDN). The IEDN is used for data paths between the z196 and the zBX, and the other ensemble members.
- ▶ 8 Gbps Fiber Channel switch modules for connectivity to an SAN
- ▶ Power Distribution Units (PDUs) and cooling fans

The zBX is configured with redundant components to provide qualities of service similar to that of System z, such as firmware management and the capability for concurrent upgrades and repairs.

The zBX provides a foundation for the future. Based on IBM's judgement of the market's needs, additional specialized or general purpose blades might be introduced¹⁰.

Statement of Direction: In the first half of 2011, IBM intends to offer a System x blade running Linux and a WebSphere DataPower Appliance, for the zBX Model 002.

IBM blades

IBM offers a selected set of IBM POWER7 blades that can be installed and operated on the zBX. These blades have been tested to ensure compatibility and manageability in the z196 environment.

The blades are virtualized and their virtual servers run the AIX operating system.

⁸ The IBM Smart Analytics Optimizer solution has a maximum of two zBX racks (B and C) and up to four BladeCenter chassis

⁹ Depending on the IBM Smart Analytics Optimizer configuration, this can be either 7 or 14 blades for the first BladeCenter chassis.

¹⁰ All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

IBM Smart Analytics Optimizer solution

The IBM Smart Analytics Optimizer solution is a defined set of software and hardware that provides a cost-optimized solution for running Data Warehouse and Business Intelligence queries against DB2 for z/OS, with fast and predictable response times, while retaining the data integrity, data management, security, availability and other qualities of service of the z/OS environment. It exploits special purpose blades, housed in a zBX.

For a more detailed description of the IBM Smart Analytics Optimizer see “IBM Smart Analytics Optimizer solution” on page 77.

1.1.5 Unified Resource Manager

Unified Resource Manager provides energy monitoring and management, goal-oriented policy management, increased security, virtual networking, and data management, consolidated in a single interface that can be tied to business requirements.

Unified Resource Manager is a set of functions that can be grouped as follows:

- ▶ Defining and managing virtual environments. This includes the automatic discovery and definition of I/O and other hardware components across z196 and zBX, as well as the definition and management of virtual servers, virtualized LANs, and ensemble members.
- ▶ Defining and managing workloads and workload policies.
- ▶ Receiving and applying corrections and upgrades to the Licensed Internal Code.
- ▶ Performing temporary and definitive z196 capacity upgrades.

The functions that pertain to an ensemble are provided by the Hardware Management Console (HMC) and Support Elements. See “Unified Resource Manager” on page 90 for more information.

1.2 Capacity On Demand

On the z196 it is possible to perform just-in-time deployment of capacity resources. This function is designed to provide more flexibility to dynamically change capacity when business requirements change. No interaction is required with IBM at the time of activation.

For example, you can:

- ▶ Define one or more flexible configurations that can be used to solve multiple temporary situations.
- ▶ Have multiple configurations active at once, and the configurations themselves have flexible selective activation of only the needed resources.
- ▶ Purchase capacity either before or after execution for On/Off Capacity on Demand. This capacity is represented by tokens that are consumed at execution time.
- ▶ Add permanent capacity to the server while temporary changes are active.

A similar capability is *not* available with the zBX.

For more information, refer to “z196 Capacity on Demand (CoD)” on page 73.

1.3 Software

The zEnterprise System is supported by a large set of software, including ISV applications. The extensive software portfolio spans, on the System z server, from IBM WebSphere®, full

support for service-oriented architecture (SOA), web services, J2EE, Linux, and open standards, to the more traditional batch and transactional environments such as Customer Information Control System (CICS®) and Information Management System (IMS™).

For instance, considering just the Linux on System z environment, more than 3,000 applications are offered by over 400 independent software vendors (ISVs). In addition, any AIX products running today on System p® servers should continue to run on the zBX blades virtualized AIX environment. There are also specialized solutions such as the IBM Smart Analytics Optimizer.

Exploitation of some features may require the latest releases. The supported operating system support for the z196 includes:

- ▶ z/OS Version 1 Release 10 or later releases
- ▶ z/OS Version 1 Release 7 with the IBM Lifecycle Extension with PTFs¹¹
- ▶ z/OS Version 1 Release 8 with the IBM Lifecycle Extension with PTFs¹¹
- ▶ z/OS Version 1 Release 9 with the IBM Lifecycle Extension with PTFs¹¹
- ▶ z/VM Version 5 Release 4 or later
- ▶ z/VSE Version 4 Release 1 or later
- ▶ z/TPF Version 1 Release 1
- ▶ Linux on System z distributions:
 - Novell SUSE: SLES 10 and SLES 11¹²
 - Red Hat: RHEL 5¹³

Operating system support for the POWER7 blades is AIX Version 5 Release 3 or later.

IBM compilers

You can empower your business applications with IBM compilers on the IBM zEnterprise™ System.

With IBM® Enterprise COBOL and Enterprise PL/I, you can leverage decades of IBM experience in application development, to integrate COBOL and PL/I with web services, XML, and Java™. Such interoperability enables you to capitalize on existing IT investments while smoothly incorporating new, web-based applications into your organization's infrastructure.

z/OS XL C/C++ helps you to create and maintain critical business applications written in C or C++, to maximize application performance and improve developer productivity. z/OS XL C/C++ can transform C or C++ source code to fully exploit System z® hardware (including the zEnterprise 196), through hardware tailored optimizations, built-in functions, performance-tuned libraries, and language constructs that simplify system programming and boost application runtime performance.

Enterprise COBOL, Enterprise PL/I and XL C/C++ are leading-edge, z/OS-based compilers that maximize middleware by providing access to IBM DB2®, CICS, and IMS systems.

More information about software support can be found in “Software support summary” on page 100.

¹¹ z/OS.e is not supported

¹² SLES is the abbreviation for Novell SUSE Linux Enterprise Server.

¹³ RHEL is the abbreviation for Red Hat Enterprise Linux



zEnterprise System hardware overview

The zEnterprise System is the next step in the evolution of the mainframe family. It continues this evolution by introducing several innovations and expanding existing functions, building upon the z/Architecture.

This chapter expands upon the overview of key hardware elements of the z196 and zBX provided in Chapter 1, “Proposing an infrastructure (r)evolution” on page 1, and compares them with the System z10 EC server, where relevant.

This chapter discusses the following topics:

- ▶ 2.1, “z196 highlights” on page 20
- ▶ 2.2, “Models and model upgrades” on page 20
- ▶ 2.3, “The frames” on page 22
- ▶ 2.4, “CPC cage and books” on page 24
- ▶ 2.5, “Multi-chip module (MCM)” on page 25
- ▶ 2.6, “z196 processor chip” on page 26
- ▶ 2.7, “Processor unit (PU)” on page 27
- ▶ 2.8, “Memory” on page 28
- ▶ 2.9, “I/O system structure” on page 30
- ▶ 2.10, “I/O cages, drawers, and features” on page 31
- ▶ 2.11, “Cryptographic functions” on page 38
- ▶ 2.12, “Coupling and clustering” on page 40
- ▶ 2.13, “Time functions” on page 42
- ▶ 2.14, “z196 HMC and SE” on page 43
- ▶ 2.15, “Power and cooling” on page 44
- ▶ 2.16, “zEnterprise BladeCenter Extension” on page 46

2.1 z196 highlights

The major z196 improvements over its predecessors include:

- ▶ Increased total system capacity in a 96-way server (with 80 characterizable PUs) and additional subcapacity settings, offering increased levels of performance and scalability to help enable new business growth
- ▶ Quad-core 5.2 GHz processor chips that can help improve the execution of processor-intensive workloads
- ▶ Implementation of out-of-order instruction execution
- ▶ Cache structure improvements and larger cache sizes that can benefit most production workloads
- ▶ Improved availability in the memory subsystem with redundant array of independent memory (RAIM)
- ▶ Up to 3 TB of available real memory per server for growing application needs (with up to 1 TB real memory per logical partition)
- ▶ Just-in-time deployment of capacity resources, which can improve flexibility when making temporary or permanent changes, and plan-ahead memory for nondisruptive memory upgrades
- ▶ A 16 GB fixed hardware system area (HSA) that is managed separately from customer-purchased memory
- ▶ Exploitation of InfiniBand technology
- ▶ Improvements to the I/O subsystem and new I/O features
- ▶ Additional security options for the CP Assist for Cryptographic Function (CPACF)
- ▶ A HiperDispatch function for improved efficiencies in hardware and z/OS software
- ▶ Hardware decimal floating point on each core on the processor unit (PU)
- ▶ Server Timer Protocol (STP) enhancements for time accuracy, availability, and systems management with message exchanges using ISC-3 or 1x InfiniBand connections

In all, these enhancements provide customers with options for continued growth, continuity, and ability to upgrade.

For an in-depth discussion of the zEnterprise 196 functions and features see the *IBM zEnterprise System Technical Guide*, SG24-7833.

2.2 Models and model upgrades

The z196 has been assigned a machine type (M/T) of 2817, which uniquely identifies the server. The server is offered in five different models. These models are named M15, M32, M49, M66, and M80. The model determines the maximum number of processor units (PUs) available for characterization. PUs are delivered in single-engine increments. The first four models utilize a 20-PU multi-chip module (MCM), of which 15 to 17 PUs are available for characterization. The fifth model, M80, utilizes 24-PU MCMs to provide a maximum of 80 configurable PUs.

Spare PUs and system assist processors (SAPs) are integral to the server. Refer to Table 2-1 for a model summary including SAPs and spare PUs for the different models. For an explanation of PU characterization see “PU characterization” on page 27.

The five z196 server orderable models are shown in Table 2-1.

Table 2-1 Model summary

Model	Books/PUs	CPs	Standard SAPs	Spares
M15	1/20	0–15	3	2
M32	2/40	0–32	6	2
M49	3/60	0–49	9	2
M66	4/80	0–66	12	2
M80	4/96	0–80	14	2

The z196 offers 125 different capacity levels, which span a range of approximately 1 to 220. This is discussed in 3.2.3, “Granular capacity and capacity settings” on page 58.

Figure 2-1 summarizes the upgrade paths to z196.

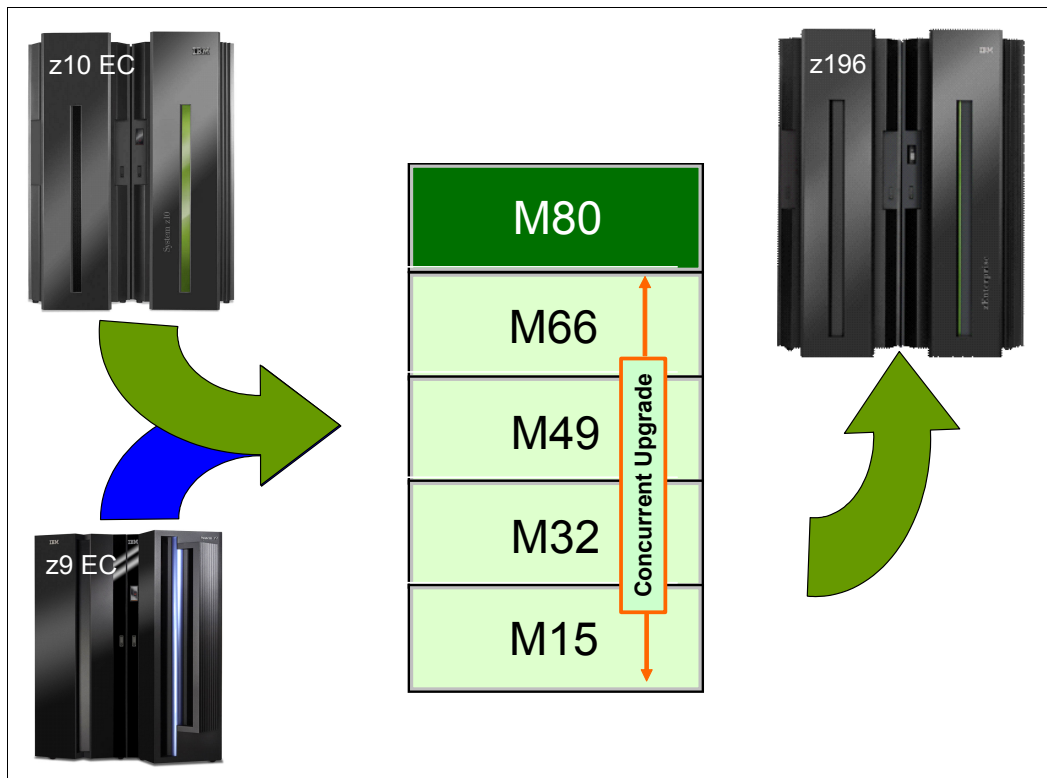


Figure 2-1 z196 upgrade paths

z196 upgrades

Model upgrades within the server family are accomplished by installing additional books. Books, being on separate power boundaries, are physically isolated from each other, thereby allowing them to be plugged and unplugged independently. Refer to Table 2-2 for upgrades available within the family.

Table 2-2 z196 to z196 upgrade paths

Model	M15	M32	M49	M66	M80
M15	—	Yes	Yes	Yes	Yes
M32	—	—	Yes	Yes	Yes
M49	—	—	—	Yes	Yes
M66	—	—	—	—	Yes

All z196 to z196 model upgrades are concurrent except when the target is the model M80. This is a non-concurrent upgrade because model M80 uses a different set of MCMs.

Upgrades to z10 EC and z196 from z9 EC

Upgrades are also available from the currently installed z10 EC and z9 EC servers. These upgrades are disruptive.

2.3 The frames

The z196 server is always a two-frame system. The frames are called the A frame and the Z frame. Several hardware elements pointed out are described later in this chapter.

The z196 can be delivered as an air cooled server or as a water cooled server. See Figure 2-2 for an internal front view of the two frames for an air cooled server.

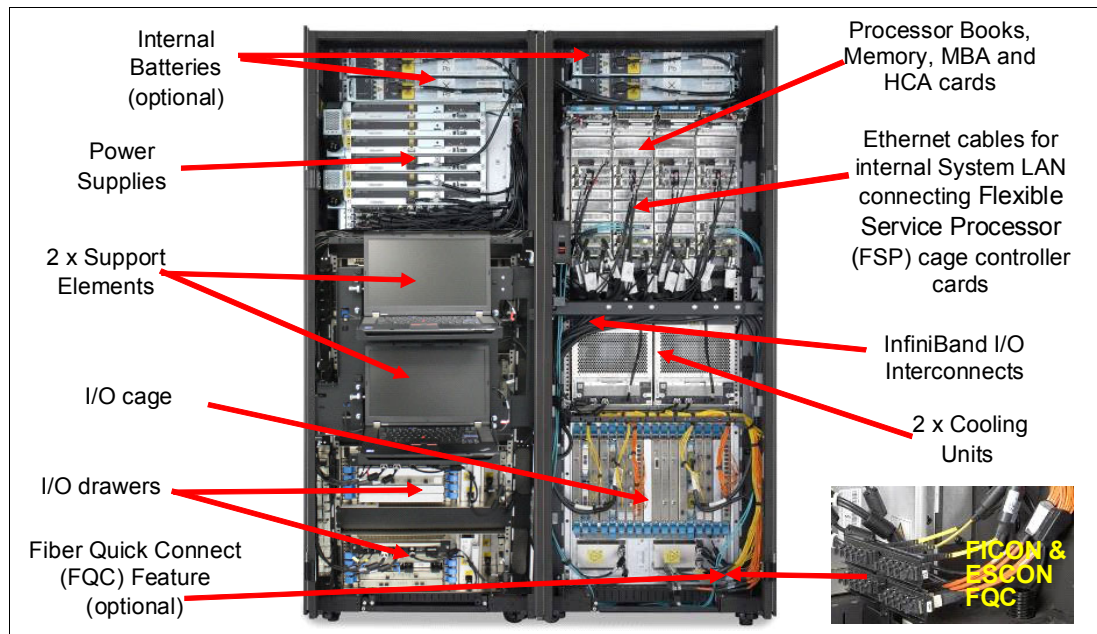


Figure 2-2 z196 internal front view - air cooled server

Figure 2-3 shows an internal front view of the two frames of a water cooled server.

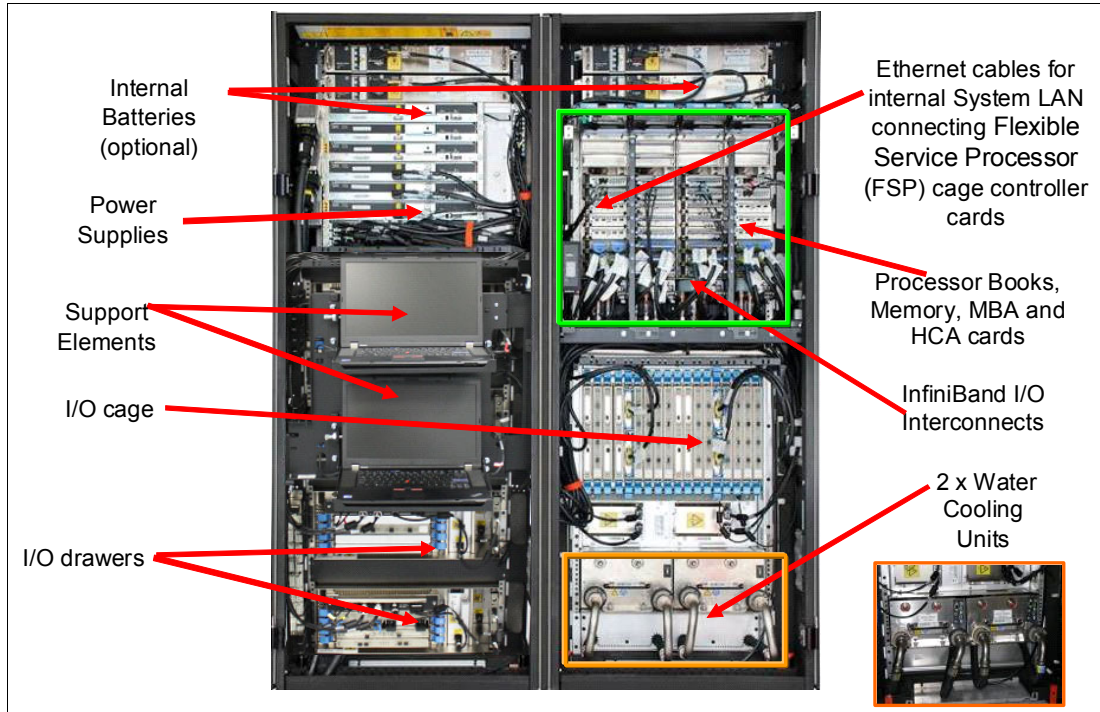


Figure 2-3 z196 internal front view - water cooled server

Table 2-3 lists the physical dimensions of the system and its frames.

Table 2-3 z196 physical dimensions

Maximum	A and Z frames without the Internal Battery - Model M80	A and Z frames with the Internal Battery - Model M80
Air cooled server		
Weight kg (lbs)	1894 (4175)	2177 (4799)
Width mm (in)	1534 (60.7)	1534 (60.7)
Depth mm (in)	1273 (50.1)	1273 (50.1)
Height mm (in)	2012 (79.2)	2012 (79.2)
Water cooled server		
Weight kg (lbs)	1902 (4193)	2185 (4817)
Width mm (in)	1534 (60.7)	1534 (60.7)
Depth mm (in)	1375 (54.1)	1375 (54.1)
Height mm (in)	2012 (79.2)	2012 (79.2)
Notes:		
1. Weight includes covers. Width, depth, and height are indicated without covers.		
2. Weight is based on maximum system configuration, not the addition of the maximum weight of each frame.		
3. Width increases to 1846 mm (72.7 in) if the top exit for I/O cables feature (FC 7942) is installed.		
4. Weight increases by 43.1 kg (95 lbs) if the I/O top exit feature is installed.		
5. If Feature code 7942 is specified the weight increases by and the width increases by		

2.3.1 Top exit I/O cabling

On the z196 you now have the option of ordering the infrastructure to support top exit of your fiber optic cables (ESCON, FICON, OSA, 12x InfiniBand, 1x InfiniBand, and ISC-3) as well as your copper cables for the 1000BASE-T Ethernet features.

Top exit I/O cabling is designed to provide you with an additional option. Instead of all of your cables exiting under the server and/or under the raised floor, you now have the flexibility to choose the option that best meets the requirements of your data center. Top exit I/O cabling can also help to increase air flow. This option is offered on new build as well as MES orders.

2.4 CPC cage and books

The z196 server has a multi-book system structure similar to z10 EC servers. Up to four books can be installed on a z196 server. A book looks like a box and plugs into one of the four slots in the central processing complex (CPC) cage of the z196 server. The CPC cage is located in the in A frame of the z196 server. Refer to Figure 2-2 on page 22 for a pictorial view of the CPC cage and the location of the four books. Each book contains:

- ▶ A multi-chip module (MCM). Each MCM includes six quad-core processor unit (PU) chips and two storage control (SC) chips. MCMs are further described in 2.5, “Multi-chip module (MCM)” on page 25. Refer to Table 2-1 on page 21 for the model summary and the relation between the number of books and number of available PUs.
- ▶ A minimum of 16 and a maximum of 752 GB of memory for customer use.
- ▶ A combination of up to eight InfiniBand Host Channel Adapter (HCA2-Optical or HCA2-Copper) fanout cards. Each of the cards has two ports, thereby supporting up to 16 connections. HCA2-Copper connections are for links to the I/O cages in the server, and the HCA2-Optical connections are to external servers (coupling links).
- ▶ Three distributed converter assemblies (DCAs) that provide power to the book. Loss of a DCA leaves enough book power to satisfy the book’s power requirements. The DCAs can be concurrently maintained.

Figure 2-4 shows a view of a z196 book without the containing box.

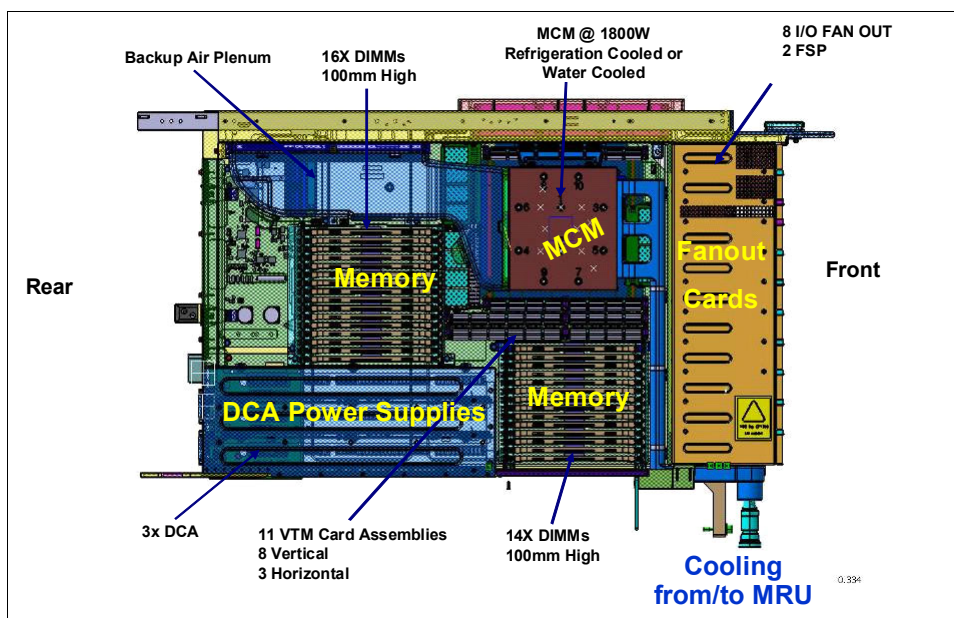


Figure 2-4 z196 book structure and components

The z196 offers a significant increase in system scalability and opportunity for server consolidation by providing a multi-book system structure. As shown in Figure 2-5, all books are interconnected in a star configuration with high-speed communications links through the L4 shared caches, which allows the system to be operated and controlled by the PR/SM facility as a symmetrical, memory-coherent multiprocessor.

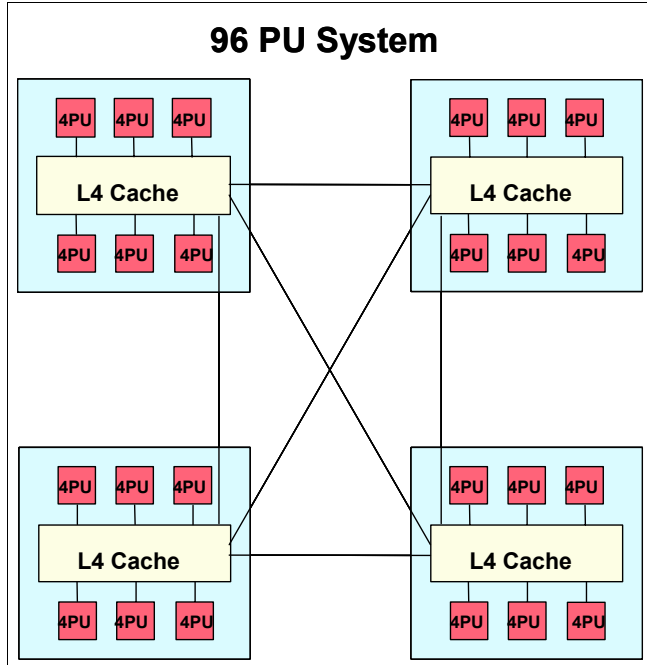


Figure 2-5 z196 inter-book communication structure

The point-to-point connection topology allows direct communication among all books.

2.5 Multi-chip module (MCM)

The multi-chip module is a high-performance, glass-ceramic module, providing the highest level of processing integration in the industry. It is the heart of the server (Figure 2-6).

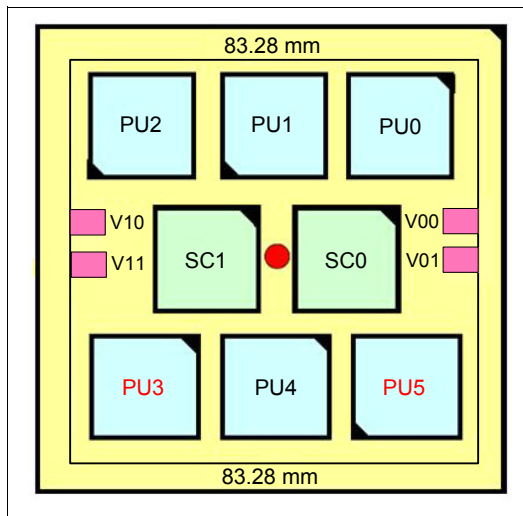


Figure 2-6 z196 multi-chip module

The z196 MCM has eight chip sites. All chip types on the MCM use Complementary Metal Oxide of Silicon (CMOS) 12s chip technology. CMOS 12s is a state-of-the-art microprocessor technology based on 13-layer copper interconnections and silicon-on insulator (SOI) technologies. The chip lithography line width is 0.045 μm (45 nm). The processor unit chip contains close to 1.4 billion transistors in a 512.3 mm^2 die.

There is one MCM per book and the MCM contains all of the processor chips and L4 cache of the book. The z196 server has six PU chips per MCM and each PU chip has up to four PUs (cores), as shown on Figure 2-7. Two MCM options are offered: with 20 or 24 PUs. All the models employ an MCM size of 20 PUs except for the model M80, which has four books with 24 PU MCMs, for a total of 96 PUs.

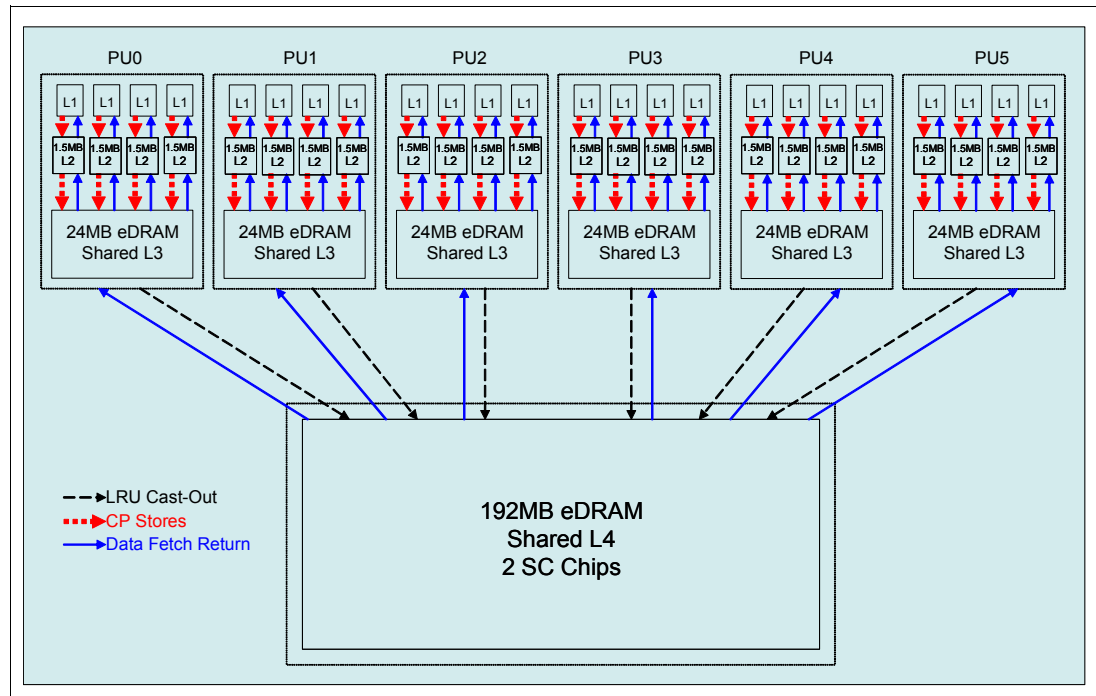


Figure 2-7 MCM chips and cache structure

The MCM also has two storage control (SC) chips. Each SC chip packs 96 MB of eDRAM cache, interface logic for 24 cores, and SMP fabric logic into 478.8 mm^2 . The two SC chips are configured to provide a single 192 MB cache shared (L4) by all 20 or 24 cores on the module, yielding outstanding SMP scalability on real-world transaction processing workloads.

There are four SEEPROM (S) chips, of which two are active and two are redundant that contain product data for the MCM, chips, and other engineering information. The clock functions are distributed across PU and SC chips.

2.6 z196 processor chip

The z196 features a high-frequency four-core processor chip, an advanced microprocessor design, a robust cache hierarchy, and an SMP design optimized for enterprise database and transaction processing workloads, as well as for workloads such as Java™ and Linux.

It leverages leading-edge technology and circuit design techniques while building on the rich heritage of mainframe system design, including industry-leading reliability, availability, and serviceability. New functional features enable increased software efficiency and scalability

while maintaining full compatibility with existing software. Further detail is given in 3.2.1, “Microprocessor” on page 56.

2.7 Processor unit (PU)

A PU is the generic term for the z/Architecture processor on the multi-chip module (MCM). A PU is imbedded in a System z chip core. Each PU is a superscalar, processor with the following attributes:

- ▶ Up to three instructions can be decoded per cycle.
- ▶ Up to five instructions can be executed (finished) per cycle.
- ▶ Instructions can be executed out of order. A high-frequency, low-latency, pipeline, providing robust performance across a wide range of workloads, is used.
- ▶ Memory accesses might not be in the same instruction order (out-of-order operand fetching).
- ▶ Most instructions flow through a pipeline with different numbers of steps for various types of instructions. Several instructions may be in progress at any moment, subject to the maximum number of decodes and completions per cycle.

Each PU has an L1 cache divided into a 64 KB cache for instructions and a 128 KB cache for data. Each PU also has a private L2 cache, with 1.5 MB in size. Each PU chip contains a L3 cache, shared by all four PUs on the chip. The shared L3 cache uses eDRAM and has 24 MB. The cache structure is shown in Figure 2-7 on page 26. This implementation optimizes performance of the system for high-frequency, very fast processors.

Each L1 cache has a translation look-aside buffer (TLB) of 512 entries associated with it. In addition, a secondary TLB is used to further enhance performance. This structure supports large working sets, multiple address spaces, and a two-level virtualization architecture.

Hardware fault detection is imbedded throughout the design and combined with comprehensive instruction-level retry and dynamic CPU sparing. Those provide the reliability and availability required for true mainframe quality.

The z196 processor provides full compatibility with existing software for ESA/390 and z/Architecture, while extending the Instruction Set Architecture (ISA) to enable enhanced function and performance. Over 110 new hardware instructions support more efficient code generation and execution.

Decimal floating-point hardware fully implements the new IEEE 754r standard, helping provide better performance and higher precision for decimal calculations, an enhancement of particular interest to financial institutions.

On-chip cryptographic hardware includes extended key and hash sizes for the Advanced Encryption Standard (AES) and Secure Hash Algorithm (SHA) algorithms.

PU characterization

Processor units are ordered in single increments. The internal server functions, based on the configuration ordered, *characterize* processor into various types during initialization of the processor—often called a power-on reset (POR) operation. Characterizing PUs dynamically without a POR is possible. A processor unit that is not characterized cannot be used.

At least one CP must be purchased with, or before, a zAAP or zIIP can be purchased. Customers can purchase one zAAP, one zIIP, or both, for each CP (assigned or unassigned)

on the server. However, a logical partition definition can contain more zAAPs or zIIPs than CPs. For example, in a server with two CPs a maximum of two zAAPs and two zIIPs can be installed. A logical partition definition for that server could contain up to two logical CPs, two logical zAAPs, and two logical zIIPs.

Converting a processor from one type to any other type is possible. These conversions happen concurrently with the operation of the system.

Notes: The addition of ICFs, IFLs, zAAP, zIIPs, and SAPs to a server does not change the server capacity setting or its MSU rating (only CPs do).

IBM does not impose any software charges on work dispatched on zAAP and zIIP processors.

2.8 Memory

Maximum physical memory sizes are directly related to the number of books in the system, and a z196 server has more memory installed than ordered. Part of the physical installed memory is used to implement the redundant array of independent memory (RAIM) design, resulting on up to 768 GB of available memory per book and up to 3072 GB (3 TB) per system. As the hardware system area (HSA) memory has a fixed amount of 16 GB and is managed separately from customer memory, you may order up to 752 GB on the one-book model and up to 3056 GB on the four-book server models.

Table 2-4 shows the maximum and minimum memory sizes for each z196 model.

Table 2-4 z196 servers memory sizes

Model	Number of books	Customer memory (GB)
M15	1	32 - 704
M32	2	32 - 1520
M49	3	32 - 2288
M66	4	32 - 3056
M80	4	32 - 3056

The minimum physical installed memory is 40 GB per book, and the minimum initial amount of memory that can be ordered is 32 GB for all z196 models. The maximum customer memory size is based on the physical installed memory minus RAIM and minus HSA memory.

On z196 servers, the memory granularity varies from 32 GB, for customer memory sizes from 32 to 256 GB, up to 256GB, for servers having from 1776 GB to 3056 GB of customer memory. Table 2-5 shows the memory granularity depending on the installed customer memory.

Table 2-5 Memory granularity

Granularity (GB)	Customer memory (GB)
32	32 - 256
64	320 - 512

Granularity (GB)	Customer memory (GB)
96	608 - 896
112	1008
128	1136 - 1520
256	1776 - 3056

Physically memory is organized as follows:

- ▶ A book always contains a minimum of 40 GB of physical installed memory.
- ▶ A book may have more memory installed than enabled. The excess amount of memory can be enabled by a Licensed Internal Code load when required by the installation.
- ▶ Memory upgrades are satisfied from already-installed unused memory capacity until exhausted. When no more unused memory is available from the installed memory cards, either the cards must be upgraded to a higher capacity or the addition of a book with additional memory is necessary.

When activated, a logical partition can use memory resources located in any book. No matter in which book the memory resides, a logical partition has access to that memory if so allocated. Despite the book structure, the z196 is still a Symmetric Multi-Processor (SMP).

A memory upgrade is concurrent when it requires no change of the physical memory cards. A memory card change is disruptive when no use is made of Enhanced Book Availability. Refer to *IBM zEnterprise System Technical Guide*, SG24-7833, for a description of Enhanced Book Availability.

For a model upgrade that results in the addition of a book, the minimum memory increment is added to the system. Remember that the minimum physical memory size in a book is 40 GB. During a model upgrade, the addition of a book is a concurrent operation. The addition of the physical memory that resides in the added book is also concurrent.

Concurrent memory upgrade

Memory can be upgraded concurrently using Licensed Internal Code - Configuration Control (LIC-CC) if physical memory is available as described previously. The plan-ahead memory function available with the z196 server provides the ability to plan for nondisruptive memory upgrades by having in the system pre-plugged memory, based on a target configuration. Pre-plugged memory is enabled through a LIC-CC order placed by the customer.

Redundant array of independent memory (RAIM)

z196 introduces the redundant array of independent memory (RAIM) on System z servers, making the memory subsystem essentially a fully fault tolerant N+1 design. RAIM design detects and recovers from DRAM, socket, memory channel or DIMM failures automatically. The RAIM design requires the addition of one memory channel that is dedicated for RAS.

Hardware system area

The hardware system area (HSA) is a reserved memory area that is used for several internal functions, but the bulk is used by channel subsystem functions. The HSA has grown with each successive mainframe generation. On previous servers, model upgrades and also new logical partition definitions or changes required pre-planning and were sometimes disruptive because of changes in HSA size. For further information and benefits see 3.2.4, "Memory" on page 59.

2.9 I/O system structure

Figure 2-8 shows a high-level view of the I/O system structure for the z196 server.

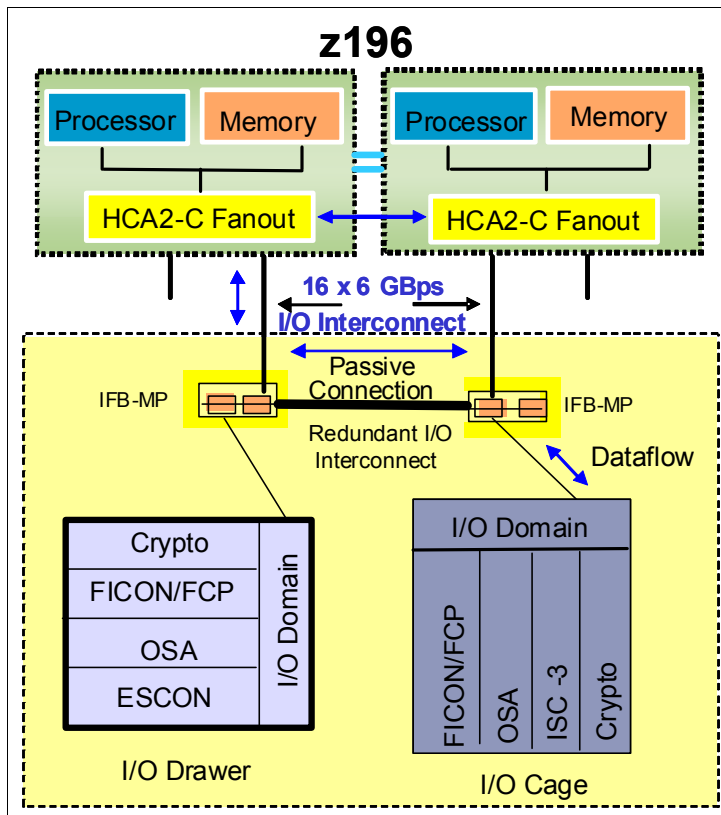


Figure 2-8 z196 system structure for I/O

The z196 has several types of fanout cards residing on the front of the book package:

- ▶ An InfiniBand HCA2-C (copper) fanout that supports ESCON, FICON, OSA, ISC-3, and Crypto Express3 features in the I/O cages
- ▶ The z196 supports up to eight fanouts (HCA-C, HCA-O, HCA2-O LR) for each book, with a maximum of 24 for a 4-book system. Each fanout comes with two ports, giving a maximum of 48 ports for I/O connectivity.

The z196 exploits InfiniBand (IFB) connections to I/O cages, driven from the Host Channel Adapter (HCA2-C) fanout cards that are located on the front of the book. The HCA2-C fanout card is designated to connect to an I/O cage or an I/O drawer by a copper cable. The two ports on the fanout card are dedicated to I/O.

The z196 server has up to eight fanout cards (numbered D1, D2, and D5 to DA) per book, each driving two IFB cables, resulting in up to 16 IFB connections per book.

In a system configured for maximum availability, alternate paths maintain access to critical I/O devices, such as disks, networks, and so on.

Refer to *System z Connectivity Handbook*, SC24-5444, for a more detailed description of the I/O interfaces.

Coupling connectivity

In addition to the HCA2-C fanout card, the z196 has two additional fanout cards, the HCA2-O, and the HCA2-O LR fanout. These cards are exclusively used for coupling link connectivity in a Parallel Sysplex configuration.

The HCA2-O provides optical connections for InfiniBand I/O interconnect (Parallel Sysplex using InfiniBand (PSIFB)) between:

- ▶ z196, z10 EC or z10 BC servers. This connection has a maximum link data rate of up to 6 GB per second.
- ▶ A z196, or a z10, and a System z9® server. This connection has a maximum data rate of up to 3 GB per second.

Note: The InfiniBand link data rate of 6 GBps or 3 GBps does not represent the performance of the link. The actual performance depends on many factors, such as latency through the adapters, cable lengths, and the type of workload. Although the link data rate can be higher with InfiniBand coupling links than with ICB links (only supported on servers prior to z196), the service times of coupling operations are greater.

The HCA-O LR card supports a maximum data link rate of 5 Gbps on a coupling link connection between two z196 or z10 servers. HCA2-O LR is exclusive to System z196 and z10 servers.

As indicated previously, the HCA2-C provides copper connections for InfiniBand I/O interconnect from book to I/O cards in I/O cages.

2.10 I/O cages, drawers, and features

Each book has up to eight dual-port fanout cards to transfer data. Each port has a bi-directional bandwidth of 6 GBps. Up to 16 IFB I/O interconnect connections provide an aggregated bandwidth of up to 96 GBps per book.

The HCA2-C IFB I/O interconnect connects to an I/O cage or an I/O drawer that may contain a variety of channels, coupling link, OSA-Express, and cryptographic features.

The z196 server supports up to four I/O drawers, two I/O cages or combinations of both. Installation of a third I/O cage requires an RPQ.

Each I/O drawer supports two I/O domains (A and B) for a total of 8 I/O card slots. Each I/O domain uses an IFB-MP card in the I/O drawer and a copper cable to connect to a Host Channel Adapter (HCA) fanout in the CPC cage. The 12 x DDR InfiniBand link between the HCA in the CPC and the IFB-MP in the I/O drawer supports a link rate of 6 GBps.

All cards in the I/O drawer are installed horizontally. The two Distributed Converter Assemblies (DCAs) distribute power to the I/O drawer. The locations of the DCAs, I/O feature cards, and IFB-MP card in the I/O drawer are shown in Figure 2-9 on page 32

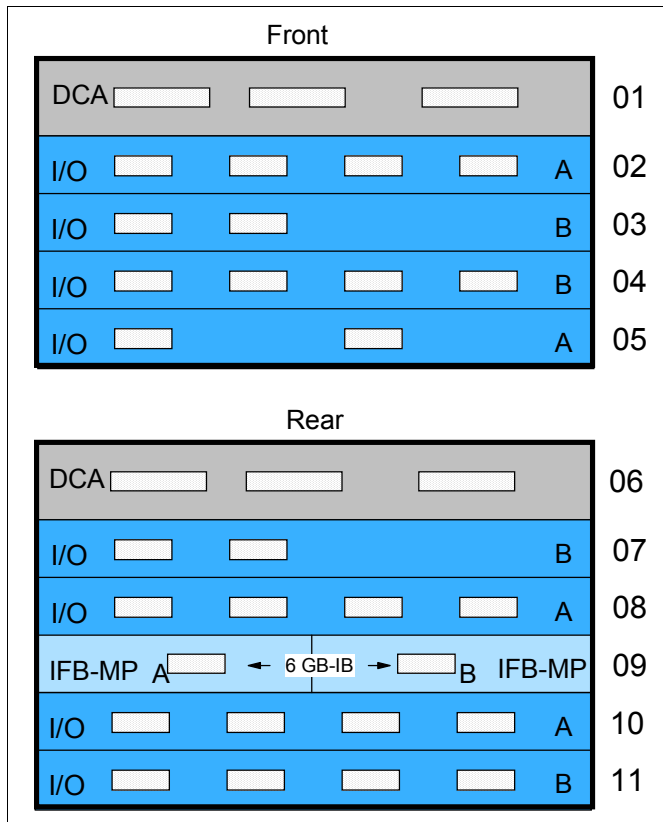


Figure 2-9 I/O feature cards plugging location in an I/O drawer

The IFB-MP cards are installed at location 09 at the rear side of the I/O drawer. The I/O cards are installed from the front and rear side of the I/O drawer. Two I/O domains (A and B) are supported. Each I/O domain has up to four I/O feature cards of any type (ESCON, FICON, ISC or OSA). The I/O cards are connected to the IFB-MP card through the backplane board.

Each I/O cage supports up to seven I/O domains and a total of 28 I/O card slots. Each I/O domain supports four I/O features (ESCON, FICON, OSA, or ISC). See Figure 2-10 on page 33 for a pictorial view of an I/O cage.

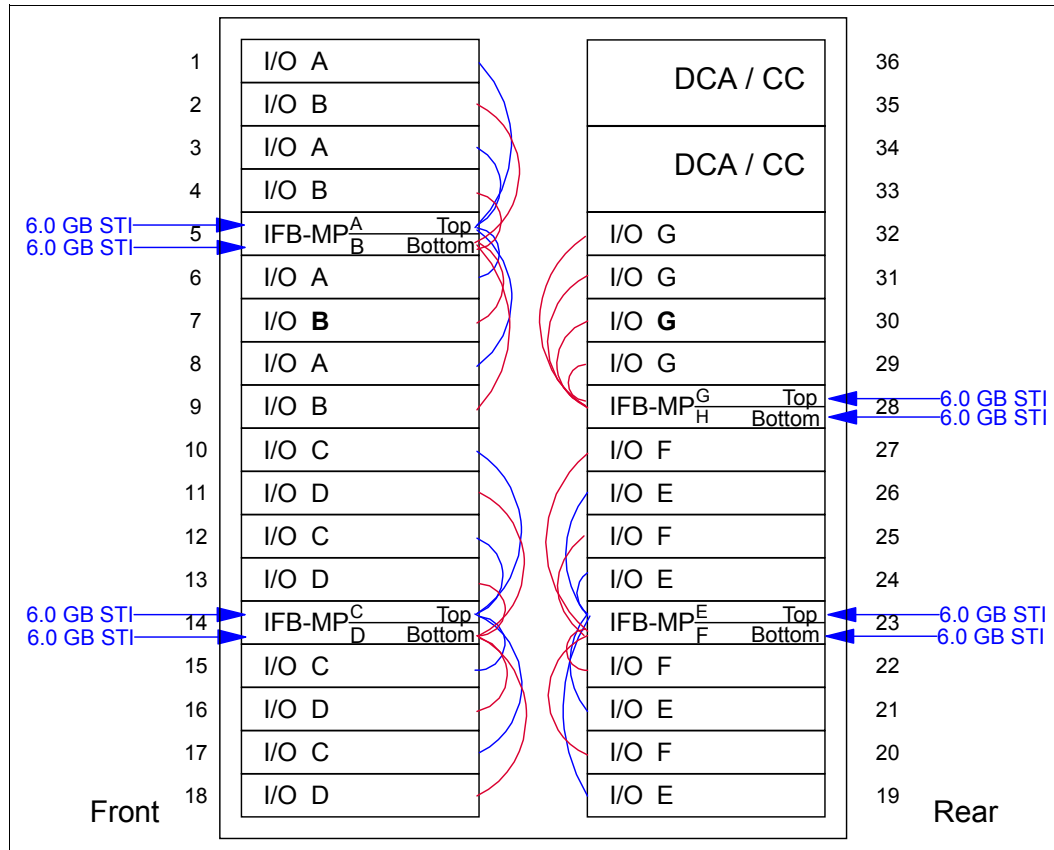


Figure 2-10 z196 I/O cage

The different I/O domains (A, B, C, D, E, F, and G) and the InfiniBand MultiPlexer (IFB-MP), which connects to the CPC cage as well as to the I/O feature itself, are shown. Up to four of the 32 slots in the I/O cage are occupied by the IFB-MB.

The following I/O features can be ordered for a new z196 server:

- ▶ ESCON
- ▶ FICON Express8 LX (long wavelength - 10 km)
- ▶ FICON Express8 SX (short wavelength)
- ▶ OSA-Express3 10 GbE LR (long reach)
- ▶ OSA-Express3 10 GbE SR (short reach)
- ▶ OSA-Express3 GbE LX (long wavelength)
- ▶ OSA-Express3 GbE SX (short wavelength)
- ▶ OSA-Express3 1000BASE-T Ethernet
- ▶ Crypto Express3
- ▶ ISC-3 (peer mode only)

The following features are not orderable for a z196, but if present in a z10 EC or z9 EC server may be *carried forward* when upgrading to a z196:

- ▶ FICON Express4 LX (4 km and 10 km)
- ▶ FICON Express4 SX
- ▶ OSA-Express2 LX (long wavelength)
- ▶ OSA-Express2 SX (short wavelength)
- ▶ OSA-Express2 1000BASE-T Ethernet

The following features are *not* supported on a z196:

- ▶ FICON Express2 (LX and SX)
- ▶ FICON Express (LX and SX)
- ▶ FICON (pre-FICON Express)
- ▶ OSA-Express2 10 GbE Long Reach
- ▶ OSA-Express
- ▶ ICB-2
- ▶ ICB-3
- ▶ ICB-4
- ▶ ISC-3 Links in Compatibility Mode
- ▶ Crypto Express2
- ▶ PCIxCC and PCICA
- ▶ Parallel channels (use an ESCON converter)

For a list of the z196 supported I/O features and their characteristics refer to Appendix D, “Channel options” on page 135.

2.10.1 ESCON channels

ESCON channels support the ESCON architecture and directly attach to ESCON-supported I/O devices.

16-port ESCON feature

The 16-port ESCON feature occupies one I/O slot in an I/O cage or I/O drawer. Each port on the feature uses a 1300 nanometer (nm) light-emitting diode (LED) transceiver, designed to be connected to 62.5 μm multimode fiber optic cables only.

Up to a maximum of 15 ESCON channels per feature are active. There is a minimum of one spare port per feature to allow for channel sparing in the event of a failure of one of the other ports.

ESCON channel port enablement feature

The 15 active ports on each 16-port ESCON feature are activated in groups of four ports through LIC-CC. Each port operates at a data rate of 200 Mbps.

The first group of four ESCON ports requires two 16-port ESCON features. This is for redundancy reasons. After the first pair of ESCON cards is fully allocated (by seven ESCON port groups, using 28 ports), single cards are used for additional ESCON ports groups.

Ports are activated equally across all installed 16-port ESCON features for high availability.

Note: The zEnterprise 196 is planned to be the last high end server to offer ordering of ESCON channels on new builds, migration offerings, upgrades, and System z exchange programs. Enterprises should begin migrating from ESCON to FICON. Alternate solutions are available for connectivity to ESCON devices.

IBM Global Technology Services, through IBM Facilities Cabling Services, offers ESCON to FICON Migration (Offering ID #6948-97D), to help facilitate migration to FICON to simplify and manage a single physical and operational environment while maximizing “green” related savings. For more information see:

<http://www-935.ibm.com/services/us/index.wss/itservice/igs/a1026000>

The PRIZM Protocol Converter Appliance from Optica Technologies Incorporated provides a FICON-to-ESCON conversion function that has been System z qualified. For more information see:

<http://www.opticatech.com/>

Note: IBM cannot confirm the accuracy of compatibility, performance, or any other claims by vendors for products that have not been System z qualified. Questions regarding these capabilities and device support should be addressed to the suppliers of those products.

2.10.2 FICON Express8

Two types of FICON Express8 transceivers are supported on new build z196 servers—one long wavelength (LX) laser version and one short wavelength (SX) laser version:

- ▶ FICON Express8 10KM LX feature
- ▶ FICON Express8 SX feature

Each port supports attachment to the following:

- ▶ FICON/FCP switches and directors that support 2 Gbps, 4 Gbps, or 8 Gbps
- ▶ Control units that support 2 Gbps, 4 Gbps, or 8 Gbps

Note: FICON Express4, FICON Express2, and FICON Express features are withdrawn from marketing.

When upgrading to a z196, replace your FICON Express, FICON Express2, and FICON Express4 features with FICON Express8 features. The FICON Express8 features offer better performance and increased bandwidth.

FICON Express8 10KM LX feature

The FICON Express8 10KM LX feature occupies one I/O slot in the I/O cage or I/O drawer. It has four ports, each supporting an LC duplex connector and auto-negotiated link speeds of 2 Gbps, 4 Gbps, and 8 Gbps up to an unrepeated maximum distance of 10 km (6.2 miles).

FICON Express8 SX feature

The FICON Express8 SX feature occupies one I/O slot in the I/O cage or I/O drawer. It has four ports, each supporting an LC duplex connector and auto-negotiated link speeds of 2 Gbps, 4 Gbps, and 8 Gbps up to an unrepeated maximum distance of up to 500 meters at 2 Gbps, 380 meters at 4 Gbps, or 150 meters at 8 Gbps.

2.10.3 FICON Express4

Three types of FICON Express4 transceivers are supported on z196 servers only if carried over during an upgrade:

- ▶ FICON Express4 10KM LX feature
- ▶ FICON Express4 4KM LX feature
- ▶ FICON Express4 SX feature

Note: FICON Express4 features will be the last features to negotiate down to 1 Gbps. It is intended that the z196 is the last server to support FICON Express4 features. We recommend that you review the usage of your installed FICON Express4 channels and where possible migrate to FICON Express8 channels.

Each port supports attachment to the following items:

- ▶ FICON/FCP switches and directors that support 1 Gbps, 2 Gbps, or 4 Gbps
- ▶ Control units that support 1 Gbps, 2 Gbps, or 4 Gbps

FICON Express4 10KM LX feature

The FICON Express4 10KM LX feature occupies one I/O slot in the I/O cage or I/O drawer. It has four ports, each supporting an LC duplex connector and link speeds of 1 Gbps, 2 Gbps, or 4 Gbps up to an unrepeated maximum distance of 10 km (6.2 miles).

FICON Express4 4KM LX feature

The FICON Express4 4KM LX feature occupies one I/O slot in the I/O cage or I/O drawer. It has four ports, each supporting an LC duplex connector and link speeds of 1 Gbps, 2 Gbps, or 4 Gbps up to an unrepeated maximum distance of 4 km (2.5 miles).

Interoperability of 10 km transceivers with 4 km transceivers is supported, provided that the unrepeated distance between the two transceivers does not exceed 4 km.

FICON Express4 SX feature

The FICON Express4 SX feature occupies one I/O slot in the I/O cage or I/O drawer. It has four ports, each supporting an LC duplex connector, and supports auto-negotiated link speeds of 1 Gbps, 2 Gbps, and 4 Gbps up to an unrepeated maximum distance of up to 860 meters operating at 1 Gbps, 500 meters operating at 2 Gbps, or 380 meters operating at 4 Gbps.

For details about all FICON features see the *IBM System z Connectivity Handbook*, SG24-5444 or *FICON Planning and Implementation Guide*, SG24-6497.

2.10.4 OSA-Express3

This section describes the connectivity options offered by the OSA-Express3 features. The following OSA-Express3 features can be installed in z196 servers:

- ▶ OSA-Express3 10 Gigabit Ethernet (GbE) Long Reach (LR)
- ▶ OSA-Express3 10 Gigabit Ethernet Short Reach (SR)
- ▶ OSA-Express3 Gigabit Ethernet long wavelength (GbE LX)
- ▶ OSA-Express3 Gigabit Ethernet short wavelength (GbE SX)
- ▶ OSA-Express3 1000BASE-T Ethernet

OSA-Express3 10 GbE LR feature

The OSA-Express3 10 GbE LR feature occupies one slot in an I/O cage or I/O drawer and has two ports that connect to a 10 Gbps Ethernet LAN through a 9 μ m single-mode fiber optic cable terminated with an LC Duplex connector. The feature supports an unrepeated maximum distance of 10 km.

The OSA-Express3 10 GbE LR feature replaces the OSA-Express2 10 GbE LR feature, which is no longer orderable.

OSA-Express3 10 GbE SR feature

The OSA-Express3 10 GbE SR feature occupies one slot in the I/O cage or I/O drawer. It has two ports that connect to a 10 Gbps Ethernet LAN through a 62.5 μ m or 50 μ m multi-mode fiber optic cable terminated with an LC Duplex connector. The maximum supported unrepeated distance is 33 m on a 62.5 μ m multi-mode fiber optic cable, and 300 m on a 50 μ m multi-mode fiber optic cable.

OSA-Express3 GbE LX feature

The OSA-Express3 GbE LX occupies one slot in the I/O cage or I/O drawer. It has four ports that connect to a 1 Gbps Ethernet LAN through a 9 μm single-mode fiber optic cable terminated with an LC Duplex connector, supporting an unrepeated maximum distance of 5 km (3.1 miles). A multimode (62.5 or 50 μm) fiber optic cable can be used with this feature. The use of these multimode cable types requires a mode conditioning patch (MCP) cable at each end of the fiber optic link. Use of the single-mode to multi-mode MCP cables reduces the supported distance of the link to a maximum of 550 meters (1084 feet).

OSA-Express3 GbE SX feature

OSA-Express3 GbE SX occupies one slot in the I/O cage or I/O drawer. It has four ports that connect to a 1 Gbps Ethernet LAN through 50 or 62.5 μm multi-mode fiber optic cable terminated with an LC Duplex connector over an unrepeated distance of 550 meters (for 50 μm fiber) or 220 meters (for 62.5 μm fiber).

OSA-Express3 1000BASE-T Ethernet feature

OSA-Express3 1000BASE-T occupies one slot in the I/O cage or I/O drawer. It has four ports that connect to a 1000 Mbps (1 Gbps), 100 Mbps, or 10 Mbps Ethernet LAN. Each port has an RJ-45 receptacle for UTP Cat5 cabling, which supports a maximum distance of 100 meters.

2.10.5 OSA-Express2

This section describes the connectivity options offered by the OSA-Express2 features. The following OSA-Express2 features are supported on z196 servers only if carried over during an upgrade:

- ▶ OSA-Express2 Gigabit Ethernet (GbE) long wavelength (LX)
- ▶ OSA-Express2 Gigabit Ethernet short wavelength (SX)
- ▶ OSA-Express2 1000BASE-T Ethernet

OSA-Express and OSA-Express2 Gigabit Ethernet 10 GbE LR features installed on previous servers are *not* supported on a z196 and *cannot* be carried forward on an upgrade.

OSA-Express2 GbE LX feature

The OSA-Express2 GbE LX feature occupies one slot in an I/O cage or I/O drawer and has two independent ports. Each port supports a connection to a 1 Gbps Ethernet LAN through a 9 μm single-mode fiber optic cable terminated with an LC Duplex connector. This feature utilizes a long wavelength laser as the optical transceiver.

A multimode (62.5 or 50 μm) fiber cable may be used with the OSA-Express2 GbE LX feature. The use of these multimode cable types requires a mode conditioning patch (MCP) cable to be used at each end of the fiber link. Use of the single-mode to multimode MCP cables reduces the supported optical distance of the link to a maximum end-to-end distance of 550 meters.

OSA-Express2 GbE SX feature

The OSA-Express2 GbE SX feature occupies one slot in an I/O cage or I/O drawer and has two independent ports. Each port supports a connection to a 1 Gbps Ethernet LAN through a 62.5 μm or 50 μm multi-mode fiber optic cable terminated with an LC Duplex connector. The feature utilizes a short wavelength laser as the optical transceiver.

OSA-Express2 1000BASE-T Ethernet feature

The OSA-Express2 1000BASE-T Ethernet occupies one slot in the I/O cage or I/O drawer. It has two ports connecting to either a 1000BASE-T (1000 Mbps), 100BASE-TX (100 Mbps), or 10BASE-T (10 Mbps) Ethernet LAN. Each port has an RJ-45 receptacle for UTP Cat5 cabling, which supports a maximum distance of 100 meters.

For details about all OSA-Express features see the *IBM System z Connectivity Handbook*, SG24-5444 or *OSA-Express Implementation Guide*, SG24-5948.

2.11 Cryptographic functions

The z196 server includes both standard cryptographic hardware and optional cryptographic features to provide flexibility and growth capability. IBM has a long history of providing hardware cryptographic solutions. Use of the cryptographic hardware function requires support by the operating system. For the z/OS operating system, the Integrated Cryptographic Service Facility (ICSF) is a base component that provides the administrative interface and a large set of application interfaces to the hardware.

Cryptographic support on the z196 includes:

- ▶ CP Assist for Cryptographic Function
- ▶ Crypto Express3 cryptographic adapter features
- ▶ Trusted key entry workstation feature

2.11.1 CP Assist for Cryptographic Function

Figure 2-11 shows the layout of the z196 Compression and cryptographic coprocessor (CoP). Each chip contains two coprocessors (CoP) for data compression and encryption functions, each one shared by two cores. Every processor in the z196 server characterized as a CP, zAAP, zIIP or an IFL has access to the CP Assist for Cryptographic Function (CPACF).

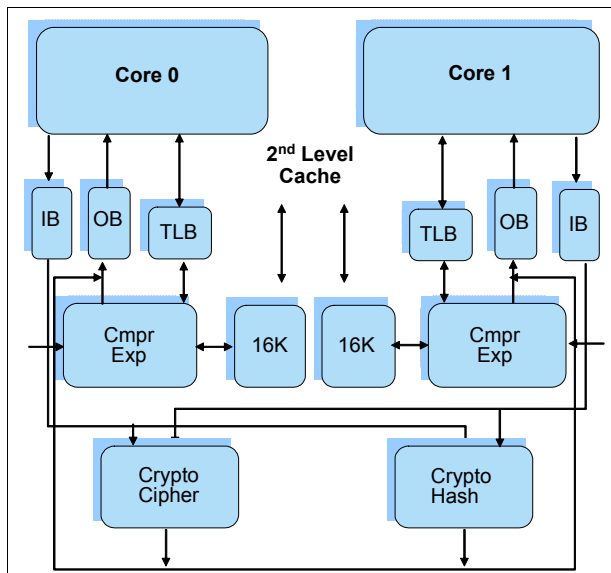


Figure 2-11 z196 Compression and cryptographic coprocessor

The assist provides high-performance hardware encrypting and decrypting support for clear key operations and is designed to scale with PU performance enhancements. Special instructions are used with the cryptographic assist function.

CPACF offers a set of symmetric cryptographic functions for high encrypting and decrypting performance of clear key operations for SSL, VPN, and data storing applications that do not require FIPS 140-2 level 4 security. The cryptographic architecture includes support for:

- ▶ Data Encryption Standard (DES) data encrypting and decrypting. It supports:
 - Single-length key DES
 - Double-length key DES
 - Triple-length key DES (T-DES)
- ▶ Advanced Encryption Standard (AES) for 128-bit, 192-bit, and 256-bit keys
- ▶ Pseudo random number generation (PRNG)
- ▶ MAC message authentication both single-length key and double-length key
- ▶ Hashing algorithms: SHA-1 and SHA-2 support for SHA-224, SHA-256, SHA-384, and SHA-512

SHA-1 and SHA-2 support for SHA-224, SHA-256, SHA-384, and SHA-512 are shipped enabled on all servers and do not require the CPACF enablement feature. The CPACF functions are supported by z/OS, z/VM, and Linux on System z.

An enhancement to CPACF is designed to facilitate the continued privacy of cryptographic key material when used for data encryption. CPACF ensures that key material is not visible to applications or operating systems during encryption operations. Protected key CPACF is designed to provide substantial throughput improvements for large-volume data encryption as well as low latency for encryption of small blocks of data.

2.11.2 Crypto Express3 feature

The Crypto Express3 feature has two PCI Express cryptographic adapters. Each of the PCI Express cryptographic adapters can be configured as a cryptographic coprocessor or a cryptographic accelerator.

The Crypto Express3 feature is the newest state-of-the-art generation cryptographic feature. Like its predecessors, it is designed to complement the functions of CPACF. This new feature is tamper-sensing and tamper-responding. It provides dual processors operating in parallel supporting cryptographic operations with high reliability.

The Crypto Express3 feature, residing in the I/O drawer or I/O cage of the z196, continues to support all of the cryptographic functions available on Crypto Express3 on System z10. When one or both of the two PCIe adapters are configured as a coprocessor, the following cryptographic enhancements introduced at z196 are supported:

- ▶ ANSI X9.8 PIN security.
- ▶ Enhance CCA key wrapping to comply with ANSI X9.24-1 key bundling requirements.
- ▶ Secure key HMAC (Keyed-Hash Message Authentication Code)
- ▶ Elliptic Curve Cryptography (ECC) Digital Signature Algorithm
- ▶ Concurrent Driver Upgrade (CDU) and Concurrent Path Apply (CPA)

Additional key features of Crypto Express3 include:

- ▶ Dynamic power management to maximize RSA performance while keeping the CEX3 within temperature limits of the tamper-responding package.
- ▶ All logical partitions (LPARs) in all Logical Channel Subsystems (LCSSs) have access the Crypto Express3 feature, up to 32 LPARs per feature.
- ▶ Secure code loading that enables the updating of functionality while installed in application systems.
- ▶ Lock-step checking of dual CPUs for enhanced error detection and fault isolation of cryptographic operations performed by a coprocessor when a PCI-E adapter is defined as a coprocessor.
- ▶ Improved RAS over previous crypto features due to dual processors and the service processor.
- ▶ Dynamic addition and configuration of the Crypto Express3 features to LPARs without an outage.

The Crypto Express3 feature is designed to deliver throughput improvements for both symmetric and asymmetric operations.

2.11.3 TKE workstation

The trusted key entry (TKE) workstation offers security-rich local and remote key management, providing authorized persons with a method of operational and master key entry, identification, exchange, separation, and update.

The TKE workstation supports connectivity to an Ethernet local area network operating at 10, 100, or 1000 Mbps.

An optional smart card reader can be attached to the TKE 7.0 workstation to allow for the use of smart cards that contain an embedded microprocessor and associated memory for data storage. Access to and the use of confidential data on the smart cards is protected by a user-defined personal identification number. The latest version of the TKE LIC is 7.0 is required to support the z196, introducing enhancements and usability features.

2.12 Coupling and clustering

In the past, Parallel Sysplex support has been provided over several types of connection; ISC, ICB, and IC, each of which involves unique development effort for the support code and for the hardware (except IC).

Coupling connectivity on z196 in support of Parallel Sysplex environments can use the InfiniBand (PSIFB) connections for Parallel Sysplex. PSIFB supports longer distances between servers compared with ICB. ICB connections are not available on z196. Customers who use ICB connections in their current environment are encouraged to migrate to PSIFB.

InfiniBand technology allows for moving all of the Parallel Sysplex support to a single type of interface that provides high-speed interconnection at short distances and longer distance fiber optic interconnection (replacing ISC).

2.12.1 ISC-3

InterSystem Channel-3 (ISC-3) links provide the connectivity required for data sharing between the Coupling Facility and the System z servers directly attached to it. The ISC-3

feature is available in peer mode only and can be used to connect to other System z servers. ISC-3 supports a link data rate of 2 Gbps. STP message exchanges can flow over ISC-3.

Statement of Direction: z196 will be the last server to offer ordering of ISC-3. Customers who use ISC-3 connections in their current environment are encouraged to migrate to PSIFB.

2.12.2 Internal Coupling (IC)

The internal coupling channel emulates the coupling facility connection in Licensed Internal Code (LIC) between images within a single system. It operates at memory speed and no hardware is required.

2.12.3 Parallel Sysplex InfiniBand (PSIFB) coupling

PSIFB coupling links are high-speed links that are available on zEnterprise 196, System z10, and System z9 servers. There are two types of Host Channel Adapter (HCA) fanouts used for PSIFB coupling links on the z196 and z10 EC:

- ▶ HCA2-O fanout supporting InfiniBand 12x Double-Data Rate (12x IB-DDR) and 12x InfiniBand Single-Data Rate (12x IB-SDR)
- ▶ HCA2-O Long Reach (LR) fanout supporting 1x IB-DDR and 1x IB-SDR

Also see “Coupling links” on page 61.

PSIFB coupling link using a HCA2-O fanout

A PSIFB coupling link using a HCA2-O fanout operates at 6 Gbps if used between two z196 or z10 servers and at 3 Gbps when connecting a z196 or a z10 to a z9 server. The link speed is auto-negotiated to the highest common rate. The HCA2-O fanout uses a fiber optical cable that is connected to a z196, z10, or z9 server. The maximum supported distance is 150 m.

PSIFB coupling link using a HCA2-O LR fanout

A PSIFB LR coupling link using HCA2-O LR operates at 2.5 Gbps or 5 Gbps between two z196 or z10 servers. HCA2-O LR uses a fiber optical cable to connect two z196 or z10 servers. The maximum unrepeated distance supported is 10 km. When using repeaters (DWDM) the maximum distance is up to 100 km.

Time source for STP traffic

PSIFB can be used to carry Server Time Protocol (STP) timekeeping information.

For details about all InfiniBand features see the *IBM System z Connectivity Handbook*, SG24-5444 or *Getting Started with InfiniBand on System z10 and System z9*, SG24-7539.

2.12.4 System-Managed CF Structure Duplexing

System-Managed Coupling Facility (CF) Structure Duplexing provides a general-purpose, hardware-assisted, easy-to-exploit mechanism for duplexing CF structure data. This provides a robust recovery mechanism for failures (such as loss of a single structure or CF or loss of connectivity to a single CF) through rapid failover to the other structure instance of the duplex pair.

Customers interested in deploying System-Managed CF Structure Duplexing should read the technical paper *System-Managed CF Structure Duplexing*, ZSW01975USEN, which you can access by selecting **Learn More** on the Parallel Sysplex Web site:

<http://www.ibm.com/systems/z/psa/index.html>

2.12.5 Coupling Facility Control Code (CFCC) level 17

Coupling Facility Control Code (CFCC) Level 17 is made available on the z196 and includes the following improvements:

- ▶ CFCC Level 17 allows an increase in the number of CHPIDs from 64 to 128. By allowing more CHPIDs, more parallel CF operations can be serviced and therefore CF link throughput can increase.
- ▶ CFCC level 17 now supports up to 2048 structures, while up to 1024 structures are supported at CFCC level 16. Growing requirements, for example due to logical grouping such as DB2, IMS, and MQ datasharing groups, or customer mergers, acquisitions and sysplex consolidations, demands for more structures than ever before.

CF structure sizing changes are expected when going from CFCC level 16 to CFCC level 17; we recommend using the CFSizer tool available at:

<http://www.ibm.com/systems/z/cfsizer/>

2.13 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 of the systems in a complex. This is critical for Parallel Sysplex operation.

2.13.1 External clock facility (ECF)

Two external clock facility cards are a standard feature of the z196 server. The ECF cards provide a dual-path interface for Pulse Per Second (PPS) support. A cable connection from the PPS port on the ECF card to the PPS output of the NTP server is required when the z196 is using STP and configured in an STP-only CTN using NTP with pulse per second as the external time source.

The two ECF cards are located in the processor cage of the z196 server.

2.13.2 Server Time Protocol (STP)

Server Time Protocol is a server-wide facility that is implemented in the Licensed Internal Code of System z. The STP presents a single view of time to PR/SM and provides the capability for multiple servers and CFs to maintain time synchronization with each other. A System z or CF may be enabled for STP by installing the STP feature.

The STP feature is the supported method for maintaining time synchronization between System z servers and CFs.

For additional information about STP, refer to *Server Time Protocol Planning Guide*, SG24-7280, and *Server Time Protocol Implementation Guide*, SG24-7281.

2.13.3 Network Time Protocol (NTP) support

NTP support is available on z196 and z10 EC servers and has been added to the STP code on System z9. This implementation answers the need for a single time source across the heterogeneous platforms in the enterprise. With this implementation the System z196, z10 and z9 servers support the use of NTP as time sources.

2.14 z196 HMC and SE

The HMC and SE are appliances which together provide hardware platform management for System z. Hardware platform management covers a complex set of setup, configuration, operation, monitoring, service management tasks and services that are essential to the use of the hardware platform product.

Please refer to 3.2.7, “Hardware Management Console functionality” on page 70 for more information about HMC functions and capabilities.

The HMC is attached to a LAN, as is the server’s support element (SE). The HMC communicates with each Central Processor Complex (CPC) and, optionally with one or more zBXs, through the CPC’s SE. When tasks are performed on the Hardware Management Console, the commands are sent to one or more support elements, which then issue commands to their CPCs and optional zBXs.

Various network connectivity options for HMCs are available, such as:

- ▶ HMC/SE LAN only
- ▶ HMC to a corporate intranet
- ▶ HMC to intranet and Internet

An HMC consists of:

- ▶ Processor or system unit, including two Ethernet LAN adapters, capable of operating at 10, 100, or 1000 Mbps, a DVD RAM, and USB Flash memory Drive (UFD) to install LIC
- ▶ Flat panel display
- ▶ Keyboard
- ▶ Mouse

The System z196 is supplied with a pair of integrated ThinkPad SEs. One is always active while the other is strictly an alternate. Power for the SEs is supplied by the server power supply, and there are no additional power requirements. The internal LANs for the SEs on the z196 server connect to the bulk power hub in the Z frame. There is an additional connection from the hub to the HMC utilizing the VLAN capability of the System z196 server.

Considerations for multiple HMCs

Customers often deploy multiple HMC instances to manage an overlapping collection of systems. Today, all of the HMCs are peer consoles to the managed systems and all management actions are possible to any of the reachable systems while logged into a session on any of the HMCs (subject to access control).

With the zEnterprise System and ensembles, this paradigm has changed with respect to resource management (see “Unified Resource Manager” on page 90 for details). In this environment, if an zEnterprise System node has been added to an ensemble, management actions targeting that system can only be done from the *primary* HMC for that ensemble (see Figure 2-12 on page 47).

2.15 Power and cooling

The power service specifications are the same, but the power consumed by the z196 server can be greater. A fully loaded z196 server maximum consumption is 31.7 KW.

2.15.1 Power consumption

The system operates with two completely redundant power supplies. Each of the power supplies have their individual line-cords or pair of line-cords depending on the configuration.

For redundancy, the server should have two power feeds. Each power feed is either 1 or 2 line cords. The number of line cords required depends on system configuration. Line cords attach to either 3 phase, 50/60 Hz, 200 to 480 V AC power or 380 to 520 V DC power. There is no impact to system operation with the total loss of one power feed.

For ancillary equipment such as the Hardware Management Console, its display, and its modem, additional single-phase outlets are required.

The power requirements depend on the cooling facility installed and on number of books, as well as the number of I/O units installed. I/O units are values for I/O cages (equals 2 I/O units) and I/O drawers (equals 1 I/O unit).

Input power in kVA is equal to the output power in kW. Heat output expressed in kBTU per hour can be derived from multiplying the table entries by a factor of 3.4.

Table 2-6 lists the maximum power requirements for the air-cooled models.

Table 2-6 Power requirements - air-cooled models

Power requirement kVA	Number of I/O units						
	0	1	2	3	4	5	6
M15	6.8	7.7	8.7	10.8	12.8	12.9	13.1
M32	11.9	13.2	14.1	16.1	18.0	18.9	20.7
M49	17.3	18.2	19.2	21.1	23.0	23.8	25.8
M66 / M80	22.7	23.6	24.6	26.4	28.4	29.2	30.1

Table 2-7 on page 44 lists the maximum power requirements water-cooled models.

Table 2-7 Power requirements - water-cooled models

Power requirement kVA	Number of I/O units						
	0	1	2	3	4	5	6
M15	6.1	6.8	7.8	9.7	11.6	11.8	12.0
M32	9.8	10.6	11.6	13.4	15.3	16.2	18.2
M49	13.7	14.4	15.5	17.3	19.2	20.1	22.0
M66 / M80	18.1	18.9	19.8	21.7	23.6	24.5	26.4

2.15.2 Hybrid cooling system

The z196 server has a hybrid cooling system that is designed to lower power consumption. It is an air-cooled system, assisted by refrigeration. Refrigeration is provided by a closed-loop liquid cooling subsystem. The entire cooling subsystem has a modular construction. Its components and functions are found throughout the cages.

Refrigeration cooling is the primary cooling source and is backed up by an air-cooling system. If one of the refrigeration units fails, backup blowers are switched on to compensate for the lost refrigeration capacity with additional air cooling. At the same time, the oscillator card is set to a slower cycle time, slowing the system down by up to 10% of its maximum capacity to allow the degraded cooling capacity to maintain the proper temperature range. Running at a slower cycle time, the MCMs produce less heat. The slowdown process is done in steps, based on the temperature in the books.

2.15.3 Water cooling

The z196 introduces water cooling unit (WCU) technology, which provides the ability to cool systems with user-chilled water. In order to allow users to remove additional heat produced by non-MCM components in the system such as power, memory, and I/O to water, z196 also supports the exhaust air heat exchange, which is standard on systems with the WCU.

Conversions from machine refrigeration unit (MRU) to WCU require a frame roll, and will not be done in the field.

2.15.4 High voltage DC power

In data centers today, many businesses are paying increasing electric bills and are also running out of power. The z196 High Voltage Direct Current power feature adds nominal 380 to 520 Volt DC input power capability to the existing System z, universal 3 phase, 50/60 hertz, totally redundant power capability (nominal 200-240VAC or 380-415VAC or 480VAC). It allows z196 to directly utilize the high voltage DC distribution in some new, green data centers. A direct HV DC data center power design can improve data center energy efficiency by removing the need for a DC to AC inversion step. The z196's bulk power supplies have been modified to support HV DC so the only difference in shipped HW to implement the option is the DC line cords. Since HV DC is a new technology there are multiple proposed standards. The z196 supports both ground referenced and dual polarity HV DC supplies, such as +/-190V or +/-260V, or +380V, and so on. Beyond the data center UPS and power distribution energy savings, a z196 run on HV DC power will draw 1 - 3% less input power. HV DC does not change the number of line cords a system requires.

2.15.5 Internal Battery Feature

The Internal Battery Feature (IBF) is an optional feature on the z196 server. Refer to Figure 2-2 on page 22 for a pictorial view of the location of this feature. This optional IBF provides the function of a local uninterrupted power source.

The IBF further enhances the robustness of the power design, increasing power line disturbance immunity. It provides battery power to preserve processor data in case of a loss of power on all four AC feeds from the utility company. The IBF can hold power briefly during a *brownout*, or for orderly shutdown in case of a longer outage. The IBF provides up to 10 minutes of full power, depending on the I/O configuration.

2.15.6 IBM Systems Director Active Energy Manager

IBM Systems Director Active Energy Manager™ (AEM) is an energy management solution building block that returns true control of energy costs to the customer. It enables you to manage actual power consumption and resulting thermal loads that IBM servers place on the data center. It is an industry-leading cornerstone of the IBM energy management framework. In tandem with chip vendors Intel® and AMD and consortiums such as Green Grid, AEM advances the IBM initiative to deliver price performance per square foot.

AEM runs on Windows®, Linux on System x, Linux on System p, and Linux on System z. Refer to its documentation for more specific information.

How AEM works

The following list is a brief overview of how AEM works:

- ▶ Hardware, firmware, and systems management software in servers and blades can take inventory of components.
- ▶ AEM adds power draw up for each server or blade and tracks that usage over time.
- ▶ When power is constrained, AEM allows power to be allocated on a server-by-server basis. Note the following information:
 - Care should be taken that limiting power consumption does not affect performance.
 - Sensors and alerts can warn the user if limiting power to this server could affect performance.
 - The z196 server does not support power capping.
- ▶ Certain data can be gathered from the HMC:
 - System name, machine type, model, serial number, firmware level
 - Ambient and exhaust temperature
 - Average and peak power (over a 1-minute period)
 - Other limited status and configuration information

2.16 zEnterprise BladeCenter Extension

The zEnterprise BladeCenter Extension (zBX) Model 002 is designed to extend the System z qualities of service and management to integrate heterogeneous systems with high redundancy.

The zBX Model 002 (2458-002) connects to the z196 to become part of a node in an ensemble. That node in turn creates an integrated multi-platform system with advance virtualization management (through the Unified Resource Manager) that supports diverse workloads.

The zBX is configured with the following key components:

- ▶ One to four standard 19 inch 42U IBM zEnterprise racks with required network and power infrastructure
- ▶ One to eight BladeCenter chassis¹ with a combination of up to 112 blades²
- ▶ Redundant infrastructure for fault tolerance and higher availability
- ▶ Management support via the z196 Hardware Management Console (HMC) and Support Element (SE)

The IBM Smart Analytics Optimizer solutions is also offered with the zBX. (See “IBM Smart Analytics Optimizer solution” on page 77 for more information.)

¹ A maximum of four BladeCenter chassis for the IBM Smart Analytics Optimizer solution.

² A maximum of 56 blades for the IBM Smart Analytics Optimizer solution.

The first rack (Rack B) in the zBX is the primary rack where one or two BladeCenter chassis reside. Four top of rack (TOR) switches are included in Rack B for intranode management network (INMN) and intraensemble data network (IEDN) connectivity. The other three racks (C, D, and E) are expansion racks with one or two BladeCenter chassis each.

The zBX is managed through a private and physically isolated 1000BASE-T network (INMN), which interconnects all components in the zEnterprise System (z196 and zBX). The OSA-Express for Unified Resource Manager (OSM) CHPID type supports the connectivity from the z196 to the primary HMC, where the Unified Resource Manager functions reside, via the Bulk Power Hubs (BPHs).

The IEDN provides private and secure 10 GbE high speed data paths between all elements of an ensemble node through the IEDN TOR switches in the zBX. The OSA-Express for zBX (OSX) CHPID type supports connectivity and access control from the z196 to the zBX.

Figure 2-12 on page 47 shows the ensemble node connections through the OSA-Express3 1000BASE-T features (CHPID type OSM) and OSA-Express 10 GbE features (CHPID type OSX) in the z196.

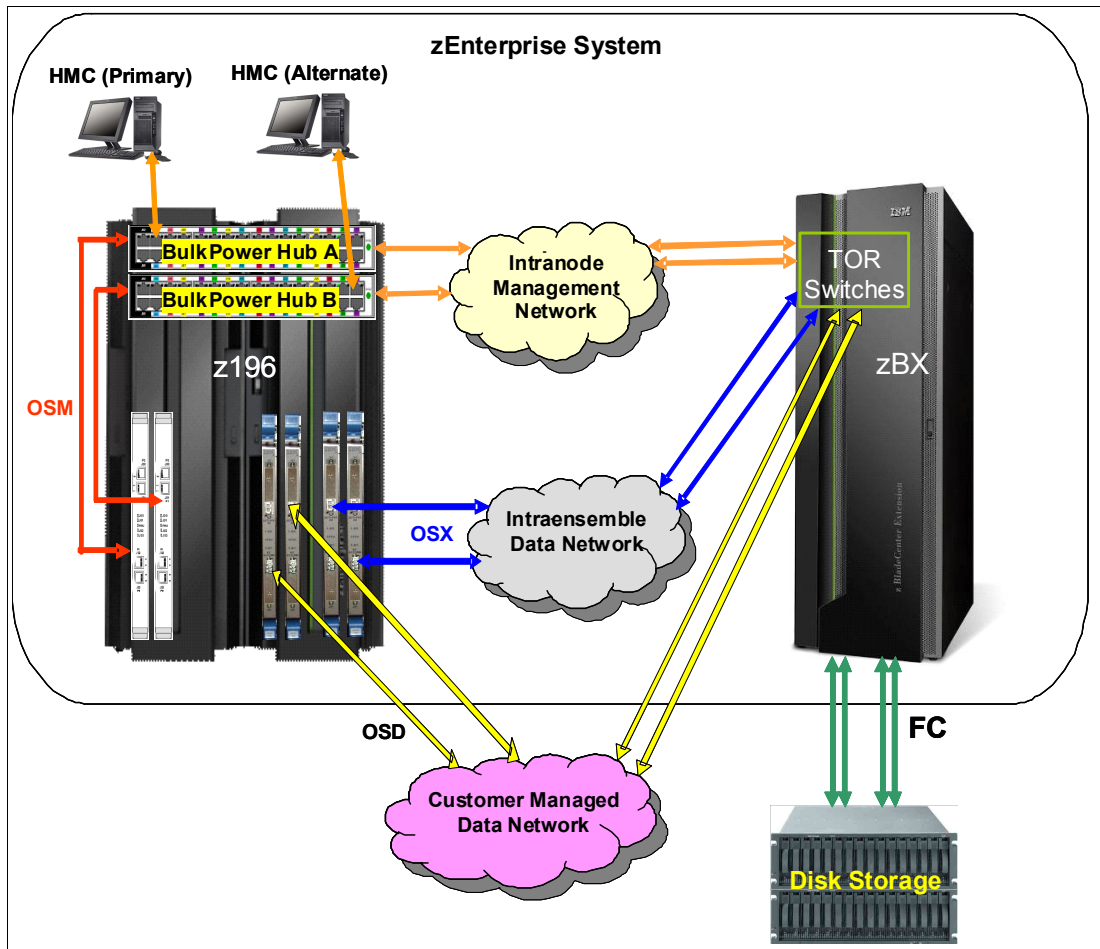


Figure 2-12 INMN, IEDN, and customer managed data networks in an ensemble

Optionally as part of the ensemble, any OSA-Express2 or OSA-Express3 features (with CHPID type OSD) in the z196 can connect to the customer managed data network. And the customer managed network can also be connected to the IEDN TOR switches in the zBX.

In addition, each BladeCenter chassis in the zBX has two Fibre Channel (FC) switch modules that connect to FC disk storage either directly or via a SAN switch³.

Figure 2-13 on page 48 shows a rear view of a two rack zBX configuration.

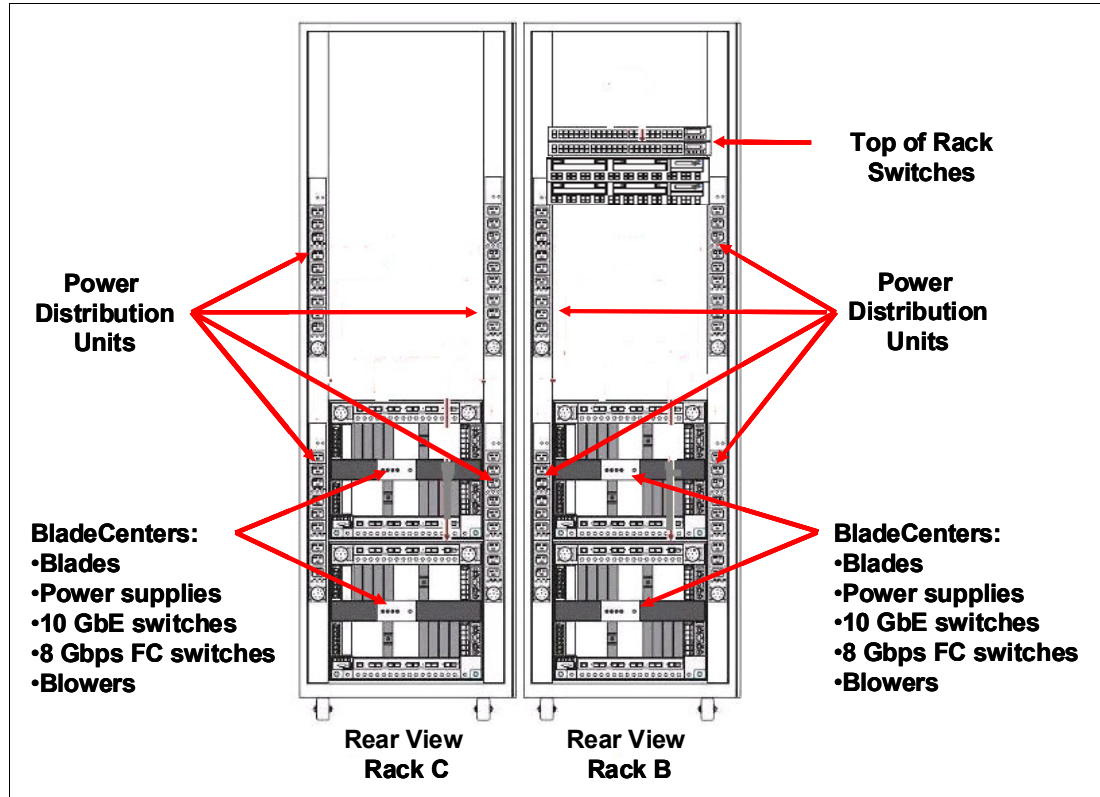


Figure 2-13 zBX rear view - two rack configuration

The zBX racks include:

- ▶ Two TOR 1000BASE-T switches (Rack B only) for the intranode management network (INMN)
- ▶ Two TOR 10GbE switches (Rack B only) for the intraensemble data network (IEDN)
- ▶ Up to two BladeCenter Chassis in each rack. Each BladeCenter consists of:
 - Up to 14 blades⁴ (POWER7 or System x)
 - Two advance management modules (AMM)
 - Two Ethernet switch modules (ESM)
 - Two 10 GbE high speed switch (HSS) modules
 - Two 8 Gbps Fiber Channel switch modules
 - Four blower modules
- ▶ Power Distribution Units (PDUs)


Client supplied external disks are required with the zBX. Supported FC disk types and vendors with IBM blades are listed on the IBM System Storage Interoperation Center (SSIC) website, at:

³ For the IBM Smart Analytics Optimizer solution, a direct connection to a client supplied IBM System Storage DS5020 is required. SAN switches are not supported.

⁴ Depending on the IBM Smart Analytics Optimizer configuration this can be either 7 or 14 blades in the first BladeCenter chassis.

http://www-03.ibm.com/systems/support/storage/config/ssic/displayessearchwithoutjs.wss?start_over=yes

For more details regarding z196 and zBX hardware, refer to *IBM zEnterprise System Technical Guide*, SG24-7833.



Key functions and capabilities of the zEnterprise System

The zEnterprise System has additional security options, improved reliability, availability, and serviceability, and enhanced virtualization capabilities compared to previous System z servers. The zEnterprise System is an integrated platform that brings mainframe and distributed technologies together. The zEnterprise System is comprised of the zEnterprise 196, the zEnterprise BladeCenter Extension, and the zEnterprise Unified Resource Manager.

The zEnterprise 196 is the follow-on to the System z10 EC (z10 EC) server. As its predecessor, it offers five hardware models, but has a more powerful uniprocessor, more processor units, and new functions and features. The z196 represents a new level of microprocessor technology, made possible through advances in the design and manufacturing processes.

The zEnterprise Unified Resource Manager is part of the System Director family. It has integrated System z management facility to unify workload management, hence extending System z qualities of service across the infrastructure.

The zEnterprise BladeCenter Extension (zBX) Model 002 with IBM blades and IBM Smart Analytics Optimizer solutions, offers specific services and components. The consistent size and shape (or form factor) of each blade allows it to fit in a BladeCenter chassis.

In this chapter we discuss the following topics:

- ▶ 3.1, “Virtualization” on page 52
- ▶ 3.2, “z196 technology improvements” on page 56
- ▶ 3.3, “z196 common time functions” on page 71
- ▶ 3.4, “z196 Capacity on Demand (CoD)” on page 73
- ▶ 3.5, “Throughput optimization with z196” on page 75
- ▶ 3.6, “zEnterprise BladeCenter Extension” on page 76
- ▶ 3.7, “z196 performance” on page 78
- ▶ 3.8, “Reliability, availability, and serviceability” on page 80
- ▶ 3.9, “High availability technology” on page 82

3.1 Virtualization

Virtualization creates the appearance of multiple concurrent servers by sharing the existing hardware. The goal of virtualization in the zEnterprise System is to fully utilize its resources, thus lowering the total amount of resources needed and their cost. Virtualization is a key strength of the zEnterprise System, it is embedded in the architecture and built into the hardware, firmware, and operating systems.

For example, the z196 is able to handle tens, hundreds, even thousands, of virtual servers, so a very high context switching rate is to be expected, and accesses to the memory, caches, and virtual I/O devices must be kept completely isolated.

Virtualization requires a hypervisor. A hypervisor is control code that manages multiple independent operating system images. Hypervisors can be implemented in software or hardware, and zEnterprise System has both. In the z196 hardware hypervisor, named Processor Resource/Systems Manager™ (PR/SM) fully virtualizes the server and is implemented in firmware. PR/SM is part of the base server and does not require any additional software to run. The z/VM operating system implements the software hypervisor. z/VM requires some PR/SM functions.

In the zBX, PowerVM™ is the hypervisor that offers a unified virtualization solution for any Power workload. It allows use of all POWER7 blade CPU cores and physical resources, providing better scalability and reduction in resource costs.

PowerVM is EAL4+ certified and PowerVM is isolated on the internal network of the zEnterprise System, providing intrusion prevention, integrity, secure virtual switches with integrated consolidation.

PowerVM is managed by the zEnterprise Unified Resource Manager, therefore it is shipped, deployed, monitored, and serviced at a single point.

Details about zBX virtualization and the Unified Resource Manager can be found in Chapter 4, “Achieving better infrastructure resource management” on page 85.

The rest of this section discusses the hardware and software virtualization capabilities of the z196.

3.1.1 z196 hardware virtualization

PR/SM was first implemented in the mainframe in the late 1980s. It allows defining and managing subsets of the server resources known as logical partitions (LPARs). PR/SM virtualizes processors, memory, and I/O features. Some features are purely virtual implementations. For example, HiperSockets works like a LAN but does not use any I/O hardware.

Up to 60 LPARs can be defined. In each, a supported operating system can be run. The LPAR definition includes a number of *logical* PUs, memory, and I/O devices.

The z/Architecture (inherent in the z196 and its predecessors) has been *designed* to meet those stringent requirements with very low overhead and the highest security certification in the industry: common criteria EAL5 with specific target of evaluation (logical partitions). This design has been proved in many customer installations in recent decades.

z/VM-mode partitions

The z/VM-mode logical partition (LPAR) mode, first supported on IBM System z10, is exclusively for running z/VM and its workloads. This LPAR mode provides increased flexibility and simplifies systems management by allowing z/VM to manage guests to perform the following tasks all in the same z/VM LPAR:

- ▶ Operate Linux on System z on IFLs.
- ▶ Operate z/OS, z/VSE and z/TPF on CPs.
- ▶ Operate z/OS while fully allowing zAAP and zIIP exploitation, by workloads such as WebSphere and DB2, for an improved economics environment.

The z/VM-mode partitions require z/VM V5R4 or later and allow z/VM to utilize a wider variety of specialty processors in a single LPAR. The processor types that can be configured to a z/VM-mode partition are:

- ▶ CPs
- ▶ IFLs
- ▶ zIIPs
- ▶ zAAPs
- ▶ ICFs

Logical processors

Logical processors are defined to and managed by PR/SM and are perceived by the operating systems as real processors. They assume the following types:

- ▶ CPs
- ▶ zAAPs
- ▶ zIIPs
- ▶ IFLs
- ▶ ICFs

SAPs are never part of an LPAR configuration.

PR/SM is responsible for honoring requests for logical processor work by dispatching logical processors on physical processors. Under certain circumstances logical zAAPs and zIIPs can be dispatched on physical CPs. Physical processors can be shared across LPARs, but can also be dedicated to an LPAR. However, an LPAR must have its logical processors either all shared or all dedicated.

PR/SM ensures that, when switching a physical processor from one logical processor to another, processor state is properly saved and restored, including all the registers. Data isolation, integrity, and coherence are strictly enforced at all times.

Logical processors can be dynamically added to and removed from LPARs. Operating system support is required in order to take advantage of this capability. Starting with z/OS V1R10, z/VM V5R4, and z/VSE V4R3, an enhanced capability, the ability to dynamically define and change the number and type of reserved PUs in an LPAR profile can be used for that purpose. No pre-planning is required.

The new resources are immediately made available to the operating systems and, in the z/VM case, to its guests. However, z/VSE, when running as a z/VM guest does not support this capability.

Memory

To ensure security and data integrity, memory cannot be shared by active LPARs. In fact, a strict isolation is maintained. When an LPAR is activated, its defined memory is allocated in

blocks, which must be a multiple of a given value. This value depends on the total allocation and varies between 256 MB and 2 GB. Thus, memory can be serially reused.

Using the plan-ahead capability, memory can be physically installed but not enabled until it is necessary. z/OS and z/VM support dynamically increasing the size of the LPAR.

LPAR memory is said to be virtualized in the sense that in all LPARs memory addresses start at zero. This should not be confused with the operating system virtualizing its LPAR memory. The z/Architecture has a robust virtual storage architecture that allows, per LPAR, the definition of an unlimited number of address spaces and the simultaneous use *by each program* of up to 1,023 of those address spaces. Each address space can be up to 16 EB (1 exabyte = 2^{60} bytes). Thus, the architecture has no real limits. Practical limits are determined by the available hardware resources, including disk storage for paging.

Isolation of the address spaces is strictly enforced by the Dynamic Address Translation hardware mechanism, which also validates the right to read or write in each page frame by comparing the page key with the key of the program requesting access. Three addressing modes, 24-bit, 31-bit, and 64-bit, are simultaneously supported. Definition and management of the address spaces is under operating system control. This mechanism has been in use since the System 370, and memory keys were part of the original System 360 design.

Operating systems may allow sharing of address spaces, or parts thereof, across multiple processes. For instance, under z/VM, a single copy of the read-only part of a kernel can be shared by all virtual machines using that operating system, resulting in large savings of real memory and improvements in performance.

A logical partition can be defined with both an initial and a reserved amount of memory. At activation time the initial amount is made available to the partition and the reserved amount can later be added, partially or totally. Those two memory zones do not have to be contiguous in real memory but appear as logically contiguous to the operating system running in the LPAR.

Until now, only z/OS was able to take advantage of this support by nondisruptively acquiring and releasing memory from the reserved area. z/VM V5R4 and later are able to acquire memory nondisruptively and immediately make it available to guests. z/VM virtualizes this support to its guests, which can also increase their memory nondisruptively. Releasing memory is still a disruptive operation.

I/O virtualization

The z196 supports four channel subsystems with 256 channels each, for a total of 1024 channels. In addition to dedicated use of channels and I/O devices to an LPAR, I/O virtualization allows concurrent sharing of channels, and the I/O devices accessed through these channels, by several active LPARs. The function is known as Multiple Image Facility (MIF). The shared channels may belong to different channel subsystems, in which case they are known as spanned channels.

Data streams for the sharing LPARs are carried on the same physical path with total isolation and integrity. For each active LPAR that has the channel configured online, PR/SM establishes one logical channel path. For availability reasons, multiple logical channel paths should exist for critical devices (for instance, disks containing vital data sets).

When additional isolation is required, configuration rules allow restricting the access of each logical partition to particular channel paths and specific I/O devices on those channel paths.

Many installations use the Parallel Access Volume (PAV) function, which allows accessing a device by several different addresses (normally one base address and three aliases), thus

increasing the throughput of the device by using more device addresses. HyperPAV takes the technology a step further by allowing the I/O Supervisor (IOS) in z/OS to dynamically create PAV structures depending on the current I/O demand in the system, thus lowering the need for manually tuning the system for PAV use.

For large installations, which usually have a large number of devices, the total number of device addresses can be very high. Thus, the concept of *channel sets* was introduced with System z9. Each channel can address three sets of 64 K device addresses, allowing the base addresses to be defined on *set 0* (IBM reserves 256 subchannels on set 0) and the aliases on *set 1* and *set 2*. In total, 196,352 subchannel addresses are available *per channel*.

Channel sets are exploited by the Peer-to-Peer Remote Copy (PPRC) function by the ability to have the PPRC primary devices defined in channel set 0, while secondary devices can be defined in channel set 1 and 2, thus providing more connectivity through channel set 0.

To further reduce the complexity of managing large I/O configurations System z introduces Extended Address Volumes (EAV). EAV is designed to build very large disk volumes using virtualization technology. By being able to extend the disk volume size a customer may potentially need less volumes to hold his data, therefore making systems and data management less complex.

The health checker function in z/OS V1R10 and later introduces a health check in the I/O Supervisor that can help system administrators identify single points of failure in the I/O configuration.

The dynamic I/O configuration function is supported by z/OS and z/VM. It provides the capability of concurrently changing the currently active I/O configuration. Changes can be made to channel paths, control units, and devices. The existence of a fixed HSA area in the z196 greatly eases the planning requirements and enhances the flexibility and availability of these reconfigurations.

3.1.2 z196 software virtualization

Software virtualization is provided by the z/VM product. Strictly, it is a function of the CP component of z/VM. Starting in 1967, IBM has continuously provided software virtualization in its mainframe servers.

z/VM uses the resources of the LPAR in which it is running to create functional equivalents of real System z servers, which are known as virtual machines (VMs) or *guests*. In addition, z/VM is able to emulate I/O peripherals including, for instance, printers by using spooling techniques and LAN switches and disks by exploiting memory.

z/VM allows very fine-grained allocation of resources, for example, in the case of processor sharing, the minimum is approximately 1/10,000 of a processor. Another example: Disks can be subdivided into independent areas, known as *minidisks*, each of which is exploited by its users as a real disk, only smaller. Minidisks are shareable, and can be used for all types of data and also for temporary space in a pool of on demand storage.

Under z/VM, virtual processors, virtual central and expanded storages, and all the virtual I/O devices of the VMs are dynamically definable (provisionable). z/VM supports the concurrent addition (but not deletion) of memory to its LPAR and immediately makes it available to guests. Guests themselves may support the dynamic addition of memory. All other changes are concurrent. To render these concurrent definitions also nondisruptive requires support by the operating system running in the guest, which is also the case when running in an LPAR.

Although z/VM imposes no limits on the number of defined virtual machines, the number of active virtual machines is limited by the available resources. On a large server, such as the z196, thousands of virtual machines can be activated.

It is beyond the scope of this book to provide a more detailed description of z/VM or other highlights of its capabilities. For a deeper discussion of z/VM see *Introduction to the New Mainframe: z/VM Basics*, SG24-7316, downloadable from:

<http://www.redbooks.ibm.com/redbooks/pdfs/sg247316.pdf>

3.2 z196 technology improvements

The technology used with the z196 fall into five categories:

- ▶ Microprocessor
- ▶ Capacity
- ▶ Memory
- ▶ Connectivity
- ▶ Cryptography

These are intended to provide a more scalable, flexible, manageable, and secure consolidation and integration platform contributing to a lower total cost of ownership.

3.2.1 Microprocessor

The zEnterprise 196 has a newly developed microprocessor chip and a newly developed infrastructure chip. Both of those chips use CMOS S12 technology and represent a major step forward in technology utilization for the System z products, resulting in increased packaging density.

Like for the z10, the microprocessor chip and the infrastructure chip for the z196 are packaged together on a new multi-chip module (MCM). The MCM contains six microprocessor chips and two infrastructure chips, while the z10 MCMs included 7 chips in total. Each microprocessor chip contains four cores with either three or four active cores. The MCM is installed inside a book, and the z196 can contain from one to four books. The book also contains the memory arrays, I/O connectivity infrastructure, and various other mechanical and power controls.

The book is connected to the I/O drawers and I/O cages through one or more cables. As new standards are making their way on to the z196, these cables are now using the standard InfiniBand protocol to transfer large volumes of data between the memory and the I/O cards located in I/O drawers and I/O cages.

z196 processor chip

The z196 chip provides more functions per chip—four cores on a single chip—thanks to technology improvements that allow designing and manufacturing more transistors per square inch. This translates into using fewer chips to implement the needed functions, which helps enhance system availability.

The z196 microprocessor chip has an improved design when compared with the z10. The System z microprocessor development has been following the same basic design set since the 9672-G4 (announced in 1997) until the z9. That basic design had been stretched to its maximum, so a fundamental change was necessary.

The processor chip, shown in Figure 3-1, includes two co-processors for hardware acceleration of data compression and cryptography, I/O bus and memory controllers, and an interface to a separate storage controller/cache chip.

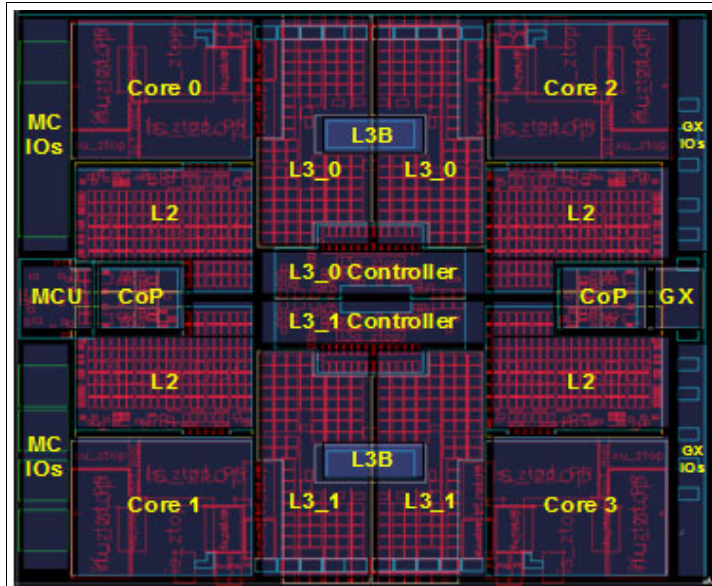


Figure 3-1 z196 Enterprise Quad-Core microprocessor chip

On-chip cryptographic hardware includes extended key and hash sizes for the AES and SHA algorithms.

Hardware decimal floating point function (HDFP)

Hardware decimal floating point support was introduced with the z9 EC. The z196, however, has a new decimal floating point accelerator feature. This facility is designed to speed up such calculations and provide the necessary precision demanded mainly by the financial institutions sector. The decimal floating point hardware fully implements the new IEEE 754r standard.

Industry support for decimal floating point is growing, with IBM leading the open standard definition. Examples of support for the draft standard IEEE 754r include Java BigDecimal, C#, XML, C/C++, GCC, COBOL, and other key software vendors such as Microsoft® and SAP.

Support and exploitation of HDFP varies with operating system and release. For a detailed description see the *IBM zEnterprise System Technical Guide*, SG24-7833. See also “Decimal floating point (z/OS XL C/C++ considerations)” on page 113.

New machine instructions

The z/Architecture offers a rich CISC Instruction Set Architecture (ISA). The z196 offers 984 instructions, of which 762 are implemented entirely in hardware. Multiple arithmetic formats are supported.

The z196 architectural extensions include over 110 new instructions, the bulk of which was designed in collaboration with software developers to improve compiled code efficiency. These should particularly be of benefit to Java-based, WebSphere-based, and Linux-based workloads. New instructions are grouped under the following categories:

- ▶ High-word facility
- ▶ Interlocked-access facility
- ▶ Load/Store on condition facility

- ▶ Distinct operands facility
- ▶ Integer to/from floating point conversions
- ▶ Decimal floating point quantum exceptions
- ▶ New crypto functions and modes
- ▶ Virtual architectural level
- ▶ Non-quiescing SSKe

3.2.2 Large system images

A single system image can control several processor units (PUs) such as CPs, zIIPs, zAAPs, and IFLs, as appropriate. See “PU characterization” on page 10 for a description.

Table 3-1 shows the maximum number of PUs supported for each operating system image.

Table 3-1 Single system image software support

Operating system	Maximum number of (CPs+zIIPs+zAAPs) ^a or IFLs per system image
z/OS V1R12	80
z/OS V1R11	80
z/OS V1R10	80
z/OS V1R9	64
z/OS V1R8	32
z/OS V1R7	32
z/VM V5R4	32 ^{b,c}
z/VSE V4	z/VSE Turbo Dispatcher can exploit up to four CPs and tolerates up to 10-way LPARs
Linux on System z	Novell SUSE SLES 11: 64 CPs or IFLs Novell SUSE SLES 10: 64 CPs or IFLs Red Hat RHEL 5: 64 CPs or IFLs
z/TPF V1R1	64 CPs
TPF V4R1	16 CPs

a. The number of purchased zAAPs and the number of purchased zIIPs cannot each exceed the number of purchased CPs. A logical partition can be defined with any number of the available zAAPs and zIIPs. The total refers to the sum of these PU characterizations.

b. z/VM guests can be configured with up to 64 virtual PUs.

c. The z/VM-mode LPAR supports CPs, zAAPs, zIIPs, IFLs, and ICFs.

3.2.3 Granular capacity and capacity settings

The z196 expands the offer on sub-capacity settings. Finer granularity in capacity levels allows the growth of installed capacity to more closely follow the enterprise growth, for a smoother, pay-as-you-go investment profile. The many performance and monitoring tools available on System z environments, coupled with the flexibility of the capacity on demand options (see 3.4, “z196 Capacity on Demand (CoD)” on page 73) provide for managed growth with capacity being available when needed.

The z196 offers four distinct capacity levels for CPs (full capacity and three sub-capacities). A processor characterized as anything other than a CP is always set at full capacity. There is,

correspondingly, a different pricing model for non-CP processors regarding purchase and maintenance prices, as well as various offerings for software licensing costs.

A capacity level is a setting of each CP to a sub-capacity of the full CP capacity. Full capacity CPs are identified as CP7. On the z196 server, 80 CPs can be configured as CP7. Besides full capacity CPs, three sub-capacity levels (CP6, CP5, and CP4), each for up to fifteen CPs, are offered. The four capacity levels appear in hardware descriptions as feature codes on the CPs. These feature codes (FC) are:

- ▶ CP7 is FC 1880
- ▶ CP6 is FC 1879
- ▶ CP5 is FC 1878
- ▶ CP4 is FC 1877

Granular capacity adds 45 sub-capacity settings to the 80 capacity settings that are available with full capacity CPs (CP7). Each of the 45 sub-capacity settings only apply to up to 15 CPs, independent of the z196 model installed.

If more than 15 CPs are configured for the server they will all be full capacity, because all CPs must be on the same capacity level. The capacity indicator numbers are:

- ▶ 701 to 780 for capacity level CP7
- ▶ 601 to 615 for capacity level CP6
- ▶ 501 to 515 for capacity level CP5
- ▶ 401 to 415 for capacity level CP4

Information about CPs in the remainder of this chapter applies to all CP capacity levels, CP7, CP6, CP5, and CP4, unless otherwise indicated.

Note: The actual throughput that a user will experience may vary, depending on considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, and the workload being processed.

To help size a System z server to fit your requirements, IBM provides a free tool that reflects the latest IBM LSPR measurements, called the IBM Processor Capacity Reference (zPCR). The tool can be downloaded from:

<http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS1381>

3.2.4 Memory

The z196 has greatly increased the available memory capacity over previous servers. The system can now have up to 3,052 GB of usable memory installed. The logical partitions can be configured with up to 1 TB of memory. In addition, the hardware systems area (HSA) is fixed in size (16 GB) and not included in the memory for which the customer orders and pays.

Note that the z/Architecture simultaneously supports 24-bit, 31-bit, and 64-bit addressing modes. This provides backwards compatibility and investment protection.

Support of large memory by operating systems is as follows:

- ▶ z/OS V1R10 and above support up to 4 TB
- ▶ z/VM V5R4 and above support up to 256 GB
- ▶ z/VSE V4R1 supports up to 8GB and V4R2 and above up to 32GB
- ▶ z/TPF V1R1 supports up to 4 TB
- ▶ Novell SUSE SLES 11 supports 4 TB and Red Hat RHEL 5 supports 64 GB

Hardware system area

The z196 has a fixed-size hardware system area. This is intended to improve the server availability. Because the HSA is big enough to accommodate all possible configurations for all the logical partitions, several operations that were disruptive on previous servers due to HSA size are now concurrent. In addition, some planning needs are eliminated.

The HSA has a fixed size of 16 GB and resides in a separately reserved area of memory separate from customer-purchased memory.

A fixed large HSA enables dynamic addition and removal of the following features without planning:

- ▶ New logical partition to new or existing channel subsystem (CSS)
- ▶ New CSS (up to four can be defined)
- ▶ New subchannel set (up to three can be defined)
- ▶ Maximum number of devices in each subchannel set
- ▶ Dynamic I/O enabled as a default
- ▶ Logical processors by type
- ▶ Cryptographic processors

Plan-ahead memory

Planning for future memory requirements and installing dormant memory in the server allows future upgrades to be done concurrently and, with appropriate operating system support, nondisruptively.

If a customer can anticipate an increase of the required memory, a target memory size can be configured along with a starting memory size. The starting memory size will be activated and the remainder will be inactive. When additional physical memory is required, it is fulfilled by activating the appropriate number of planned memory features. This activation is concurrent and can be nondisruptive to the applications depending on the operating system support. z/OS and z/VM support this function.

Plan-ahead memory should not be confused with flexible memory support. Plan-ahead memory is for a permanent increase of installed memory, whereas flexible memory provides a temporary replacement of a part of memory that becomes unavailable.

Flexible memory

Flexible memory was first introduced on the z9 EC as part of the design changes and offerings to support enhanced book availability. Flexible memory is used to temporarily replace the memory that becomes unavailable when performing maintenance on a book. On z196, the additional resources required for the flexible memory configurations are provided through the purchase of planned memory features along with the purchase of memory entitlement. Flexible memory configurations are available on multi-book models M32, M49, M66, and M80 and range from 32 GB to 2288 GB, depending on the model.

Contact an IBM representative to help determine the appropriate configuration.

Large page support

The size of pages and page frames has been 4 KB for a long time. Starting with System z10, System z servers have the capability of having large pages with the size of 1 MB, in addition to supporting pages of 4 KB. This is a performance item addressing particular workloads and relates to large main storage usage. Both page frame sizes can be simultaneously used.

Large pages cause the translation lookaside buffer (TLB) to better represent the working set and suffer fewer misses by allowing a single TLB entry to cover more address translations.

Exploiters of large pages are better represented in the TLB and are expected to perform better.

This support is primarily of benefit for long-running applications that are memory access intensive. Large pages are not recommended for general use. Short-lived processes with small working sets are normally not good candidates for large pages and would see little to no improvement. The use of large pages must be decided based on knowledge obtained from measurement of memory usage and page translation overhead for a specific workload.

The large page support function is not enabled without the required software support. Without the large page support, page frames are allocated at the current 4 KB size. Large pages are treated as fixed pages and are never paged out. They are only available for 64-bit virtual private storage such as virtual memory located above 2 GB.

3.2.5 Connectivity

In addition to I/O cages the z196 supports I/O drawers, introduced with z10 BC. I/O drawers complement the I/O cages available. z196 supports most of the I/O cards available on z10 EC, with some exceptions. No longer supported are: FICON Express2 (all features), OSA-Express2 10 GbE LR feature, and Crypto Express2. In addition, ICB-4 links are not supported. See Table D-1 on page 135 for more details.

All five OSA-Express3 cards continue to be supported, see 2.10.4, “OSA-Express3” on page 36.

The physical connection between the processor and memory and the I/O cages uses InfiniBand technology. Prior to and including the z9, STI cables were specifically developed for this function. With new InfiniBand cables the bandwidth per cable increases from 2.7 GB per second to 6 GB per second.

Standard InfiniBand cables and protocol can be used for Parallel Sysplex coupling links for servers that are up to 150 meter apart. InfiniBand offers a longer distance than the no longer available ICB cable, and increased bandwidth, similar to the bandwidth obtained with the cable used internally in the server. The System z9 servers can be upgraded to use the new coupling link and can participate in a Parallel Sysplex with z196, using this technology.

The Server Time Protocol (STP) can also benefit from this coupling technology. STP timing signals can be transported over PSIFB coupling links.

Coupling links

The four coupling link options for communication in a Parallel Sysplex environment are:

- ▶ Internal Coupling links (ICs), which are used for internal communication between Coupling Facilities (CFs) defined in LPARs and z/OS images on the same server.
- ▶ InterSystem Channel-3 (ISC-3), which supports a link data rate of 2 Gbps and is used for z/OS-to-CF communication at unrepeated distances up to 10 km (6.2 miles) using 9 μ m single mode fiber optic cables and repeated distances up to 100 km (62 miles) using System z-qualified DWDM equipment. ISC-3s are supported exclusively in peer mode.
- ▶ InfiniBand (HCA2-O) coupling links (12x IB-SDR or 12x IB-DDR) are used for z/OS-to-CF communication at distances up to 150 meters (492 feet) using industry standard OM3 50 μ m fiber optic cables.
 - 2x InfiniBand coupling links support single data rate (SDR) at 3 GBps when System z196 or z10 is connected to System z9.

- 2x InfiniBand coupling links support double data rate (DDR) at 6 GBps for a System z196 or z10-to-System z196 or z10 connection.
- ▶ InfiniBand (HCA2-O LR) coupling links (1x IB-SDR or 1x IB-DDR) for z/OS-to-CF communication at unrepeated distances up to 10 km (6.2 miles) using 9 µm single mode fiber optic cables and repeated distances up to 100 km (62 miles) using System z-qualified DWDM equipment.

Note: The InfiniBand coupling link data rate (6 GBps, 3 GBps, 5 Gbps, or 2.5 Gbps) does *not* represent the performance of the link. The actual performance depends on many factors, including latency through the adapters, cable lengths, and the type of workload.

When comparing coupling links data rates, InfiniBand (12x IB-SDR or 12x IB-DDR) might be higher than ICB-4¹ and InfiniBand (1x IB-SDR or 1x IB-DDR) might be higher than that of ISC-3, but with InfiniBand the service times of coupling operations are greater and the actual throughput might be less than with ICB-4¹ links or ISC-3 links.

Refer to the *Coupling Facility Configuration Options* white paper for a more specific explanation regarding the use of ICB-4¹ or ISC-3 technology versus migrating to InfiniBand coupling links. The white paper is available at:

<http://www.ibm.com/systems/z/advantages/ps0/whitepaper.html>

For details about all InfiniBand features see the *IBM System z Connectivity Handbook*, SG24-5444 or *Getting Started with InfiniBand on System z10 and System z9*, SG24-7539.

Third subchannel set

To help facilitate storage growth a third subchannel set is available to extend the amount of addressable storage capacity an additional 64K subchannels. This complements other functions such as large or extended addressing volumes and HyperPAV. This may also help facilitate consistent device address definitions, simplifying addressing schemes for congruous devices.

The first subchannel set (SS0) allows definitions of any type of device (such as bases, aliases, secondaries, and those other than disk that do not implement the concept of associated aliases or secondaries). The second and third subchannel sets (SS1 and SS2) can be designated for use for disk alias devices (of both primary and secondary devices) and/or Metro Mirror secondary devices only.

The third subchannel set applies ESCON, FICON, and zHPF protocols, and is supported by z/OS and Linux on System z.

FICON Express8

FICON Express8 is the newest generation of FICON features. They provide a link rate of 8 Gbps, with autonegotiation to 4 or 2 Gbps, for compatibility with previous devices and investment protection. Both 10KM LX and SX connections are offered (in a given feature all connections must have the same type).

With FICON Express8 customers may be able to consolidate existing FICON, FICON Express2, and FICON Express4 channels while maintaining and enhancing performance.

¹ ICB-4 is not supported on z196

z/OS discovery and autoconfiguration

z/OS discovery and autoconfiguration for FICON channels (zDAC) is designed to automatically perform a number of I/O configuration definition tasks for new and changed disk and tape controllers connected to a switch or director, when attached to a FICON channel.

Customers can define a policy, using the hardware configuration dialog (HCD). Then, when new controllers are added to an I/O configuration or changes are made to existing controllers, the system is designed to discover them and propose configuration changes based on that policy. This policy can include preferences for availability and bandwidth including parallel access volume (PAV) definitions, control unit numbers, and device number ranges.

zDAC is designed to perform discovery for all systems in a sysplex that support the function. The proposed configuration will incorporate the current contents of the I/O definition file (IODF) with additions for newly installed and changed control units and devices. zDAC is designed to help simplify I/O configuration on z196 servers running z/OS and reduce complexity and setup time.

zDAC applies to all FICON features supported on z196 when configured as CHPID type FC.

High performance FICON for System z (zHPF)

High performance FICON (zHPF), first provided on System z10, is a FICON architecture for protocol simplification and efficiency, reducing the number of information units (IUs) processed. Enhancements have been made to the z/Architecture and the FICON interface architecture to provide optimizations for on line transaction processing (OLTP) workloads.

When exploited by the FICON channel, the z/OS operating system, and the control unit (new levels of Licensed Internal Code are required) the FICON channel overhead can be reduced and performance can be improved. Additionally, the changes to the architecture provide end-to-end system enhancements to improve reliability, availability, and serviceability (RAS).

The zHPF channel programs can be exploited, for instance, by z/OS OLTP I/O workloads; DB2, VSAM, PDSE and zFS.

At announcement zHPF supported the transfer of small blocks of fixed size data (4 K). This has been extended on z10 EC to multitrack operations (limited to 64k bytes) and z196 removes the 64k byte data transfer limit on multitrack operations. zHPF requires matching support by the DS8000® series or similar devices from other vendors.

The zHPF is exclusive to z196 and System z10. The FICON Express8 and FICON Express4² (CHPID type FC) concurrently support both the existing FICON protocol and the zHPF protocol in the server Licensed Internal Code.

For more information about FICON channel performance, see the technical papers on the System z I/O connectivity Web site at:

http://www-03.ibm.com/systems/z/hardware/connectivity/ficon_performance.html

Extended distance FICON

Exploitation of an enhancement to the industry standard FICON architecture (FC-SB-3) can help avoid degradation of performance at extended distances by implementing a new protocol for *persistent* information unit (IU) pacing. Control units that exploit the enhancement to the architecture can increase the pacing count (the number of IUs allowed to be in flight from channel to control unit). Extended distance FICON allows the channel to remember the last pacing update for use on subsequent operations to help avoid degradation of performance at the start of each new operation.

² FICON Express4 10KM LX, 4KM LX and SX features are withdrawn from marketing.

Improved IU pacing can help optimize the utilization of the link (for example, help keep a 4 Gbps link fully utilized at 50 km) and allows channel extenders to work at any distance, with performance results similar to those experienced when using emulation.

The requirements for channel extension equipment are simplified with the increased number of commands in flight. This may benefit z/OS Global Mirror (Extended Remote Copy, XRC) applications, as the channel extension kit is no longer required to simulate specific channel commands. Simplifying the channel extension requirements may help reduce the total cost of ownership of end-to-end solutions.

Extended Distance FICON is transparent to operating systems and applies to all the FICON Express4, and FICON Express8 features carrying native FICON traffic (CHPID type FC). For exploitation, the control unit must support the new IU pacing protocol.

Exploitation of extended distance FICON is supported by the IBM System Storage® DS8000 series with an appropriate level of Licensed Machine Code (LMC).

FICON name server registration

The FICON channel now provides the same information to the fabric as is commonly provided by open systems, registering with the name server in the attached FICON directors. This enables a quick and efficient management of storage area network (SAN) and performing problem determination and analysis.

Platform registration is a standard service defined in the Fibre Channel - Generic Services 3 (FC-GS-3) standard (INCITS (ANSI) T11.3 group). It allows a platform (storage subsystem, host, and so on) to register information about itself with the fabric (directors).

This z196 exclusive function is transparent to operating systems and applicable to all FICON Express8 and FICON Express4 features (CHPID type FC).

FCP enhancements for small block sizes

The Fibre Channel Protocol (FCP) Licensed Internal Code has been modified to help provide increased I/O operations per second for small block sizes.

This FCP performance improvement is transparent to operating systems and applies to all the FICON Express8 and FICON Express4 features when configured as CHPID type FCP, communicating with SCSI devices.

For more information about FCP channel performance, see the performance technical papers on the System z I/O connectivity Web site at:

http://www-03.ibm.com/systems/z/hardware/connectivity/fcp_performance.html

SCSI IPL base function

The SCSI Initial Program Load (IPL) ennoblement feature, first introduced on z990 in October of 2003, is no longer required. The function is now delivered as a part of the server Licensed Internal Code. SCSI IPL allows an IPL of an operating system from an FCP-attached SCSI disk.

N_Port ID Virtualization (NPIV)

NPIV is designed to allow the sharing of a single physical FCP channel among operating system images, whether in logical partitions or as z/VM guests in virtual machines. This is achieved by assigning a unique World Wide Port Name (WWPN) for each operating system connected to the FCP channel. In turn, each operating system appears to have its own distinct WWPN in the SAN environment, hence enabling separation of the associated FCP traffic on the channel.

Access controls based on the assigned WWPN can be applied in the SAN environment, using standard mechanisms such as zoning in SAN switches and logical unit number (LUN) masking in the storage controllers.

Worldwide portname prediction tool

A part of the installation of your zEnterprise 196 server is the planning of the SAN environment. IBM has made available a standalone tool to assist with this planning prior to the installation.

The tool, known as the worldwide port name (WWPN) prediction tool, assigns WWPNs to each virtual Fibre Channel Protocol (FCP) channel/port using the same WWPN assignment algorithms that a system uses when assigning WWPNs for channels utilizing N_Port Identifier Virtualization (NPIV). Thus, the SAN can be set up in advance, allowing operations to proceed much faster once the server is installed.

The WWPN prediction tool takes a .csv file containing the FCP-specific I/O device definitions and creates the WWPN assignments that are required to set up the SAN. A binary configuration file that can be imported later by the system is also created. The .csv file can either be created manually or exported from the Hardware Configuration Definition/Hardware Configuration Manager (HCD/HCM).

The WWPN prediction tool on z196 requires at a minimum:

- ▶ z/OS V1R8 and later releases with PTFs
- ▶ z/VM V5R4 with PTFs and V6R1

The WWPN prediction tool is available for download at the Resource Link™ and is applicable to all FICON channels defined as CHPID type FCP (for communication with SCSI devices) on z196. See:

<http://www.ibm.com/servers/resourceLink/>

Fiber Quick Connect for FICON LX

Fiber Quick Connect (FQC), an optional feature on z196, is now being offered for all FICON LX (single-mode fiber) channels, in addition to the current support for ESCON (62.5 µm multimode fiber) channels. FQC is designed to significantly reduce the amount of time required for on-site installation and setup of fiber optic cabling.

FQC facilitates adds, moves, and changes of ESCON and FICON LX fiber optic cables in the data center, and may reduce fiber connection time by up to 80%. FQC is for factory installation of IBM Facilities Cabling Services - Fiber Transport System (FTS) fiber harnesses for connection to channels in the I/O cage. FTS fiber harnesses enable connection to FTS direct-attach fiber trunk cables from IBM Global Technology Services.

FQC supports all of the ESCON channels and all of the FICON LX channels in all of the I/O cages of the server.

MIDAW facility

The Modified Indirect Data Address Word (MIDAW) facility is a system architecture and software exploitation designed to improve FICON performance. This facility was introduced with System z9 servers and is exploited by the media manager in z/OS.

The MIDAW facility provides a more efficient structure for certain categories of data-chaining I/O operations:

- ▶ MIDAW can significantly improve FICON performance for extended format (EF) data sets. Non-extended data sets can also benefit from MIDAW.

- ▶ MIDAW can improve channel utilization and can significantly improve I/O response time. This reduces FICON channel connect time, director ports, and control unit overhead.

From IBM laboratory tests it is expected that applications that use EF data sets (such as DB2, or long chains of small blocks) gain significant performance benefits using the MIDAW facility.

For more information about FICON, FICON channel performance, and MIDAW, see the I/O Connectivity Web page:

<http://www.ibm.com/systems/z/connectivity/>

An excellent paper called *How does the MIDAW Facility Improve the Performance of FICON Channels Using DB2 and other workloads?*, REDP-4201, is available at:

<http://www.redbooks.ibm.com/redpapers/pdfs/redp4201.pdf>

Also see *IBM TotalStorage DS8000 Series: Performance Monitoring and Tuning*, SG24-7146

OSA-Express3

The z196 offers five OSA-Express3 features. Other capabilities are mentioned that can help consolidate or simplify the data center environment.

When compared with similar OSA-Express2 features, which they replace, the new features provide important benefits, such as:

- ▶ Doubling the density of ports: This reduces the number of CHPIDs to manage and the number of required I/O slots, which may reduce the number of I/O cages or I/O drawers. Up to 96 LAN ports versus 48 are offered.
- ▶ Designed to reduce the minimum round-trip networking time between systems (reduced latency): Improvements of up to 45% on round-trip time at the TCP/IP application layer may be realized with the OSA-Express3 10 GbE and OSA-Express3 GbE features.
- ▶ Designed to improve throughput (mixed inbound/outbound) for standard and jumbo frames.

These enhancements are because of a new function present in all OSA-Express3 features: the data router. With OSA-Express3, what was previously done in firmware is now performed in hardware. Additional logic in the IBM ASIC handles packet construction, inspection, and routing, thereby allowing packets to flow between host memory and the LAN at line speed without firmware intervention.

With the data router, the *store and forward* technique in DMA is no longer used. The data router enables a direct host memory-to-LAN flow. This avoids a *hop* and is designed to reduce latency and to increase throughput for standard frames (1492 bytes) and jumbo frames (8992 bytes).

With z196 two CHPID types are introduced: OSM for the intranode management network; and OSX for the intraensemble data network.

Intranode management network (INMN)

The intranode management network (INMN) is one of the ensemble's two private and secure internal networks. INMN is used by the Unified Resource Management functions in the primary HMC.

The z196 introduces the OSA-Express for Unified Resource Manager (OSM) CHPID type. The OSM connections are via the Bulk Power Hubs (BPHs) in the z196. The BPHs are also connected to the INMN TOR switches in the zBX. The INMN requires two OSA Express3 1000BASE-T ports from separate features.

Intraensemble data network (IEDN)

The intraensemble data network (IEDN) is the ensemble's other private and secure internal networks. IEDN is used for communications across the virtualized images (LPARs and virtual machines on z/VM and blades).

The z196 introduces the OSA-Express for zBX (OSX) CHPID type. The OSX connection is from the z196 to the IEDN TOR switches in zBX. The IEDN requires two OSA Express3 10 GbE ports from separate features.

Open Systems Adapter for NCP

The Opens Systems Adapter for NCP (OSN) support available with OSA-Express3 Gigabit Ethernet, OSA-Express3 1000BASE-T Ethernet, OSA-Express2 Gigabit Ethernet, and OSA-Express2 1000BASE-T Ethernet features has the capability to provide channel connectivity from System z operating systems to IBM Communication Controller for Linux on System z (CCL) using the Open Systems Adapter for the Network Control Program (OSA for NCP) supporting the Channel Data Link Control (CDLC) protocol.

If SNA solutions that require NCP functions are required, CCL can be considered as a migration strategy to replace IBM Communications Controllers (374x). The CDLC connectivity option enables z/TPF environments to exploit CCL.

VLAN support

Virtual local area network (VLAN) is a function of OSA features that takes advantage of the IEEE 802.q standard for virtual bridged LANs. VLANs allow easier administration of logical groups of stations that communicate as though they were on the same LAN. In the virtualized environment of System z many TCP/IP stacks can exist, potentially sharing OSA features. VLAN provides a greater degree of isolation by allowing contact with a server from only the set of stations comprising the VLAN.

VLAN is supported by z/OS, z/VM, and Linux on System z.

VMAC support

When sharing OSA port addresses across LPARs, VMAC support enables each operating system instance to have a unique virtual MAC (VMAC) address. All IP addresses associated with a TCP/IP stack are accessible using their own VMAC address, instead of sharing the MAC address of the OSA port. Advantages include a simplified configuration setup and improvements to IP workload load balancing and outbound routing.

This support is available for Layer 3 mode and is exploited by z/OS and by z/VM for guest exploitation.

QDIO data connection isolation for the z/VM environment

New workloads increasingly require multi-tier security zones. In a virtualized environment, an essential aspect is to protect workloads from intrusion or exposure of data and processes from other workloads.

The Queued Direct Input/Output (QDIO) data connection isolation enables:

- ▶ Adherence to security and HIPPA-security guidelines and regulations for network isolation between the instances sharing physical network connectivity
- ▶ Establishing security zone boundaries that have been defined by the network administrators
- ▶ A mechanism to isolate a QDIO data connection (on an OSA port) by forcing traffic to flow to the external network, ensuring that all communication flows only between an operating system and the external network

Internal *routing* can be disabled on a per-QDIO connection basis. This support does not affect the ability to share an OSA-Express port. Sharing occurs as it does today, but the ability to communicate between sharing QDIO data connections can be restricted through the use of this support.

QDIO data connection isolation applies to the z/VM environment, when using the Virtual Switch (VSWITCH) function, and to all of the OSA-Express3 features (CHPID type OSD) on z196. z/OS supports a similar capability. See “QDIO interface isolation for z/OS”.

QDIO interface isolation for z/OS

Some environments require strict controls for routing data traffic between servers or nodes. In certain cases, the LPAR-to-LPAR capability of a shared OSA port can prevent such controls from being enforced. With interface isolation, internal routing can be controlled on an LPAR basis. When interface isolation is enabled, the OSA will discard any packets destined for a z/OS LPAR that is registered in the OAT as isolated.

QDIO interface isolation is supported by Communications Server for z/OS V1R11 and above, and all OSA-Express3 features on z196.

QDIO optimized latency mode

QDIO optimized latency mode (OLM) can help improve performance for applications that have a critical requirement to minimize response times for inbound and outbound data. OLM optimizes the interrupt processing as follows:

- ▶ For inbound processing, the TCP/IP stack looks more frequently for available data to process, ensuring that any new data is read from the OSA-Express3 without requiring additional program controlled interrupts (PCIs).
- ▶ For outbound processing, the OSA-Express3 also looks more frequently for available data to process from the TCP/IP stack, thus not requiring a Signal Adapter (SIGA) instruction to determine whether more data is available.

QDIO inbound workload queuing

Inbound Workload Queuing (IWQ) has been designed to help reduce overhead and latency for inbound z/OS network data traffic, as well as implement an efficient way for initiating parallel processing. This is all achieved by using an OSA-Express3 feature in QDIO mode (CHPID types OSD and OSX) with multiple input queues and by processing network data traffic based on workload types. The data from a specific workload type is placed in one of four input queues (per device) and a process is created and scheduled to execute on one of multiple processors, independent from the other three queues. This greatly improves performance, because IWQ can exploit the SMP architecture of the z196.

Network management - query and display OSA configuration

As additional complex functions have been added to OSA, the ability for the system administrator to display, monitor, and verify the specific current OSA configuration unique to each operating system has become more complex. OSA-Express3 has the capability for the operating system to directly query and display the current OSA configuration information (similar to OSA/SF). z/OS exploits this OSA capability with a TCP/IP operator command called Display OSAINFO. Display OSAINFO allows the operator to monitor and verify the current OSA configuration, which helps improve the overall management, serviceability, and usability of OSA-Express3.

Display OSAINFO is exclusive to OSA-Express3 (CHPID types OSD, OSM, and OSX), the z/OS operating system, and z/VM for guest exploitation.

HiperSockets

HiperSockets has been called the *network in a box*. z196 supports 32 HiperSockets. One HiperSocket can be shared by up to 60 LPARs. Up to 4096 communication paths support a total of 12288 IP addresses across all 32 HiperSockets.

HiperSockets Layer 2 support

With this support the HiperSockets internal networks on z196 can support two transport modes: Layer 2 (link layer) as well as the current Layer 3 (network or IP layer). Traffic can be Internet Protocol (IP) Version 4 or Version 6 (IPv4, IPv6) or non-IP (such as AppleTalk, DECnet, IPX, NetBIOS, SNA, or others). HiperSockets devices are now protocol-independent and Layer 3 independent. Each HiperSockets device has its own Layer 2 Media Access Control (MAC) address, which is designed to allow the use of applications that depend on the existence of Layer 2 addresses such as Dynamic Host Configuration Protocol (DHCP) servers and firewalls.

Layer 2 support can help facilitate server consolidation. Complexity can be reduced, network configuration is simplified and intuitive, and LAN administrators can configure and maintain the mainframe environment the same as they do a non-mainframe environment.

HiperSockets Layer 2 support is supported by Linux on System z, and by z/VM for guest exploitation.

HiperSockets Multiple Write Facility

HiperSockets performance has been enhanced to allow for the streaming of bulk data over a HiperSockets link between logical partitions (LPARs). The receiving LPAR can now process a much larger amount of data per I/O interrupt. This enhancement is transparent to the operating system in the receiving LPAR. HiperSockets Multiple Write Facility, with fewer I/O interrupts, is designed to reduce CPU utilization of the sending and receiving LPAR.

The HiperSockets Multiple Write Facility is supported in the z/OS environment.

zIIP-Assisted HiperSockets for large messages

In z/OS, HiperSockets has been enhanced for zIIP exploitation. Specifically, the z/OS Communications Server allows the HiperSockets Multiple Write Facility processing for outbound large messages originating from z/OS to be performed on a zIIP.

zIIP-Assisted HiperSockets can help make highly secure and available HiperSockets networking an even more attractive option. z/OS application workloads based on XML, HTTP, SOAP, Java, and traditional file transfer can benefit from zIIP enablement by lowering general-purpose processor utilization for such TCP/IP traffic.

When the workload is eligible, the TCP/IP HiperSockets device driver layer (write) processing is redirected to a zIIP, which will unblock the sending application.

zIIP Assisted HiperSockets for large messages is available with z/OS V1R10 (plus service) and later releases on z196.

HiperSockets Network Traffic Analyzer

HiperSockets Network Traffic Analyzer (HS NTA) is a function available in the z196 Licensed Internal Code (LIC). It can make problem isolation and resolution simpler, by allowing Layer 2 and Layer 3 tracing of HiperSockets network traffic.

HS NTA permits Linux on System z to control tracing of the internal virtual LAN. It captures records into host memory and storage (file systems) that can be analyzed by system

programmers and network administrators, using Linux on System z tools to format, edit, and process the trace records.

A customized HiperSockets NTA rule enables you to authorize an LPAR to trace messages only from LPARs that are eligible to be traced by the NTA on the selected IQD channel.

3.2.6 Cryptography

z196 provides two major groups of cryptographic functions which, from an application program perspective, are both synchronous and asynchronous cryptographic functions:

- ▶ Synchronous cryptographic functions are provided by the CP Assist for Cryptographic Function (CPACF).
- ▶ Asynchronous cryptographic functions are provided by the Crypto Express3 feature.

CPACF

The cryptographic functions include improvements designed to facilitate continued privacy of cryptographic keys. CPACF helps to ensure that keys are not visible to applications and operating systems when used for encryption (protected key).

CPACF supports the full standard for AES (symmetric encryption) and SHA (hashing). Support for this function is provided through Web-delivered code.

CPACF is designed to provide significant throughput improvements for encryption of large volumes of data as well as low latency for encryption of small blocks of data. Furthermore, enhancements to the information management tool, IBM Encryption Tool for IMS and DB2 Databases, is designed to improve performance for protected key encryption applications.

Crypto Express3

The Crypto Express3 features can be configured as a coprocessor or as an accelerator. Support of Crypto Express3 functions varies by operating system and release and with the base and enhanced functions. Several functions require software support, which can be downloaded from the Web (see “ICSF web deliverables” on page 113).

Web deliverables

For z/OS downloads see the z/OS Web site:

<http://www-03.ibm.com/systems/z/os/zos/downloads/>

3.2.7 Hardware Management Console functionality

HMC/SE Version 2.11.0 is the current version available for the zEnterprise System servers.

The HMC and SE are appliances that together provide hardware platform management for System z. Hardware platform management covers a complex set of setup, configuration, operation, monitoring, service management tasks and services that are essential to the use of the hardware platform product.

The HMC also allows viewing and managing multi-nodal servers with virtualization, I/O networks, service networks, power subsystems, cluster connectivity infrastructure, and storage subsystems through the Unified Resource Manager. A task called *Create Ensemble* will allow the Access Administrator to create an ensemble that contains CPCs, Images, workloads, virtual networks and storage pools, either with or without an optional zBX.

The ensemble starts with a pair of HMCs that are designated as the primary and alternate HMCs, and are assigned an ensemble identity. The HMC has a global (ensemble) management function, whereas the SE has local node management responsibility. When tasks are performed on the HMC, the commands are sent to one or more SEs, which then issue commands to their CPCs and zBXs.

These Unified Resource Manager feature codes must be ordered to equip an HMC to manage an ensemble:

- ▶ 0025 - Ensemble Membership Flag
- ▶ 0019 - Manage Firmware Suite
- ▶ 0020 - Automate Firmware Suite (optional)

Refer to “Unified Resource Manager” on page 90 for more information about these HMC functions and capabilities.

HMC enhancements

The HMC application has several enhancements in addition to the Unified Resource Manager, such as:

- The “Monitors Dashboard” supersedes the “System Activity Display” (SAD) and provides a tree-based view of resources and allows an aggregated activity view when looking at large configurations. It also allows for details for objects with smaller scope. Multiple graphical ways of displaying data are available, such as history charts.
- The “Environmental Efficiency Statistic Task” provides historical power consumption and thermal information for z196 on the HMC. This task provides similar data along with a historical summary of processor and channel utilization. The data is presented in table form, graphical (“histogram”) form and it can also be exported to a .csv formatted file so that it can be imported into tools such as Excel or Lotus® 1-2-3®.
- Security and User ID Management enhancements
 - HMC audit reports can be generated, viewed, saved, and offloaded. These actions can be scheduled. The data can be offloaded in the HTML and XML formats. Security Log events can result in e-mail notifications to be sent.
 - LDAP User Authentication and HMC User ID templates enable adding/removing HMC users according to your own corporate security environment, by using an LDAP server as the central authority.
 - View Only User IDs/Access for HMC/SE allows you to create users IDs who have View Only access to selected tasks.
- The ability to export the “Change LPAR Controls” table data to an .csv formatted file. This support is available to a user when connected to the HMC remotely via a web browser.
- Partition capping values can be scheduled and is specified on the “Change LPAR Controls scheduled operation” support. Viewing of Details about an existing “Change LPAR Controls schedule operation” is available on the SE.

3.3 z196 common time functions

Each server must have an accurate time source to maintain a time-of-day value. Logical partitions use their server’s time. When servers participate in Sysplex, coordinating the time across all the systems in a complex is critical to its operation.

The z196 supports the Server Time Protocol and can participate in a common time network.

3.3.1 Server Time Protocol (STP)

Server Time Protocol is a message-based protocol in which timekeeping information is passed over data links between servers. The timekeeping information is transmitted over externally defined coupling links.

The STP feature is the supported method for maintaining time synchronization between the z196 and Coupling Facilities (CFs) in a Sysplex environments.

The STP design uses a concept called Coordinated Timing Network (CTN). A CTN is a collection of servers and CFs that are time synchronized to a time value called Coordinated Server Time (CST). Each server and CF planned to be configured in a CTN must be STP-enabled. STP is intended for servers that are configured to participate in a Parallel Sysplex or in a sysplex (without a CF), as well as servers that are not in a sysplex, but must be time synchronized.

STP is implemented in LIC as a server-wide facility of z196 (and other System z servers and CFs). STP presents a single view of time to PR/SM and provides the capability for multiple servers and CFs to maintain time synchronization with each other. A System z server or CF may be enabled for STP by installing the STP feature.

STP provides the following additional value over the Sysplex Timer®:

- ▶ STP supports a multi-site timing network of up to 100 km (62 miles) over fiber optic cabling, without requiring an intermediate site. This allows a Parallel Sysplex to span these distances and reduces the cross-site connectivity required for a multi-site Parallel Sysplex.
- ▶ The STP design allows more stringent synchronization between servers and CFs using short communication links, such as PSIFB links, compared with servers and CFs using long ISC-3 links across sites. With the z196 server, STP will support coupling links over InfiniBand.
- ▶ STP helps eliminate infrastructure requirements, such as power and space, needed to support the Sysplex Timers.
- ▶ STP helps eliminate maintenance costs associated with the Sysplex Timers.
- ▶ STP may reduce the fiber optic infrastructure requirements in a multi-site configuration. Dedicated links may not be required to transmit timing information.

The CTN concept is used to help meet two key goals of z196 and System z customers:

- ▶ Concurrent migration from an existing External Time Reference (ETR) network to a timing network using STP.
- ▶ Capability of servers to be synchronized in the timing network that contains a collection of servers and has at least one STP-configured server stepping to timing signals provided by the Sysplex Timer (z10 or previous System z servers). Such a network is called a Mixed CTN.

STP supports dial-out time services to set the time to an international time standard, such as Coordinated Universal Time (UTC), as well as to adjust to the time standard on a periodic basis. In addition, setting local time parameters, such as time zone and Daylight Saving Time (DST), and automatic updates of DST are supported.

Network Time Protocol (NTP) client support

The use of Network Time Protocol servers as an external time source (ETS) usually fulfills a requirement for a time source or common time reference across heterogeneous platforms. In most cases, this fulfillment is an NTP server that obtains the exact time through satellite.

NTP client support is added to the support element (SE) code of the z196. The code interfaces with the NTP servers. This allows an NTP server to become the single time source for z196 servers, and for other servers that have NTP clients. NTP can be used only for an STP-only CTN environment.

Pulse per second (PPS) support

Some NTP servers also provide a PPS output signal. The PPS output signal is more accurate (within 10 microseconds) than that from the HMC dial-out function or an NTP server without PPS (within 100 milliseconds).

Two External Clock Facility (ECF) cards ship as a standard feature of the z196 server and provide a dual-path interface for the Pulse Per Second (PPS) signal. The redundant design allows continuous operation in case of failure of one card and concurrent maintenance.

The two z196 server ECF cards are located in the processor cage of the z196 server. Each of the standard ECF cards have a PPS port (for a coaxial cable connection) that can be used by STP in conjunction with the NTP client.

NTP server on HMC

NTP server capability on the HMC addresses the potential security concerns that users may have for attaching NTP servers directly to the HMC/SE LAN. Note that when using the HMC as the NTP server, there is no pulse per second capability available.

For a more in-depth discussion of STP, refer to the *Server Time Protocol Planning Guide*, SG24-7280, and the *Server Time Protocol Implementation Guide*, SG24-7281.

3.4 z196 Capacity on Demand (CoD)

The z196 servers continue to deliver on demand offerings. The offerings provide flexibility and control to the customer in order to ease the administrative burden in the handling of the offerings and to give the customer finer control over resources needed to meet the resource requirements in various situations.

The z196 servers have the capability of *concurrent* upgrades, providing additional capacity with no *server* outage. In most cases, with prior planning and operating system support, a concurrent upgrade can also be *nondisruptive* to the operating system.

It is important to note that these upgrades are based on the enablement of resources already physically present in the z196 servers.

Capacity upgrades cover both permanent and temporary changes to the installed capacity. The changes can be done using the Customer Initiated Upgrade (CIU) facility, without requiring IBM service personnel involvement. Such upgrades are initiated through the Web, using IBM Resource Link. Use of the CIU facility requires a special contract between the customer and IBM, through which terms and conditions for online CoD buying of upgrades and other types of CoD upgrades are accepted. For more information consult the IBM Resource Link site:

<http://www.ibm.com/servers/resourceLink>

For more information regarding the Capacity on Demand offerings, refer to the *IBM zEnterprise System Technical Guide*, SG24-7833.

Permanent upgrades

Permanent upgrades of processors (CPs, IFLs, ICFs, zAAPs, zIIPs, and SAPs) and memory, or changes to a server's Model-Capacity Identifier, up to the limits of the installed books on an existing z196 server, can be performed by the customer through the IBM On-line Permanent Upgrade offering, using the CIU facility. These permanent upgrades require a special contract between the customer and IBM, through which the terms and conditions of the offering are accepted.

Temporary upgrades

Temporary upgrades of a z196 server can be done by On/Off CoD, Capacity Backup (CBU) or Capacity for Planned Event (CPE) ordered from the CIU facility. These temporary upgrades require a special contract between the customer and IBM, through which the terms and conditions of the offering are accepted.

On/Off Capacity on Demand

On/Off CoD is a function available on the z196 server that enables *concurrent* and *temporary* capacity growth of the server. On/Off CoD *can* be used for customer peak workload requirements, for any length of time, and has a daily hardware charge and may have an associated SW charge. On/Off CoD offerings can be pre-paid or post-paid. Capacity tokens are available on z196 servers. Capacity tokens are always present in pre-paid offerings and can be present in post-paid if the customer so desires. In both cases capacity tokens are being used to control the maximum resource and financial consumption.

Using On/Off CoD, the customer can concurrently add processors (CPs, IFLs, ICFs, zAAPs, zIIPs, and SAPs), increase the CP capacity level, or both.

Capacity Backup (CBU)

CBU allows the customer to perform a *concurrent* and *temporary* activation of additional CPs, ICFs, IFLs, zAAPs, zIIPs, and SAPs, an increase of the CP capacity level, or both, in the event of an unforeseen loss of System z capacity within the customer's enterprise, or to perform a test of the customer's disaster recovery procedures. The capacity of a CBU upgrade cannot be used for peak workload management.

CBU features are optional and require unused capacity to be available on installed books of the backup server, either as unused PUs or as a possibility to increase the CP capacity level on a sub-capacity server, or both. A CBU contract must be in place before the LIC-CC code that enables this capability can be loaded on the server. An initial CBU record provides for at least five tests (each up to 10 days in duration) and one disaster activation (up to 90 days in duration) and can be configured to be valid for up to five years.

Capacity for Planned Event (CPE)

Capacity for Planned Event allows the customer to perform a *concurrent* and *temporary* activation of additional CPs, ICFs, IFLs, zAAPs, zIIPs, and SAPs, an increase of the CP capacity level, or both, in the event of a planned outage of System z capacity within the customer's enterprise (for example, data center changes or system maintenance). CPE cannot be used for peak workload management and is available for up to a maximum of three days.

The CPE feature is optional and requires unused capacity to be available on installed books of the back-up server, either as unused PUs or as a possibility to increase the CP capacity level on a sub capacity server, or both. A CPE contract must be in place before the LIC-CC that enables this capability can be loaded on the server.

z/OS capacity provisioning

Capacity provisioning helps customers manage the CP, zAAP, and zIIP capacity of z196 servers that are running one or more instances of the z/OS operating system. Based on On/Off CoD, temporary capacity may be activated and deactivated under control of a defined policy. Combined with functions in z/OS, the z196 provisioning capability gives the customer a flexible, automated process to control the configuration and activation of On/Off CoD offerings.

3.5 Throughput optimization with z196

The z990 was the first server to use the concept of books. Despite the memory being distributed through the books and books having individual Level 3 caches (level 1 and level 2 caches were private to each core), all processors had access to all the Level 4 caches and memory. Thus, the server was managed as a memory coherent symmetric multi-processor (SMP).

Processors within the z196 book structure have different *distance to memory* attributes. As described on 2.4, “CPC cage and books” on page 24, books are connected in a star configuration, which helps to minimize the distance.

Other non-negligible effects result from data latency when grouping and dispatching work on a set of available logical processors. In order to minimize latency, one can aim to dispatch and later re-dispatch work to a group of physical CPUs that share the same Level 3 cache.

PR/SM manages the utilization of physical processors by logical partitions by dispatching the logical processors on the physical processors. But PR/SM is not aware of which workloads are being dispatched by the operating system in which logical processors. The Workload Manager (WLM) component of z/OS has the information at the task level, but is unaware of physical processors. This disconnect is solved by enhancements on z196 that allow PR/SM and WLM to work more closely together. They can cooperate to create an affinity between task and physical processor rather than between logical partition and physical processor. This is known as HiperDispatch.

HiperDispatch

HiperDispatch, introduced with z10 and enhanced in z196, combines two functional enhancements, one in the z/OS dispatcher and one in PR/SM. This is intended to improve efficiency both in the hardware and in z/OS.

In general, the PR/SM dispatcher assigns work to a minimum number of logical processors needed for the priority (weight) of the LPAR. PR/SM also attempts to group the logical processors into the same book and, if possible, the same chip. The end result is to reduce the multi-processor effects, maximize use of shared cache and lower the interference among multiple partitions.

The z/OS dispatcher is enhanced to operate with multiple dispatching queues, and tasks are distributed among these queues. The current implementation operates with an average of four logical processors per queue. Specific z/OS tasks may then be dispatched to a small subset of logical processors, which PR/SM will tie to the same physical processors, thus improving the hardware cache re-use and locality of reference characteristics such as reducing the rate of cross-book communication.

To use the correct logical processors, the z/OS dispatcher obtains the necessary information from PR/SM through interfaces implemented on the z196. The entire z196 stack (hardware, firmware, and software) now tightly collaborates to obtain the hardware’s full potential.

The HiperDispatch function can be switched on and off dynamically without requiring an IPL.

3.6 zEnterprise BladeCenter Extension

The zEnterprise System represents a new height for mainframe functionality and qualities of service. It has been rightly portrayed as a corner stone for the IT infrastructure, especially when the need for flexibility on rapidly changing environments is called for.

zEnterprise System characteristics make it especially valuable for mission critical workloads and, today, most of these applications have multi-tiered architectures and logically and physically span several hardware and software platforms. However, there are differences in the qualities of service offered by the platforms as well as in the configuration procedures for their hardware and software, operational management, software servicing, failure detection and correction, and so on. These, in turn, require personnel with several separate skills sets, several sets of operational procedures and an integration effort which is far from trivial or negligible and, therefore, not often achieved. Failure in achieving integration translates to lack of flexibility and agility, which can impact the bottom line.

IBM mainframe systems have been providing specialized hardware and dedicated computing capabilities for a long time. Not counting the machine instruction assists, one can recall for instance, the vector facility of the IBM 3090 (in its separate frame), back in the mid-1980s. Other such specialty hardware include the System Assist Processor, for I/O handling, which implemented the 370-XA architecture, the Coupling Facility, and the Cryptographic processors. And of course, all the I/O cards which are nothing less than specialized dedicated hardware with sophisticated software, which offloads processing from the System z processor units (PUs).

The common theme with these specialized hardware components is their seamless integration within the mainframe. The zEnterprise Blade Extension Model 002 (zBX) components are configured, managed, and serviced the same way as the other components of the System z server. Despite the fact that the zBX processors may not be System z PUs, the zBX is in fact, handled by System z firmware called zEnterprise Unified Resource Manager. The zBX hardware features are part of the mainframe, not an add-ons.

System z has long been an integrated heterogeneous platform. With zBX Model 002, that integration reaches a new level. zBX provides within the zEnterprise System a solution for running AIX workloads with IBM POWER7 blades. Also zBX supports a cost optimized solution for running Data Warehouse and Business Intelligence queries against DB2 for z/OS. This solution is known as the IBM Smart Analytics Optimizer.

3.6.1 IBM blades

IBM offers a selected set of IBM POWER7 blades that can be installed and operated on the zBX Model 002. These blades are virtualized by PowerVM. The virtual servers in PowerVM run the AIX operating system.

PowerVM handles all the access to the hardware resources. PowerVM provides the user with a Virtual I/O Server (VIOS) function and the ability to create logical partitions. The logical partitions can be either Dynamic Logical Partitions (DLPARs), which require a minimum of 1 core per partition or Micro-Partitions, which can be as small as 0.1 core per partition.

Statement of Direction: In the first half of 2011, IBM intends to offer a System x blade running Linux and a WebSphere DataPower Appliance, for the zBX Model 002.

Refer to “zEnterprise ensembles and virtualization” on page 86 for an overview of virtualized resources in the zBX.

3.6.2 IBM Smart Analytics Optimizer solution

The IBM Smart Analytics Optimizer solution is designed to execute queries typically found in business intelligence (BI) and data warehousing (DW) applications with fast and predictable response times, thus offering a comprehensive Business Intelligence solution on zEnterprise.

The offering is comprised of hardware and software. The software, IBM Smart Analytics Optimizer for DB2 for z/OS, Version 1.1 (Program Product 5697-AQT), exploits the zBX to provide a comprehensive Business Intelligence solution on System z that can deliver:

- ▶ Up to 10x performance gains on certain types of queries
- ▶ Accelerated business insight for faster and better decision making
- ▶ Reduced administrative time and cost associated with database tuning
- ▶ Higher qualities of service across fit-for-purpose, workload optimized infrastructure
- ▶ Improved economics through better infrastructure management and resiliency

To simplify the whole process, from ordering to exploitation and administration, five solution offerings are available. These zBX Model 002 offerings that can be selected by the client are based on the amount of raw DB2 data (DB2 tables, number of indexes, number of AQT³s) to be queried. After completing a workload assessment as part of the solution assurance process, the required number of blades may be ordered in quantities of 7, 14, 28, 42, or 56.

More information on Smart Analytics Optimizer and the solution assurance process can be found at:

<http://www.ibm.com/software/data/infosphere/smart-analytics-optimizer-z/>

The zBX hardware and the IBM Smart Analytics Optimizer software are separately ordered.

A single zBX can be accessed from up to eight z196 servers, but is controlled from a single z196 server, called the *owning server*. A DB2 data sharing group spanning the z/OS LPARS of the up to eight z196 servers can exploit a single IBM Smart Analytics Optimizer (on the zBX). Other data sharing groups, running in the same servers, could exploit *different* IBM Smart Analytics Optimizer (on zBXs).

This is an integrated solution offering a centralized environment that extends System z legendary availability and security to heterogeneous Business Intelligence (BI) and Data Warehouse (DW) workloads. These benefits come without change to current applications, since DB2 for z/OS transparently exploits the special purpose hardware and software for query execution by sending qualified queries to the IBM Smart Analytics Optimizer running on zBX.

For further discussion on the benefits and usage of the IBM Smart Analytics Optimizer solution, refer to *Using IBM System z as the foundation for your information management architecture*, REDP-4606.

³ Eligible queries for the IBM Smart Analytics Optimizer solutions will be executed on data marts specified as Accelerator Query Table (AQT) in DB2 for z/OS. An AQT is based on the same principles as a Materialized Query Table (MQT). MQTs are tables whose definitions are based on query results. The data in those tables is derived from the table or tables on which the MQT definition is based. See the article at: <http://www.ibm.com/developerworks/data/library/techarticle/dm-0509me1nyk>

3.7 z196 performance

The z196 Model M80 is designed to offer approximately 1.6 times more capacity than the z10 EC Model E64 system. Uniprocessor performance has also increased significantly. A z196 Model 701 offers, on average, performance improvements of about 1.35 to 1.5 times the z10 EC Model 701.

On average, the z196 can deliver up to 40% more performance in an n-way configuration than a System z10 EC n-way. However, variations on the observed performance increase are dependent upon the workload type.

IBM continues to measure performance of the systems by using a variety of workloads and publishes the results in the Large Systems Performance Reference (LSPR) report. The LSPR is available at:

<https://www-304.ibm.com/servers/resourceink/lib03060.nsf/pages/lsprindex?OpenDocument>

The MSU ratings are available at:

<http://www.ibm.com/servers/eserver/zseries/library/swpriceinfo>

LSPR workload suite - This has changed with the z196

Historically, LSPR capacity tables, including “pure” workloads and mixes, have been identified with application names or some software characteristic. Examples are CICS, IMS, OLTP-T, CB-L, Low I/O Content Mix Workload (LoIO-mix) and Transaction Intensive Mix Workload (TI-mix). However, capacity performance is, in fact, more closely associated with how a workload uses and interacts with a particular processor hardware design. Of particular significance are: instruction path length, instruction complexity, and memory hierarchy.

On System z10, IBM made available the CPU measurement facility (MF) which provides insight on how a production workload interacts with the hardware design, in particular the processor access to caches and memory. This is known as “nest” activity intensity. MF data can be collected by z/OS System Measurement Facility on SMF 113 records.

With the availability of this data, LSPR is adjusting workload capacity curves based on the underlying hardware sensitivities. Thus the latest LSPR introduces three new workload capacity categories which replace all prior primitives and mixes.

The new categories are based on the relative nest intensity, which is influenced by many variables such as application type, I/O rate, application mix, CPU usage, data reference patterns, LPAR configuration, and software configuration running, among others.

The three new workload categories represented in the LSPR tables are as follows:

- ▶ Low: A workload category representing light use of the memory hierarchy. This would be similar to past high scaling primitives.
- ▶ Average: A workload category representing average use of the memory hierarchy. This would be similar to the past LoIO-mix workload and is expected to represent the majority of production workloads.
- ▶ High: A workload category representing heavy use of the memory hierarchy. This would be similar to the past TI-mix workload

Guidance in converting LSPR previous categories to the new ones is provided, as well as built-in support on the zPCR tool.

The previous LSPR workload suite comprised the following workloads:

- ▶ Traditional online transaction processing workload OLTP-T (formerly known as IMS)
- ▶ Web-enabled online transaction processing workload OLTP-W (also known as Web/CICS/DB2)
- ▶ A heavy Java-based online stock trading application WASDB (previously referred to as Trade2-EJB).
- ▶ Batch processing, represented by the CB-L (commercial batch with long-running jobs or CBW2)
- ▶ A new ODE-B Java batch workload, replacing the CB-J workload

The traditional Commercial Batch Short Job Steps (CB-S) workload (formerly CB84) was dropped.

The previous LSPR provided performance ratios for individual workloads and for the default mixed workload, which was composed of equal amounts of four of the workloads described above (OLTP-T, OLTP-W, WASDB, and CB-L).

The latest zPCR provides, in addition to Low, Average and High categories the Low-Average and Average-High categories, which allow better granularity for workload characterization.

The z196 LSPR tables continue to rate all z/Architecture processors running in LPAR mode and 64-bit mode. The single-number values are based on a combination of the default mixed workload ratios, typical multi-LPAR configurations, and expected early-program migration scenarios. In addition to z/OS workloads used to set the single-number values, the z196 LSPR tables contain information pertaining to Linux and z/VM environments.

Capacity ratio estimates

Figure 3-2 on page 79 shows the estimated capacity ratios for z196 and z10 EC. The capacity estimate is based on the LSPR workload suite described previously.

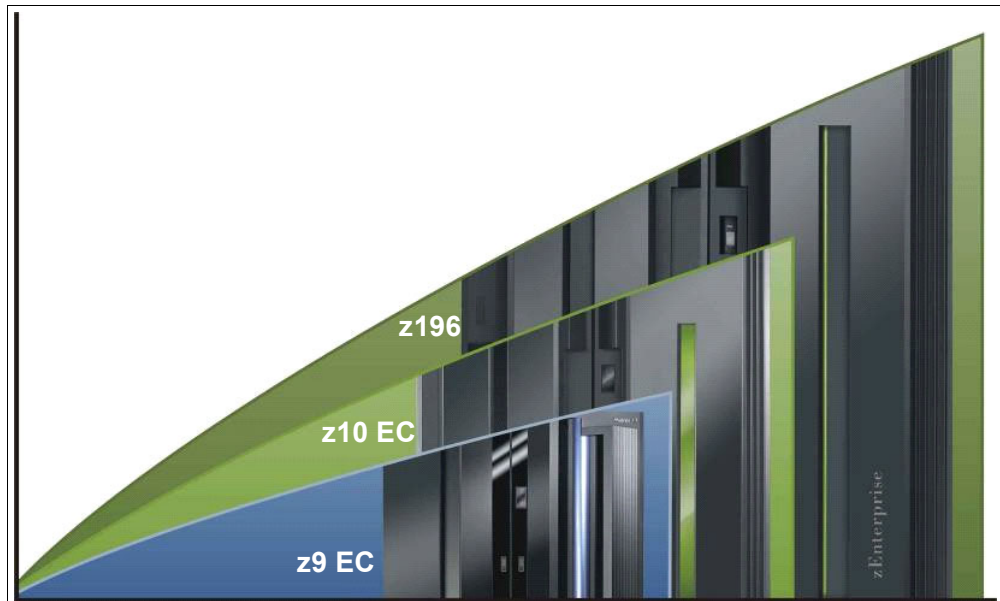


Figure 3-2 z196 to z10 EC and z9 EC performance comparison

The LSPR contains the internal throughput rate ratios (ITRRs) for the z196 and the previous generations of processors based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user might

experience varies depending on factors such as the amount of multiprogramming in the user's job stream, the I/O configuration, and the workload processed.

Workload performance variation

Because of the nature of the z196 multi-book system and resource management across those books, performance variability similar to what occurred with the z9 EC and z10 EC is expected. This variability can be observed in several ways. The range of performance ratings across the individual workloads is likely to have some spread, but not as large as with the z10 EC.

The new memory and cache designs affect different workloads with a certain variability. All workloads are improved, with cache-intensive loads benefiting the most. When comparing moving from z9 EC to z10 EC with moving from z10 EC to z196, it is likely that the relative benefits per workload are different. Those workloads which benefited more than the average when moving from z9 EC to z10 EC will benefit less than the average when moving from z10 EC to z196, and vice-versa.

The customer impact of this increased variability is seen as increased deviations of workloads from single-number metric-based factors such as MIPS, MSUs, and CPU time charge back algorithms.

Experience demonstrates that System z servers can be run at up to 100% utilization levels, sustained, although most clients prefer to leave a bit of white space and run at 90% or slightly under.

3.8 Reliability, availability, and serviceability

The z196 server presents numerous enhancements in the reliability, availability, and serviceability areas. In the availability area focus was given to reduce the planning requirements, while continuing to improve the elimination of planned, scheduled, and unscheduled outages.

Enhanced driver maintenance (EDM) helps reduce the necessity and the eventual duration of a scheduled outage. One of the contributors to scheduled outages is LIC Driver updates performed in support of new features and functions. When properly configured, the System z196 can concurrently activate a new LIC Driver level. Concurrent activation of the select new LIC Driver level is supported only at specifically released synchronization points. However, there are certain LIC updates where a concurrent update/upgrade may not be possible.

Availability enhancements include single processor core checkstop and sparing, enhanced cooling system including a water-cooled option, humidity and altimeter sensors, point-to-point fabric for SMP, and fixed size HSA.

The z196 introduces a new way to increase memory availability - Redundant Array of Independent Memory - where a fully redundant memory system can identify and correct memory errors without stopping. This new concept is unique to the z196; the implementation is similar to the RAID concept used in storage systems for a number of years.

If an additional system assist processor (SAP) is required on a z196 server (for example, as a result of a disaster recovery situation), the SAPs can be concurrently added to the server configuration.

It is possible to concurrently add CPs, zAAPs, zIIPs, IFLs, and ICFs processors to an LPAR. This is supported by z/VM V5R4 and later with appropriate PTFs, by z/OS, and z/VSE V4R3.

Previously, proper planning was required in order to concurrently add CPs, zAAPs, and zIIPs to a z/OS LPAR.

Concurrently adding memory to an LPAR is also possible and is supported by z/OS and z/VM.

z196 supports dynamically adding Crypto Express features to an LPAR by providing the ability to change the cryptographic information in the image profiles without outage to the LPAR. Users can also dynamically delete or move Crypto Express features. This enhancement is supported by z/OS, z/VM, and Linux on System z.

The System Activity Display (SAD) screens now include energy efficiency displays.

RAS capability for the HMC

The HMC for the zEnterprise is where parts of the Unified Resource Manager routines are executing.

The Unified Resource Manager is an active part of the zEnterprise System infrastructure. The HMC is therefore a stateful environment that needs high availability features to assure survival of the system in case of an HMC failure. Each zEnterprise comes equipped with two HMC workstations - a primary and a backup. The contents and activities of the primary are kept synchronously updated on the backup HMC, so that the backup can automatically take over the activities of the primary, should the primary fail. While the primary HMC can do the classic HMC activities in addition to the Unified Resource Manager activities, the backup can only be the backup; no additional tasks or activities can be performed at the backup HMC.

3.8.1 RAS capability for zBX

The zBX has been built with the traditional System z QoS to include RAS capabilities. The zBX offering provides extended service capability via the z196 hardware management structure. The HMC/SE functions of the z196 server provide management and control functions for the zBX solution.

Apart from a zBX configuration with one chassis installed, the zBX is configured to provide $N + 1$ components. The components are designed to be replaced concurrently. In addition zBX configuration upgrades can be performed concurrently.

The zBX has 2 top of rack switches (TORs). These switches provide $N + 1$ connectivity for the private networks between the z196 server and the zBX for monitoring, controlling, and managing the zBX components.

zBX Firmware

The testing, delivery, installation, and management of the zBX firmware is handled exactly the same way as for the z196 server. The same z196 server process and controls are used. Any fixes to the zBX machine are downloaded on to the owning z196 server's SE and are applied to the zBX.

The MCLs for the zBX are designed to be concurrent and their status can be viewed at the z196 server's HMC.

These and additional features are further described *IBM zEnterprise System Technical Guide*, SG24-7833.

3.9 High availability technology

Parallel Sysplex technology is a clustering technology for logical and physical servers, allowing the highly reliable, redundant, and robust System z technology to achieve near-continuous availability. Both hardware and software tightly cooperate to achieve this result. The hardware components comprise:

- ▶ **Coupling Facility (CF):** This is the cluster center. It can be implemented either as an LPAR of a stand-alone System z server or as an additional LPAR of a System z server where other loads are running. Processor units characterized as either CPs or ICFs can be configured to this LPAR. ICFs are often used because they do not incur any software license charges. Two CFs are recommended for availability.
- ▶ **Coupling Facility Control Code (CFCC):** This IBM Licensed Internal Code is both the *operating system* and the *application* that executes in the CF. No other code executes in the CF.⁴
- ▶ **Coupling links:** These are high-speed links connecting the several system images (each running in its own logical partition) that participate in the Parallel Sysplex. At least two connections between each physical server and the CF should exist. When all of the system images belong to the same physical server, internal coupling links are used.

On the software side, the z/OS operating system exploits the hardware components to create a Parallel Sysplex⁵. Normally, two or more z/OS images are clustered to create a Parallel Sysplex, although it is possible to have a configuration setting with a single image, called a *monoplex*. Multiple clusters can span several System z servers although a specific image (logical partition) can belong to only one Parallel Sysplex.

A z/OS Parallel Sysplex implements a shared-all access to data. This is facilitated by System z I/O virtualization capabilities such as Multiple Image Facility (MIF). MIF allows several logical partitions to share I/O paths in a totally secure way, maximizing utilization and greatly simplifying the configuration and connectivity.

In short, a Parallel Sysplex comprises one or more z/OS operating system images coupled through one or more Coupling Facilities. A properly configured Parallel Sysplex cluster is designed to maximize availability at the application level.

The major characteristics of a Parallel Sysplex are:

- ▶ **Data sharing with integrity:** The CF is key to the implementation of a share-all access to data. Every z/OS system image has access to all the data. Subsystems in z/OS declare resources to the CF. The CF accepts and manages lock and unlock requests on those resources, guaranteeing data integrity. A duplicate CF further enhances the availability. Key exploiters of this capability are DB2, WebSphere MQ, WebSphere ESB, IMS, and CICS.
- ▶ **Continuous (application) availability:** Changes, such as software upgrades and patches, can be introduced one image at a time, while the remaining images continue to process work. For additional details see the manual *Parallel Sysplex Application Considerations*, SG24-6523.
- ▶ **High capacity:** scales from two to 32 images. Remember that each image can have from 1 to 80 processor units. CF scalability is near linear. This contrasts with other forms of

⁴ CFCC can also execute in a z/VM Virtual Machine (as a z/VM guest system). In fact, a complete Sysplex can be set up under z/VM allowing, for instance, testing and operations training. This setup is not recommended for production environments.

⁵ z/TPF can also exploit the CF hardware components. However, the term Sysplex exclusively applies to z/OS exploitation of CF.

clustering that employ n-to-n messaging, leading to rapidly degrading performance with growth of the number of nodes.

- ▶ Dynamic workload balancing: Viewed as a single logical resource, work can be directed to any of the Parallel Sysplex cluster operating system images where capacity is available.
- ▶ Systems management: The architecture provides the infrastructure to satisfy a customer requirement for continuous availability, while enabling techniques for achieving simplified systems management consistent with this requirement.
- ▶ Resource sharing: A number of base z/OS components exploit Coupling Facility shared storage. This exploitation enables sharing of physical resources with significant improvements in cost, performance, and simplified systems management.
- ▶ Single system image: The collection of system images in the Parallel Sysplex appears as a single entity to the operator, the user, the database administrator, and so on. A single system image ensures reduced complexity from both operational and definition perspectives.

Figure 3-3 on page 83 illustrates the components of a Parallel Sysplex as implemented within the System z architecture. It shows one of many possible Parallel Sysplex configurations.

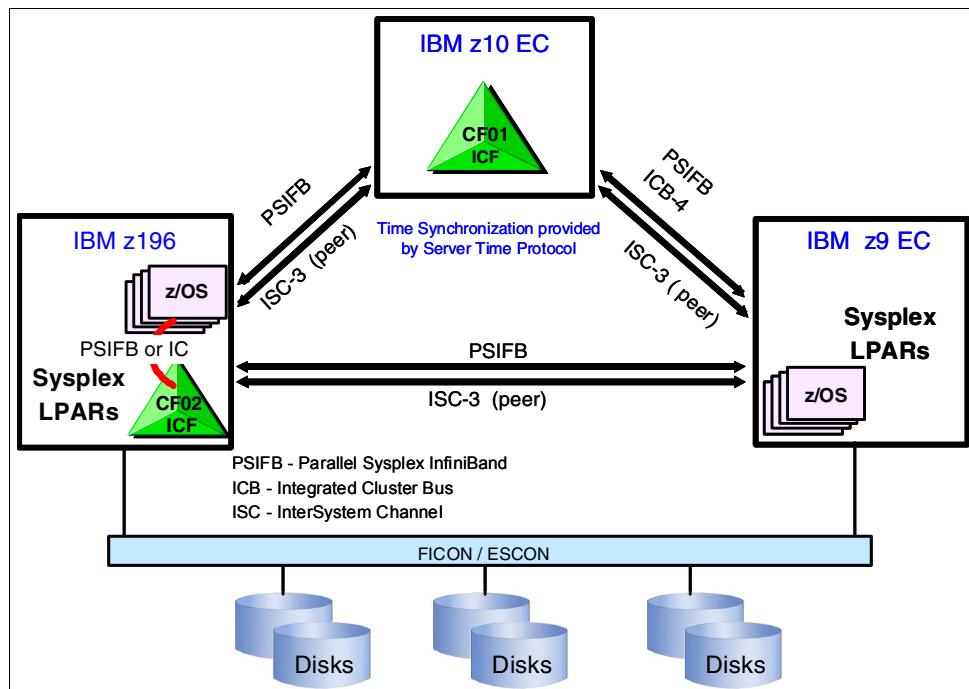


Figure 3-3 Sysplex hardware overview

Figure 3-3 shows a z196 server containing multiple z/OS sysplex partitions and an internal coupling facility (CF02), a z10 EC server containing a stand-alone CF (CF01), and a System z9 EC containing multiple z/OS sysplex partitions. STP over coupling links provides time synchronization to all servers. Appropriate CF link technology (PSIFB or ISC-3) selection depends on server configuration. Note the ICB-4 coupling link is no longer supported on z196.

Through this state-of-the-art cluster technology, the power of multiple z/OS images can be harnessed to work in concert on shared workloads and data. The System z Parallel Sysplex cluster takes the commercial strengths of the z/OS platform to improved levels of system management, competitive price/performance, scalable growth, and continuous availability.



Achieving better infrastructure resource management

Business processes and the workloads supporting them are becoming more service oriented, modular in their construction, and integrated. Commonly, the components of these services are implemented on a variety of architectures and hosted on diverse infrastructures. This creates inherent inefficiencies, underutilized assets, and problems with availability, integrity, and security.

However, attempts to manage these infrastructures along the lines of platform architecture boundaries have not produced the desired alignment of IT with business objectives, agility to change, improved resource utilization, or reduced cost of ownership.

What is needed is an infrastructure resource management approach that has the ability to seamlessly manage diverse infrastructures and provide visibility, control, and automation across different architectures. Resource management capabilities should drive up new levels of efficiency and optimization, while mitigating operational risk by simplification. Such capabilities also improve the quality of service (QoS).

In this chapter we discuss infrastructure resource management and how it can be improved with the zEnterprise System to support diverse workloads. The sections in this chapter include:

- ▶ 4.1, “zEnterprise ensembles and virtualization” on page 86
- ▶ 4.2, “How can I tell if my business will benefit” on page 87
- ▶ 4.3, “Unified Resource Manager” on page 90
- ▶ 4.4, “Physical resource management” on page 92
- ▶ 4.5, “Virtualization management” on page 93
- ▶ 4.6, “Performance management” on page 96

4.1 zEnterprise ensembles and virtualization

To support a workload, an infrastructure can have dedicated, shared, or even better, virtualized resources. Through virtualization, utilization of hardware and software is increased, while the cost of ownership is decreased. Other areas of potential savings with virtualized resources are reduced risk, less idle time, acceleration of innovative and enhanced client services, and the ability to significantly reduce time to introduce new technologies.

Typically, virtualized resources include processor units, memory, networking, storage, and the hypervisors that manage them. To execute the workload, an image (containing the relevant operating system, the middleware stack, and the business applications) is also required.

The zEnterprise System provides all those capabilities. It is a workload-optimized system that spans mainframe, UNIX®, and in the future, x86 technologies. The zEnterprise System also has an integrated System z management facility to unify workload management, by delivering workload-based resource allocation and provisioning.

In addition, the zEnterprise System is capable of acting as a *node* in an *ensemble*. An ensemble is a collection of one or more nodes that are managed as a single logical virtualized system. Each node is comprised of a zEnterprise 196 (z196) and its optionally attached zEnterprise BladeCenter Extension (zBX).

An ensemble can have from one to eight zEnterprise nodes, and the zEnterprise Unified Resource Manager enables provisioning and management of the ensemble.

Figure 4-1 shows a logical view of a single node ensemble and the Unified Resource Manager. Unified resource management is provided by zEnterprise firmware, also known as Licensed Internal Code (LIC), executing in the Hardware Management Console (HMC) and Support Element (SE).

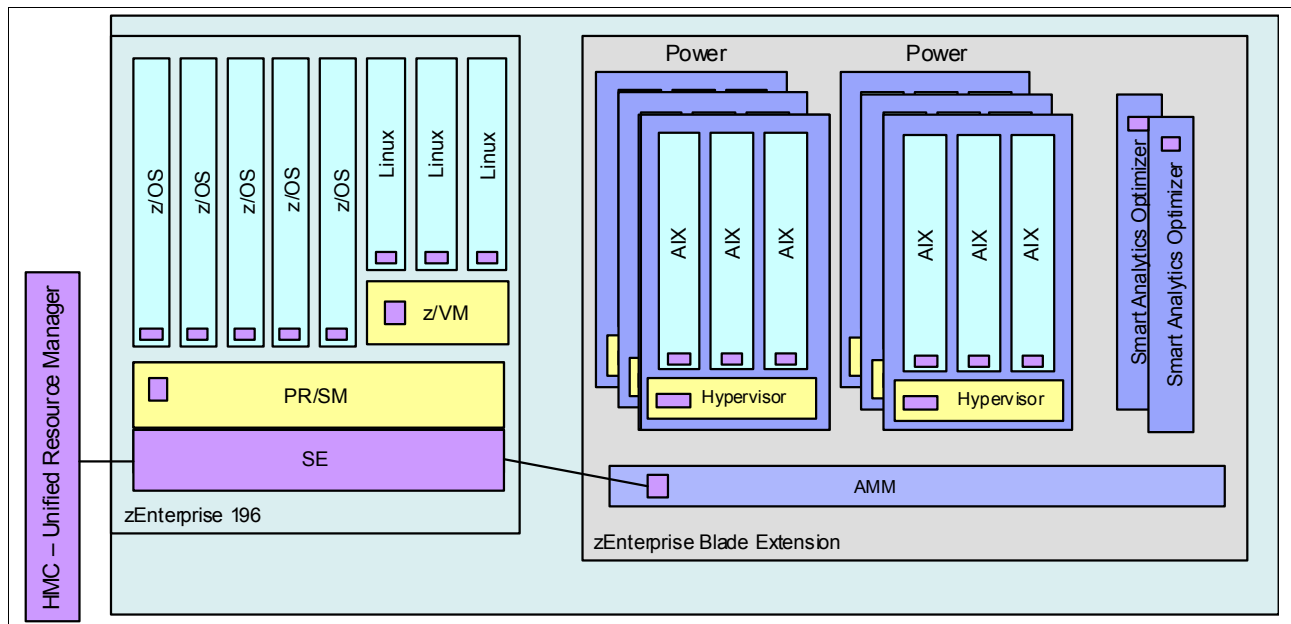


Figure 4-1 A logical view of an ensemble with unified resource management

The zEnterprise System has been designed to optimize and better manage diverse workloads. The ensemble components all contribute toward meeting workload performance objectives.

An optional suite of applications called the guest platform management provider (GPMP) can be installed in specific z/OS, Linux, and AIX operating system images to support resource management functions. GPMP collects and aggregates performance data for virtual servers and workloads. Reports can be viewed through the HMC that manages the ensemble.

4.2 How can I tell if my business will benefit

Multitier workloads and their deployment on diverse infrastructures are common today. The zEnterprise System presents a highly optimized environment for deploying such workloads with improved monitoring and management.

In this section, we introduce a few workload patterns to describe some common examples. You might recognize these patterns in your own workloads and be able to see how they can benefit from the zEnterprise System. Maybe in one of these examples you will find a great idea that can be applied to your business.

4.2.1 Mainframe workloads

It is impossible to discuss workloads that can benefit from the zEnterprise System without including the traditional mainframe workloads. The innovations in the z196 continue IBM's long-standing history of continuous improvements in mainframe processing. IBM's well-known transaction processing systems, such as CICS and IMS, can handle even greater volumes with the additional capabilities that the z196 provides.

There are software solutions that leverage z/OS capabilities to enhance SOA architectures and provide significant improvements for Java programs. With z/OS, z/TPF, z/VSE, z/VM, and Linux on System z, there are plenty of opportunities to create powerful application solutions.

When you consider how to deploy multitier applications, keep in mind that they can leverage the unmatched reliability and security of the zEnterprise 196. The z196 is ideal for data and transaction serving for mission-critical applications.

4.2.2 Heterogeneous platform deployments

The introduction of the zBX means that new and existing workloads will be running on the zEnterprise System. Here we outline a few types of applications that can benefit from this new environment.

World Wide Web application - a three-tier architecture

One deployment model that has become common with the explosive growth of the World Wide Web is a three-tier web server application. This commonly involves an http server to present web pages to the users and accept their input. The http server identifies the application function requested by the user and sends the request to the next tier, the application server.

The application server will accept the request from the http server, select the appropriate business logic, and begin processing the request. This may imply calling upon business logic implemented in CICS or IMS, or even externally through web services. When the application server needs to retrieve data to supply the user's request, or when it needs to store the information provided by the user, it calls on a database server.

The database server maintains the organized data structures required by the application and invokes the necessary storage (disk) requests to retrieve or store data as requested by the application server. An example of such an architecture is depicted in Figure 4-2 on page 88.

This three-tier architecture is often deployed with the http server outside a firewall to protect the customer's network from unwanted intrusions from the public Internet. The http server may be implemented on multiple servers to insure sufficient capacity for a large volume of requests and to maintain availability. The http server then communicates through the firewall to the application servers.

The application servers, also likely implemented as a cluster, communicate with the database server and that data flow might be encrypted or flow through another firewall, depending on data sensitivity and industry or government privacy regulations.

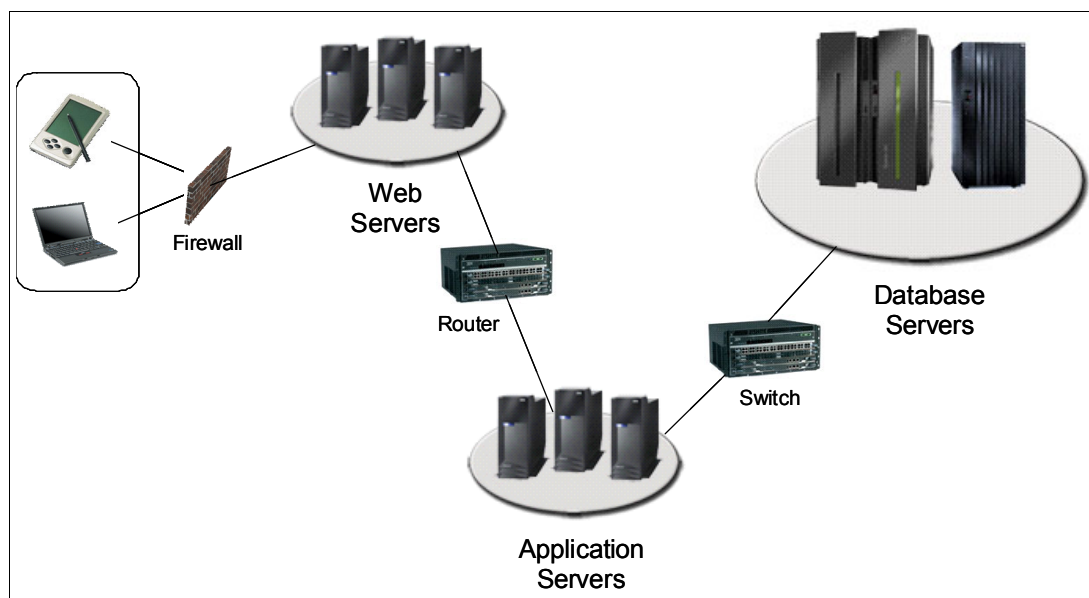


Figure 4-2 An example three-tier application architecture implementation

This application architecture may benefit significantly from being deployed across a zEnterprise System. For example, the http server can be virtualized and deployed across several blades in the zBX. The public “Internet” communications could be isolated across one VLAN in the intraensemble data network (IEDN) to which no other virtual servers are allowed access. VLAN isolation is considered to be as secure as physical isolation by many networking industry groups.

The internal (“intranet”) communications could then be directed to the application server cluster deployed in the zBX blades or z196 (z/OS or Linux on System z under z/VM) through a separate VLAN.

Lastly, the application server's communications with the database server, which might be a DB2 for z/OS running in the z196, also flow over the IEDN. Because the IEDN is privately managed in the zEnterprise ensemble and may be configured without physical connections between the servers that might be compromised, the application server and database communications are highly secure. Figure 4-3 on page 89 illustrates how this could be implemented with the zEnterprise System.

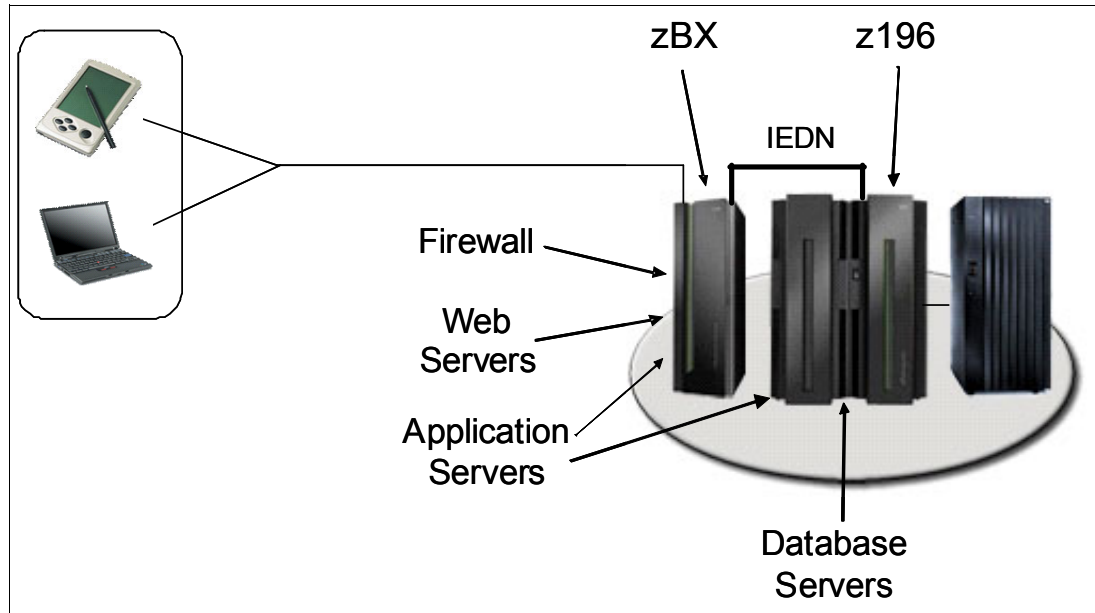


Figure 4-3 Three-tier application architecture on the zEnterprise System

Grid-like cooperative processing architectures

The evolution of information processing solutions has grown to include special purpose environments, tailored to perform some tasks very efficiently. One example in this category is the IBM Smart Analytics Optimizer. The IBM Smart Analytics Optimizer is delivered as a packaged solution that installs directly in the zBX. It provides a number of blades that are managed as a query processor for DB2 for z/OS. With appropriate software features installed, DB2 can use the IBM Smart Analytics Optimizer as an optimizer to build large data structures that can be queried very quickly, in parallel, to produce results much faster than traditional DB2 queries might perform.

Another example might be a high performance grid. In the grid concept, numerous independent but coordinated processors are given part of a complex task and allowed to proceed through the necessary calculations independently. After all the components have completed, the coordinator assembles the final result and returns it to the requestor. An application of this technology might involve investment portfolio analysis that provides recommendations for future investments based on recent market trends. Another such application is used in biomedical research where the search for possible medicines and vaccines involves the testing of so many combinations that linear processing simply would not complete in a feasible timeframe. Such grid implementations can sometimes take advantage of the computing capacity that is built for peak periods and apply it to streamlining other processes during off-peak times.

Virtualization and server consolidation

The zEnterprise System differentiates itself by including monitoring and management functions that are designed to get the most from the virtualized environment.

Because the zEnterprise System is built with virtualization as a central theme, it presents an attractive solution for server consolidation. This is a topic of current interest in many data centers. The past years of adding multiple distributed server platforms for every new application has created “server sprawl” which has contributed to power, cooling, and floorspace problems in many data centers. With so many separate servers running at very low average utilization rates, a tremendous opportunity exists to consolidate servers onto fewer hardware instances through virtualization.

The features of the zEnterprise System are designed to take full advantage of the resources available to meet your business needs. System z operating systems have a history of managing disparate work within a single system using a set of business objectives. This workload management is the inspiration for the similar functions included in the Unified Resource Manager suites.

The Performance Management component includes the same kind of workload classification rules to define class of work by hostname, the virtual server's name, or other criteria. By applying these classification rules to different workloads running under the same hypervisor, the Performance Management component can help allocate resources. When a more important workload needs more processor resource, the entitlement to CPU resources can be altered to move the available resources to the most important workload.

With this kind of management capability, server consolidation becomes a strategy that can improve more than simply the hardware costs of your servers; it can make your workload perform better with fewer resources.

Quality of service improvements

With the introduction of the zEnterprise System, the qualities for which the System z is renowned are extended to other components that are part of the ensemble to provide support for mission-critical workloads running on the heterogeneous infrastructure of the ensemble. Compared to a diverse infrastructure, the zEnterprise System provides:

- ▶ A single management and policy framework across web serving, transaction, database and servers to lower the cost of enterprise computing
- ▶ Integration of multiplatform management capabilities through extended functionality in the well-known mainframe Hardware Management Console (HMC)
- ▶ Mainframe systems management and service extended to the zBX environment
- ▶ Dynamic resource management of the mainframe to all devices within a multitier architecture to improve service
- ▶ Monitoring and management of a heterogeneous solution as a single, logical virtualized solution
- ▶ Management of the platform's resources in accordance with specified business service level objectives
- ▶ Management of virtual servers as part of the overall deployed business workload
- ▶ A secure and managed Layer 2 network, connecting the zBX blades with the z196 CPC of the zEnterprise System

Mainframe QoS characteristics are extended to accelerators and application servers to mitigate risk of operational failures.

4.3 Unified Resource Manager

The Unified Resource Manager is an integral part of the zEnterprise System. It provides end-to-end virtualization and management of z196 and zBX resources with the ability to align those resources according to individual workload requirements.

Through virtualization, the physical resources can be shared among multiple workloads because they likely have different policies with different objectives to meet. The Unified Resource Manager's goal is to fulfill the objectives of the workload policies in the most optimal and efficient way.

4.3.1 Resource management suites

The zEnterprise System has resource management functions as part of the Unified Resource Manager, which are accessed through the HMC. The functions delivered by the Unified Resource Manager are implemented in two operational *suites* and provide the following capabilities:

- ▶ Integrated hardware management across all elements of the system, the z196 and the zBX
- ▶ Fully automatic and coherent integrated resource discovery and inventory for all elements of the system without requiring user configuration, deployment of libraries or sensors, or user scheduling
- ▶ Hypervisors that are shipped, serviced, and deployed as System z LIC; booted automatically at power-on reset and isolated on the internal platform management network
- ▶ Virtual server lifecycle management, enabling directed and dynamic virtual server provisioning across all hypervisors from a single uniform point of control
- ▶ Representation of the physical and virtual resources that are used in the context of a deployed business function as a named workload
- ▶ Monitoring and trend reporting of CPU energy efficiency which can be helpful in managing the costs of deployed workloads
- ▶ Delivery of system activity by using a new user interface, the Monitors Dashboard (which augments the existing System Activity Display), thereby enabling a broader and more granular view of system resource consumption

The Unified Resource Manager offers the ability to optimize technology deployment according to individual workload requirements. To achieve this, the Unified Resource Manager is delivered in two suites of tiered functionality, namely the Manage suite and the Automate suite.

Manage suite

The Manage suite provides the following functionality:

- ▶ Operational Controls
- ▶ Virtual Server provisioning and management
- ▶ Virtual Network Management
- ▶ Hypervisor Management
- ▶ Storage Virtualization Management
- ▶ Energy Controls
- ▶ Energy Monitoring
- ▶ Monitors Dashboard
- ▶ Default Workload Performance Context, Monitoring, and Reporting

The Manage suite provides the interface for defining the components of the ensemble. It provides monitoring capabilities for the ensemble components. It also enables definition of the default workload, and a performance policy for its execution. In addition, it provides the means to monitor the workload against your defined policy objectives.

Although this is provided for a default workload, it is unlikely that all workloads can be managed with only that level of capability. Therefore, the Automate suite should be considered to differentiate between multiple workloads and manage them to separate business objectives.

Automate suite

The Automate suite provides the following functionality:

- ▶ Energy Management
- ▶ Workload Performance Context, Monitoring, and Reporting
- ▶ Performance Management

The Automate suite expands on the capabilities of workloads and performance management delivered in the Manage suite. With the Automate suite you can define your own custom workloads (by name). The performance management capabilities are improved, as well. By creating your own named workload definitions, you can differentiate between multiple workloads in an ensemble. Monitoring and reporting is provided for each workload, and energy management capabilities are added.

4.4 Physical resource management

The Hardware Management Console (HMC) and Support Elements (SEs) are appliances that together provide hardware platform management for System z. Hardware platform management covers a complex set of setup, configuration, operation, monitoring, and service management tasks and services that are essential to the use of the System z hardware.

The HMC allows viewing and managing multinodal servers with virtualization, I/O networks, support networks, power subsystems, cluster connectivity infrastructure, and storage subsystems. The HMC has a management responsibility for the entire ensemble, while the SE has a management responsibility at the node level. When tasks are performed on the HMC, the commands are sent to one or more SEs, which then issue commands to their CPCs and zBXs. This represents a well-layered structure that supports the components of the ensemble.

The Unified Resource Manager suite (Manage or Automate) determines the functions that are available on the HMC. The HMC is used to manage, monitor, and operate one or more nodes configured as members of an ensemble. An ensemble is managed by a primary/alternate HMC pair.

The HMC presents a highly interactive and dynamic web-based user interface. The HMC user interface views, management, and monitoring tasks provide everything needed for complete management of the virtual machine life cycle across the PR/SM, z/VM, and PowerVM (hypervisors), from its inception all the way through monitoring, migration, and policy-based administration during its deployment.

The HMC is the authoritative owning (stateful) component for Unified Resource Manager configuration and policies that have a scope that spans all of the managed nodes in the ensemble. In addition, the HMC will have an active role in ongoing system monitoring and adjustment.

Defining ensembles and new virtual servers and assigning workloads

Typical functions that can be performed from an HMC are not only the ordinary operational start/stop actions on virtual servers, but also include instantiating a new ensemble, defining new virtual servers and workloads, and assigning those virtual servers to one or more workloads. This requires both of the Unified Resource Manager suites (Manage and Automate) to be available, and they are orderable features of the zEnterprise System.

Management examples

With the Scheduled Operations task you can schedule particular performance policy activations, such as policies for day shifts and night shifts or for seasonal peaks. The new Monitors Dashboard provides links to several reports. This allows you to view performance characteristics such as the processor usage of entitled blades, or processor usage of virtual servers. Workload Reports are available that provide information such as met or missed objectives, Service Class Performance Index per Workload, or CPU utilization per Workload.

Event Monitoring allows you to set up triggers (with the option to send notification emails), for example, when a Service Class Performance Index is below a certain threshold.

4.4.1 Serviceability

The serviceability function for the components of the ensemble is delivered through the HMC/SE constructs, as for earlier System z servers. From a serviceability point of view all the components of the ensemble, including the zBX, are treated as System z features, similar to the treatment of I/O and other System z features.

All the functions for managing the components will be delivered as they were delivered on previous System z servers, including management of change, configuration, operations, and performance. The zBX receives all of its serviceability and problem management through the HMC/SE infrastructure, and all service reporting, including call-home functions, will be delivered in a similar fashion.

The physical zBX components are duplicated for redundancy purposes as dictated by System z QoS. The blades are standard blades provided by the customer, or by a solution, depending on the configuration.

There are several possibilities for blade deployment:

- ▶ Blades can be deployed as part of a solution delivered by IBM; for example, the IBM Smart Analytics Optimizer.
- ▶ In addition, blades can be acquired and deployed by the customer. For this solution, IBM will provide a list of blade products that can participate in an ensemble.

4.5 Virtualization management

The purpose of virtualization management is to allow the definition and management of the virtualized resources in the ensemble. Functions to define the virtualization are necessary to create the virtual servers, the virtual network components, and virtual storage volumes.

Functions to manage the hypervisors and other virtual resources are provided by the Unified Resource Manager through the HMC.

4.5.1 Network virtualization

Networking is a pervasive component of an ensemble. Figure 4-4 on page 94 shows a representation of the important networks contained in a single node of an ensemble.

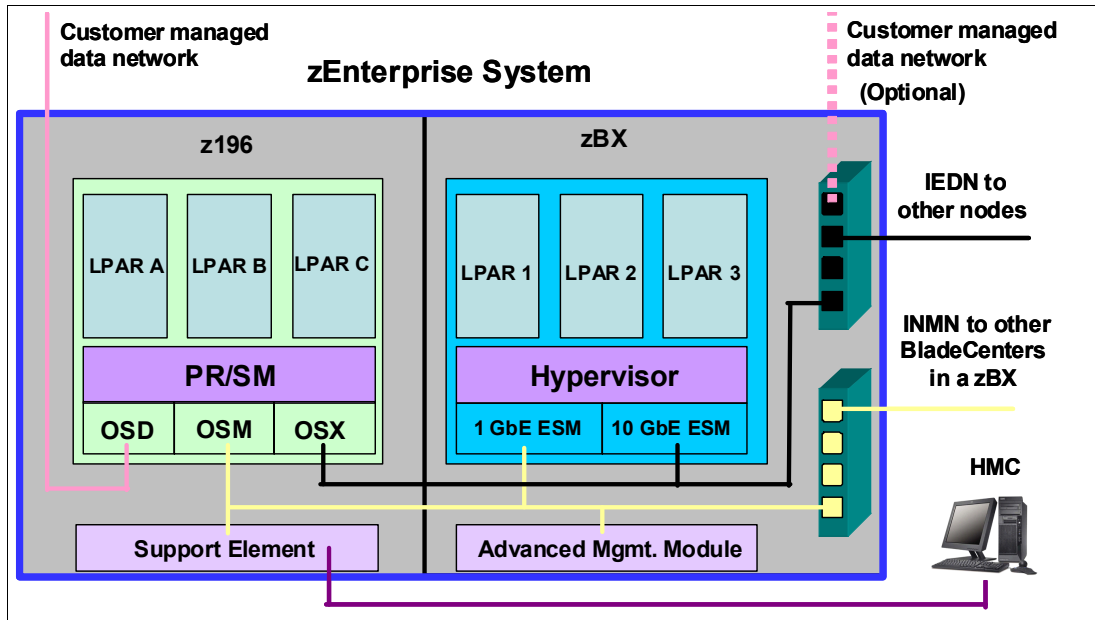


Figure 4-4 Networks contained within an ensemble

- ▶ Intranode management network (INMN): This private internal network provides the connections necessary to monitor and control components of the node, such as virtual servers or physical switches.
- ▶ Intraensemble data network (IEDN): This is the network for system and application data communications within the ensemble. This network connects together all nodes of the ensemble, including all z196 and zBX frames.
- ▶ Customer management network: Also known as the HMC LAN, this network provides the communication link between Hardware Management Consoles (HMCs) and the nodes of the ensemble.
- ▶ Customer managed data network: This network represents the existing enterprise data communication network. In addition, this network may optionally be connected directly to the IEDN, depending on your configuration requirements.

The IEDN is the network used for application communications within an ensemble. It exists only within an ensemble, although it can also be connected to the customer data network outside of the ensemble. It is implemented as a flat Layer 2 network, which means that all the network interfaces can communicate directly with each other as if they were all connected to a single network switch. No routers are necessary to communicate across the IEDN. Although there are physical network switches that are part of the IEDN, the appearance of a single network is maintained through virtualization. Several VLANs can be defined on top of this infrastructure.

The physical construction of the IEDN contributes to the security and reliability of the ensemble. All the network switches are inside the frames of the z196 and zBX frames and all network cables are point-to-point between the frames. With no intervening switches or routers, the opportunity to compromise network integrity is greatly reduced. The switches are managed and configured only from the Unified Resource Manager.

By virtualizing the network definitions, you can isolate the virtual servers from the physical definitions of the network interfaces and devices. This allows the virtual servers to be placed anywhere within the ensemble without changing the network definitions inside the virtual

server. This virtualization also helps you to fully utilize the physical network capacity while still meeting your organization's security requirements.

As mentioned, in some cases it might be appropriate to connect the customer-managed data network to the IEDN. The network configuration tasks allow specific ports on the IEDN TOR switches to be configured for attachment to the existing data network, external to the ensemble, and can impose restrictions on the attaching network.

4.5.2 Hypervisor management

A *hypervisor* is control code that manages multiple independent operating system images. Hypervisors can be implemented in software or hardware.

The Unified Resource Manager works with a set of hypervisors (PR/SM, z/VM, and PowerVM) to support deployment of workloads on the various hardware platforms of an ensemble. These hypervisors virtualize resources and provide basic management functions for virtual servers. Hypervisor management tasks are provided by the firmware installed through the Manage suite. Functions such as deploying and initializing hypervisors, virtual switch management, and monitoring hypervisors are provided.

4.5.3 Virtual server management

A virtual server could be described as a container for the operating system required to support a given workload. Virtual server management is provided by the Manage and Automate suites.

The hypervisor provides virtual resources to the server. When provisioning a virtual environment to support a workload, the relevant platform hypervisors will provision the virtual servers and their associated resources.

The resources that are assigned to a virtual server are the hypervisor, the number of processors, the amount of memory, the network devices, the storage devices and the "boot" options for the operating system.

Virtual server life-cycle management offers directed and dynamic virtual server provisioning across all hypervisors through a single uniform point of control. It includes integrated storage and network configuration and ensemble membership integrity. Functions to create virtual servers, start and stop them, and modify their configuration are also provided.

4.5.4 Storage virtualization

With a zEnterprise BladeCenter Extension (zBX), additional storage connectivity requirements arise. Storage Virtualization Management (SVM) provides the functionality to define virtualized storage to workloads. The zBX will require access to Fibre Channel Protocol (FC) storage (SCSI) disks. FICON connectivity for System z workloads is unchanged.

The Storage Administrator role is responsible for allocating storage from physical storage pools to support an ensemble of virtual servers. The allocation is performed based on input from the Server Administrator role, which provides the workload requirements. The Storage Administrator defines and assigns resources and access rights, and also provides separate storage access lists (SALs) for each hypervisor required to support the ensembles. (A SAL contains the accessible storage resources for a hypervisor). Figure 4-5 illustrates the relationship of the SAL to the hypervisor and virtual server.

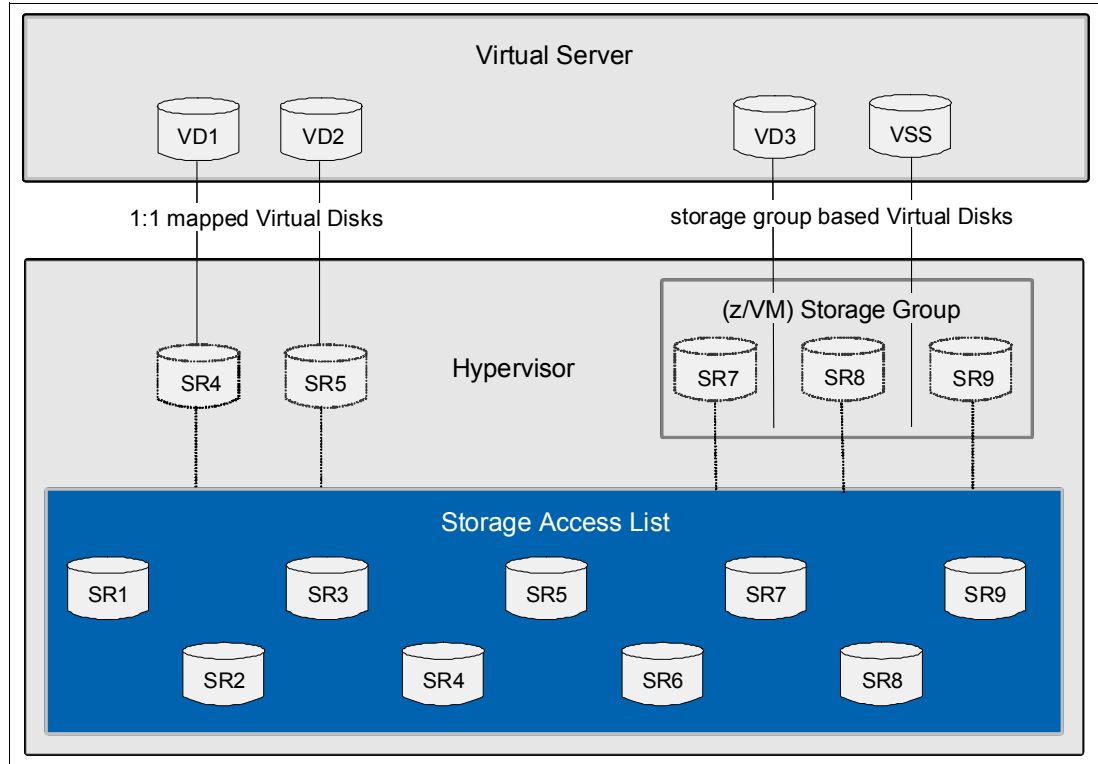


Figure 4-5 Storage virtualization

The Unified Resource Manager provides the abstraction from the underlying technologies. For storage virtualization, this means a simplified storage management interface with common steps across the various hypervisors.

4.6 Performance management

The performance manager is a component responsible for monitoring, managing, and reporting goal-oriented resources across the zEnterprise System. The primary objective of the performance manager is to extend the goal-oriented performance management approach of the z/OS Workload Manager to both System z and to the zEnterprise BladeCenter Extension environments. Workload monitoring and reporting are based on a goal-oriented management policy defined and implemented by the performance manager. It is functionality that is integrated in Unified Resource Manager. The code is structured and packaged as System z firmware, and the intercomponent communication is performed over the trusted intranode management network.

The role of the ensemble performance manager is to:

- ▶ Provide a set of performance monitoring functions that allow an administrator to understand if the performance goals of the deployed workloads are being achieved.

If a performance goal is not being achieved, this monitoring support will help the administrator to understand which virtual servers, partitions, or accelerators are contributing to the performance problem.
- ▶ Dynamically manage resources to achieve workload goals.

- Manage CPU allocations across virtual servers hosted by the same hypervisor instance. This extends today's concept of IRD CPU management function from z/OS to other environments supported by the zEnterprise System.
- Provide recommendations to load balancers on how to distribute incoming work across multiple virtual servers/partitions.

4.7 Energy monitoring

Energy monitoring and management can help you to better understand the power and cooling demands of the zEnterprise System by providing complete monitoring and trending capabilities.

When a workload spans multiple infrastructures, attempting to understand the total energy utilization of all components supporting that workload can be challenging. The Unified Resource Manager has capabilities to monitor power consumption across the ensemble through the Monitors Dashboard.

For more information about the Unified Resource Manager functions, refer to *IBM zEnterprise System Technical Guide*, SG24-7833.

4.8 Technical support services

IBM Global Technology Services (GTS) can help you assess and design a zEnterprise System that aligns IT strategy and business priority. This includes developing the business case and high level transition plan, and a roadmap for an adaptable and efficient infrastructure. GTS can also enable you to build and run a smarter zEnterprise System environment.

With these services, you can migrate effectively and efficiently to a zEnterprise System, create a more cost-effective and manageable computing environment with server, storage, and network optimization, integration, and implementation, and effectively run and manage the zEnterprise System with maintenance and technical support services.

For details about available services for IBM zEnterprise System, contact your IBM representative or visit:

<http://www.ibm.com/services/us/gts/html/services-for-zenterprise.html>

**A**

Operating Systems support and considerations

This appendix contains operating system requirements and support considerations for the z196 and its features.

This chapter discusses the following topics:

- ▶ “Software support summary” on page 100
- ▶ “Support by operating system” on page 103
- ▶ “References” on page 110
- ▶ “Software support for zBX” on page 110
- ▶ “z/OS considerations” on page 110
- ▶ “Coupling Facility and CFCC considerations” on page 113
- ▶ “IOCP considerations” on page 114
- ▶ “ICKDSF considerations” on page 114

Support of the zEnterprise 196 functions is dependent on the operating system version and release. This information is subject to change. Therefore, for the most current information, refer to the Preventive Service Planning (PSP) bucket for 2817DEVICE.

Software support summary

The software portfolio for the z196 server includes a large variety of operating systems and middleware that support the most recent and significant technologies. Continuing the mainframe-rich tradition, five major operating systems are supported on the z196:

- ▶ z/OS
- ▶ z/VM
- ▶ z/VSE
- ▶ z/TPF
- ▶ Linux on System z

For zBX software support see, “Software support for zBX” on page 110.

Operating systems summary

Table A-1 summarizes the current and minimum operating system levels required to support the z196. Note that operating system levels that are no longer in service are not covered in this publication. These older levels may provide support for some features.

Table A-1 z196 operating system requirements

Operating system	ESA/390 (31-bit mode)	z/Architecture (64-bit mode)	End of service	Notes
z/OS V1 R12	No	Yes	Not announced	Refer to the z/OS, z/VM, z/VSE, and z/TPF subsets of the 2817DEVICE Preventative Service Planning (PSP) bucket prior to installing the IBM System z196.
z/OS V1R11	No	Yes	Not announced	
z/OS V1R10	No	Yes	September 2011 ^a	
z/OS V1R9 ^b	No	Yes	September 2010 ^a	
z/OS V1R8 ^c	No	Yes	September 2009 ^c	
z/OS V1R7 ^d	No	Yes	September 2008 ^d	
z/VM V6R1 ^e	No ^f	Yes	April 2013	
z/VM V5R4	No ^f	Yes	September 2013 ^a	
z/VSE V4R2	No ^g	Yes ^h	Not announced	
z/TPF V1R1	Yes	Yes	Not announced	
Linux on System z	See Table A-5 on page 108.	See Table A-5 on page 108.	See footnote ⁱ	Novell SUSE SLES 11 Novell SUSE SLES 10 Red Hat RHEL 5

a. Planned date. All statements regarding IBM plans, directions, and intent are subject to change or withdrawal without notice. Any reliance on these Statements of General Direction is at the relying party's sole risk and will not create liability or obligation for IBM.

b. With the announcement of IBM Lifecycle Extension for z/OS V1.9 fee-based corrective service can be ordered for up to two years after the withdrawal of service for z/OS V1R9.

c. With the announcement of IBM Lifecycle Extension for z/OS V1.8 fee-based corrective service can be ordered for up to two years after the withdrawal of service for z/OS V1R8.

d. With the announcement of IBM Lifecycle Extension for z/OS V1.7 fee-based corrective service can be ordered for up to two years after the withdrawal of service for z/OS V1R7.

e. z/VM V6R1 requires an architectural level set exclusive to z10 and z196.

f. z/VM supports both ESA/390 mode and z/Architecture mode virtual machines.

g. ESA/390 is not supported. However, 31-bit mode is supported.

- h. z/VSE V4R2 and later support 64-bit real addressing only. They do not support 64-bit addressing for user, system, or vendor applications.
- i. For information about support-availability of Linux on System z distributions, see:
Novell SUSE:
<http://support.novell.com/lifecycle/lcSearchResults.jsp?st=Linux+Enterprise+Server&x=32&y=11&sl=-1&sg=-1&pid=1000>
Red Hat:
<http://www.redhat.com/security/updates/errata/>

Note: Exploitation of several features depends on a particular operating system. In all cases, PTFs might be necessary with the operating system level indicated. PSP buckets are continuously updated and should be reviewed regularly when planning for installation of a new server. They contain the latest information about maintenance.

PSP buckets contain installation information, hardware and software service levels, service recommendations, and cross-product dependencies.

Middleware

Middleware offerings for the z196 environments include:

- ▶ Transaction processing
 - WebSphere Application Server and WebSphere Extended Deployment
 - CICS Transaction Server
 - CICS Transaction Gateway
 - IMS DB and IMS DC
 - IMS Connect
- ▶ Application integration and connectivity
 - WebSphere Message Broker
 - WebSphere MQ
 - WebSphere ESB
- ▶ Process integration
- ▶ WebSphere Process Server
- ▶ WebSphere MQ Workflow
- ▶ WebSphere Business Integration Server
- ▶ Database
 - ▶ DB2 for z/OS
 - ▶ DB2 for Linux
 - ▶ DB2 Connect™

Operations

The Tivoli® brand has a large product set that includes:

- ▶ Tivoli Service Management Center
- ▶ Tivoli Information Management for z/OS
- ▶ Tivoli Workload Scheduler
- ▶ Tivoli OMEGAMON® XE
- ▶ Tivoli System Automation

Security

A highly secure System z environment can be implemented at various levels using the following products:

- ▶ Security Server feature of z/OS (includes Resource Access Control Facility (RACF) and LDAP server)
- ▶ Tivoli Access Manager
- ▶ Tivoli Federated Identity Manager
- ▶ z/OS Communications Server and Policy Agent (for policy-based network security)

Application development and languages

Many languages are available for the z196 environments. Because the Linux environment is similar to Linux on other servers, we focus on the z/OS environment.

In addition to the traditional COBOL, PL/I, FORTRAN, and Assembler languages, C, C++, and Java, including J2EE and batch environments, are available.

Development can be conducted using the latest software engineering technologies and advanced IDEs. The extensive tool set uses a workstation environment for development and testing, with final testing and deployment performed on z/OS. Application development tools, many of which have components based on the Eclipse platform, include:

- ▶ Rational® Application Developer for WebSphere
- ▶ Rational Developer for System z
- ▶ WebSphere developer for System z
- ▶ Rational Rose® product line
- ▶ Rational Software Architect and Software Modeler

The following Web site is organized by category and has an extensive set of links to information about software for System z:

<http://www-306.ibm.com/software/sw-bycategory/systemz>

IBM compilers

Each new version of IBM z/OS compilers (Enterprise COBOL, Enterprise PL/I, XL C/C++) underscores the continuing IBM commitment to the COBOL, PL/I, and C/C++ programming languages on the z/OS platform.

The latest version of Enterprise COBOL delivers enhanced XML parsing support, facilitates compiler message severity customization, exploits system-determined block size for QSAM files, supports the underscore (_) character in COBOL user-defined words, provides compiler listings that display CICS options in effect, and supports Java 5 and Java 6 SDKs for Java interoperability.

The latest version of Enterprise PL/I delivers enhanced XML parsing support, exploits the latest z/Architecture® for application performance improvements, improves application debugging with the IBM Debug Tool through compiler enhancements, improves SQL preprocessing, and leverages productivity with new programming features.

The latest version of z/OS XL C/C++ delivers application performance improvements by exploiting the latest advancements in optimization and hardware technology, leverages system programming capabilities with METAL C, matches the behavior of interprocedural analysis (IPA) on other platforms, provides new compiler options for deeper pointer analysis and message severity customization, and reduces application development effort with new compiler suboptions, macros and pragma directives.

IBM Enterprise COBOL and Enterprise PL/I support are strategic components (separately orderable products) for IBM Rational® Developer for IBM System z software-providing a robust, integrated development environment (IDE) for COBOL and PL/I and connecting web services; Java Platform, Enterprise Edition (Java EE) applications; and traditional business processes.

z/OS XL C/C++ programmers can also tap into Rational Developer for System z to boost their productivity by easily editing, compiling and debugging z/OS XL C and XL C++ applications right from their workstation.

Support by operating system

In this section we list the support of new System z196 functions by the current operating systems. See the companion manual *IBM zEnterprise System Technical Guide*, SG24-7833, for a detailed description of z196 and its features. For an in-depth description of all I/O features refer to the *IBM System z Connectivity Handbook*, SG24-5444.

z/OS

z/OS Version 1 Release 9 is the earliest service release supporting the z196. Although service support for z/OS Version 1 Release 8 ended in September of 2009, a fee-based extension for defect support (for up to two years) can be obtained by ordering the IBM Lifecycle Extension for z/OS V1.8. Similarly, IBM Lifecycle Extension for z/OS V1.7 provides fee-based support for z/OS V1.7 up to September 2010. Service support for z/OS Version 1 Release 6 ended on September 30, 2007. Also note that z/OS.e is not supported on z196 and that the last release of z/OS.e was z/OS.e Version 1 Release 8.

Table A-2 on page 104 summarizes the z196 function support requirements for the currently supported z/OS releases. It uses the following conventions:

- Y** The function is supported.
- N** The function is not supported.

Table A-2 z/OS support summary

Function	V1R12	V1R11	V1R10	V1R9	V1R8 ^a	V1R7 ^a
z196	Y	Y	Y	Y	Y	Y
Greater than 54 PUs single system image	Y	Y	Y	Y	N	N
Dynamic add of logical CPs	Y	Y	Y	N	N	N
zAAP on zIIP	Y	Y	Y ^d	Y ^d	N	N
Large memory > 128 GB	Y	Y	Y	Y	Y ^d	N
Large page support	Y	Y	Y ^d	Y ^d	N	N
Hardware decimal floating point	Y ^b	Y ^b	Y ^b	Y ^{bd}	Y ^{bd}	Y ^{bd}
CPACF protected public key	Y ^c	Y ^c	Y ^c	Y ^c	N	N
Enhanced CPACF	Y	Y	Y	Y ^c	Y ^c	Y ^c
Personal account numbers of 13 to 19 digits	Y ^c	Y ^c	Y ^c	Y ^c	Y ^c	Y ^c
Crypto Express3	Y ^c	Y ^c	Y ^c	Y ^c	N	N
Capacity Provisioning Manager	Y	Y	Y ^d	Y ^d	N	N
HiperDispatch	Y	Y	Y ^d	Y ^d	Y ^d	Y ^{de}
HiperSockets multiple write facility	Y	Y	Y	Y ^d	N	N
High Performance FICON	Y	Y	Y ^d	Y ^d	Y ^d	Y ^d
FICON Express8	Y ^f	Y ^f	Y ^f	Y ^f	Y ^f	Y ^f
OSA-Express3 10 Gigabit Ethernet LR CHPID type OSD	Y	Y	Y	Y	Y	Y
OSA-Express3 10 Gigabit Ethernet SR CHPID type OSD	Y	Y	Y	Y	Y	Y
OSA-Express3 Gigabit Ethernet LX using four ports CHPID types OSD and OSN	Y	Y	Y	Y ^d	Y ^d	N
OSA-Express3 Gigabit Ethernet LX using two ports CHPID types OSD and OSN	Y	Y	Y	Y	Y	Y
OSA-Express3 Gigabit Ethernet SX using four ports CHPID types OSD and OSN	Y	Y	Y	Y ^d	Y ^d	N

Function	V1R12	V1R11	V1R10	V1R9	V1R8 ^a	V1R7 ^a
OSA-Express3 Gigabit Ethernet SX using two ports CHPID types OSD and OSN	Y	Y	Y	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet using four ports CHPID types OSC, OSD, and OSN ^g	Y	Y	Y	Y ^d	Y ^d	N
OSA-Express3 1000BASE-T Ethernet using two ports CHPID types OSC, OSD, OSE and OSN ^g	Y	Y	Y	Y	Y	Y ^d
Coupling using InfiniBand CHPID type CIB	Y	Y	Y	Y	Y	Y
InfiniBand coupling links (12x IB-SDR or 12x IB-DDR) at a distance of 150 m	Y	Y	Y	Y	Y	Y ^d
InfiniBand coupling links (1x IB-SDR or 1x IB-DDR) at an unrepeated distance of 10 km	Y	Y	Y ^d	Y ^d	Y ^d	Y ^d
CFCC Level 16	Y	Y	Y ^d	Y ^d	Y ^d	Y ^d
Assembler instruction mnemonics	Y	Y	Y	Y ^d	Y ^d	Y ^d
C/C++ exploitation of hardware instructions	Y	Y	Y	Y ^d	Y ^d	N
Layer 3 VMAC	Y	Y	Y	Y	Y ^d	N
Large dumps	Y	Y	Y	Y ^d	Y ^d	N
CPU measurement facility	Y	Y	Y	Y ^d	Y ^d	N

- a. With the announcement of IBM Lifecycle Extension for z/OS V1.8, fee-based corrective service can be ordered for up to two years after the withdrawal of service for z/OS V1R8. Similarly, IBM Lifecycle Extension for z/OS V1.7 provides fee-based support for z/OS V1.7 up to September 2010.
- b. The level of decimal floating-point exploitation varies with z/OS release and PTF level.
- c. FMIDs are shipped in a Web deliverable.
- d. PTFs are required.
- e. Requires Web deliverable support for zIIP.
- f. Support varies with operating system and level.
- g. CHPID type OSN does not use ports. LPAR-to-LPAR communication is used.

z/VM

At general availability, z/VM V6R1 provides compatibility exploitation support of some features and z/VM V5R4 provides compatibility support only. Table A-3 lists the z196 functions currently supported for z/VM releases. It uses the following conventions:

- Y** The function is supported.
N The function is not supported.

Table A-3 z/VM support summary

Function	V6R1	V5R4
z196	Y	Y
Greater than 32 PUs for single system image	N ^b	N ^b
Dynamic add of logical CPs	Y	Y
zAAP on zIIP	Y ^a	Y ^a
Large memory > 128 GB	Y ^b	Y ^b

Function	V6R1	V5R4
Large page support	N ^c	N ^c
Hardware decimal floating point	Y ^d	Y ^d
CPACF protected public key	N ^c	N ^c
Enhanced CPACF	Y	Y ^e
Personal account numbers of 13 to 19 digits	Y ^d	Y ^d
Crypto Express3	Y ^d	Y ^d
Execute relative guest exploitation	Y ^d	Y ^d
Capacity provisioning	N ^c	N ^c
HiperDispatch	N ^c	N ^c
Restore subchannel facility	Y	Y
HiperSockets multiple write facility	N ^c	N ^c
High Performance FICON	N ^c	N ^c
FICON Express8	Y ^f	Y ^f
OSA-Express QDIO data connection isolation for z/VM environments	Y	Y ^e
OSA-Express3 10 Gigabit Ethernet LR CHPID type OSD	Y	Y
OSA-Express3 10 Gigabit Ethernet SR CHPID type OSD	Y	Y
OSA-Express3 Gigabit Ethernet LX using four ports CHPID types OSD and OSN	Y	Y
OSA-Express3 Gigabit Ethernet LX using two ports CHPID types OSD and OSN	Y	Y
OSA-Express3 Gigabit Ethernet SX using four ports CHPID types OSD and OSN	Y	Y
OSA-Express3 Gigabit Ethernet SX using two ports CHPID types OSD and OSN	Y	Y
OSA-Express3 1000BASE-T Ethernet using four ports CHPID types OSC, OSD ^g , OSE, and OSN ^h	Y	Y
OSA-Express3 1000BASE-T Ethernet using two ports CHPID types OSC, OSD, OSE, and OSN ^h	Y	Y
Dynamic I/O support for InfiniBand CHPIDs	Y	Y
InfiniBand coupling links (1x IB-SDR or 1x IB-DDR) at an unrepeated distance of 10 km	N	N
CFCC Level 16	Y ^d	Y ^d

a. Available for z/OS on virtual machines without virtual zAAPs defined when the z/VM LPAR does not have zAAPs defined.

b. 256 GB of central memory are supported by z/VM V5R3 and later. z/VM V5R3 and later support more than 1 TB of virtual memory in use for guests.

- c. Not available to guests.
- d. Supported for guest use only.
- e. PTFs are required.
- f. Support varies with operating system and level.
- g. PTFs are required for CHPID type OSD.
- h. CHPID type OSN does not use ports, it uses LPAR-to-LPAR communication.

Notes: We recommend that the capacity of z/VM logical partitions and any guests, in terms of the number of IFLs and CPs, real or virtual, be adjusted in face of the PU capacity of the z196.

z/VSE

Table A-4 lists z196 support requirements for the currently supported z/VSE releases. It uses the following conventions:

- Y** The function is supported.
- N** The function is not supported.

Table A-4 z/VSE support summary

Function	V4R2	V4R1
z196	Y ^a	Y ^a
CPACF protected public key	N	N
Enhanced CPACF	Y	Y
Crypto Express3	Y ^b	N
FICON Express8	Y ^c	Y ^c
OSA-Express3 10 Gigabit Ethernet LR CHPID type OSD	Y	Y
OSA-Express3 10 Gigabit Ethernet SR CHPID type OSD	Y	Y
OSA-Express3 Gigabit Ethernet LX using four ports CHPID types OSD	Y ^e	Y ^e
OSA-Express3 Gigabit Ethernet LX using two ports CHPID types OSD and OSN	Y	Y
OSA-Express3 Gigabit Ethernet SX using four ports CHPID types OSD	Y ^e	Y ^e
OSA-Express3 Gigabit Ethernet SX using two ports CHPID types OSD and OSN	Y	Y
OSA-Express3 1000BASE-T Ethernet using four ports CHPID types OSC, OSD ^e , OSE, and OSN ^d	Y	Y ^e
OSA-Express3 1000BASE-T Ethernet using two ports CHPID types OSC, OSD ^e , OSE, and OSN ^d	Y	Y

- a. z/VSE V4 is designed to exploit z/Architecture, specifically 64-bit real-memory addressing, but does not support 64-bit virtual memory addressing.
- b. PTFs are required.
- c. Support varies with operating system and level.

- d. CHPID type OSN does not use ports. All communication is LPAR to LPAR.
- e. Exploitation of two ports per CPHID type OSD requires a minimum of z/VSE V4R1 with PTFs.

Linux on System z

Linux on System z distributions are built separately for the 31-bit and 64-bit addressing modes of the z/Architecture. The newer distribution versions are built for 64-bit only. You can run 31-bit applications in the 31-bit emulation layer on a 64-bit Linux on System z distribution.

None of the current versions of Linux on System z distributions (Novell SUSE SLES 10, SLES 11, and Red Hat RHEL 5) require z196 toleration support, so that any release of these distributions can run on z196 servers.

Table A-5 lists the most recent service levels of the current SUSE and Red Hat releases at the time of writing.

Table A-5 Current Linux on System z distributions as of October 2009, by z/Architecture mode

Linux distribution	ESA/390 (31-bit mode)	z/Architecture (64-bit mode)
Novell SUSE SLES 11	No	Yes
Novell SUSE SLES 10 SP3	No	Yes
Red Hat RHEL 5.4	No	Yes

Table A-6 lists selected z196 features, showing the minimum level of Novell SUSE and Red Hat distributions that support each feature.

Table A-6 Linux on System z support summary

Function	Novell SUSE	Red Hat
z196	SLES 10	RHEL 5
Large page support	SLES 10 SP2	RHEL 5.3
Hardware decimal floating point	SLES 11	N ^b
CPACF protect public key	N	N
Enhanced CPACF	SLES 10 SP2	RHEL 5.3
Crypto Express3	SLES 10 SP3 ^a	RHEL 5.4 ^a
HiperSockets Layer 2 support	SLES 10 SP2	RHEL 5.3
FICON Express8	SLES 10	RHEL 5
High Performance FICON	Note ^b	Note ^b
FICON Express4 ^b CHPID type FCP	SLES 10	RHEL 5
OSA-Express3 using four ports CHPID type OSD	SLES 10	RHEL 5.2

Function	Novell SUSE	Red Hat
OSA-Express3 using two ports CHPID type OSD	SLES 10 SP2	RHEL 5
OSA-Express3 CHPID type OSN	SLES 10	RHEL 5

- a. Toleration support only.
- b. FICON Express4 10KM LX, 4KM LX, and SX features are withdrawn from marketing. All FICON Express2 and FICON features are withdrawn from marketing

IBM is working with its Linux distribution partners so that exploitation of further z196 functions will be provided in future Linux on System z distribution releases. We recommend that:

- ▶ SUSE SLES 11 or Red Hat RHEL 5 be used in any new projects for the z196
- ▶ Any Linux distributions be updated to their latest service level before migration to z196
- ▶ The capacity of any z/VM and Linux logical partitions, as well as any z/VM guests, in terms of the number of IFLs and CPs, real or virtual, be adjusted in face of the PU capacity of the z196

z/TPF

Table A-7 lists the z196 function' support requirements for the currently supported z/TPF release. It uses the following conventions:

- Y** The function is supported.
- N** The function is not supported.

Table A-7 TPF and z/TPF support summary

Function	z/TPF V1R1
z196	Y
Greater than 54 PUs for single system image	Y
Large memory > 128 GB (4 TB)	Y
CPACF protected public key	N
Enhanced CPACF	Y
Crypto Express3 (accelerator mode only)	Y
HiperDispatch	N
FICON Express8	Y
OSA-Express3 10 Gigabit Ethernet LR CHPID type OSD	Y
OSA-Express3 10 Gigabit Ethernet SR CHPID type OSD	Y
OSA-Express3 Gigabit Ethernet LX using four ports CHPID types OSD and OSN	Y
OSA-Express3 Gigabit Ethernet LX using two ports CHPID types OSD and OSN	Y
OSA-Express3 Gigabit Ethernet SX using four ports CHPID types OSD and OSN	Y

Function	z/TPF V1R1
OSA-Express3 Gigabit Ethernet SX using two ports CHPID types OSD and OSN	Y
OSA-Express3 1000BASE-T Ethernet using four ports CHPID types OSC, OSD, and OSN	Y
OSA-Express3 1000BASE-T Ethernet using two ports CHPID types OSC, OSD, and OSN	Y
Coupling over InfiniBand CHPID type CIB	Y ^a
CFCC Level 16	Y

a. Compatibility is supported.

Software support for zBX

The zEnterprise BladeCenter Extension offers two types of application environments:

- ▶ Special purpose, dedicated environments such as the IBM Smart Analytics Optimizer. In this case support is dictated by the solution.
- ▶ POWER7 blades. These blades offer a virtualized environment. AIX is supported in the blades's virtual servers.

References

Planning information for each operating system is available on the following support Web pages:

- ▶ z/OS
<http://www.ibm.com/systems/support/z/zos>
- ▶ z/VM
<http://www.ibm.com/systems/support/z/zvm>
- ▶ z/TPF
<http://www.ibm.com/tpf/maint/supportgeneral.htm>
- ▶ z/VSE
<http://www.ibm.com/servers/eserver/zseries/zvse/>
- ▶ Linux on System z
<http://www.ibm.com/systems/z/os/linux>

z/OS considerations

z196 base processor support is required in z/OS. With that exception, software changes do not require the new z196 functions and, equally, the new functions do not require functional software. The approach has been to, where applicable, automatically decide to enable or

disable a function based on, respectively, the presence or absence of the required hardware and software.

General recommendations

The z196 introduces the latest System z technology. Although support is provided by z/OS starting with z/OS V1R7, exploitation of z196 is dependent on the z/OS release. z/OS.e is *not* supported on z196.

In general, we recommend that you:

- ▶ Do not migrate software releases and hardware at the same time.
- ▶ Keep members of the sysplex at the same software level other than during brief migration periods.
- ▶ Review z196 restrictions and considerations prior to creating an upgrade plan.

zAAP on zIIP capability

This new capability, first made available on System z9 servers, enables, under defined circumstances, workloads eligible to run on Application Assist Processors (zAAPs) to run on Integrated Information Processors (zIIP). It is intended as a means to optimize the investment on existing zIIPs, not as a replacement for zAAPs. The rule of at least one CP installed per zAAP and zIIP installed still applies. Exploitation of this capability is by z/OS only, and is only available in these situations:

- ▶ When there are zIIPs but no zAAPs installed in the server.
- ▶ When z/OS is running as a guest of z/VM V5R4 or later, and there are no zAAPs defined to the z/VM LPAR. The server may have zAAPs installed. Because z/VM can dispatch both virtual zAAPs and virtual zIIPs on real CPs¹, the z/VM partition does not require any real zIIPs defined to it, although we recommend the use of real zIIPs due to software licensing reasons.

HCD

When using HCD on z/OS V1R6 to create a definition for z196, *all* subchannel sets must be defined or the VALIDATE will fail. On z/OS V1R7, HCD or HCM will assist in the definitions.

Large page support

Memory reserved for large page support can be defined with the following new parameter in the IEASYSxx member of SYS1.PARMLIB:

```
LFAREA=xx%|xxxxxxM|xxxxxxG
```

This parameter *cannot* be changed dynamically.

HiperDispatch

There is a new HIPERDISPATCH=YES/NO parameter in the IEAOPTxx member of SYS1.PARMLIB and on the SET OPT=xx command to control whether HiperDispatch is enabled or disabled for a z/OS image. It can be changed dynamically (without an IPL or any outage).

To effectively exploit HiperDispatch, adjustment of defined WLM goals and policies may be required. We recommend that you review WLM policies and goals and update them as necessary. You may want to run with the new policies and HiperDispatch on for a period, turn it off and use the older WLM policies while analyzing the results of using HiperDispatch,

¹ The z/VM system administrator can use the SET CPUAFFINITY command to influence the dispatching of virtual specialty engines on CPs or real specialty engines.

re-adjust the new policies, and repeat the cycle as needed. In order to change WLM policies, turning HiperDispatch off then on is not necessary.

A health check is provided to verify whether HiperDispatch is enabled on a z196 system.

Capacity provisioning

Installation of the capacity provision function on z/OS requires:

- ▶ Setting up and customizing z/OS RMF, including the Distributed Data Server (DDS)
- ▶ Setting up the z/OS CIM Server (a z/OS base element with z/OS V1R9)
- ▶ Performing capacity provisioning customization as described in the *z/OS MVS Capacity Provisioning User's Guide*, SA33-8299

Exploitation of the capacity provisioning function requires:

- ▶ TCP/IP connectivity to observed systems
- ▶ TCP/IP connectivity from the observing system to the HMC of observed systems
- ▶ RMF Distributed Data Server must be active
- ▶ CIM Server must be active
- ▶ Security and CIM customization
- ▶ Capacity Provisioning Manager customization

In addition, the Capacity Provisioning Control Center must be downloaded from the host and installed on a PC workstation. This application is only used to define policies. It is not required to manage operations.

Customization of the capacity provisioning function is required on the operating system that will observe other z/OS systems in one or multiple sysplexes. For a description of the capacity provisioning domain refer to the *z/OS MVS Capacity Provisioning User's Guide*, SA33-8299. Also see *IBM System z10 Enterprise Class Capacity on Demand*, SG24-7504, which discusses capacity provisioning in more detail.

ICSF

Integrated Cryptographic Service Facility (ICSF) is a base component of z/OS and is designed to transparently use the available cryptographic functions, whether CPACF or Crypto Express features, to balance the workload and help address the bandwidth requirements of the customer's applications.

Specific support is available as a Web download for z/OS V1R7 (FMID HCR7730) and for z/OS V1R8 (FMID HCR7731) in support of the cryptographic coprocessor and accelerator functions as well as the CPACF AES, PRNG, and SHA support. The z/OS V1R9 has this support (FMID HCR7740) integrated in the base, so no download is necessary.

For support of the SHA-384 and SHA-512 function on z/OS V1R7 and later, download and installation of FMID HCR7750 is required.

Support for the most recent functions, which include Secure Key AES, new Crypto Query Service, enhanced IPv6 support, and enhanced SAF Checking and Personal Account Numbers with 13 to 19 digits, is provided by FMID HCR7751, which is available for z/OS V1R8 and later.

Support for the Crypto Express3, Crypto Express3-1P, and CPACF protected key is provided for z/OS V1R9 and later by FMID HCR7770. Planned availability for FMID HCR7770 is November 2009.

ICSF web deliverables

Consider the following points regarding the version of Web-delivered ICSF code:

- ▶ Increased size of the PKDS file: This is required to allow 4096-bit RSA keys to be stored. If you use the PKDS for asymmetric keys you must copy your PKDS to a larger VSAM data set before using the new version of ICSF. The ICSF options file must be updated with the name of the new data set. ICSF can then be started.

A toleration PTF must be installed on any system that is sharing the PKDS with a system running HCR7750 ICSF. The PTF allows the PKDS to be larger and prevents any service from accessing 4096-bit keys stored in a HCR7750 PKDS.

- ▶ Reduced support for retained private keys. Applications that make use of the retained private key capability for key management will no longer be able to store the private key in the crypto coprocessor card. The applications will continue to be able to list the retained keys and to delete them from the crypto coprocessor cards.

InfiniBand coupling links

Each system can use, or not use, InfiniBand coupling links independently of what other systems are doing, and do so in conjunction with other link types.

InfiniBand coupling connectivity is only available when other systems also support InfiniBand coupling. We recommend that you consult the *Coupling Facility Configuration Options* white paper when planning to exploit the InfiniBand coupling technology, available at:

<http://www.ibm.com/systems/z/advantages/pso/whitepaper.html>

Decimal floating point (z/OS XL C/C++ considerations)

The two new options for the C/C++ compiler are ARCHITECTURE and TUNE. They require z/OS V1R9.

The ARCHITECTURE C/C++ compiler option selects the minimum level of machine architecture on which your program will run. Note that certain features provided by the compiler require a minimum architecture level. ARCH(8) exploits instructions available on the z196.

The TUNE compiler option allows optimization of the application for a specific machine architecture within the constraints imposed by the ARCHITECTURE option. The TUNE level must not be lower than the setting in the ARCHITECTURE option.

For more information about the ARCHITECTURE and TUNE compiler options refer to the *z/OS V1R9.0 XL C/C++ User's Guide*, SC09-4767. See the Authorized Program Analysis Report, APAR PK60051, which provides guidance to installation of the z/OS V1.9 XL C/C++ compiler on a z/OS V1.8 system.

Note: A C/C++ program compiled with the ARCH() or TUNE() options run only on z196 servers, otherwise an operation exception will result. This is a consideration for programs that might have to run on different level servers during development, test, production, and fallback or DR.

Coupling Facility and CFCC considerations

Coupling Facility connectivity to a z196 server is supported on the z10 BC, z9 EC, z9 BC, or another z196 server. The logical partition running the Coupling Facility Control Code (CFCC) can reside on any of the supported servers previously listed.

Because coupling link connectivity to z990, z890, and previous servers is *not* supported, this might affect the introduction of z196 into existing installations and require additional planning. For more information refer to the *IBM zEnterprise System Technical Guide*, SG24-7833.

z196 servers support CFCC Level 17. To support migration from one CFCC level to the next, different levels of CFCC can be run concurrently as long as the Coupling Facility logical partitions are running on different servers. (CF logical partitions running on the same server share the same CFCC level.)

For additional details about CFCC code levels, see the Parallel Sysplex Web site at:

<http://www.ibm.com/systems/z/psocftable.html>

IOCP considerations

All System z servers require a description of their I/O configuration. This description is stored in input/output configuration data set (IOCDS) files. The input/output configuration program (IOCP) allows creation of the IOCDS file from a source file known as the input/output configuration source (IOCS).

The IOCS file contains detailed information for each channel and path assignment, each control unit, and each device in the configuration.

The required level of IOCP for the z196 is V2 R1 L0 (IOCP 2.1.0). See the *Input/Output Configuration Program User's Guide*, SB10-7037, for details.

ICKDSF considerations

The ICKDSF Release 17 device support facility is required on all systems that share disk subsystems with a z196 server.

ICKDSF supports a modified format of the CPU information field, which contains a 2-digit logical partition identifier. ICKDSF uses the CPU information field instead of CCW reserve/release for concurrent media maintenance. It prevents multiple systems from running ICKDSF on the same volume, and at the same time allows user applications to run while ICKDSF is processing. In order to prevent any possible data corruption, ICKDSF must be able to determine all sharing systems that can potentially run ICKDSF. Therefore, this support is required for the z196.

Important: The need for ICKDSF Release 17 applies even to systems that are not part of the same sysplex or that are running other than the z/OS operating system, such as z/VM.

**B**

Frequently asked questions

Q: What is System z?

A: IBM System z is a brand name for IBM mainframe computers. It is the line of computers that started in 1964 with S/360 and evolved over the decades. It still preserves backward compatibility with previous systems while bringing new features and technologies.

Q: What did IBM announce on July 22, 2010 for IBM System z?

A: On July 22nd, IBM introduced the IBM zEnterprise System - a system that combines unprecedented innovations to the gold standard of enterprise computing with new built-in functions that extend IBM's leading mainframe-like governance and qualities of service even further to special-purpose workload optimizers and general-purpose application serving blades, to simplify operations across all these application environments. IBM also introduces new end-to-end management of this heterogeneous environment.

The IBM zEnterprise System includes the IBM zEnterprise 196 (z196), the IBM zEnterprise Unified Resource Manager, the IBM zEnterprise BladeCenter Extension (zBX), and integrated optimizers and/or IBM blades. IBM also announced strategic software and services innovations for the IBM zEnterprise System that help many more customers solve critical problems that were impossible or unaffordable to solve with any previous technologies, in areas such as business analytics and insight, accelerating new business function delivery, auditing and risk reduction, and business process optimization, among others. The Unified Resource Manager provides energy monitoring and management, goal-oriented policy management, increased security, virtual networking, and information management, all consolidated into a single easy-to-use interface firmly grounded in real-world business requirements.

Q: What part does the IBM zEnterprise System play in a “Smarter Planet”?

A: Businesses and governments need smarter systems and software for enterprise computing and for robust cloud environments. They must be able to unify and optimize multiple systems to work as a single, integrated service delivery platform to address real-world business problems in real time. They need to scale without adding complexity to meet ever-growing demands on the infrastructure. They need to offer vastly simplified data center management that slashes space, power, and cooling requirements to respect both our planet and our budgets. They need to move well beyond business-as-usual to transform IT into the leading catalyst for continuous business growth and innovation. And they must do all that while improving service qualities: eliminating interruptions for any reasons (including

software and application upgrades), preventing security breaches, assuring customer privacy, and reducing enterprise risk.

With these accomplishments we can all live on a "Smarter Planet." IBM is announcing the smartest technologies ever created that will help customers meet and exceed these challenges: the new IBM zEnterprise System - a new dimension in computing. The zEnterprise System is a "System of Systems," integrating IBM's leading technologies to dramatically improve the productivity of today's multi-architecture data centers and tomorrow's clouds. The z196 is the world's fastest and most scalable enterprise system with unrivaled reliability, security, and manageability as well as the industry's most efficient platform for large-scale data center simplification and consolidation.

Q: What are the three layers of management function that can be used to show the architectural construct of hardware, software and services?

A: The new management functions IBM is introducing fit into three categories, or layers. Each layer enhances the business value of workloads whether they are running in one or in some combination of application environments. All the layers are transparent to the applications and do not require application programming or other special action to exploit. The three layers are:

1. Hardware management – Capabilities designed to discover, configure, virtualize and manage the basic system hardware and networking resources of multiple processor families from a single, consistent point of control.
2. Platform management - Capabilities designed to manage the lifecycle and operational characteristics of virtualized runtime environments in support of several application architectures, driven by a workload context that is independent of the underlying architectures.
3. Service management – Capabilities designed to align the management of IT with business goals, provide IT service management and automation, incorporate software and process management, handle multi-site IT operations and multi-ensemble management scope.

Q: Is IBM positioning the zEnterprise System to replace all servers in the entire data center? Should I consider moving all my applications that run on UNIX® to the zEnterprise?

A: Enterprise customers with one or more applications that are currently deployed on a complex, heterogeneous, multi-tiered environment, including z/OS, now have the opportunity to upgrade that infrastructure with zEnterprise and reap the management benefits that the Unified Resource Manager brings. The zEnterprise BladeCenter Extension supports scores of blades running hundreds or even a thousand virtual servers, and the zEnterprise 196 delivers unprecedented opportunities for enterprise simplification that build upon the well-known strengths of System z.

Many of the largest data centers already have far more blades and/or rack-mounted servers in their inventory that could realistically fit into the zBX. However, IBM expects that current System z customers will (and should) start to bring particular end-to-end enterprise applications onto zEnterprise, particularly those where there are affinities between the application components and System z-based applications and information. Customers can then manage these end-to-end applications in common ways and achieve higher service qualities and reduced costs. Also, IBM is targeting new zEnterprise customers that may never have had System z in the past. They, too, will benefit from zEnterprise's end-to-end application management as they move their end-to-end application components onto the zBX and the z196. For both new and existing customers, zEnterprise represents a breakthrough in IT simplification and, combined with IBM's other server offerings, the most potent and sophisticated data center consolidation and simplification portfolio in the industry for even the world's largest data centers

Q: What kind of cloud computing options do I have with zEnterprise?

A: IBM System z has provided superior levels of cloud infrastructure support for decades.

This includes extreme levels of resource sharing, sophisticated virtualization technology, rapid provisioning of virtual servers and applications, highest levels of system availability and efficient operational and management support. With z/VM and zEnterprise, clients can host virtual servers for less than \$1 a day when hosting workloads that achieve high levels of consolidation and scale. IBM Tivoli products like Tivoli Provisioning Manager, Tivoli Service Automation Manager, and OMEGAMON XE for z/VM and Linux provide even greater levels of automation, control, and service management, helping clients increase the productivity of their staff and improve the quality of service offered by a zEnterprise cloud. Also, hosting virtual Linux servers on zEnterprise IFLs is a very efficient and reliable way to leverage the industry-leading data serving capabilities of z/OS within a cloud infrastructure.

IBM offers the System z Solution Edition for Cloud Computing and the Smart Analytics Cloud for System z. These zEnterprise offerings provide a very cost-attractive packaging of hardware, software, and services to help clients deploy cloud infrastructures for general purpose virtual server hosting and business intelligence respectively.

The zEnterprise provides even greater flexibility for cloud computing than “one architecture fits all” alternatives. The inclusion of Power and System x blades in a zEnterprise System allows clients to optimize workload placement in order to more closely align IT spending with business goals. The zEnterprise Unified Resource Manager helps the IT staff integrate and manage a multi-architecture environment at a platform level, enabling businesses to unlock the value of workload-optimized systems without suffering the operational complexity that might be experienced in a non-integrated environment.

Q:What operating system software releases are supported on the zEnterprise System?

A: The following are the minimum levels of the operating systems planned to run on z196:

- ▶ **z/OS:**
 - z196: z/OS V1.9 (for toleration only), exploitation starts with z/OS V1.10, full exploitation starts with z/OS V1.12
 - zEnterprise Unified Resource Manager ensemble support : z/OS V1.10 or later
- ▶ **Linux® on System z distributions:**
 - Novell SUSE SLES 10 and SLES 11
 - Red Hat RHEL 5
- ▶ **z/VM®**
 - z196: z/VM V5.4 or higher

For enablement of virtual server lifecycle management and support for managing real and virtual networking resources by the Unified Resource Manager: z/VM V6.1

- ▶ **z/VSE™** V4.1 or higher
- ▶ **z/TPF** V1.1 or higher
- ▶ On the zBX blades we support:
 - AIX® 5.3, 6.1
 - Linux on System x (Statement of Direction)

Q: Will the zEnterprise System support for Microsoft Windows or IBM i OS operating systems?

A: No, Windows or i OS are not supported at this time. IBM will initially support AIX, Linux on System x (SOD available 1H 2011) and optimizers through the zBX. Support for additional operating systems will be evaluated over time based on demand from our clients.

Q: What about the z196 makes it so adept at for consolidation of server farms?

A: Scale is the first big factor making the z196 so compelling for consolidating of distributed servers. The 5.2 GHz chip means more processing power per core, and there are lots more cores available in each frame – to 80 “visible” cores per server but many more helping to keep those 80 focused on application work. The z196 has new RAIM (redundant array of

independent memory) structure that delivers double the amount of memory (3.0 TB) over the z10 EC and greater system reliability.

The use of integrated blades offers an added dimension for workload consolidation and optimization. And Unified Resource Manager governs Linux on System z and blade resources for greater command and control.

Q: What is different about the Model M80 compared to M66?

A: The Model M80 is an enhanced capacity model which contains a different configuration of MCMs than other models. The z196 is fully populated with four high density books and 80 orderable cores. You can configure the M80 machine to be a 1 to 80-way. Like the other four-book model, the M66, the M80 can be ordered with a minimum of 32 GB of memory and up to a maximum of 3 TB.

Upgrading from any other model of the z196 to a Model M80 will require a planned outage of that machine, but you can still avoid application service interruptions if you exploit Parallel Sysplex which can automatically and dynamically shift workloads to another machine.

Q: Will the z196 be physically bigger than the System z10 EC?

A: No, the z196 will have the same floor cutouts as the z10 EC. If you choose optional water cooling or top-exit I/O cabling the clearance will be different (4" additional depth for water and 12" width for top-exit I/O cabling) from the standard z196.

Q: Can you help me understand the firmware implementation of z196?

A: In general, z196 and zBX firmware is implemented in the traditional System z manner. What is different is that firmware (for example, the hypervisors) for blades installed in a zBX will be loaded from firmware delivered with the attached z196, to help optimize the integration of the zBX with the attached CPC, as well as treating the blade firmware as an "always there" component of the zEnterprise System. Also, there are some changes to licensing of certain firmware in a zEnterprise System. First, firmware for the z196 cryptography feature includes technology that requires license terms in addition to the standard license terms governing use of LIC, so an addendum to the LIC license will be included when a z196 is configured with a cryptography feature. Second, certain zBX components and features include firmware that is licensed under non-IBM terms (called "separately licensed code" or "SLC"). IBM will deliver the license agreements governing use of SLC along with the IBM license agreement for LIC when a zBX is ordered.

Q: What is the maximum allowable distance between the z196 frame and the zBX frame?

A: The controlling z196 must be connected to the zBX with a 26 meter (85') cable. This requirement is for serviceability reasons.

Q: What cooling options are available for the z196 and the zBX.

A: For the z196, there is optional water cooling available. For the zBX there is an optional rear door heat exchanger available.

Q: Is it possible to update driver code on the zBX separately from the z196, on a different schedule?

A: No. Once the zBX is installed firmware is common and updated as a unified zEnterprise system.

Q: Can I buy a z196 that has only IFL or ICF processors without including a general-purpose processor (CP)?

A: Yes. Similar to the System z10 EC, you can order only IFLs or ICFs in a z196, using a model capacity of 700 with 1 to 80 IFLs or a maximum of 16 ICFs.

Q: Why are there only two BladeCenter chassis per zBX frame?

A: The maximum power that a rack can support is 240amps. Two BladeCenters, fully

configured in accordance with the zBX infrastructure and the redundancy it provides, will reach this limit. Additionally, as the zBX is shipped as a unit, more than two BladeCenters would exceed shipment tilt/weight limitations.'

Q: Is there a limit on the number of virtual servers per blade?

A: There is no limit per se, but you could not have more than 72 virtual servers with more than one shared CPU/core apiece. You are limited to fewer virtual servers if you have several dedicated cores. Each shared virtual CPU requires at least 0.1 processing units (PUs), each core has 1.0 processing units, so there can be at most 10 virtual servers per core. On the PS701, VIOS requires 0.1 PUs for each core, for a total of 0.8 PUs.

Q: What workloads are good candidates running on the zEnterprise System?

A: The zEnterprise provides the opportunity to bring together most or all of the application and information components of important end-to-end business applications into one unified, simplified, and easily managed server environment. Strong candidates for zBX hosting include application components with affinities to System z application and information services. Such applications are found in every industry, including banking, insurance, retail, government, and manufacturing. Some of the candidate workloads include business intelligence, data warehousing, business analytics, ERP (including SAP applications), CRM, infrastructure services (such as monitoring, storage management, and security services), Web serving, and other multi-tier application architectures. The zEnterprise also offers new opportunities for recentralizing enterprise information (for improved governance and customer privacy protection), master data management, and enterprise reporting.

IBM System z representatives can meet with prospective zEnterprise customers and use a new assessment tool to identify the strongest application candidates and determine the business value that zEnterprise delivers. Please contact an IBM representative for more information.

Q: I am a user of a large ERP application and have been running it on System z for a long time. I have migrated everything in my entire shop over to System z and now run DB2 for z/OS, several applications on z/OS, all my ERP application servers on Linux on System z, and some other applications on Linux on z. The few exceptions to our general deployment strategy consist of some third party applications. The application vendors do not support their products on System z, but the applications are critical to our business. I think these applications would be good candidates for a zBX blade solution. Correct?

A: Yes. The zBX's Power blades provide an excellent option for such applications, and customers can use them to consolidate applications that do not yet run on Linux on System z or on z/OS.

Customers should also ask their vendors for the solution attributes they need, including support for their preferred platforms. Through its PartnerWorld program, IBM welcomes and assists vendors in bringing their applications to more operating systems, including Linux on System z and z/OS. In this way, vendors can expand their market opportunities, increase customer satisfaction and retention, and grow their revenues and profits. Typically the effort required is minimal, and IBM is happy to work with vendors to help them achieve these goals.

Q: What is the benefit of zBX frames with blades compared to separate BladeCenters?

A: There are several benefits of using the zBX. One is the 10Gb network that potentially speeds up all connections between the distributed systems and the System z operating systems. The connectivity is provided by a fast Layer 2 network, which can reduce latency and overhead. Another benefit is the Unified Resource Manager, which provides uniformity of management tasks, independent of server type or operating system. The system administrator uses Unified Resource Manager to set up new virtual servers in the same way, independently of the operating system and the hardware.

Q: Can an IFL only server (z196 Model 700) run Unified Resource Manager with a zBX attached?

A: Yes. There is no need for CPs or z/OS. The Unified Resource Manager does not require a specific z196 configuration or operating system.

Q: What is the configuration of the BladeCenter chassis in the zBX?

A: All the blades are virtualized, so the underlying hardware is never material for management and provisioning. The BladeCenter chassis is part of the configured hardware that comes with the zBX order. When an order is placed, depending on the number of blades specified in the configurator you plan to install, the required hardware is placed into the zBX. This also means that when you want to upgrade the zBX to add additional blades, or optimizers, the z configurator (econfig) will do that planning work again and the resulting configuration will have the necessary hardware.

Q: Is the Unified Resource Manager the mechanism for managing the BladeCenter chassis and its I/O modules?

A: Yes. Using the Unified Resource Manager (part of the enhanced management suite) you will define the virtual server. All aspects of a virtual server can be defined – CPU, memory, network, console, disk storage, and virtual DVD – taking into account differences in the underlying capabilities (e.g., a virtual DVD is not supported for a z/VM-based virtual server). In addition, virtual guests can be listed, started and stopped, reconfigured, and deleted when no longer required. A virtual server definition can also be moved from one hypervisor to another of the same type.

You will define a workload -- a workload being representation of physical and virtual resources in the context of named business processes. When you set up the workload you will assign the virtual servers, storage, network (VLANs) that you want to make available to the workload.

There are two top-of-rack (TOR) switches for the INMN and two TOR switches for the IEDN in the first rack. All BladeCenters are connected to these TOR switches to provide redundant access to the internal networks.

Q: Will we have Linux available to run on the POWER7 blades?

A: No. The POWER7 blades will run AIX 5.3 (Technology Level 12) and later in POWER6™ and POWER6+™ compatibility mode and AIX 6.1 (Technology Level 5) and later.

Q: Which version of PowerVM™ do I need?

A: You will need to get a license for PowerVM Enterprise Edition (EE). You don't need to get a copy of the software as it will be loaded as System z firmware. Note that when the blade is installed in the zBX and 'discovered' by Unified Resource Manager some of PowerVM functions will not be exploited – such as Live Partition Mobility.

Q: What if the POWER7 blade has more memory on it than the supported configuration?

A: The blade would be rejected by Unified Resource Manager.

Q: What applications can run on the POWER7 blade in a zBX?

A: Applications that are certified for POWER7 and PowerVM EE should run as because the same software and hardware environment is provided on the blade in the zBX.

Q: Will there be separate certification and support statements required for the various middleware products run on the supported AIX?

A: If middleware or applications run on PowerVM on a POWER7 blade today, then existing support statements should be sufficient. IBM has been working closely with its ISV partners and does not believe that there will be a need to provide separate certification statements.

Q: If I have a rack with POWER7 blades installed in it (existing BladeCenter H) integrate those existing blades into a zBX so they can become part of zEnterprise?

A: If the blades meet the specifications defined by System z for integration into the zBX, those blades can be installed into the zBX through the defined processes. Note: You can not integrate an existing BladeCenter H chassis into the zBX.

Q: What hypervisor will be used for the System x blades?

A: We have only provided a Statement of Direction for the System x blades. When we do the full announcement we will provide the information about the hypervisor that will be used.

Q: What about VMware – are their plans to make this available?

A: No.

Q: Will zBX frames have the same reliability as the classic System z machine?

A: No, the classic part of the new machine has outstanding availability that is the result of collaboration among different components like system hardware, firmware implementation and operating systems (z/OS, z/VM, z/VSE). Blades do not have the same qualities of service just because they are attached to a System z machine. However, the blades benefit from the capabilities of the robust System z management environment such as first-failure data capture, call home, and guided repair, as well as from the redundancy that is built in to the zBX hardware infrastructure.

Q: What availability connection exists between the z196 and the zBX? Can the zBX continue to run if there is no host (for example, you have to do a POR)?

A: Yes the zBX can continue to run; however, it may lose the "guidance" provided by the z196. Other z196 servers in the ensemble that includes the zBX can access the zBX even if the host isn't available, because their connection to the zBX is via the private data network (IEDN).

Q: Can I put storage or other things in the zBX rack?

A: No. The zBX is managed by the Unified Resource Manager, which ensures that blades in the zBX are of supported types. Unsupported blades would not be powered on or configured

Q: How does z/OS work with IBM zEnterprise Unified Resource Manager?

A: z/OS integrates with the new Unified Resource Manager environment. A new agent, Guest Platform Management Provider (GPMP), in z/OS V1.12 communicates with z/OS WLM and provides basic data (such as system resource utilization, system delays, and paging delays) back to the Unified Resource Manager over the INMN network. The Unified Resource Manager is designed to add additional workload relationships from the ensemble components to your z/OS workload; for example, linking a transaction that may have started on the zBX back to DB2 on z/OS data.

Q: What's the difference between what you're announcing and IBM's existing management software, such as Tivoli?

A: The new IBM Unified Resource Manager, part of the IBM Systems Director family, monitors and adjusts hardware resources to meet changes in demand. IBM Tivoli software extends the benefits of the Unified Resource Manager to manage software resources including diagnosing, isolating and repairing potential software problems.

The Unified Resource Manager operates at the hardware and platform management level on HW resources, virtual images, and virtualization levels, while Tivoli operates at the Service Management level for applications, transactions, databases.

Q: Where is the code to set up an ensemble?

A: The customer interfaces of the Unified Resource Manager that are used to create an ensemble are on the HMC.

Q: Is Tivoli® software mandatory when setting up an ensemble?

A: No. Tivoli software is not mandatory and will be dictated by what functions of the zEnterprise System and how much of them you want to manage from Tivoli.

Q: Can or must a z196 node be a member of a Parallel Sysplex®?

A: Remember that a Parallel Sysplex is an availability design and the ensemble is a workload design. Accordingly, logical partitions on a z196 may, but do not have to be, members of a Parallel Sysplex.

Q: Which System z server can I attach an IBM Smart Analytics Optimizer?

A: An IBM Smart Analytics Optimizer can be attached to either a z10 or a z196, a zBX Model 001 connects to a z10, while a zBX Model 002 connects to a z196.

Q: What level of DB2 on z/OS do I need to have to be able to attach an IBM Smart Analytics Optimizer?

A: The minimum level of DB2 is DB2 9 for z/OS with PTFs that need to be installed via SMP/E

Q: What are the new technologies in Smart Analytics Optimizer and where did they come from?

A: The product is a commercialization of technologies developed in the IBM Almaden Research Center and the IBM Research Development Lab Boeblingen, Germany. The research project called Blink, proved the concepts of; frequency based compression, the blending of row and columnar store technologies, and compressed data Predicate Evaluation all in an in-memory, massively parallel solution.

Q: Do I have to care about the software version and operating system that are installed on the IBM Smart Analytics Optimizer blades?

A: No, this is all managed by DB2 for z/OS.

Q: Do I have to change my applications to use IBM Smart Analytics Optimizer?

A: No, the use of IBM Smart Analytics Optimizer is transparent to the end-user application. DB2 manages the IBM Smart Analytics Optimizer. The only thing the end user has to do is to define the data mart (the subset of information, generally a department or a division of the business), that should be copied into the optimizer.

Q: What data is targeted for acceleration on the IBM Smart Analytics Optimizer?

A: Data that is normally organized in DB2 as a star schema is the data the Smart Analytics Optimizer is tuned to accelerate. This data is usually organized in this format to perform multi-dimensional analysis. This is the way sales data would be organized if users wanted to see sales presented over multiple periods, in multiple locations for multiple products.

Q: Must the primary HMC and backup HMC that manage the ensemble be exclusive to z196 or can they also be used as 'regular' HMCs.

A: The code to manage the ensemble is an application running on an HMC. That HMC can be used for other 'regular' HMC functions. But the backup HMC has no function other than to mirror the ensemble state information from the primary. Accordingly, no other function can execute on the backup HMC.

Q: Can the different functions of the ensemble HMC be secured by using LDAP services?

A: Yes, all functions can be secured by the use of an LDAP server. It is recommended that z/OS be used as the LDAP server for the ensemble HMC.

Q: Will I be able to upgrade from a full capacity z196 to a subcapacity z196?

A: Yes. Each of the first fifteen general purpose processors on the z196 can be divided into one full capacity and three subcapacity units. This creates a 15 by 4 matrix of settings. As long as upgrades are positive capacity growth, you can move around anywhere within the

matrix when adding capacity. When your number of general-purpose processors exceeds fifteen, then all of the general-purpose processors must be full capacity.

Q: If I have a System z10 EC and a zBX Model 001 and I want to upgrade to a z196 do I need to upgrade the zBX at the same time?

A: The zBX Model 001 cannot be managed by the z196. You will need to upgrade to a zBX Model 002.

Q: When should I look at having water cooling on z196?

A: If you have a problem with hot spots in your data center, water cooling may help eliminate them.

If you are limited on power in your data center, a water-cooled system is a way to increase server capacity without increasing power requirements.

Your savings will vary based on the server configuration in terms of the number of processor books and I/O cards, as well as on the power and cooling used in your data center. For a well-utilized (not maximum) four-book system, you can expect to see savings of:

About 2.5 kw of server input power. Depending on your data center power/cooling that will save 3.8 – 6.3 kw of data center input power.

Approximately 13 kw less heat load delivered to the air in the data center. Depending on your data center power/cooling that will save over 1kw of data center input power in the most modern, efficient data centers and far more than that in typical data centers.

Simply said, if you have a data center that is bounded by limited power capacity or if you want to reduce the cost to remove server heat load, you should look at the water cooling option. In addition, you'll should explore new capabilities such as high voltage DC input.

If you are building a new data center water cooling may be one way to get a reduction in energy use. When considering water cooling it's important to look at your entire data center strategy. System z is one component but even more significant improvements in removal of heat load can be achieved by implementing water cooling across your other server platforms.

Q: What happens to power consumption of the z196 if I add a zBX and blades into the infrastructure?

A: The z196 and zBX are two separate products so they will each have their own individual power use. Both offer water-cooling options to help improve energy use.

Q: What are the major changes to the z/OS V1R11 LSPR?

A: The LSPR ratios reflect the range of performance between System z servers as measured using a wide variety of application benchmarks. The latest release of LSPR contains a major change to the workloads represented in the tables. In the past, workloads have been categorized by their application type or software characteristics (for example, CICS, OLTP-T, LoIO-mix). With the introduction of CPU Measurement Facility (SMF 113) data on z10, greater insight into the underlying hardware characteristics that influence performance is now possible. Thus, the LSPR defines three new workload categories - LOW, AVERAGE, HIGH - based on a newly defined metric called "Relative Nest Intensity (RNI)" that reflects a workload's use of a processor's memory hierarchy. For details on RNI and the new workload categories, refer to:

<https://www-304.ibm.com/servers/resourceLink/lib03060.nsf/pages/lsprindex>

Q: I notice you now have an I/O drawer, in addition to the I/O cage. Why is that?

A: The I/O drawer, first offered on z10 BC, has 8 I/O slots, compared to the I/O cage which has 28 I/O slots. By allowing an I/O drawer to exist in a configuration your infrastructure can

be optimized. The configuration tool selects the appropriate packaging based upon your requirements. In addition, the I/O drawer can be concurrently added and deleted in the field. The I/O cage cannot be concurrently added or deleted. This design allows installation of I/O features based on application growth and connectivity growth, obviating the requirement to plan ahead for I/O cages and I/O drawers.

Q: Tell me about IBM XIV® Storage System that runs with Linux on IBM System z

A: The IBM XIV Storage System is a revolutionary open disk system that represents a generation of high-end disk storage, offering self-tuning and self-healing for consistently high performance and reliability as well as management simplicity and low total costs.

The benefits of Linux on System z can be combined with the phenomenal capabilities of XIV – storage reinvented to support today’s fast growing, smart business infrastructures.

Q: Does z/VM provide support for the IBM XIV Storage System?

A: XIV Storage Systems can be directly attached to z/VM for system use (e.g., paging, spooling, IPL). This support provides the ability to define system and guest volumes as emulated devices (EDEVICs) on IBM XIV Storage.

XIV Storage can also be directly accessed by z/VM guests through guest-attached FCP subchannels. This support requires the Linux on System z guest to provide the applicable FCP multi-path driver.

Note: XIV storage does not support ECKD on System z.

For the most current information visit: <http://www.vm.ibm.com/storman/xiv>

Q: How does System z fit into the IBM Dynamic Infrastructure® initiative?

A: In the IBM vision, a Dynamic Infrastructure drives a new scale of efficiency and service excellence for businesses, helping to align IT with business goals.

The Dynamic Infrastructure, being an evolutionary model for efficient IT delivery, provides a highly dynamic, efficient and shared environment which allows IT to better manage costs, improve service levels, improve operational performance and resiliency, and more quickly respond to business needs. Operational issues are addressed through consolidation, virtualization, energy offerings, and service management. The existence of virtualized resource pools for server platforms, storage systems, networks, and applications enables IT delivery to users in a more fluid way.

System z is considered to be the most robust, secure, and virtualized platform in the industry. The z196 server has just-in-time deployment of additional resources, known as Capacity on Demand (CoD). CoD provides flexibility, granularity, and responsiveness by allowing the user to dynamically change capacity when business requirements change. The IBM zEnterprise System further extends these capabilities to a heterogeneous infrastructure. Considering the other elements of the stack (software and services), zEnterprise System is an up-to-date evolutionary platform that truly is the cornerstone for implementation of a Dynamic Infrastructure.

Q: What security classifications do the System z servers have?

A: System z servers are certified at the highest security level in the industry: EAL 5 Common Criteria for Logical Partitions. System z10 EC received its certification on October 29, 2008, and System z10 BC on May 4, 2009. The z196 certification is currently pending.

Q: What is HiperDispatch?

A: HiperDispatch is a name for several improvements in interaction between PR/SM and z/OS. It is a mechanism that recognizes the physical processor where the work was started and then dispatches subsequent work to the same physical processor. This helps to reduce

the movement of cache and data and improves overall system throughput. HiperDispatch is available only with z196 and z10 PR/SM and z/OS functions.

Q: What are the consequences when I switch off HiperDispatch in z/OS?

A: For some workloads that are dispatched on many processors across multiple books it may decrease the performance because processor caches must be reloaded more often.

Q: Can I run Linux on z196?

A: Yes. Major Linux on System z distributions include Novell SUSE and Red Hat. IBM is working with these Linux distribution partners to provide Linux with appropriate functionality on all of its hardware platforms.

Q: Can I run MS Windows on z196?

A: No, because no Windows operating system is available for z196.

Q: Can I run Sun Solaris on z196?

A: No. There is no Solaris on z196 offering. However, as a result of a collaboration project between IBM and the Open Source community to investigate the feasibility of bringing OpenSolaris to System z, a distribution of OpenSolaris for System z is now available to run as a guest of z/VM. Note that IBM does not warrant, and is not responsible for support of, a non-IBM operating system. For the announcement details see:

http://www-01.ibm.com/common/ssi/rep_ca/5/897/ENUS108-875/ENUS108-875.PDF

Q: Can any z9 EC or z10 EC model be upgraded to a z196?

A: Yes.

Q: Can I do all upgrades within the z196 concurrently?

A: Most upgrades are concurrent. For example, if memory is already installed in the book, enabling it with Licensed Internal Code (LIC) is a concurrent action. But if the memory must be installed physically first, the action may not be concurrent. It also depends on the number of available books and their configuration. With proper planning, the user may be able to avoid planned outages.

Upgrades to model M80 are not concurrent. However, if the z/OS LPARs running on a z196 are part of a Sysplex with multiple servers, applications will continue to run on z/OS images of other servers, and no lack of service will be incurred.

Q: What is the difference between concurrent and nondisruptive upgrade?

A: In general, concurrency addresses the continuity of operations of just the hardware part of an upgrade, for instance, whether a server (as a box) is required to be switched off during the upgrade. Disruptive versus nondisruptive refers to whether the running software or operating system must be restarted for the upgrade to take effect. Thus, even concurrent upgrades can be disruptive to those operating systems or programs that do not support them while at the same time being nondisruptive to others.

Q: What is meant by the plan-ahead memory function on the z196?

A: Memory can be upgraded concurrently using LIC-CC if physical memory is available on the z196 server. The plan-ahead memory function provides the ability to plan for nondisruptive memory upgrades by having the system pre-plugged based on a target configuration. Pre-plugged memory will be enabled through a LIC-CC order placed by the customer.

Q: What is the difference between the plan-ahead memory and the flexible memory option?

A: Plan-ahead memory should not be confused with flexible memory support. Plan-ahead memory is for a permanent increase of installed memory, while flexible memory provides a temporary replacement of a part of memory that becomes unavailable.

Q: What is the benefit of 1 MB page size? Should I switch from 4 KB to 1 MB?

A: For workloads with large memory requirements, large pages cause the Translation Lookaside Buffer (TLB) to better represent the working set and suffer fewer misuses by allowing a single TLB entry to cover more address translations. Exploiters of large pages are better represented in the TLB and are expected to perform better. Long-running memory access-intensive applications especially benefit. Short processes with small working sets see little or no improvement. The use of large pages must be decided based on knowledge obtained from measurement of memory usage and page translation overhead for a specific workload.

Under z/OS, large pages are treated as fixed pages and are never paged out. They are only available for 64-bit virtual private storage such as virtual memory located above 2 GB.

IBM is working with its Linux distribution partners to include support in future Linux on System z distribution releases, where the benefits will be similar to those described above.

Q: How can I control and limit the costs related to On/Off CoD?

A: Customers can limit the financial exposure towards On/Off CoD and order an exact amount of temporary On/Off CoD processing capacity in the form of capacity tokens. The actual balance of capacity tokens can be checked anytime.

Q: What is a capacity token?

A: A capacity token is a representation of resources available for a given period of time. The measurement units used are MSU days for CP capacity and specialty engine days for specialty engines. One MSU token is worth one MSU day and one specialty engine token is worth one specialty engine day.

Q: What is a pre-paid capacity?

A: Pre-paid capacity is temporary processing capacity that can be ordered, paid for, and kept in reserve for future consumption.

Q: What is CPE?

A: Capacity for Planned Event is a Capacity on Demand offering. It delivers a replacement capacity for a planned event-like data center planned outage, server move, and so on. When it is activated, a specific pre-ordered configuration can be set online and used for up to three days. Upon activation, it does not incur additional charges from IBM.

Q: Can I use NTP instead of STP?

A: No. However, z196 servers can be synchronized against external time sources through NTP.

Q: Mainframes are known for EBCDIC code pages. How can Linux, which is ASCII, run on z196?

A: There is no requirement for EBCDIC defined in the z/Architecture. For example, z/OS supports EBCDIC, ASCII, and Unicode. Linux on System z uses ASCII and Unicode.

Q: How many new machine instructions were added to z196?

A: More than 110 instructions were added.

Q: What are the new machine instructions used for?

A: The instructions added to z/Architecture are the result of active collaboration between hardware and software designers, specifically with the compiler teams. Hardware and software are being co-optimized, while maintaining full upward compatibility. Most of these instructions are specifically targeted to be used by compilers to improve the efficiency of generated code. Examples are combining two simple functions into a single instruction or reducing the number of active general registers needed for a sequence. Providing access to the high order part of a register, effectively doubling the number of registers available.

Another group of new instructions improves software/hardware synergy by enabling software to give hints to the hardware on caching of specific blocks of memory and by communicating the effective SMP topology so that processes can be kept close to their cached data. The remaining instructions provide minor extensions to existing functions.

Q: When using a sub-capacity model, is it possible to mix different CP feature codes?

A: No, all CP feature codes must be the same. In other words, all CPs must be at the same capacity level.

Q: Are subcapacity versions of specialty engines available?

A: No, specialty engines are always at maximum capacity.

Q: Is it possible to have more than one book although purchased processors would fit into fewer books?

A: Yes, there is no restriction that prevents this. Enhanced book availability uses this approach in order to avoid planned outages.

Q: I read in the z/OS announcement that you can run zAAP workload on a zIIP. Does that mean the zAAP is going away?

A: z/OS has a new function that can enable System z Application Assist Processor (zAAP) eligible workloads to run on System z Integrated Information Processors (zIIPs). This new capability is ideal for customers without enough zAAP- or zIIP-eligible workload to justify both specialty engines today; the combined eligible TCB and enclave SRB workloads might make the acquisition of a zIIP cost effective. This new capability is also intended to provide more value for customers having only zIIP processors by making Java- and XML-based workloads eligible to run on existing zIIPs.

zAAPs are not going away and are available on z196. Customers who have already invested in zAAP, or have invested in both zAAP and zIIP processors, should continue to do so as this maximizes the potential for new workload on the platform. The zAAP on zIIP capability is not available for z/OS LPARs if zAAPs are installed on the server.

This capability is available with z/OS V1.11 (and also on z/OS V1.9 and z/OS V1.10 with the PTFs for APAR OA27495) and all System z9® and System z10 servers; some additional restrictions apply. This capability does not provide an overflow that directs additional zAAP-eligible workload to run on a zIIP, but enables the zAAP-eligible work to run on zIIP when no zAAP is defined. This new capability does not remove the requirement to purchase and maintain one or more general purpose processors for every zIIP processor on the server. This part of the IBM terms and conditions surrounding the IBM System z specialty engines is unchanged.

Q: I cannot have more zAAPs than CPs and I cannot have more zIIPs than CPs. Combined, can I have more zAAPs plus zIIPs than CPs?

A: Yes. For each CP you can have one zAAP and one zIIP.

Q: I have a question about the 1:1 ratio between CP processor cores and zAAPs/zIIPs when I upgrade my processor. I have a z10 EC-E26-712 with 12 CP processor cores and 12 zAAPs and am planning to upgrade to a z196-M32-708 with 8 CP processor cores. Due to the 1:1 will I only be able to upgrade 8 zAAPs and lose my investment in the 4 other zAAPs?

A: No. Our System z configurator will NOT reduce the numbers of zAAPs since you have already bought them. In the example above, you will get 8 CP processor cores and 12 zAAPs. At the time of the upgrade, all 12 zAAPs will be available for use, but the configurator will not allow you to order any additional zAAPs until you come back within the 1:1 ratio. That means that if you order 1, 2, 3 or 4 CPs in future you can NOT order another zAAP. When you order the fifth additional CP, you could then order another zAAP.

Q: Can I mix dedicated and shared processors in one logical partition?

A: No. Dedicated and shared processors cannot be mixed in one logical partition, regardless of their type.

Q: Is it possible to order an IFL-only server?

A: Yes.

Q: Why can I not use zAAP for Java workload in Linux?

A: zAAP is designed to offload Java workload from CPs in z/OS to keep MSU values lower while providing more processing capacity for Java workload. Because there is no MSU measurement for Linux, it makes no sense to use zAAP there.

Q: What is a z/VM-mode LPAR?

A: A z/VM-mode LPAR is a special type of partition designed to allow z/VM guests to utilize a broader range of specialty processors. This new LPAR mode allows z/VM and its guests to utilize CPs, IFLs, zAAPs, zIIPs, and ICFs in the same logical partition.

Q: What is the Capacity Provisioning Manager?

A: The Capacity Provisioning Manager is software delivered with the z/OS BCP feature. It is a component that allows you to switch OOCoD records on and off automatically according to defined policies. It monitors RMF™ metrics to decide when to activate OOCoD and when to deactivate it.

Q: How much memory should I plan for HSA?

A: With z196 no planning of memory for HSA is required. Each z196 server by default contains a 16 GB memory that is fixed and used for HSA. This memory is *not* part of the memory purchased by the customer. HSA never occupies additional memory outside of its 16 GB.

Q: Can I run out of HSA space?

A: No. The HSA size is large enough to hold all possible definitions.

Q: What are the increments for ordering the memory?

A: Up to 256 GB, the increment is 32 GB. From 256 GB to 512 GB it is 64 GB. From 608 GB to 896 GB it is 96 GB. For 1008 GB it is 112 GB. From 1136 GB to 1520 GB it is 129 GB, For 1776 GB to 3056 GB it is 256 GB.

Q: Can I connect z196 to a SAN?

A: Yes. FICON cards support both FICON and FC protocols. If the operating system supports the FC protocol, it can participate in the same environment as any other operating system supporting the FC protocol.

Q: Can I use the 2-port FICON Express4 adapter with a z196 server?

A: No. It is unique to the z10 BC and z9 BC.

Q: Can I carry forward my older OSA-Express2 adapters?

A: In general, yes. Check Appendix D, "Channel options" on page 135 for details.

Q: What is the High Performance FICON for System z (zHPF)?

A: zHPF is an extension of the FICON channel architecture compatible with FC-FS, FC-SW, FC-SB-4 Fibre Channel standards. If fully exploited by the FICON channel, z/OS, and control unit, it reduces the FICON channel overhead between z/OS and the control unit, thus improving the channel performance.

Q: What are the maximum supported distances of InfiniBand coupling links?

A: The HCA2-O fanout (FC 0163) provides the distance up to 150 m, while HCA2-O LR fanout (FC0168) supports up to 10 km unrepeated (up to 100 km with repeaters).

Q: What CFCC level is supplied with the z1Enterprise server?

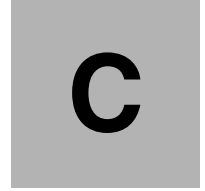
A: The current level is CFCC 17.

Q: CFCC 17 is a new level. What new functionality does it provide?

A: CFCC Level 17 includes the following improvements:

- Increased number of CHPIDs from 64 to 128. By allowing more CHPIDs, more parallel CF operations can be serviced and therefore CF link throughput can increase.

Support for up to 2048 structures, while up to 1024 structures are supported at CFCC level 16. Growing requirements, for example due to logical grouping such as DB2, IMS, and MQ datasharing groups, or customer mergers, acquisitions and sysplex consolidations, demands for more structures than ever before.



Software licensing

This appendix briefly describes the software licensing options available for zEnterprise 196.

Software licensing considerations

The zEnterprise 196 mainframe IBM software portfolio includes operating system software (that is, z/OS, z/VM, z/VSE, and z/TPF) and middleware that runs on these operating systems. It also includes middleware for Linux on System z environments.

Two major metrics for software licensing are available from IBM, depending on the software product:

- ▶ Monthly License Charge (MLC)
- ▶ International Program License Agreement (IPLA)

MLC pricing metrics have a recurring charge that applies each month. In addition to the right to use the product, the charge includes access to IBM product support during the support period. MLC metrics, in turn, include a variety of offerings. Those applicable to the zEnterprise 196 are:

- ▶ Workload License Charges (WLC)
- ▶ System z New Application License Charges (zNALC)
- ▶ Parallel Sysplex License Charges (PSLC)
- ▶ Midrange Workload License Charges (MWLC)

IPLA metrics have a single, up-front charge for an entitlement to use the product. An optional and separate annual charge called *subscription and support* entitles customers to access IBM product support during the support period and also receive future releases and versions at no additional charge. For details, consult the *IBM System z Software Pricing Reference Guide*, Web page:

http://www-03.ibm.com/servers/eserver/zseries/library/refguides/sw_pricing.html

Workload License Charges (WLC)

Workload License Charges require z/OS or z/TPF operating systems in 64-bit mode. Any mix of z/OS, z/VM, Linux, z/VSE, and z/TPF images is allowed.

The two WLC license types are:

- ▶ Flat WLC (FWLC): Software products licensed under FWLC are charged at the same flat rate, independently of the capacity installed on the server, measured in Millions of Service Units (MSUs).
- ▶ Variable WLC (VWLC): This type applies to products such as z/OS, DB2, IMS, CICS, MQSeries®, and Lotus Domino®. VWLC software products can be charged as:
 - Full-capacity: The server's total number of MSUs is used for charging. Full-capacity is applicable when the server is not eligible for sub-capacity.
 - Sub-capacity: Software charges are based on the logical partition's usage where the product is running.

WLC sub-capacity allows software charges based on logical partition usage instead of the server's total number of MSUs. Sub-capacity removes the dependency between software charges and server (hardware) installed capacity.

Sub-capacity is based on the logical partition's rolling 4-hour average usage. It is *not* based on the usage of each product,¹ but on the usage of the logical partitions where it runs. The VWLC licensed products running on a logical partition will be charged by the maximum value of this partition's rolling 4-hour average usage within a month.

The logical partition's rolling 4-hour average usage can be limited by a *defined capacity* definition on the partition's image profiles. This activates the *soft capping* function of PR/SM, limiting 4-hour average partition usages above the defined capacity value. Soft capping controls the maximum rolling 4-hour average usage (the *last* 4-hour average value at every 5-minute interval), but does *not* control the maximum *instantaneous* partition use.

Also available is an LPAR group capacity limit, which allows you to set soft capping of PR/SM for a group of logical partitions running z/OS.

Even using the soft capping option, the partition's use can reach up to its maximum share based on the number of logical processors and weights in the image profile. Only the rolling 4-hour average use is tracked, allowing usage peaks above the defined capacity value.

As with the Parallel Sysplex License Charges (PSLC) software license charge type, the aggregation of the servers' capacities within the same Parallel Sysplex is also possible in WLC following the same prerequisites.

The Entry Workload License Charges (EWLC) charge type is not offered for IBM System z10 EC.

For further information about WLC and details about how to combine logical partitions usage, see the publication *z/OS Planning for Workload License Charges*, SA22-7506, available from:

http://www-03.ibm.com/systems/z/os/zos/bkserv/find_books.html

¹ With the exception of products licensed using the SALC pricing metric

System z New Application License Charges (zNALC)

System z New Application License Charges offers a reduced price for the z/OS operating system on logical partitions running a qualified *new workload* application such as Java language business applications running under WebSphere Application Server for z/OS, Domino, SAP, PeopleSoft, and Siebel.

z/OS with zNALC provides a strategic pricing model available on the full range of System z servers for simplified application planning and deployment. zNALC allows for aggregation across a qualified Parallel Sysplex, which can provide a lower cost for incremental growth across new workloads that span a Parallel Sysplex.

For additional information see the zNALC Web page:

<http://www-03.ibm.com/servers/eserver/zseries/swprice/znalc.html>

Select Application License Charges (SALC)

Select Application License Charges applies to WebSphere MQ for System z only. It allows a WLC customer to license MQ under product utilization rather than the sub-capacity pricing provided under WLC.

WebSphere MQ is typically a low-usage product that runs pervasively throughout the environment. Clients who run WebSphere MQ at a very low usage may benefit from SALC. Alternatively, you can still choose to license WebSphere MQ under WLC.

A reporting function, which IBM provides in the operating system IBM Software Usage Report Program, is used to calculate the daily MSU number. The rules to determine the billable SALC MSUs for WebSphere MQ use the following algorithm:

1. Determines the highest daily usage of a program² family, which is the highest of 24 hourly measurements recorded each day
2. Determines the monthly usage of a program² family, which is the fourth highest daily measurement recorded for a month
3. Uses the highest monthly usage determined for the next billing period

For additional information about SALC, see the Other MLC Metrics Web page:

<http://www.ibm.com/servers/eserver/zseries/swprice/other.html>

Midrange Workload Licence Charges

Midrange Workload Licence Charges (MWLC) applies to z/VSE V4 when running on z196, System z10 and System z9 servers. The exceptions are the z10 BC and z9 BC servers at capacity setting A01 to which zELC applies.

Similarly to Workload Licence Charges, MWLC can be implemented in full-capacity or sub-capacity mode. MWLC applies to z/VSE V4 and several IBM middleware products for z/VSE. All other z/VSE programs continue to be priced as before.

The z/VSE pricing metric is independent of the pricing metric for other systems (for instance, z/OS) that might be running on the same server. When z/VSE is running as a guest of z/VM, z/VM V5R4 or later is required.

² Program refers to all active versions of MQ

To report usage, the sub-capacity report tool is used. One SCRT report per server is required.

For additional information see the MWLC Web page:

<http://www.ibm.com/servers/eserver/zseries/swprice/mwlc.html>

System z International Program License Agreement (IPLA)

On the mainframe, the following types of products are generally in the IPLA category:

- ▶ Data management tools
- ▶ CICS tools
- ▶ Application development tools
- ▶ Certain WebSphere for z/OS products
- ▶ Linux middleware products
- ▶ z/VM Versions 5 and 6

Generally, three pricing metrics apply to IPLA products for System z:

- VU** Value unit pricing, which applies to the IPLA products that run on z/OS. Value unit pricing is typically based on the number of MSUs and allows for lower cost of incremental growth. Examples of eligible products are IMS tools, CICS tools, DB2 tools, application development tools, and WebSphere products for z/OS.
- EBVU** Engine-based value unit pricing enables a lower cost of incremental growth with additional engine-based licenses purchased. Examples of eligible products include z/VM V5 and certain z/VM middleware, which are priced based on the number of engines.
- PVU** Processor value units. Here the number of engines is converted into processor value units under the Passport Advantage® terms and conditions. Most Linux middleware is also priced based on the number of engines.

For additional information see the System z IPLA Web page at:

<http://www.ibm.com/servers/eserver/zseries/swprice/zipla/>



D

Channel options

Table D-1 lists the attributes of the channel options supported on z196 servers, the required connector and cable types, the maximum unrepeated distance, and the bit rate.

At least one ESCON, FICON, ISC, or PSIFB feature is required.

Statement of Direction: z196 will be the last high-end server to offer ordering of ESCON channels. IBM intends not to offer ESCON channels on future servers

Table D-1 System z196 channel feature support

Channel feature	Feature codes	Bit rate	Connector	Cable type	Maximum unrepeated distance ^a
Enterprise Systems CONnection (ESCON)					
16-port ESCON	2323	200 Mbps	MT-RJ	MM 62.5 μ m	3 km (800)
Fiber Connection (FICON)					
FICON Express4 SX ^b	3322	4 Gbps	LC Duplex	MM 62.5 μ m MM 50 μ m	70 m (230) 380 m (1247) 150 m (492)
		2 Gbps	LC Duplex	MM 62.5 μ m MM 50 μ m	150 m (492) 500 m (1640) 300 m (984)
		1 Gbps	LC Duplex	MM 62.5 μ m MM 50 μ m	300 m (984) 860 m (2822) 500 m (1640)
FICON Express4 ^b 4KM LX	3324	1, 2, or 4 Gbps	LC Duplex	SM 9 μ m	4 km
FICON Express4 ^b 10KM LX	3321	1, 2, or 4 Gbps	LC Duplex	SM 9 μ m	10 km/20 km ^c

Channel feature	Feature codes	Bit rate	Connector	Cable type	Maximum unrepeat distance ^a
FICON Express8 SX	3326	8 Gbps	LC Duplex	MM 62.5 μm MM 50 μm	21 m (69) 150 m (492) 50 m (164)
		4 Gbps	LC Duplex	MM 62.5 μm MM 50 μm	70 m (230) 380 m (1247) 150 m (492)
		2 Gbps	LC Duplex	MM 62.5 μm MM 50 μm	150 m (492) 500 m (1640) 300 m (984)
FICON Express8 10KM LX	3325	2, 4, or 8 Gbps	LC Duplex	SM 9 μm	10 km
Open Systems Adapter (OSA)					
OSA-Express2 GbE LX ^b	3364	1 Gbps	LC Duplex	SM 9 μm	5 km
				MCP	550 m (500)
OSA-Express2 GbE SX ^b	3365	1 Gbps	LC Duplex	MM 62.5 μm	220 m (166) 275 m (200)
				MM 50 μm	550 m (500)
OSA-Express2 1000BASE-T Ethernet ^b	3366	10/100/1000	RJ45	UTP Cat5	100 m
OSA-Express3 GbE LX	3362	1 Gbps	LC Duplex	SM 9 μm	5 km
				MCP	550 m (500)
OSA-Express3 GbE SX	3363	1 Gbps	LC Duplex	MM 62.5 μm	220 m (166) 275 m (200)
				MM 50 μm	550 m (500)
OSA-Express3 1000BASE-T Ethernet	3367	10/100/1000	RJ45	UTP Cat5	100 m
OSA-Express3 10 GbE LR	3370	10 Gbps	LC Duplex	SM 9 μm	10 km
OSA-Express3 10 GbE SR	3371	10 Gbps	LC Duplex	MM 62.5 μm	33 m (200)
				MM 50 μm	300 m (2000) 82 m (500)
Parallel Sysplex					
IC	n/a		N/A	N/A	N/A
ISC-3 (peer mode)	0217 0218 0219	2 Gbps	LC Duplex	SM 9 μm MCP 50 μm	10 km/20 km 550 m (400)
ISC-3 (RPQ 8P2197 Peer mode at 1 Gbps) ^c		1 Gbps		SM 9 μm	20 km
PSIFB	0163	6 GBps	MPO	OM3 MM 50 μm	150 m
PSIFB LR	0168	5 Gbps	LC Duplex	SM 9 μm	10 km/100 km ^d

Channel feature	Feature codes	Bit rate	Connector	Cable type	Maximum unrepeat- ed distance ^a
Cryptography					
Crypto Express3	0864	N/A	N/A	N/A	N/A

- a. Minimum fiber bandwidth in MHz/km for multi-mode fiber optic links are included in parentheses were applicable.
- b. Feature is only available if carried forward by an upgrade from a previous server.
- c. RPQ 8P2197 enables the ordering of a different daughter card supporting 20 km unrepeat- ed 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 must be activated.
- d. Up to 100 km at 2.5 Gbps, with repeater (System z qualified DWDM vendor product that supports 1x IB-SDR)

Related publications

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

IBM Redbooks publications

For information about ordering these publications, see “How to get IBM Redbooks publications” on page 140. Note that some of the documents referenced here may be available in softcopy only.

- ▶ *IBM zEnterprise System Technical Guide*, SG24-7833
- ▶ *IBM System z Strengths and Values*, SG24-7333
- ▶ *Getting Started with InfiniBand on System z10 and System z9*, SG24-7539
- ▶ *IBM System z Connectivity Handbook*, SG24-5444
- ▶ *Server Time Protocol Planning Guide*, SG24-7280
- ▶ *Server Time Protocol Implementation Guide*, SG24-7281
- ▶ *IBM System z10 Enterprise Class Capacity On Demand*, SG24-7504
- ▶ *IBM TotalStorage DS8000 Series: Performance Monitoring and Tuning*, SG24-7146
- ▶ *How does the MIDAW Facility Improve the Performance of FICON Channels Using DB2 and other workloads?*, REDP-4201
- ▶ *FICON Planning and Implementation Guide*, SG24-6497
- ▶ *OSA-Express Implementation Guide*, SG24-5948
- ▶ *Introduction to the New Mainframe: z/OS Basics*, SG24-6366
- ▶ *Introduction to the New Mainframe: z/VM Basics*, SG24-7316

Online resources

These Web sites are also relevant as further information sources:

- ▶ ResourceLink Web site
<http://www.ibm.com/servers/resourceLink>
- ▶ Large Systems Performance Reference (LSPR)
<http://www-03.ibm.com/servers/eserver/zseries/lspr/>
- ▶ MSU ratings
<http://www-03.ibm.com/servers/eserver/zseries/library/swpriceinfo/hardware.html>

Other publications

These publications are also relevant as further information sources:

- ▶ *Hardware Management Console Operations Guide Version 2.11.0*, SC28-6867
- ▶ *Support Element Operations Guide V2.11.0*, SC28-6868
- ▶ *IOCP User's Guide*, SB10-7037
- ▶ *Stand-Alone Input/Output Configuration Program User's Guide*, SB10-7152
- ▶ *Planning for Fiber Optic Links*, GA23-0367
- ▶ *CHPID Mapping Tool User's Guide*, GC28-6825
- ▶ *Common Information Model (CIM) Management Interfaces*, SB10-7154
- ▶ *IBM System z Functional Matrix*, ZSW0-1335
- ▶ *z/Architecture Principles of Operation*, SA22-7832
- ▶ *z/OS Cryptographic Services Integrated Cryptographic Service Facility Administrator's Guide*, SA22-7521
- ▶ *z/OS Cryptographic Services Integrated Cryptographic Service Facility Overview*, SA22-7519

How to get IBM Redbooks publications

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

<http://ibm.com/redbooks>

Help from IBM

IBM support and downloads

<http://ibm.com/support>

IBM Global Services

<http://ibm.com/services>

Index

Symbols

???????????????? 140

Numerics

50.0 μm 37

62.5 μm 37

A

Active Energy Manager (AEM) 46

Advanced Encryption Standard (AES) 11, 39

AES 11, 39

B

Business Intelligence (BI) 77

C

cage, CPC and I/O 13

Capacity Backup (CBU) 74

Capacity for Planned Event 126

Capacity for Planned Events (CPE) 74

Capacity provisioning 75

capacity ratios 79

capacity token 126

CBU 74

CF 41

CFCC level 129

Channel path 12

chip lithography 26

CHPID 41

cloud computing 116

Commercial Batch Short (CB-S) 79

Compression Unit 38

cooling 45

Coupling Facility (CF) 41

Coupling Facility Control Code level 129

Coupling Link 31

coupling links 14

CP characterization 10

CP Cryptographic Assist Facility 38

CPACF

cryptographic capabilities 6, 11

description 38

PU design 38

CPACF enhancements 70

Crypto Express2 13

accelerator 14

coprocessor 14

Crypto Express3

accelerator 39

coprocessor 39

Cryptographic Accelerator (CA) 39

Cryptographic Coprocessor (CC) 39

cryptographic hardware 27

Customer Initiated Upgrade (CIU) facility 73

D

data connection isolation 67

Data Encryption Standard (DES) 11, 39

Data Warehouse (DW) 77

Decimal Floating Point 27

DES 11, 39

Dynamic Infrastructure 124

E

EAL 5 124

EAL5 52

Enhanced driver maintenance (EDM) 80

ESA/390 27

Extended Address Volumes (EAV) 55

F

FCP 13

enhancements for small block sizes 64

switch 13

Federal Information Processing Standard (FIPS) 14

Fibre Channel Protocol (FCP) 13, 64

FICON

extended distance 63

name server registration 64

FICON Express4 13, 35

FICON Express8 35

FICON to ESCON

conversion function 35

flexible memory 60

H

Hardware system area (HSA) 11

high voltage DC power 45

HiperDispatch 80, 124

HiperSockets 14–15

IPv6 14

multiple write facility 69

zIIP-Assisted 69

HiperSockets Layer 2 support 69

HiperSockets Multiple Write Facility 69

HiperSockets Network Traffic Analyzer 69

HMC 43

HMC applications 71

HyperPAV 55

I

I/O cage 13

I/O device 30

I/O operation 65

I/O virtualization 54
 IBM Enterprise racks 46
 IBM Smart Analytics Optimizer 77
 IC3 41
 ICF characterization 10
 ICSF 112
 IFL characterization 10
 InfiniBand 12, 61
 Input/Output Configuration Dataset (IOCDS) 41
 Internal Battery Feature (IBF) 45
 intraensemble data network (IEDN) 67
 intranode management network (INMN) 66
 ISC-3 40
 ITRR 79

L

L1 cache 27
 LAN 40
 large page support 60
 Licensed Internal Code (LIC) 29
 Linux for System z 100, 108
 Local area network (LAN) 40
 logical processors 53

M

MCM 10
 memory
 card 29
 size 28
 MIDAW facility 114
 mode conditioner patch (MCP) 37
 MSU value 78

N

Network Control Program 67
 Network Time Protocol (NTP) 72
 NPIV 64
 NTP server 73

O

ODIO inbound workload queuing 68
 On/Off CoD 74
 On-line Permanent Upgrade 74
 operating system 73, 100–101, 103
 support 103
 support Web page 110
 OSA for NCP 67
 OSA-Express 14
 OSA-Express2 37
 Gigabit Ethernet 67
 OSN 67
 OSA-Express3 36

P

Parallel Access Volume (PAV) 54
 Parallel Sysplex 42, 132
 License Charge 132

Web site 114
 PCI-e
 cryptographic adapter 39
 cryptographic coprocessor 39
 Peer-to-Peer Remote Copy (PPRC) 55
 permanent upgrade 74
 personal identification number (PIN) 40
 physical memory 28–29
 plan-ahead memory 60
 power-on reset (POR) 27
 PR/SM 52
 prediction tool 65
 pre-paid capacity 126
 processor unit (PU) 27
 Pseudo Random Number Generation (PRNG) 11, 39,
 112
 PSIFB 41
 PSP buckets 101
 PU characterization 27
 pulse per second (PPS) support 73

Q

QDIO interface isolation 68
 QDIO optimized latency mode 68

R

Redbooks Web site
 Contact us xiv
 refrigeration 45
 Resource Link 73

S

SALC 133
 SAP characterization 10
 SE 43
 Secure Hash Algorithm (SHA-1) 11
 Secure Sockets Layer (SSL) 14, 39
 Select Application License Charges 133
 SHA-1 39
 SHA-2 39
 SHA-256 11
 single system image 111
 soft capping 132
 software licensing 110, 114
 software support 58
 Subchannels 12
 support requirement
 z/OS 103, 109
 System Assist Processor (SAP) 11
 system image 58
 System z BladeCenter Extension (zBX) 15
 System z9 Integrated Information Processor 10

T

temporary upgrades 74
 TKE workstation 78
 top exit I/O cabling 46
 Top of Rack (TOR) 48

Triple Data Encryption Standard (TDES) 11, 39
Trusted Key Entry (TKE) 40

U

user 14
User Defined Extensions (UDX) 14

V

VMAC support 67

W

water cooling 45
WebSphere MQ 133
Workload License Charge (WLC) 132
 Flat WLC (FWLC) 132
 sub-capacity 132
 Variable WLC (VWLC) 132

Z

z/Architecture 7, 27, 100, 108
z/OS 105, 112
z/TPF 17
z/VM 55
z196 78
zAAP characterization 10
zBX 15
zHPF 63, 128
zIIP 10

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(->Hide:)>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine:fm still open and **File>Import>Formats the Conditional Text Settings (ONLY)** to the book files.

Draft Document for Review July 28, 2010 1:45 pm

7832spine.fm 145



IBM zEnterprise System Technical Introduction

(0.2"spine)

0.17"<->0.473"

90<->249 pages

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(->Hide:)>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine:fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.

Draft Document for Review July 28, 2010 1:45 pm

7832spine.fm 146



IBM zEnterprise System Technical Introduction



Addresses increasing complexity, rising costs, and infrastructure constraints

Describes key functional elements and features

Discusses a smart infrastructure for the data center of the future

Recently we have seen an explosion in applications, architectures, and platforms. With the generalized availability of the Internet and the appearance of commodity hardware and software, several patterns have emerged that have gained center stage. Workloads have changed. Many applications, including mission-critical ones, are deployed in heterogeneous infrastructures and the System z design has adapted to this change. IBM has a holistic approach to System z design, which includes hardware, software and procedures. It takes into account a wide range of factors, including compatibility and investment protection, thus ensuring a tighter fit with the IT requirements of the entire enterprise.

This IBM® Redbooks® publication introduces the revolutionary scalable IBM zEnterprise System, which consists of the IBM zEnterprise 196 (z196) and the IBM zEnterprise BladeCenter® Extension (zBX). IBM is taking a bold step by integrating heterogeneous platforms under the well-proven System z hardware management capabilities, while extending System z qualities of service to those platforms. The z196 is a general-purpose server that is equally at ease with compute-intensive workloads and with I/O-intensive workloads. The integration of heterogeneous platforms is based on IBM's BladeCenter® technology, allowing improvements in price and performance for key workloads, as well as enabling a new range of heterogeneous platform solutions. The z196 is at the core of the enhanced System z platform that is designed to deliver technologies that businesses need today along with a foundation to drive future business growth.

This book provides basic information about z196 and zBX capabilities, hardware functions and features, and its associated software support.

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**