IBM

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
Octavian Lascu
Erik Bakker
Parwez Hamid
Fernando E Nogal
Frank Packheiser
Vicente Ranieri Jr.
Karl-Erik Stenfors
Esra Ufacik
Chen Zhu

Redbooks

ibm.com/redbooks

International Technical Support Organization

IBM zEnterprise System Technical Introduction

August 2011

SG24-7832-01

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

Second Edition (August 2011)

This edition applies to the IBM zEnterprise System, which includes the z196, z114, zBX, and Unified Resource Manager.

Contents

Notices

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

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

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which 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®
AIX®
BladeCenter®
CICS®
DataPower®
DB2 Connect™
DB2®
Domino®
DRDA®
DS8000®
ESCON®
FICON®
HACMP™
HiperSockets™
IBM Systems Director Active Energy Manager™
IBM®
IMS™
Lotus®
MQSeries®
OMEGAMON®
Parallel Sysplex®
Passport Advantage®
Power Systems™
POWER7™
PowerHA™
PowerVM™
POWER®
PR/SM™
Processor Resource/Systems Manager™
RACF®
Rational Rose®
Rational®
Redbooks®
Redpaper™
Redbooks (logo) ®
Resource Link™
RMF™
Sysplex Timer®
System p®
System Storage®
System x®
System z10®
System z9®
System z®
Tivoli®
WebSphere®
z/Architecture®
z/OS®
z/VM®
z/VSE™
z10™
z9®
zSeries®

The following terms are trademarks of other companies:

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.

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

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

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, the IT industry has 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. 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, which ensures a tighter fit with IT requirements of an enterprise.

This IBM Redbooks® publication introduces the revolutionary scalable IBM zEnterprise System, which consists of the IBM zEnterprise 196 (z196) or the IBM zEnterprise 114 (z114), the IBM zEnterprise BladeCenter® Extension (zBX), and the IBM zEnterprise Unified Resource Manager. IBM is taking a bold step by integrating heterogeneous platforms under the proven System z hardware management capabilities, while extending System z qualities of service to those platforms. The z196 and z114 are general-purpose servers that are equally at ease with compute-intensive workloads and with I/O-intensive workloads. The integration of heterogeneous platforms is based on IBM BladeCenter technology, allowing improvements in price and performance for key workloads, while enabling a new range of heterogeneous platform solutions. The z196 and z114 are at the core of the enhanced System z platforms, which are designed to deliver technologies that businesses need today along with a foundation to drive future business growth.

The changes to this edition are based on the System z hardware announcement dated July 12, 2011.

This book provides basic information about z196, z114, zBX, and Unified Resource Manager capabilities, hardware functions and features, and associated software support. It is intended for IT managers, architects, consultants, and anyone else who wants to understand the elements of the zEnterprise System. For this introduction to the zEnterprise System, readers are not expected to be 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 System z Networking and Connectivity Architect at the International Technical Support Organization, Poughkeepsie Center.

Octavian Lascu is a Project Leader at the International Technical Support Organization, Poughkeepsie Center. He writes extensively and teaches IBM classes worldwide. Before joining the ITSO, Octavian worked in Global Technology Services, Romania, as a Senior IT Consultant.

Erik Bakker is a Senior IT Specialist working for IBM Server and Technology Group in the Netherlands. During the past 25 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.

Parwez Hamid is an Executive IT Consultant working for the IBM Server and Technology Group. During the past 38 years he has worked in various IT roles within IBM. Since 1988 he has worked with a large number of IBM mainframe customers and spent much of his time introducing new technology. Currently, he provides pre-sales technical support for IBM System z product portfolio and is the lead System z technical specialist for the UK and Ireland. Parwez has co-authored a number of IBM Redbooks publications and prepares technical material for the worldwide announcement of System z Servers. Parwez works closely with System z product development in Poughkeepsie and provides input and feedback for future product plans. Additionally, Parwez is a member of the IBM IT Specialist professional certification board in the UK and is a Technical Staff member of the IBM UK Technical Council, which is made of senior technical specialists representing all IBM client, consulting, services, and product groups. Parwez teaches and presents at numerous IBM user group and IBM conferences.

Fernando E 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 29 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 includes presenting and consulting in architectures and infrastructures. He provides 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 publications 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.

Frank Packheiser is a Senior IT Specialist in System z at the Field Technical Sales Support office in Germany. He has over 20 years of experience with mainframe servers. He has worked for 10 years for the IBM Education Center in Germany, developing and providing professional training. He also provides professional services to System z clients in Europe, the Middle East, and North Africa.

Vicente Ranieri Jr. is an Executive IT Specialist at STG Advanced Technical Skills (ATS) team supporting System z in Latin America. He has more than 30 years of experience working for IBM. Ranieri is a member of the zChampions team, a worldwide IBM team who participates in the creation of System z technical roadmap and value proposition materials. Besides co-authoring several Redbooks publications, he has been an ITSO guest speaker since 2001, teaching the System z security update workshops worldwide. Vicente 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 Ranieri is certified as Distinguished IT Specialist by the Open group. Vicente is a member of the Technology Leadership Council – Brazil (TLC-BR), and he is also a member of the IBM Academy of Technology.

Karl-Erik Stenfors is a Senior IT Specialist in the PSSC Customer Center in Montpellier, France. He has more than 42 years of working experience in the mainframe environment, as a systems programmer, as a consultant with IBM 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 workgroup. His current responsibility is to execute System z Early Support Programs in Europe and Asia.

Esra Ufacik is an STG System z Client Technical Specialist. She holds a B.Sc. degree in Electronics and Telecommunication Engineering from Istanbul Technical University. Her IT career started with a Turkish Bank as a z/OS systems programmer. Her responsibilities were maintaining a parallel sysplex environment with DB2® data sharing, planning and executing hardware migrations, installing new operating system releases, middleware releases, and system software maintenance, and generating performance reports for capacity planning. Esra joined IBM in 2007 as a Software Support Specialist in Integrated Technology Services where she work with mainframe customers within the country. Acting as their account advocate, Esra was assigned to a Software Support Team Leader position in ITS and got involved in several projects. Since 2010 she has been with STG, where her role covers pre-sales technical consultancy, conducting System z competitive assessment studies, presenting the value of System z to various audiences, and assuring technical feasibility of proposed solutions. Esra is also a guest lecturer for *System z and large scale computing* classes, which are given to undergraduate computer engineering students within the scope of Academic Initiative.

Chen Zhu is a Senior System Service Representative at the IBM Global Technical Services in Shanghai, China. He joined IBM in 1998 to support and maintain System z products for clients throughout China. Chen has been working in the Technical Support Group (TSG) providing second-level support to System z clients since 2005. His areas of expertise include System z hardware, Parallel Sysplex, and FICON® connectivity.

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

Connie Beuselinck, Patty Driever, Tom Dewkett, Steve Fellenz, Jeff Frey, Les Geer, Michael Jordan, Gary King, Bill Kostenko, Jeff Kubala, Charles Reimers, Kelly Ryan, Lisa Schloemer, Jaya Srikrishnan, Frank Wisnewski, Peter Yocom, and 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

Laurent Boudon, Gerard Laumay
IBM Montpellier, France

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 email 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

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/pages/IBM-Redbooks/178023492563?ref=ts

- Follow us on twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

1

Proposing an IT infrastructure (r)evolution

A great deal of experimentation has occurred in the marketplace in the last decades, due to the explosion in applications, architectures, and platforms. Out of these experiments, and with the generalized availability of the internet and 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 use modes by a growing number of users, and the most rapidly growing type of user is not people, but devices. New services are being offered and new business models are being implemented.

1.1 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 is not 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, new acquisitions, new software licences, and recertification becomes necessary. 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 is required for greater levels of workload performance and scalability, which results in additional system heterogeneity.

1.2 Future infrastructures

There is a growing awareness that the 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 therefore necessary to define the target infrastructure and how to affect the change.

As infrastructure changes, 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 run a business cost-effectively, while supporting growth and innovation.

In the IBM vision of the future, transformation of the IT delivery model is based on new levels of efficiency and service excellence for businesses, driven by and from the data center. This evolution prepares systems to:

- Handle massive scale and integration.
- Capture, store, manage, and retrieve vast amounts of data
- 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. Freedom to select the best placement for the applications' component is a requirement for overall efficiency and IT optimization. One size really does *not* fit all.

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 demands on the infrastructure, which needs to be dynamic, automated, and possess policy-based resource provisioning, deployment, reallocation, and optimization. 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. Infrastructures and the systems in the infrastructure need to become smarter.

A smarter infrastructure allows organizations to better position themselves, 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 seems then that the key to solving today's problems and unlocking the road to the future is based on a smart IT infrastructure, composed of diverse systems that are flexible, highly virtualized, automated, and tightly managed.

The IBM zEnterprise System implements a revolutionary concept designed to help overcome fundamental problems of today's IT infrastructures and simultaneously provide a foundation for the future. The zEnterprise System, with its diverse platform management capabilities, provides many of these answers, offering great value in a scalable solution that integrates and simplifies hardware and firmware management and support, and the definition and management of a network of virtualized servers, across multiple diverse platforms.

1.3 Introducing the zEnterprise System

The zEnterprise System brings about a revolution by pulling together diversified systems and unifying them under a centralized end-to-end infrastructure management capability, while offering expanded and evolved traditional System z capabilities.

With zEnterprise, a system of systems can be created where the virtualized resources of the IBM zEnterprise 196 (z196) or IBM zEnterprise 114 (z114) and select IBM blade-based servers, housed in the zEnterprise BladeCenter Extension (zBX), are pooled together and jointly managed through the zEnterprise Unified Resource Manager.

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. This embodies "Freedom by Design" into the system architecture.

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 collocate all of the workload components under the same management platform and thereby benefit from uniformly high qualities of service is quite appealing and provide tangible benefits and a rapid ROI.

Terminology: In the remainder of the book we use the designation *zEnterprise CPC* when referring to the z196 and the z114, unless there is a need to distinguish between them.

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

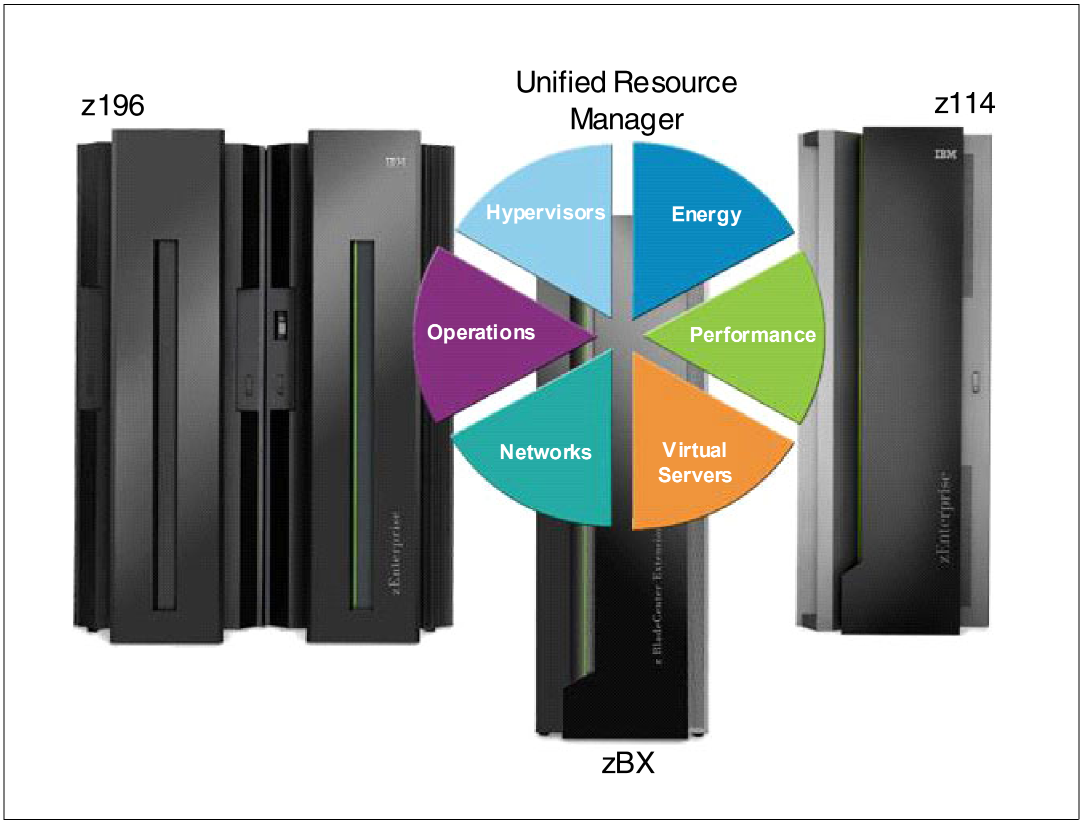

Figure 1-1 zEnterprise System and management capabilities

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

The z114 is built on the same technology as the z196 designed specifically for small and medium enterprises. It provides the same level of qualities of service as the z196 in a granular packaging technique that allows for right sizing your z114 to meet the needs of your business. It uses the same chip technology as the z196, running at 3.8 GHz. Up to 10 PUs can be configured. For example, a five-way general purpose CP system with a mixture of specialty engines, or a system with 10 Integrated Facility for Linux (IFLs). With up to 256 GB of RAIM memory, the z114 is well placed as an entry system for a variety of workloads.

The workload support capabilities of the zEnterprise CPCs are complemented by the zBX. Each zBX can have up to four frames housing up to eight BladeCenters, each with up to 14 select IBM blades, for a maximum of 112 blades.

The zBX supports the following types of blades:

- Select IBM POWER7™ blades

  These blades run a wide variety of applications on the AIX® operating system.

- Select IBM System x® blades

  These blades run Linux applications. IBM has issued a Statement of Direction regarding the support of the Microsoft Windows operating system.

- Special-purpose blades

  These blades are dedicated to specific tasks. Examples are the IBM WebSphere® DataPower® XI50 for zEnterprise (DataPower XI50z) blade and the IBM Smart Analytics Optimizer solution.

High-speed virtualized networks support data communication between applications running in the zEnterprise CPC and the zBX, and management of 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. This ensures a tighter fit with the IT requirements of the entire enterprise and gives better total cost of ownership.

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

1.4 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 that they need.

With the zEnterprise System a new concept in IT infrastructures is being introduced: zEnterprise *ensembles*. 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. 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.

1.4.1 zEnterprise ensemble management

IBM is widening the enterprise role and application domain of System z by integrating additional heterogeneous infrastructure components 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-1980s, 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 zEnterprise CPC, with its optional zBX, makes up a *node* of a zEnterprise ensemble. A zEnterprise ensemble is composed of up to eight members, with up to eight zEnterprise CPCs and up to 896 blades housed in up to eight zBXs, plus dedicated integrated networks for management and data, and the Unified Resource Manager function. With the Unified Resource Manager, the zEnterprise System provides advanced end-to-end management capabilities for the diverse environments within the zBX.

The zBX components are configured, managed, and serviced the same way as the other components of the zEnterprise CPCs. 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, while retaining the freedom to place each workload where it fits best.

It is worth mentioning that the concept of an ensemble is similar to that of a cloud. An ensemble provides 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 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 policy requirements.

1.4.2 z196 and z114 virtualized environments

The zEnterprise CPCs support advanced server consolidation and offer the best virtualization in the industry. Virtualization is built into the hardware, and software virtualization is also available.

Processor Resource/Systems Management

Processor Resource/Systems Management (PR/SM™) is responsible for hardware virtualization of the system. It is always active and has been enhanced to provide additional performance and platform management benefits. PR/SM technology on previous System z systems has received Common Criteria EAL5[1] security certification. Each logical partition is as secure as an isolated system.

Up to 60 logical partitions (LPARs) can be deployed on a z196 and up to 30 on a z114. Each one can run any of the supported operating systems:

- z/OS
- z/VM
- z/VSE™

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

- z/TPF
- Linux on System z

Software virtualization

The zEnterprise CPCs offer software virtualization through z/VM. z/VM's virtualized z/Architecture systems, known as virtual machines, support all operating systems and other software supported in a logical partition. 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 zEnterprise CPC.

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 are:

- Isolating production, test, training, and development environments
- Supporting back-level applications
- Testing new hardware configurations without buying the actual hardware
- Enabling parallel migration to new system or application levels, and providing easy back-out capabilities

The Unified Resource Manager exploits z/VM's management API to provide a set of resource management functions for the z/VM environment.

1.4.3 zBX virtualized environments

On the zBX, the IBM POWER7 blades run PowerVM™ Enterprise Edition to create a virtualized environment that is similar to the one found in IBM Power Systems™ servers. The POWER7-based LPARs run the AIX operating system. The IBM System x blades are also virtualized. The integrated System x hypervisor uses Kernel-based virtual machines. Support is provided for Linux, and IBM has also announced the intent to support Microsoft Windows. Management of the zBX environment is done as a single logical virtualized environment by the Unified Resource Manager.

1.4.4 Flexibility and security

On the z196, the subcapacity settings offering is expanded to up to 15 central processors (CPs). These processors deliver 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 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 system 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 14.

The z114 offers fine granularity, with 130 different capacity settings. This allows smooth growth from entry-level to medium-sized systems on one-way to five-way systems. The faster microprocessors and new technologies allow it to deliver up to 12% more capacity than the z10 BC.

Most hardware upgrades can be installed concurrently. The z196 reaches new availability levels by eliminating various pre-planning needs and other disruptive operations. The z114 also has higher availability than its predecessor.

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

IBM has further enhanced and extended the zEnterprise leadership with improved secured 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 stronger algorithm implementation for 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 number of enterprises are reaching the limits of available physical space and electrical power at their data centers. The extreme virtualization capabilities of the zEnterprise enable the creation of dense and simplified infrastructures that are highly secure and can lower operational costs.

Further simplification is possible by exploiting the zEnterprise HiperSockets™[2] and z/VM virtual switch functions. These can 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 actual simplification cases, cables have been reduced by 97%.

A paper, written by atsec Information Security Corp, on Payment Card Industry compliance, recognizes the inherent qualities of the mainframe and the simplification in the infrastructure that it can provide. The paper can be found at the following web page:

http://www.atsec.com/downloads/white-papers/PCI_Compliance_for_LCS.pdf

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

[2] For a description of HiperSockets, see "HiperSockets" on page 20. The z/VM virtual switch is a z/VM system function that uses memory to emulate switching hardware.

1.4.5 A cornerstone of a smart IT infrastructure

Summing up the following 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
+  Secure access to data (and the network)
+  Additional platforms and the Unified Resource Manager
-------------------------------------------------------------------------------------
=  A flexible infrastructure based on a integrated heterogeneous environment, on which a
   wide range of workloads can be seamlessly deployed and managed
```

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 system. 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. It includes the following functions:

- Access management to authenticate and authorize who can access specific business services and associated IT resources.
- Use management to drive maximum use of the system. Unlike other classes of servers, System z systems are designed to run at 100% utilization all 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 zEnterprise brings all this innovation while helping customers make use of their current investment in the mainframe, and 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, therefore, the most powerful tool available to reduce cost, energy, and complexity in enterprise data centers.

The zEnterprise CPCs continue the evolution of the mainframe, building upon the z/Architecture definitions. IBM mainframes traditionally provide an advanced combination of reliability, availability, security, scalability, and virtualization. The zEnterprise CPCs have been designed to extend these capabilities into heterogeneous resources and are 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.

1.5 zEnterprise CPCs technical description

In this section we briefly review the most significant characteristics of the z196 and the z114. Chapter 3, “zEnterprise System hardware overview” on page 45, provides further technical details.

1.5.1 IBM zEnterprise 196

The z196 employs leading-edge silicon-on-insulator (CMOS 12s-SOI) and other technologies, such as InfiniBand and Ethernet. The z196 provides benefits such as 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:

- M15
- M32
- M49
- M66
- 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 options are factory installed, and it is not possible to convert in the field from one to the other, so careful consideration is given to current and future needs.

The IBM commitment to, and sustained investment in, the System z is well portrayed in Figure 1-2, which provides a comparison of z196 with previous System z systems according to 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 z10 only exploits a subset of its designed I/O capability.)

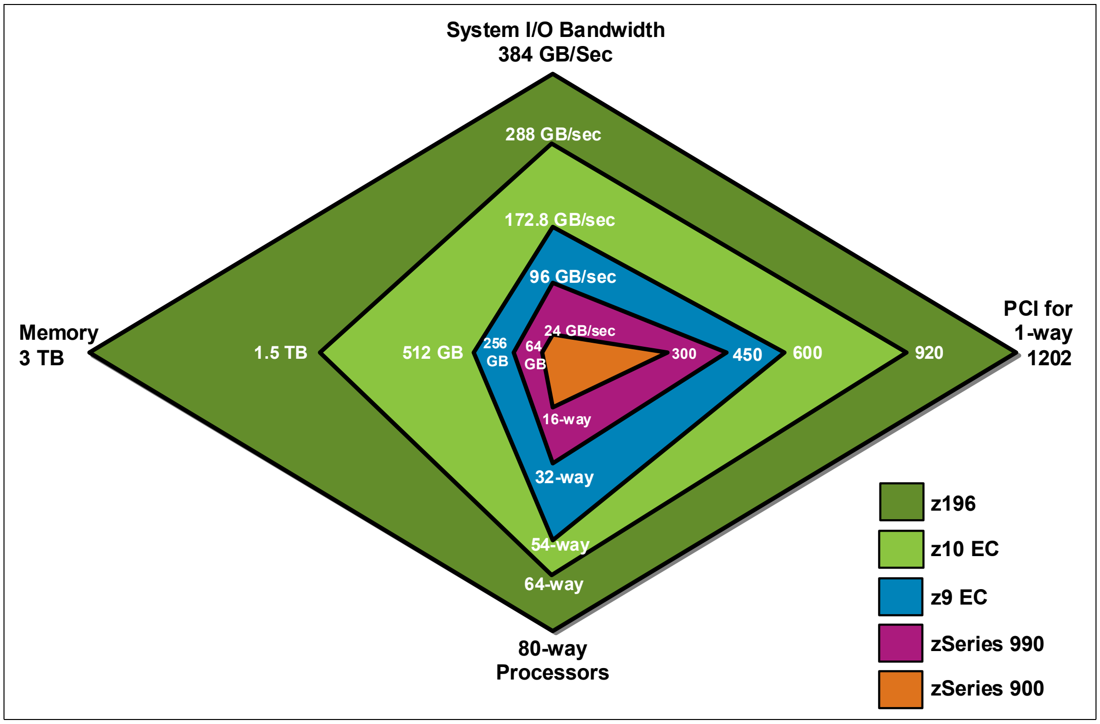

Figure 1-2 System z design comparison for high end systems

The z196 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 following elements:

- The processor cage
- Modular cooling units (different for water and air cooling)
- I/O cages, I/O drawers, and PCIe I/O drawers, available in several combinations
- Power supplies
- An optional internal battery feature (IBF)

The Z frame contains the following elements:

- Two system Support Elements (SEs)
- I/O cages, I/O drawers, and PCIe I/O drawers, available in several combinations
- Power supplies
- An optional IBF

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

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 inches of depth to the back of both system frames).

1.5.2 IBM zEnterprise 114

The z114 employs the same technologies as the z196. Compared to the z10 BC, the z114 provides benefits such as higher frequency chips, out-of-order execution, improved availability, and enhanced on demand options. In addition, it supports the latest offerings for data encryption.

Figure 1-3 shows a comparison of z114 with previous System z systems along four major attributes:

- Single engine processing capacity (based on Processor Capacity Index (PCI))
- Number of engines (Up to five PUs can be added as speciality engines.)
- Memory
- I/O bandwidth

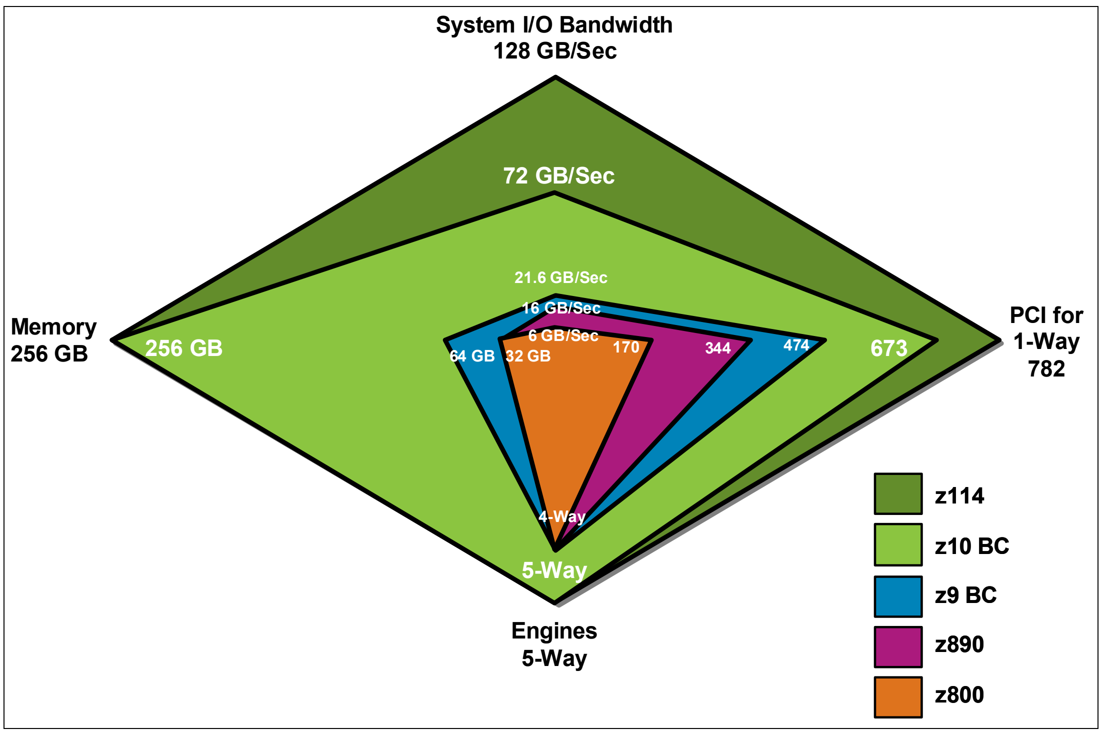

Figure 1-3 System z design comparison for mid-sized system

Two models of the z114 are offered:

- M05
- M10

The names represent the maximum number of PUs that can be configured in the model.

The z114 system architecture ensures continuity and upgradeability from the z10 BC and z9 BC designs. The z114 Model M10 can be upgraded to a z196 M15.

The z114 is a single-frame system. It has a machine type designation of 2818. The single frame of the z114 is known as the A frame. The A frame contains the following elements:

- The processor drawers (up to two)
- I/O drawers and PCIe I/O drawers, available in several combinations
- Power supplies
- Two system SEs
- An optional internal battery feature (IBF)

1.5.3 Processor cages and drawers

The same technologies are used in the z196 and z114. However, the differences in capacity and scalability between the family members justify the use of different packaging for certain elements. One example is the housing of the PUs.

z196 processor cage

On a z196 the PUs are 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 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
- Chilled water

In any model of the system, two PUs are designated as spares, and each individual PU can be transparently spared, as with the z10. 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 this number 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.

z114 processor drawer

The z114 employs a processor drawer. The drawer was introduced with the z10 BC. The z114 Model M05 has one processor drawer and the Model M10 has two. The processor drawer is air-cooled.

Each processor drawer houses three single-chip modules (SCMs), memory, I/O drawer connectors, and (optionally) coupling link connectors. The z114 uses the same chip technology as the z196.

There are two PU SCMs and the third is a storage control single-chip module (SC SCM). One PU SCM has three PUs, while the other has four PUs, for a total of seven, running at 3.8 GHz each.

On the M05 any unassigned PU can be used as a spare. The M10 has two PUs designated as spares. Each individual PU can be transparently spared, as with the z10. This contrasts with previous systems where the chip was the sparing unit.

Up to 256 GB of memory is available, the same as the z10 BC. It is now implemented as a RAIM for enhanced availability. In addition, 8 GB are part of the base system and are reserved for the hardware system area (HSA), making the maximum amount of customer purchasable memory 248 GB. 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 system initialization time, each purchased PU is characterized as one of a variety of types. It is also possible to characterize PUs dynamically. A PU that is not characterized cannot be used. A PU can be characterized in the following ways:

- Central processor (CP)

  The standard processor. For use with any supported operating system and user applications.

- Internal Coupling Facility (ICF)

  Used for z/OS clustering. ICFs are dedicated to this function and exclusively run the Coupling Facility Control Code (CFCC).

- Integrated Facility for Linux (IFL)

  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.

- System Assist Processor (SAP)

  Offloads and manages I/O operations. Several are standard with the z196. More can be configured if additional I/O processing capacity is needed.

- IBM System z Application Assist Processor (zAAP)[3]

  Exploited under z/OS for designated workloads, which include the IBM JVM and XML System Services functions.

- IBM System z Integrated Information Processor (zIIP)[3]

  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.

> **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 if no zAAPs are installed on the system. This capability is offered to enable optimization and maximization of investment on zIIPs.

CP Assist for Cryptographic Function

The zEnterprise CPCs continue to use the cryptographic assist implementation, first deployed in 2003. Further enhancements have been made to the CP Assist for Cryptographic Function (CPACF).

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

CPACF is physically implemented in the quad-core chip by the compression and cryptography accelerators. Each of the two accelerators set is shared by two cores. CPACF-supported functions are:

- Data Encryption Standard (DES)
- Triple Data Encryption Standard (TDES)
- Advanced Encryption Standard (AES) for 128-bit, 192-bit, and 256-bit keys
- Secure Hash Algorithm (SHA):
  - SHA-1: 160 bits
  - SHA-2: 224 bits, 256 bits, 384 bits, and 512 bits
- Message authentication code (MAC)
- Pseudo Random Number Generation (PRNG)

  PRNG is also a standard function supported on the Crypto Express3 features.
- Protected key capabilities

The CPACF functions are supported by z/OS, z/VM, z/VSE, z/TPF, and Linux on System z

1.5.4 I/O subsystem

As with its predecessors, the zEnterprise CPCs have a dedicated subsystem to manage all input/output operations. Known as the channel subsystem, it is made up of the following elements:

- SAP

  System Assist Processor (SAP) is a specialized processor that uses the installed PUs. (Each z196 or z114 PU can be characterized as one of six configurations. For more information see "PU characterization" on page 14.) 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 machine model. These are part of the base configuration of the system.
- HSA

  Hardware system area (HSA) is a reserved part of the system memory containing the I/O configuration. It is used by SAPs. On the z196, a fixed amount of 16 GB is reserved (8 GB on the z114), which is not part of the customer-purchased memory. This provides for greater configuration flexibility and higher availability by eliminating 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 sharing of the channel between multiple logical partitions, with each partition having a unique channel path. The function that allows sharing I/O paths across logical partitions is known as the multiple image facility (MIF).

- 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. The z196 has three subchannel sets[4] and the z114 has two.

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 zEnterprise CPCs I/O subsystem direction is evolutionary, expanding on developments from the z9 and z10, and includes both PCIe and InfiniBand infrastructures (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.

The PCIe I/O bus connects the processor cage (z196) or processor drawers (z114) to the PCIe I/O drawer. The InfiniBand I/O bus connects the processor cage or processor drawers to I/O drawers or an I/O cage (z196 only). PCIe I/O drawers house PCIe I/O cards. I/O cages and I/O drawers house I/O cards.

Peripheral Component Interconnect Express

Peripheral Component Interconnect Express (PCIe) is a standard for computer expansion cards. It includes a serial bus standard used by a large variety of computer platforms. The bus operates at 8 GBps.

The PCI Special Interest Group is responsible for developing and maintaining format specifications.

PCIe in the zEnterprise CPCs provides an internal I/O infrastructure that positions the system for continued support of the industry's direction for high-performance I/O.

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 makes use of 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 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.

[4] Subchannel set 0 can have up to 63.75 K devices (256 devices are reserved), and subchannels 1 and 2 can have up to 64 K devices each.

InfiniBand is being exploited by the zEnterprise CPCs. Within the system the cables from the processor cage to the I/O cages and I/O drawers (*not* the PCIe I/O drawers) carry the InfiniBand protocol. For external usage, InfiniBand (IFB) links are available, which can completely replace the ISC-3 and ICB-4 offerings available on previous systems. They are used to interconnect System z mainframes in a Parallel Sysplex.

1.5.5 I/O connectivity

The zEnterprise CPCs generation of the I/O platform, particularly through the exploitation of PCIe, InfiniBand, enhanced cards and protocols (High Performance FICON for System z — zHPF), is intended to provide significant performance improvements over the previous I/O platform used on previous systems. The offerings include I/O infrastructure elements, the PCIe I/O drawer, and the I/O drawer. On the z196 these are available along the I/O cage supported on predecessor systems. The z114 supports the PCIe I/O drawer and the I/O drawer.

I/O cage

The I/O cage accommodates I/O features. The z196 has a processor 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. I/O cages are not supported on the z114.

I/O drawer

I/O drawers provide increased I/O granularity and capacity flexibility, as compared with the I/O cage. I/O drawers can be concurrently added and removed in the field, an advantage over I/O cages, which also eases pre-planning. The z196 system 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 each can accommodate up to eight I/O features, in any combination. The z114 system supports a maximum of two I/O drawers.

PCIe I/O drawer

PCIe I/O drawers were introduced with the z114 and the z196. They provide for a higher number of cards (four times as much as the I/O drawer and a 14% increase over the I/O cage) and increased port granularity. The PCIe drawers can be concurrently installed and repaired in the field. Each drawer can accommodate up to 32 PCIe I/O features in any combination. Up to five PCIe I/O drawers can be installed on the z196 and up to two on the z114.

I/O features and PCIe I/O features

The zEnterprise CPCs support the following I/O feature types, which can be installed in both the I/O drawers and I/O cages:

- ESCON®
- FICON Express8
- FICON Express4
- OSA-Express3
- OSA-Express2
- Crypto Express3
- ISC-3 coupling links

Both the z196 and the z114 support the following PCIe features, which can be installed only in the PCIe I/O drawers:

- FICON Express8S
- OSA-Express4S

For a description of each I/O feature supported by the zEnterprise CPCs, refer to 3.9, “I/O features” on page 65.

Enterprise Systems Connection (ESCON) channels

The 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. The maximum number of supported ESCON features on the zEnterprise CPCs is 16 (up to 240 ports).

Statement of General Direction[a]: The IBM zEnterprise 196 (z196) and IBM zEnterprise 114 (z114) are the last System z servers to support ESCON channels. IBM plans not to offer ESCON channels as an orderable feature on System z servers that follow the z196 (machine type 2817) and z114 (machine type 2818). In addition, ESCON channels cannot be carried forward on an upgrade to such follow-on servers.

Alternate solutions are available for connectivity to ESCON devices.

a. 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.

Fibre Connection (FICON) channels

FICON channels follow the Fibre Channel (FC) standard and support data storage and access requirements and 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 (no zHPF), 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.

There are some restrictions on combining the FICON Express8S, 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 and FICON Express8S).

FICON Express8 provides significant improvements in start I/Os and data throughput over previous cards. FICON Express8S are PCIe cards and offer better port granularity and improved capabilities. FICON Express8S is the preferred technology.

Statement of General Direction: The IBM zEnterprise 196 and the IBM zEnterprise 114 are the last System z systems to support FICON Express4 features.

Open Systems Adapter (OSA)

The 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 normally are performed by the processor. This can provide significant performance benefits.

There are some restrictions on combining the OSA-Express4S, OSA-Express3, and OSA-Express2 features. The OSA-Express2 10 GbE LR feature is not supported.

> **Statement of General Direction:** The IBM zEnterprise 196 and the IBM zEnterprise 114 are the last System z systems to support OSA-Express2 features.

Cryptography

The Crypto Express3 and Crypto Express3-1P (z114 only) features provide tamper-proof, high-performance cryptographic operations. Each Crypto Express3 feature has two PCI Express adapters. Each Crypto Express3-1P feature has one PCI Express adapter. Each of the adapters can be configured as either a coprocessor or an accelerator:

- Crypto Express3 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 Express3 Accelerator: for Secure Sockets Layer/Transport Layer Security (SSL/TLS) 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/TLS protocol

The features have specialized hardware to perform DES, TDES, AES, RSA, SHA-1, and SHA-2 cryptographic operations. The tamper-resistant hardware security module, which is contained in the Crypto Express3 and Crypto Express3-1P features, is designed to meet the FIPS 140-2 Level 4 security requirements for hardware security models.

The configurable Crypto Express3 features are supported by z/OS, z/VM, z/VSE, and z/TPF (accelerator mode only) and Linux on System z.

Coupling links

Coupling links are used when clustering System z systems running the z/OS operating system. A clustered configuration is known as a Parallel Sysplex and can have up to 32 member nodes. The links provide high-speed bidirectional communication between members of a sysplex. The zEnterprise CPCs support internal coupling links for memory-to-memory transfers (between LPARs), 12x InfiniBand links for distances up to 150 meters (492 feet), and InterSystem Channel-3 (ISC-3) and 1x InfiniBand links for unrepeated distances up to 10 km (6.2 miles).

InfiniBand (HCA3-O) coupling links are being introduced to provide a performance improvement that reduces overhead and latency. Refer to 3.11.3, “InfiniBand coupling links” on page 75, for a technical description.

> **Statement of Direction:** The IBM zEnterprise 196 and the IBM zEnterprise 114 are the last System z systems to offer ordering of the ISC-3 features. Migrating to Parallel Sysplex InfiniBand links should be considered.

HiperSockets

The HiperSockets function is an integrated function of the zEnterprise CPCs 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 use I/O subsystem operations and having to traverse an external network connection to communicate between logical partitions in the same zEnterprise CPC. Therefore, HiperSockets offers significant value in server consolidation by connecting virtual servers.

1.5.6 zEnterprise BladeCenter Extension

The zEnterprise BladeCenter Extension Model 002 (zBX) is available as an option with the z196 and z114 systems and consists of the following:

- Up to four IBM Enterprise racks
- Up to eight BladeCenter[5] chassis (two per rack), with up to 14 blades[6] each
- Select IBM blades, up to 112
- Two Top the Rack (TOR) 1000BASE-T switches for the intranode management network (INMN)

  The INMN provides connectivity for management purposes between the zEnterprise CPC's SE and zBX.
- Two TOR 10 GbE switches for the intraensemble data network (IEDN)

  The IEDN is used for data paths between the zEnterprise CPCs and the zBX, and the other ensemble members.
- 8 Gbps Fibre 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 judgement of the market's needs, additional specialized or general-purpose blades might be introduced[7].

IBM blades

IBM offers a selected set of IBM blades that can be installed and operated on the zBX. These blades have been tested to ensure compatibility and manageability in the zEnterprise CPC environment. The following blades are available:

- Select IBM POWER7 PS701 Express blades
- Select IBM System x blades (HX5 7873 dual-socket 16-core)
- DataPower XI50z blades (double-width)
- Blades for the IBM Smart Analytics Optimizer solution offering

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

[6] Depending on the IBM Smart Analytics Optimizer configuration, this can be either seven or 14 blades for the first BladeCenter chassis.

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

The POWER7 blades offer a virtualized environment through the PowerVM Enterprise Edition hypervisor. The virtual servers run the AIX operating system. The System x blades have an integrated hypervisor using Kernel-based virtual machines, which provides a virtualized environment and runs the Linux operating system.

> **Statement of General Direction:** In the future, IBM intends to offer select IBM System x blades running Microsoft Windows in the IBM zEnterprise BladeCenter Extension Model 002.

IBM DataPower X150z integration appliance

The IBM WebSphere DataPower Integration Appliance XI50 for zEnterprise (DataPower XI50z) has been integrated into the zEnterprise infrastructure. DataPower XI50z is a multifunctional appliance that can help provide multiple levels of XML optimization, streamline and secure valuable service-oriented architecture (SOA) applications, and provide drop-in integration for heterogeneous environments by enabling core Enterprise Service Bus (ESB) functionality, including routing, bridging, transformation, and event handling. It can help to simplify, govern, and enhance the network security for XML and web services.

For a more detailed description of the IBM DataPower X150z integration appliance visit:

http://www-01.ibm.com/software/integration/datapower/xi50z

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 4.9.2, “IBM Smart Analytics Optimizer solution” on page 115.

1.5.7 Unified Resource Manager

The zEnterprise System is a perfect fit in this world of smarter computing, being both the next step in the evolution of System z leadership and a premier solution for centrally managed enterprise cloud environments. It is a true hybrid computing system comprised of virtualized heterogeneous resources that are integrated and managed as a single system by the zEnterprise 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 zEnterprise CPC and zBX, and 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 permanent zEnterprise CPC capacity upgrades.

The functions that pertain to an ensemble are provided by the Hardware Management Console (HMC) and Support Elements. See 2.2, “Unified Resource Manager” on page 28, for more information.

> **Statement of General Direction:** IBM intends to offer Application Program Interfaces (APIs) for IBM zEnterprise Unified Resource Manager. These APIs provide access to the same underlying functions that support the Unified Resource Manager user interface.

1.6 Capacity On Demand

On the zEnterprise CPCs 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.

You can perform the following tasks:

- 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 system while temporary changes are active.

A similar capability is not available with the zBX. For more information, see 4.6, “zEnterprise CPC Capacity on Demand (CoD)” on page 109.

1.7 Software

The zEnterprise System is supported by a large set of software, including independent software vendor (ISV) applications. The extensive software portfolio spans, on the zEnterprise CPCs, 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 ISVs. In addition, any AIX products running today on System p® servers continue to run on the zBX POWER7 blades’ virtualized AIX environment, and System x blades support Linux (see "Statement of General Direction", on page 21). There are also specialized solutions such as the DataPower XI50z appliance.

IBM Exploitation of features might require the latest releases. The supported operating systems for the z196 and z114 include:

- z/OS Version 1 Release 10 or later releases[8]
- z/OS Version 1 Release 9 with the IBM Lifecycle Extension with PTFs
- z/OS Version 1 Release 8 with the IBM Lifecycle Extension with PTFs[9]

[8] The z/OS V1 Release 10 will require IBM Lifecycle Extension after September 2011.
[9] z/OS.e is not supported.

- z/VM Version 5 Release 4 or later
- z/VSE Version 4 Release 2 or later
- z/TPF Version 1 Release 1
- Linux on System z distributions:
  - Novell SUSE: SLES 10 and SLES 11[10]
  - Red Hat: RHEL 5[11]

Operating system support for the IBM blades on the zBX includes:

- For the POWER7 blades, AIX Version 5 Release 3 or later, with PowerVM Enterprise Edition
- For the System x blades, Linux on System x (see "Statement of General Direction", on page 21.)

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 make use of 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 124.

[10] SLES is the abbreviation for Novell SUSE Linux Enterprise Server.
[11] RHEL is the abbreviation for Red Hat Enterprise Linux.

2

Achieving a (r)evolutionary IT infrastructure

We started this book by reviewing the current IT infrastructures and the increasing and diversifying demands that are being placed upon them. This growth has, so far, been sustained by "throwing hardware at it," but that way is no longer tenable. The resulting complexity, caused by the large number of components and interconnections, and the lack of common operational tools able to provide global and unified views of the enterprise's IT infrastructure, is reaching the point of being unmanageable.

At the same time, new infrastructure paradigms have emerged, namely, cloud computing. The ability to provision a piece of infrastructure (a virtual server, for instance) at will, or order a service (and implicitly the virtual server where it runs) demands the utmost flexibility and governance from the underlying infrastructure. It requires very good insight into the interplay between the infrastructure and workloads, to achieve flexibility and manageability.

Where "the cloud" offers standardized elements and patterned services, enterprises, bound to their history, have largely heterogeneous environments and many tailored or purposefully built applications and computing islands.

Yet, to continue to meet users' requests and align with business guidelines, both need to minimize costs without compromising flexibility, security, and availability. Both require the use of management disciplines with end-to-end views of the infrastructure and the applications deployed across it.

These requirements are best met by environments where resources can be finely grained and easily shared, yet subject to tight security and governance controls. Large servers excel in these areas and specialized application needs can be met by appliances.

Having recognized the problem, and commanding a large spectrum of technologies, IBM is uniquely positioned to provide a solution. Management of large, heterogeneous, workloads and providing finely grained and shared-all resources is the defining principle of IBM System z and is sustained by its built-in extreme virtualization.

The second part, embracing and enhancing heterogeneous environments, can also be achieved by totally virtualizing those environments and extending to them the infrastructure management capabilities first pioneered and developed on System z.

That was the reason to conceive of and offer the IBM zEnterprise System, and its most visible value is the flexible, manageable, integrated, heterogeneous, virtualized infrastructure that provides freedom of choice, by design.

As stated previously, the IBM zEnterprise System has three main components:

- The zEnterprise CPC, implementing the System z platform environments
- The zEnterprise BladeCenter Extension (zBX), implementing POWER® and System x environments, and specialized solutions and appliances
- The Unified Resource Manager, providing the overall management capabilities for the other components

In the remainder of this chapter we focus on the infrastructure provided by zEnterprise by:

- Addressing the role of virtualization in zEnterprise
- Describing the Unified Resource Manager's major functions
- Describing virtualization management and its components involved in the definition and management of virtualized servers, networks, and storage
- Describing other management-related topics such as performance and energy management, and the role of the Hardware Management Console (HMC)

To provide you with an overview of how zEnterprise might be beneficial to your organization, we:

- Review workload concepts.
- Characterize workloads and their requirements.
- Discuss cloud computing concepts and how zEnterprise provides value to a cloud implementation.
- Review zEnterprise qualities of service and the business value that they provide.

2.1 zEnterprise ensembles and virtualization

zEnterprise offers an integrated environment where heterogeneous resources are fully virtualized. This allows sharing of the real hardware resources while maximizing its utilization, which also contributes to lowering the cost of ownership. Other areas of potential savings with virtualized resources are reductions in software licensing, lowered risk, less idle time, and the ability to significantly reduce the time required to introduce new technologies, accelerating the introduction of innovative and enhanced client services.

Typically, virtualized resources include processor units, memory, networking, storage, and the hypervisors that manage them. To execute a workload, an image (containing the relevant operating system, the middleware stack, and the business applications) is also required. To get the most from the infrastructure, good monitoring and responsive management is required.

The zEnterprise System provides all those capabilities. It provides enterprise-class heterogeneous platform deployment spanning mainframe, UNIX, and x86 technologies and also offers an integrated resource management capability able to unify workload management by delivering workload-based resource allocation and provisioning.

In addition to its large scalability characteristics, a zEnterprise System is capable of acting as a node in an ensemble. An ensemble is a collection of up to eight nodes that are managed as a single logical virtualized system. Each node comprises a zEnterprise CPC and, optionally, an attached zEnterprise BladeCenter Extension (zBX). The zEnterprise Unified Resource Manager enables provisioning and management of the ensemble.

Figure 2-1 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 HMC and Support Element (SE).

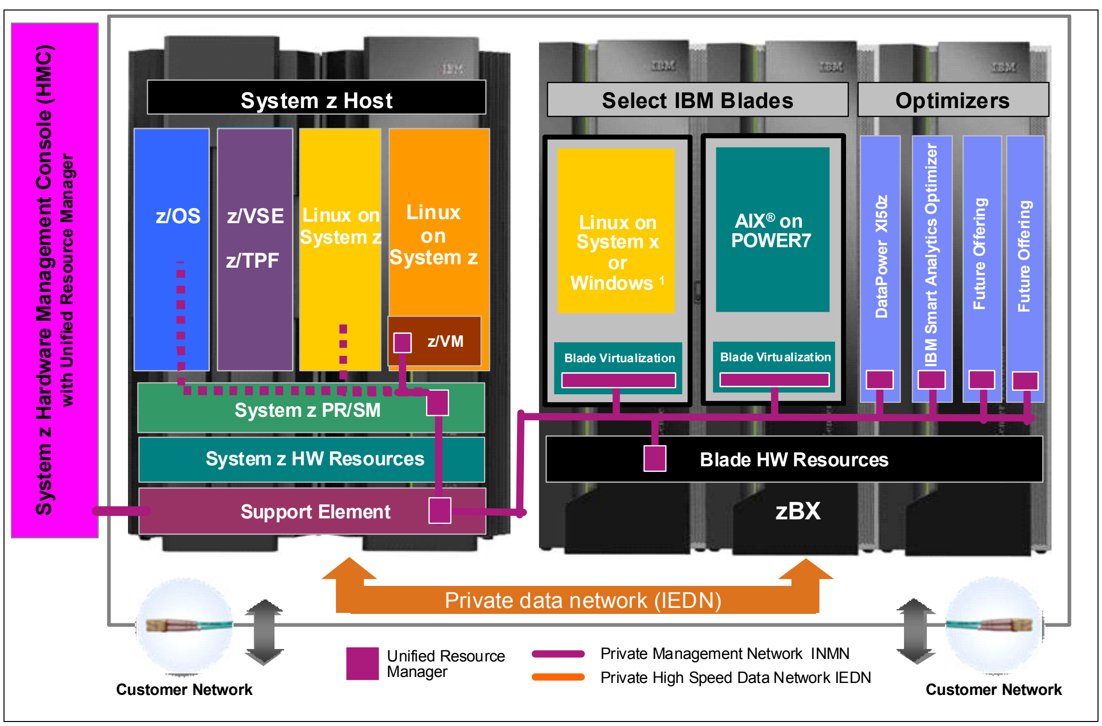

Figure 2-1 A logical view of a one-node ensemble with Unified Resource Manager

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.

The zEnterprise System has been designed to optimize and better manage diversified workloads. All of the ensemble's components contribute towards providing a flexible infrastructure and management capabilities to allow meeting the enterprise's objectives.

2.2 Unified Resource Manager

The Unified Resource Manager is an integral part of the zEnterprise System. It provides end-to-end management of zEnterprise CPCs and zBX resources, and virtualized environments, with the ability to align those resources according to individual workload requirements.

Through virtualization, the physical resources can be shared among multiple workloads. Most likely, the workloads have varying policies with different objectives. The Unified Resource Manager's goal is to fulfill the objectives of the workload policies in the most optimal and efficient way.

The zEnterprise System offers resource management functions as part of the Unified Resource Manager. The functions are delivered in tiers, by two operational suites. Within the Unified Resource Manager several roles are defined. This promotes security through task isolation and authorization.

Resource management suites

The functions delivered by the Unified Resource Manager are accessed through the HMC and provide the following capabilities:

- Integrated hardware management across all elements of the system, the zEnterprise CPCs, and the zBX and integrated networks
- 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 managed through the isolated internal platform management network
- Virtual server life-cycle management, enabling uniform directed and dynamic virtual server provisioning across all hypervisors from a single 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:

- Manage
- Automate/Advanced Management

Refer to the following IBM Redbooks publications to read more about Unified Resource Manager functions and capabilities:

- *IBM zEnterprise Unified Resource Manager,* SG24-7921
- *IBM zEnterprise 196 Technical Guide,* SG24-7833
- *IBM zEnterprise 114 Technical Guide,* SG24-7954

2.3 Virtualization management

The purpose of virtualization management is to define and manage the virtualized resources in the ensemble. Functions are available to create the virtual servers, virtual network components, and virtual storage volumes. Some of the virtualization is provided by the hypervisors and some by the other components of the ensemble. Functions to manage the hypervisors and other virtual resources are provided by the Unified Resource Manager firmware through the HMC.

2.3.1 Hypervisor management

A *hypervisor* is control code that manages multiple independent operating system images. Hypervisors can be implemented in software or hardware. The IBM Unified Resource Management works with set of hypervisors (PR/SM, z/VM, PowerVM Enterprise Edition, and the System x integrated hypervisor) 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. With the exception of z/VM installation and maintenance procedures, which have not changed, 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. See Figure 2-1 on page 27 for a representation of the links between hypervisors and the Unified Resource Manager.

2.3.2 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.

Virtual server life-cycle management enables 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.

Virtual servers and their resources are defined through the HMC. The resources available to be provisioned for a virtual server are the:

- Hypervisor type to be used
- Number of processors

- Memory size
- Assigned network devices
- Assigned storage devices
- "Boot" options for the operating system

When provisioning a virtual environment to support a workload, the relevant platform hypervisors will provision the virtual servers and their associated resources as defined in *virtual server* definitions. You describe a given workload and its associated resource requirements using the HMC. Refer to Figure 2-1 on page 27 to see how the HMC can communicate with the hypervisors and manage virtual servers.

2.3.3 Virtual network management

One of the design objectives of zEnterprise is to simplify the infrastructure without compromising qualities of service such as manageability and security. Networking is an area in which dramatic simplification is possible. Networking is a pervasive component of an ensemble. Figure 2-2 shows a representation of the important networks contained in one node of an ensemble.

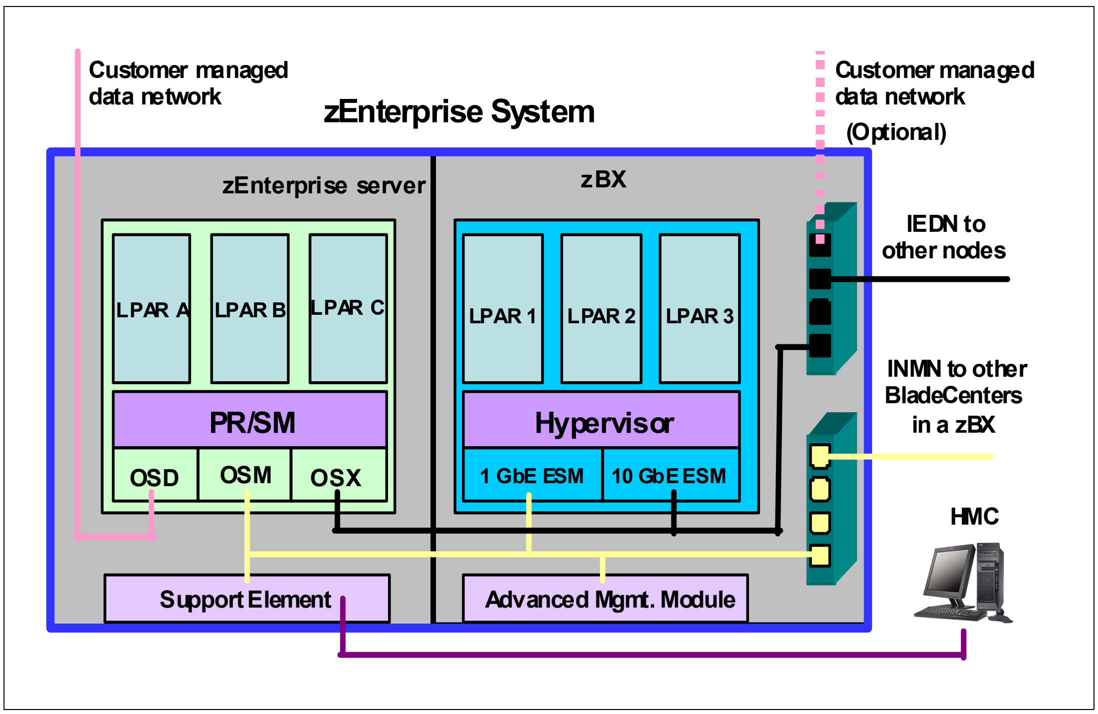

Figure 2-2 Networks contained within an ensemble

Networks in an ensemble node include:

- 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 interconnects all nodes of the ensemble, including all z196 or z114, and zBX.

- Customer-managed management network: Also know as the HMC LAN, this network provides the communication link between Hardware Management Consoles (HMCs) and the nodes of the ensemble.
- Customer data network: This network represents your existing enterprise data communication network. In addition, this network can optionally be connected directly to the IEDN, depending our your configuration requirements.

The IEDN is the network used for system and application communication within an ensemble. It exists only within an ensemble, although it might also have a connection to the customer data network outside the ensemble. It is implemented as a 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. While there are physical network switches that are part of the IEDN, the appearance of a single network is maintained through virtualization.

The physical construction of the IEDN contributes to the security and reliability of the ensemble. All the physical network switches are inside the 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 by the Unified Resource Manager.

By virtualizing the network definitions it is possible to 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 fully utilize the physical network capacity while still meeting your organization's security requirements.

As mentioned, in some cases in might be appropriate to connect the customer data network to the IEDN. The network configuration tasks allow specific ports on the TOR switches to be configured for attachment to your existing data network, external to the ensemble, and can impose restrictions on the attaching network.

2.3.4 Virtual storage management

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 requires access to Fibre Channel Protocol (FC) storage (SCSI) disks. FICON connectivity for System z workloads is unchanged.

The Storage Administrator role of Unified Resource Manager is responsible for allocating storage from physical storage pools to support the ensemble's virtual servers. The allocation is performed based on input from the Server Administrator role (of Unified Resource Manager). The Storage Administrator defines and assigns resources and access rights, and provides separate Storage Access Lists (SALs) for each hypervisor required to support the virtual servers. (A SAL contains the accessible storage resources for a hypervisor.)

Figure 2-3 illustrates the relationship of the SAL to the hypervisor and virtual server.

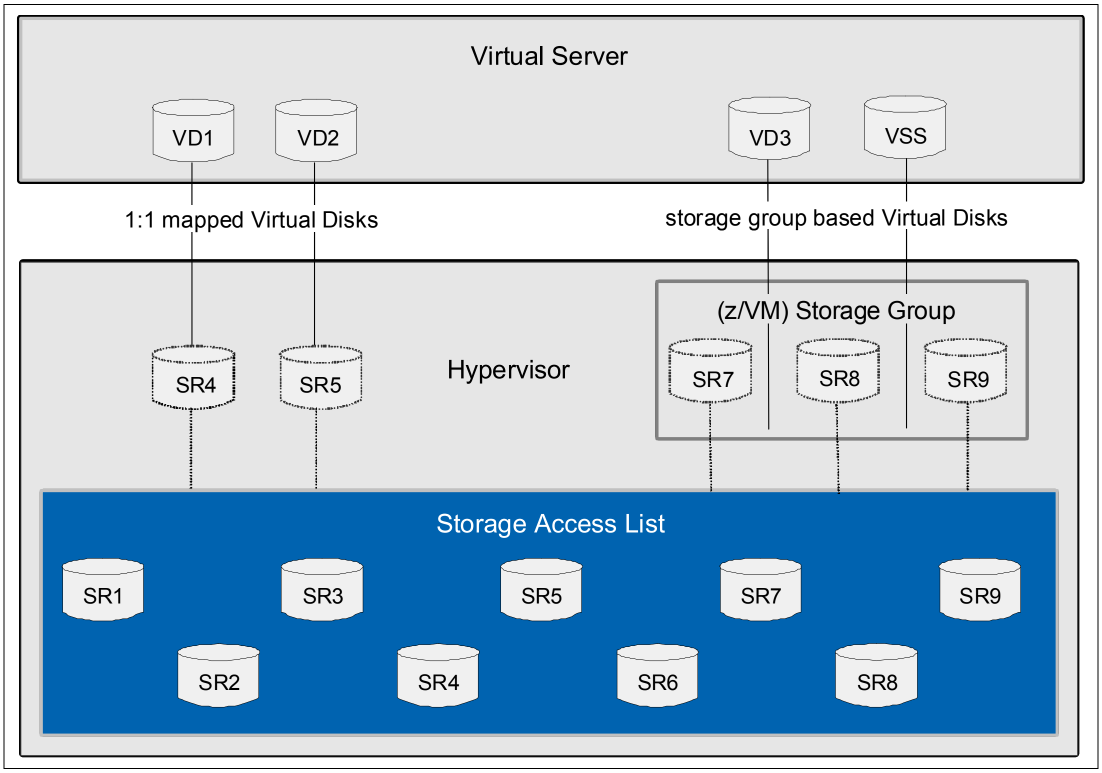

Figure 2-3 Storage virtualization

The Unified Resource Manager provides an abstraction from the underlying technologies. For Storage Virtualization this means providing a simplified storage management interface with common steps across different hypervisors.

2.4 Performance management

Built-in, end-to-end views of the infrastructure and of the workloads in it are one of the differentiating points of a zEnterprise System. Performance is one of such views.

The performance manager is the component of Unified Resource Manager responsible for monitoring, managing, and reporting goal-oriented resources across the zEnterprise System. The code is structured and packaged as System z firmware, and the intercomponent communication is performed over the trusted intranode management network (INMN).

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 goal-oriented management policies defined to and implemented by the performance manager. The ensemble performance manager is designed to:

- Provide a set of performance monitoring functions that allow an administrator to understand whether the performance goals of the deployed workloads are being achieved.

  If a performance goal is not being achieved, this monitoring support helps 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.

2.5 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 use 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 energy management functions of the Unified Resource Manager, see *IBM zEnterprise 196 Technical Guide,* SG24-7833, or *IBM zEnterprise 114 Technical Guide,* SG24-7954.

2.6 Physical resources 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.

2.6.1 Role of the HMC in an ensemble

The Unified Resource Manager is installed in the HMC and, alongside other functionality, enables extending those tasks to an ensemble.

The HMC allows viewing and managing multinodal configurations 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 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 local CPCs and zBXs. This represents a well-layered structure that supports the components of the ensemble.

The functions that are available on the HMC are determined by the Unified Resource Manager suite deployed: Manage, Advanced Management, or Automate. 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 possesses 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 zEnterprise hypervisors:

- PR/SM
- z/VM
- PowerVM Enterprise Edition
- System x blades integrated hypervisor (using Kernel-based virtual machines)

Virtual machines can be managed from their inception all the way through monitoring, migration, and policy-based administration during their deployment.

The HMC is the authoritative owning (stateful) component for Unified Resource Manager configuration and policies the scope of which spans all of the managed nodes in the ensemble. In addition, the HMC has an active role in ongoing system monitoring and adjustment.

Typical functions that can be performed from an HMC go beyond the simple operational start/stop actions on virtual servers, and also include instantiating an ensemble, defining virtual servers and workloads, and assigning those virtual servers to one or more workloads. This requires both the Manage and Automate suites of Unified Resource Manager to be available. The suites are orderable features of the zEnterprise System.

2.6.2 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 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 use 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.

2.6.3 Serviceability

The serviceability function for the components of the ensemble is delivered through the HMC/SE constructs, as for previous System z servers. From a serviceability point of view, all the components of the ensemble, including the zBX, are treated as System z features, similarly to the treatment of I/O and other System z features.

All the functions for managing the components are 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. All service reporting, including call-home functions, is 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).
- Blades can be acquired and deployed by the customer. IBM provides a list of selected blade products that can participate in an ensemble.

2.7 Benefiting from a (r)evolutionay infrastructure

When distilled to a single central theme, the zEnterprise System ensemble has been created simply to provide an optimized infrastructure in which multiple workloads can be deployed across heterogeneous environments and managed under a common umbrella. The computing resources of different HW platforms are managed as a single system that is:

- Tuned for the task and optimized across the infrastructure
- Managed end-to-end for flexible delivery of high value services, similarly to a cloud
- Designed for enterprise-wide real-time data modeling

Adopting the zEnterprise System and extracting benefit from it can be done gradually, and does not require disruptive moves such as those popularized under the names *lift and shift* and *rip and replace*. In classic IBM style, protection of the client’s investment is maximized.

Client's IT infrastructures are all unique and different in areas such as virtualization, monitoring, and automation. A single approach does *not* fit all, and there are multiple points of entry to the adoption of an integrated, heterogeneous, virtualized infrastructure. Figure 2-4 shows a simplified view of a progressive adoption process, where building blocks are added to achieve an enterprise class infrastructure for heterogeneous workload deployment. Again, one can notice a resemblance with cloud computing.

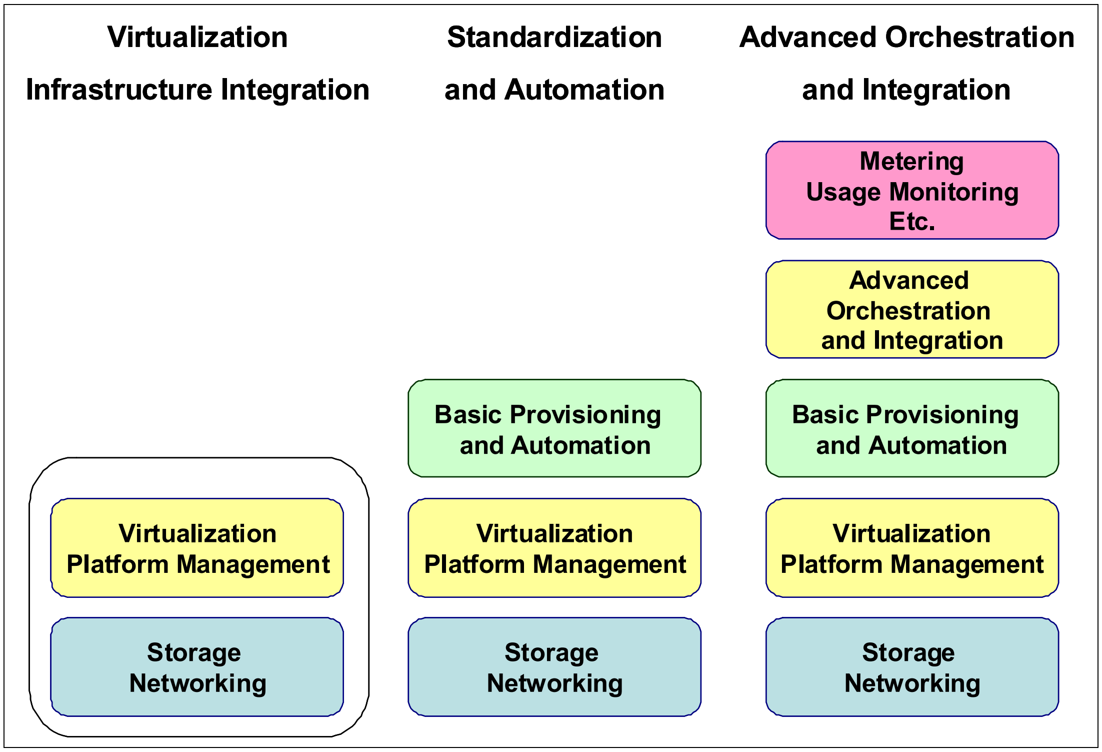

Figure 2-4 Enterprise class heterogeneous platform deployment building blocks

2.7.1 Workloads

Successfully deploying workloads across such an infrastructure requires not only knowledge of application requirements and behavior, but also could impact the enterprise's operational procedures and methodologies, and even the organization.

For the sake of the present discussion let us define the terms application and workload:

Application A computer program or a set of computer programs dedicated to perform a defined computational work.

Workload An application (load) runs on a computer to perform a set of functions (work), and requires computing resources to perform these functions.

Workloads require computing resources, which can basically be classified as:

- CPU
- Memory
- I/O (networking, storage, and other devices, for instance, graphics devices)

I/O handling also consumes a certain amount of CPU and memory resources, but the functions are highly specialized for the type of actions (I/O) performed.

Each application function requires a certain amount of the previously mentioned resources. In principle, every type of computer, regardless of the architecture employed (for instance, System z, POWER, or System x) provides the three types of resources and can (in principle) perform the same tasks. Thus, a workload is the work performed consuming a certain amount of computing resources to perform the functions of an application.

Each application has a specific workload *profile* determined by the type resources needed computational (CPU, memory) or data movement (I/O) and their variation with time. Throughout the evolution of computing platforms, applications have been developed to exploit the platform characteristic features (CPU, memory, I/O). Historically, there also have been cases where a platform has been changed or adapted to better suit the application needs (for example, the continuous evolution of the System z platform and the design of the RISC architecture).

The two basic types of application-required resources result from the two most important requirements:

- Response time (ability to return results in a specified time)
- The amount of data that can be processed in a specified time interval (throughput)

In addition, data and service reliability have driven the platforms' evolution by adding specific mechanisms and tools to achieve the desired results. These characteristics are the measurable parameters enabling the establishment of a service level agreement (SLA).

Platform design has been improved to serve application requirements more effectively and safely. Thus, specialized engines that can carry out specific tasks have been developed. Such examples include (but are not limited to):

- I/O co-processors, designed to offload the main processors from the I/O-related tasks.
- Specialized cryptographic co-processors, designed to offload the computational intensive mathematical functions required to encrypt and decrypt data
- Specialized units inside the processor, such as BCD Arithmetic Logic Unit or vector units

These improvements have been driven by the necessity of reserving the CPU execution capabilities for the core work of the application (main data computation). The diversity of the business needs and platforms also has determined two approaches for application design:

- Custom code
- Commercial Off-The-Shelf (COTS)

Across industries, we see a variety of applications with their specific workload profiles running on a variety of platforms. Historically, the choice of a platform has been determined by two major aspects:

- Platform availability (Development costs always must consider this aspect.)
- Platform's fit-for-the-purpose

The use of programming tools, such as compilers, has greatly contributed to application portability (the ability to run on several platforms). However, due to the diversity of application workload profiles, not all platforms can run the same workload with the same efficiency.

Moreover, depending on the industry and business requirements, an application that might in essence perform the same core functionality, might have additional (non-functional) requirements, such as specific data security and availability, that cannot be obtained in a cost-effective manner on all available platforms. For example:

- In *banking* there are components across retail and wholesale banking that rely on several architectures to execute, but the core of most banking applications relies on System z and z/OS.
- *Insurance* typically maintains claims processing on System z but reaches out to the internet for interaction with consumers, utilizing Linux, UNIX, Power, and x86.
- The *public sector* is relying more and more on the web-based capability to reach out to citizens and improve the rate of return for taxes, accurate payment of social benefits, election process, and even census-based reporting.
- The *retail industry* can effectively exploit zEnterprise, for instance, by benefiting from System z large I/O capacity to implement large databases, exploiting Business Intelligence to characterize their clients.
- *Airline reservation systems* are one example of extreme online transaction processing. zEnterprise provides the z/TPF operating system and application environment specifically for this situation. Another example of z/TPF utilization is in banking with credit card and ATM processing.

The workloads tend to follow well-established technology and infrastructure patterns, such as:

- Core applications (for example, database engines)
- Multi-tier web serving
- Data warehouse/data mining

Workload component attributes

Each of the workload patterns is typically made up of components that have distinct characteristics and requirements. Their components are woven together with application programs and middleware to enable a business process to achieve the desired business objectives.

Based on the workload attributes (characteristics), we can identify the following main types:

- Transaction processing and database (OLTP)
  - High transaction rates
  - High quality of service
  - Peak workloads
  - Resiliency and security
- Analytics and high performance
  - Compute or I/O intensive
  - High memory bandwidth
  - Floating point and vector processing (SIMD[1])
  - Scale-out capable (horizontal scalability)

[1] SIMD - Single Instruction Multiple Data - same instruction executed on a vector of data

- Business applications
  - Scale up (vertical scalability)
  - High quality of service
  - Large memory footprint
  - Responsive infrastructure
- Web, collaboration, and infrastructure management
  - Highly threaded
  - Throughput-oriented
  - Scale out capable
  - Lower quality of service

From an architectural perspective, it is critical to deploy workload components on the server technology that is the best fit and most effective in satisfying their requirements. Thus, multiple platforms might be appropriate. In a more synthetic approach, the workloads might be characterized as:

- Shared data and multiple work queues (OLTP, for example, or large batch jobs)
- Parallel data structures (HPC, Analytics)
- Highly threaded (for example, business applications)
- Small discrete applications

In addition, today's applications can rarely be classified as only one of the above-mentioned types (Figure 2-5). In most cases, an application consists of mixed workload components. Thus, proper platform choice is key in obtaining the desired results (SLA). For example, a batch job requires fast movement of data through the processor, and does not, of itself, exploit any multi-threading capabilities that a platform might have. Thus, the higher the CPU speed, the faster the job is processed. Conversely, a highly threaded application (web services, for example) performs better on a multi-threaded capable platform.

Figure 2-5 Workload characterization

2.7.2 Cloud computing

"Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction."[2] A cloud can be public, private, or a hybrid of both. With cloud computing, the application can be running on a server anywhere in the world. That is why it is changing the way companies provide services to their clients and suppliers.

The IBM Cloud Computing Reference Architecture (CC RA[3], Figure 2-6) defines the fundamental architectural elements constituting a cloud computing environment. It is required that all of these infrastructure components be managed from a single, central Common Cloud Management Platform with the ability to place instances of each cloud service on the corresponding infrastructure. This requirement perfectly fits the zEnterprise System, with its end-to-end management capabilities for flexible delivery of high-value services.

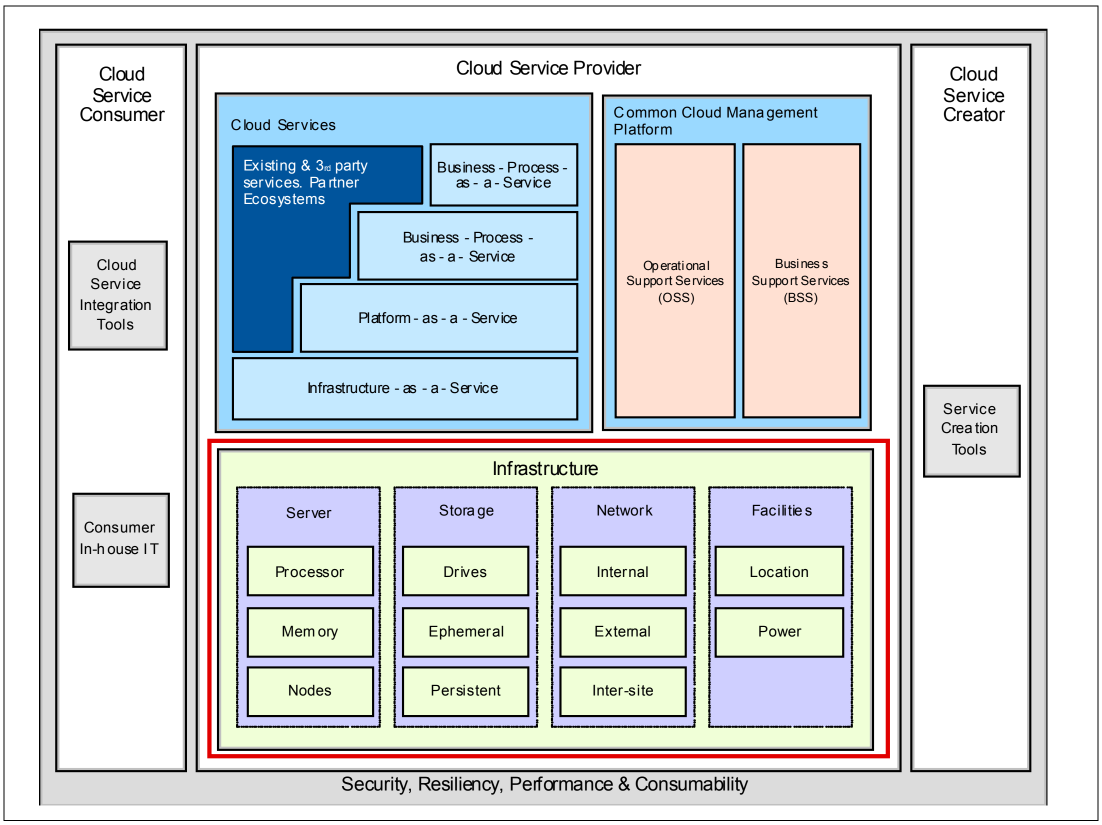

Figure 2-6 The IBM Cloud Computing Reference Architecture infrastructure details

Virtualization is the foundation for "cloud," and the benefits of consolidation and virtualization are widely accepted by the IT community. Adding standardization and automation to a virtualized environment will enable IT optimization for cloud computing. Workflow orchestration, monitoring, and metering for accounting are other major components of cloud computing.

[2] U.S. National Institute of Standards and Technology. Read more at: `http://csrc.nist.gov/groups/SNS/cloud-computing/index.html`

[3] `https://www.opengroup.org/cloudcomputing/uploads/40/23840/CCRA.IBMSubmission.02282011.doc`

Deploying a cloud infrastructure is not a simple process, but there is a defined path that can be followed. Figure 2-7 depicts the path from Standard Managed Services to Cloud.

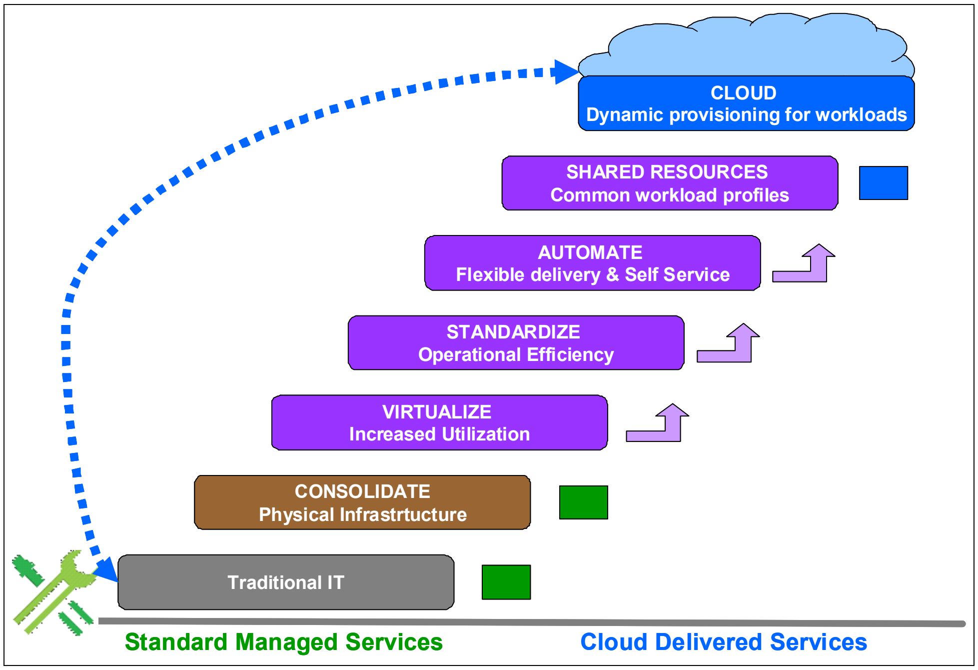

Figure 2-7 Cloud delivered services

As zEnterprise offers a fully virtualized system with its "shared everything infrastructure," it becomes easier to integrate a cloud computing deployment as part of the existing IT optimization strategy and roadmap. Table 2-1 summarizes potential benefits provided through Cloud Computing.

Table 2-1 Benefits with Cloud Computing

Virtualization	Standardization	Automation
Higher utilization Economy-of-scale benefits Lower capital expense Lower operating expense	Easier access Flexible pricing Reuse and sharing Easier integration	Faster cycle times Lower support costs Optimized utilization Improved compliance Optimized security Better user experience

Cloud computing on zEnterprise builds on the industry's leading virtualization technology that leverages virtualization, standardization, and automation to free operational budget for new investments and allow you to optimize new investments for direct business benefits. zEnterprise:

- Provides a highly scalable heterogeneous pool of virtualized resources managed in a single system
- Activates, allocates, prioritizes, and retires resources on demand, and automates service delivery
- Maximizes utilization of resources for improved ROI and lower cost of service delivery
- Brings new levels of security, resiliency, and manageability to create a cloud environment that is enterprise ready

As depicted in Figure 2-8, System z enables the following attributes, being the IT industry's only multi-architecture cloud solution:

- Higher utilization
  - Up to 100% CPU utilization
  - *Shared everything* architecture
  - Hosting of thousands of mixed workloads
- Increased productivity
  - Single point of control for a heterogeneous infrastructure at a platform level, with the Unified Resource Manager
  - Efficient, rapid provisioning
  - Superior workload management enabled with Unified Resource Manager
  - Workload optimization with fit for purpose approach
  - Collocating applications where industry-leading z/OS transaction and data services run
- More efficient data center
  - Less power and cooling
  - Less floor space
  - Fewer parts to manage
- Greater reliability and availability
  - Built-in hardware redundancy
  - Decades of RAS innovation
  - Capacity and backup on demand
  - Decades-proven virtualization security protecting sensitive data and critical business processes in the cloud
  - Resiliency management and fewer points of failure
- Security
  - Extending System z security to a private network across heterogeneous resources

Figure 2-8 IBM zEnterprise for IT optimization, consolidation, and cloud computing

Cloud computing is one of the key ways to address the challenges of today and build a smarter planet. As more and more companies embrace cloud computing, zEnterprise becomes more obviously the perfect platform for delivering large-scale SasS application software services.[4]

Quality of service improvements

With the introduction of the zEnterprise System, the qualities for which System z is renowned are extended to other components of the ensemble, providing support for mission-critical workloads running on the ensemble's heterogeneous infrastructure. Compared to a diverse infrastructure, the zEnterprise System provides the following benefits:

- A single management and policy framework across web serving, transaction, database, and servers can lower the cost of enterprise computing.
- Integration of multiplatform management capabilities through extended functionality in the well-known mainframe HMC simplifies operations.
- Mainframe systems management and service extended to the zBX environment improves the reliability of the infrastructure.
- Dynamic resource management of the mainframe to all devices within a multi-tier architecture allows improving service.
- Monitoring and management of a heterogeneous solution as a single, logical virtualized solution simplifies and improves infrastructure management.
- Management of the platform's resources in accordance with specified business-service-level objectives allows better alignment of IT with business objectives.
- Management of virtual servers as part of the overall deployed business workload improves manageability of the infrastructure.
- A secure and managed Layer 2 network connecting the zBX blades with the zEnterprise CPC can dramatically simplify the infrastructure, improve application performance, and simplify management.

Mainframe QoS characteristics are extended as appropriate to accelerators and distributed application servers to mitigate risk of operational failures.

4 More information about cloud on System z can be found in Deploying a Cloud on IBM System z, REDP-4711. Available at http://w3.itso.ibm.com/abstracts/redp4711.html?Open

3

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 zEnterprise CPCs and zBX provided in Chapter 1, "Proposing an IT infrastructure (r)evolution" on page 1, and compares them with the System z10 servers, where relevant.

This chapter discusses the following topics:

3.1 zEnterprise CPC highlights, models, and upgrades

z196 and z114 models, as well as the improvements and upgrades over their predecessors, are discussed in this section.

3.1.1 z196 highlights

The major z196 improvements over its predecessors include the following features:

- 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 a 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, also introducing enhanced HCA3 fanouts with up to 40% better service time
- 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 Time Protocol (STP) enhancements for time accuracy, availability, recovery, and systems management with message exchanges using ISC-3 or 1x InfiniBand connections

In all, these enhancements provide 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 196 Technical Guide,* SG24-7833.

3.1.2 z196 models

The z196 has been assigned a machine type (M/T) of 2817, which uniquely identifies the CPC. The z196 is offered in five models:

- M15
- M32
- M49
- M66
- 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 use 20-PU multi-chip modules (MCM), of which 15 to 17 PUs are available for characterization. The fifth model, M80, uses 24-PU MCMs to provide a maximum of 80 configurable PUs.

Spare PUs and system assist processors (SAPs) are integral to the system. Table 3-1 provides a model summary including SAPs and spare PUs for the various models. For an explanation of PU characterization, see “PU characterization” on page 59.

Table 3-1 z196 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 capacity levels, which span a range of approximately 1 to 220. This is discussed in 4.2.2, “Capacity settings” on page 91.

z196 upgrades

Figure 3-1 summarizes the upgrade paths to z196.

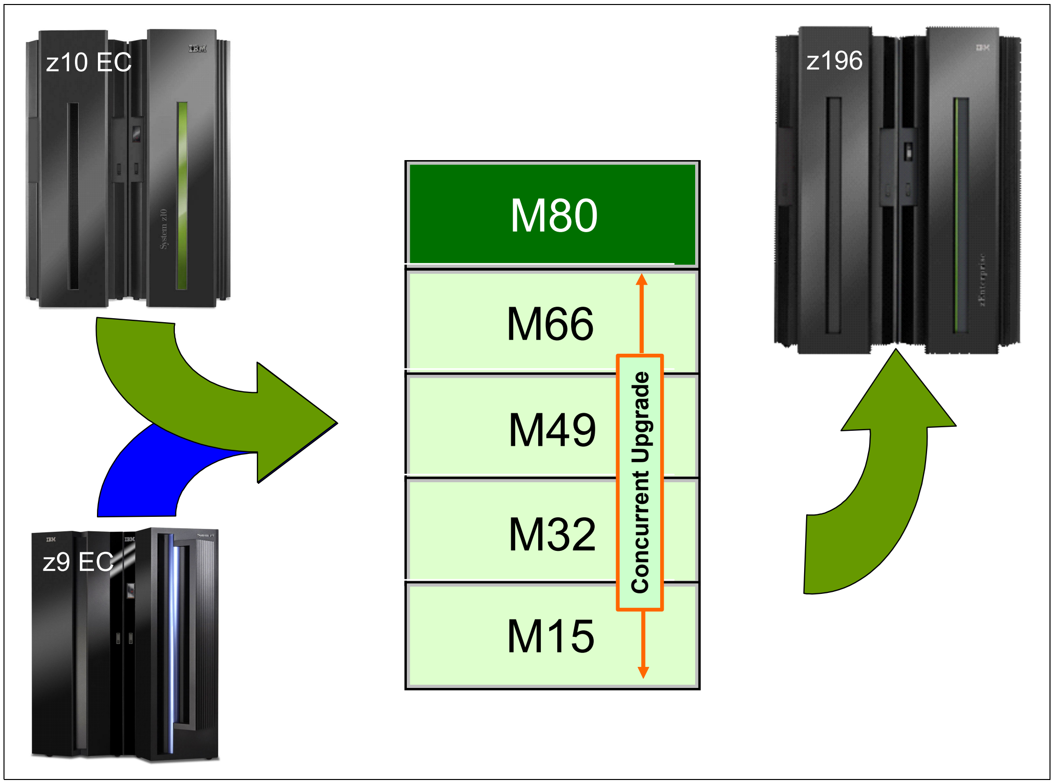

Figure 3-1 z196 upgrade paths

Model upgrades within the same CPC 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. Table 3-2 lists upgrades available within the family.

Table 3-2 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 z196 from z10 EC and z9 EC

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

3.1.3 z114 highlights

The major z114 improvements over its predecessors include the following features:

- Increased total system capacity in a 14-way server (with 10 characterizable PUs) and additional subcapacity settings, offering increased levels of performance and scalability to help enable new business growth
- Quad-core 3.8 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 256 GB of available real memory per server for growing application needs
- 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 8 GB fixed hardware system area (HSA) that is managed separately from customer-purchased memory
- Exploitation of InfiniBand technology, also introducing enhanced HCA3 fanouts with up to 40% better service time
- 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 Time Protocol (STP) enhancements for time accuracy, availability, recovery, and systems management with message exchanges using ISC-3 or 1x InfiniBand connections

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

For an in-depth discussion of the zEnterprise 114 functions and features, see the *IBM zEnterprise 114 Technical Guide,* SG24-7954.

3.1.4 z114 models

The z114 has been assigned a machine type (M/T) of 2818, which uniquely identifies the CPC. The z114 is offered in two models:

- M05
- M10

The digits in the model name refer to the maximum number of processor units (PUs) available for characterization. PUs are delivered in single-engine increments. The M05 has one processor drawer, and the M10 has two processor drawers. Each drawer contains three Single Chip Modules (SCM), two for processor chips and one for a Storage Controller chip.

Both z114 models can have up to five PUs characterized as CPs. Each CP has 26 capacity settings, resulting in 130 capacity settings for the z114. Table 3-3 provides a summary of possible PU configuration options for the z114.

Table 3-3 z114 model summary

Model	CPs	IFLs - unassigned IFLs	zAAPs	zIIPs	ICFs	SAPs	Additional SAPs	Spares
M05	0 - 5	0 - 5	0 - 2	0 - 2	0 - 5	2	0 - 2	0
M10	0 - 5	0 - 10	0 - 5	0 - 5	0 - 10	2	0 - 2	2

z114 upgrades

Figure 3-2 shows the upgrade paths for the z114.

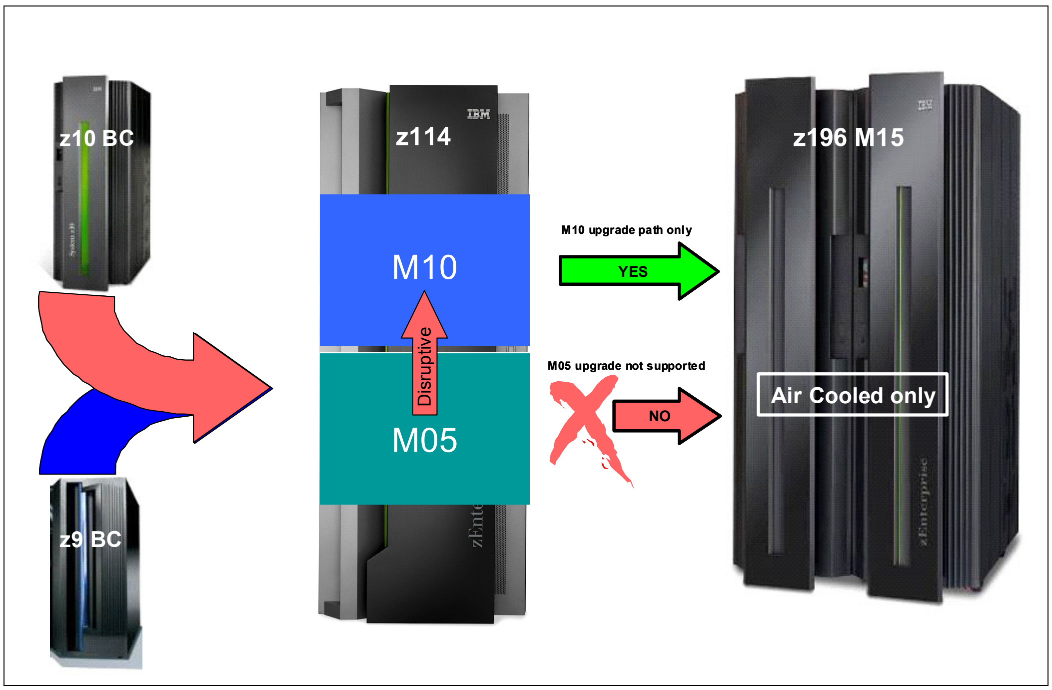

Figure 3-2 z114 upgrade paths

A model upgrade from an M05 to an M10 is accomplished by installing a second processor drawer. That upgrade is disruptive.

Capacity upgrades within the two models are concurrent and can be accomplished by increasing the capacity settings of the existing processors, increasing the number of processors, or both. In addition, a permanent capacity upgrade within one of the models can also be accomplished by increasing the capacity settings of the existing processors and at the same time reducing the number of active processors. The z114 offers 130 capacity settings ranging from a one-way processor at three MSUs to a five-way at 388 MSUs.

Upgrades from z10 BC and z9 BC to the z114

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

Upgrades from the z114 to the z196

The z114 model M10 can be upgraded to a z196 model M15 air-cooled model. That upgrade is disruptive.

3.2 The frames

The frames for the z196 and z114 are described in this section.

3.2.1 z196 frames

The z196 server is always a two-frame system. The frames are called the *A frame* and the *Z frame*. The z196 can be delivered as an air-cooled system or as a water-cooled system. Figure 3-3 shows an internal front view of the two frames for an air-cooled CPC.

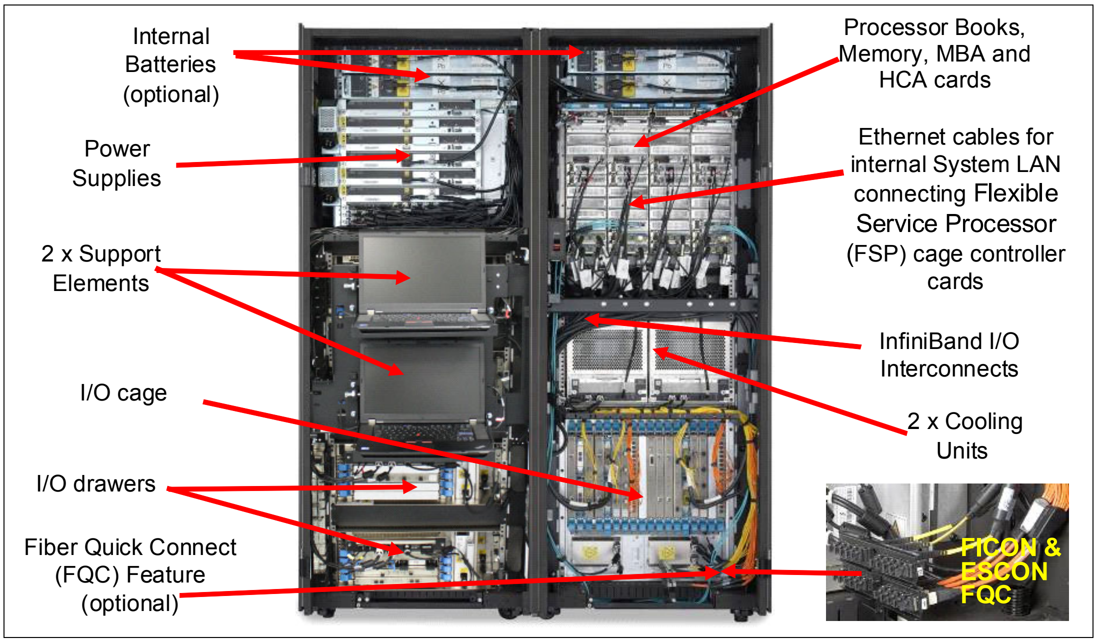

Figure 3-3 z196 internal front view: air-cooled CPC

Figure 3-4 shows an internal front view of the two frames of a water-cooled CPC.

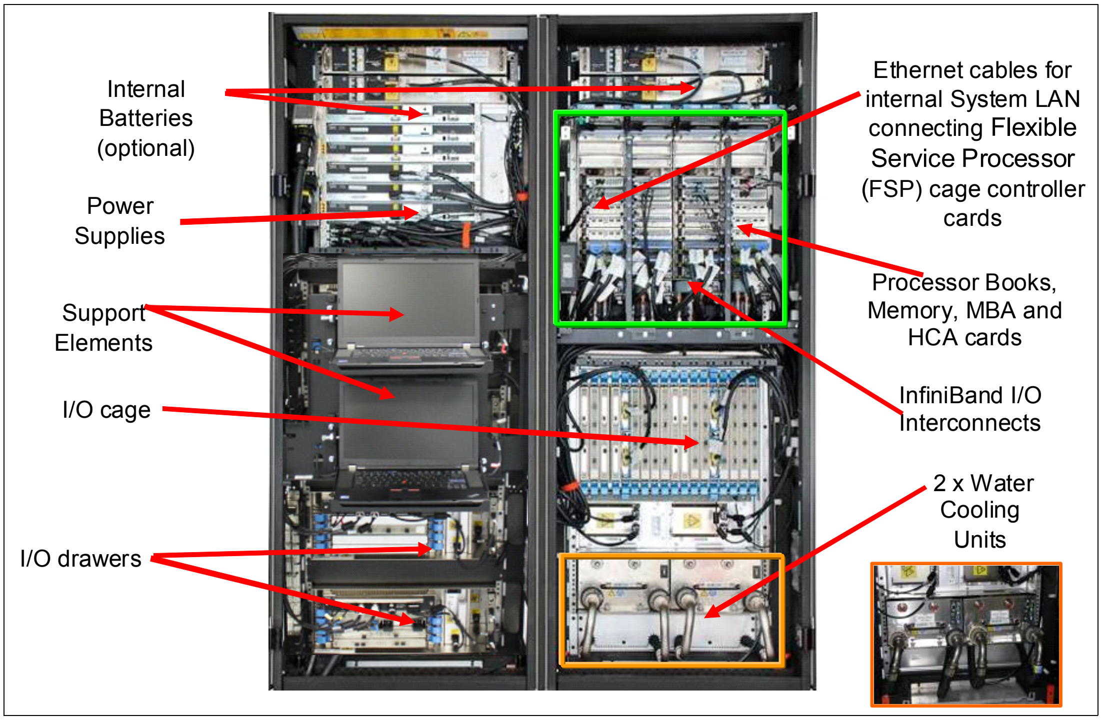

Figure 3-4 z196 internal front view: water-cooled CPC

Figure 3-5 shows an internal front view of an air-cooled z196 with PCIe I/O drawers installed.

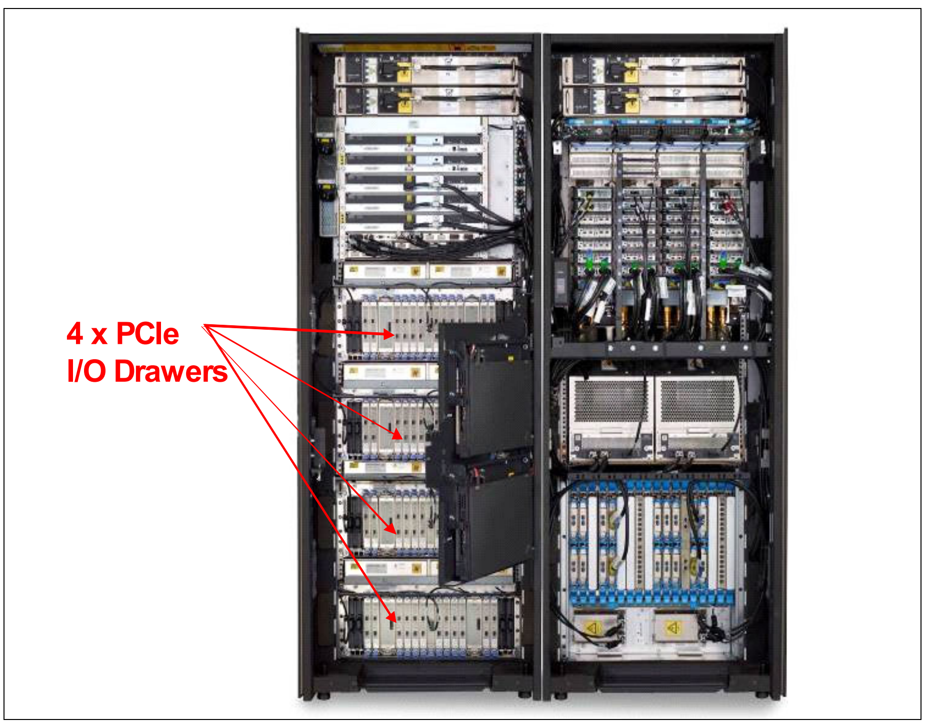

Figure 3-5 z196 internal front view with PCIe I/O drawers

Top exit I/O cabling

On the z196 you have the option of ordering the infrastructure to support the top exit of your fiber optic cables (ESCON, FICON, OSA, 12x InfiniBand, 1x InfiniBand, and ISC-3) and 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 CPC to 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 and MES orders.

3.2.2 z114 frame

The z114 is a one-frame system that is air cooled. Figure 3-6 shows an internal front view of the z114 CPC.

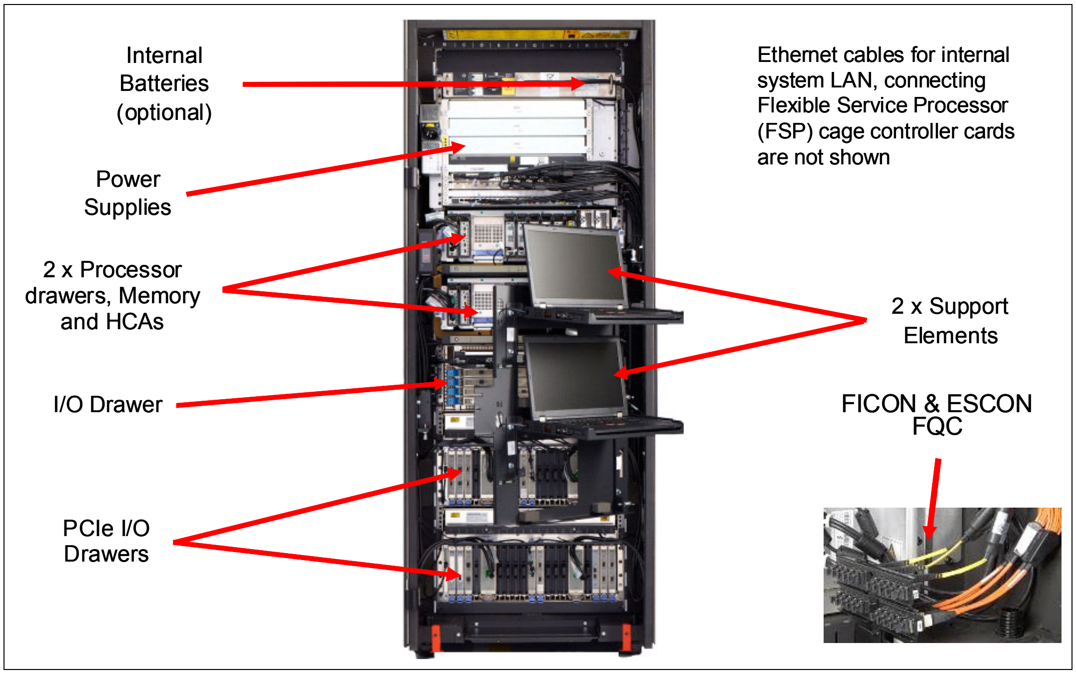

Figure 3-6 z114 internal front view

Cabling options for the z114 server

The z114 can be delivered with a number of cabling options for both raised and non-raised floor situations. Both bottom and top exits for power and I/O cables are available in various combinations.

Top exit I/O cabling and power cords are designed to provide you with an additional option. Instead of all of your cables and power cords exiting under the server 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 and power cords can also help to increase air flow.

3.3 z196 processor cage, books, and MCM

The z196 server has a multi-book system structure similar to the z10 EC. Up to four books can be installed on a z196 CPC. A book looks like a box and plugs into one of the four slots in the processor cage of the z196 CPC. The processor cage is located in the in A frame of the z196 CPC. Figure 3-3 on page 51 provides a pictorial view of the processor cage and the location of the four books.

Each book contains the following elements:

- A multi-chip module (MCM)

  Each MCM includes six quad-core PU chips and two storage control (SC) chips. MCMs are described in “z196 MCM” on page 55. See Table 3-1 on page 47 for the model summary and the relation between the number of books and number of available PUs.

- A minimum of 32 and a maximum of 704 GB of memory for customer use

  See Table 3-4 on page 60 for details.

- A combination of up to eight InfiniBand Host Channel Adapter (HCA2-Optical, HCA3-Optical, HCA2-Copper) fanouts and PCIe fanouts

  Each fanout has two or four ports, thereby supporting up to 32 connections. HCA2-Copper connections are for links to the I/O cages in the server. The HCA2-Optical and HCA3-Optical connections are to external systems (coupling links). PCIe fanouts are for links to the PCIe I/O drawers

- 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 3-7 shows a view of a z196 book without the containing box.

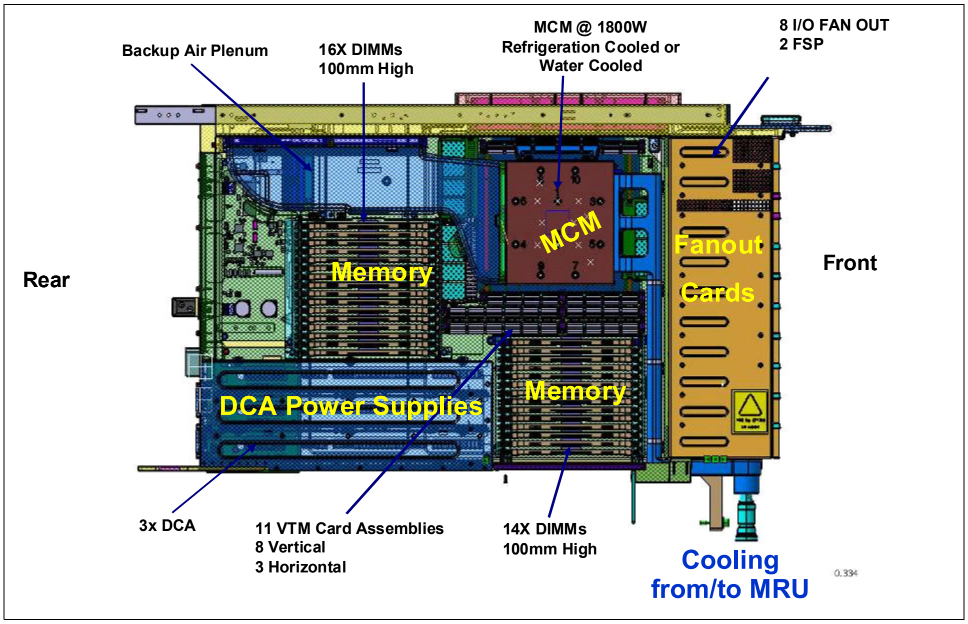

Figure 3-7 z196 book structure and components

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 3-8, 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.

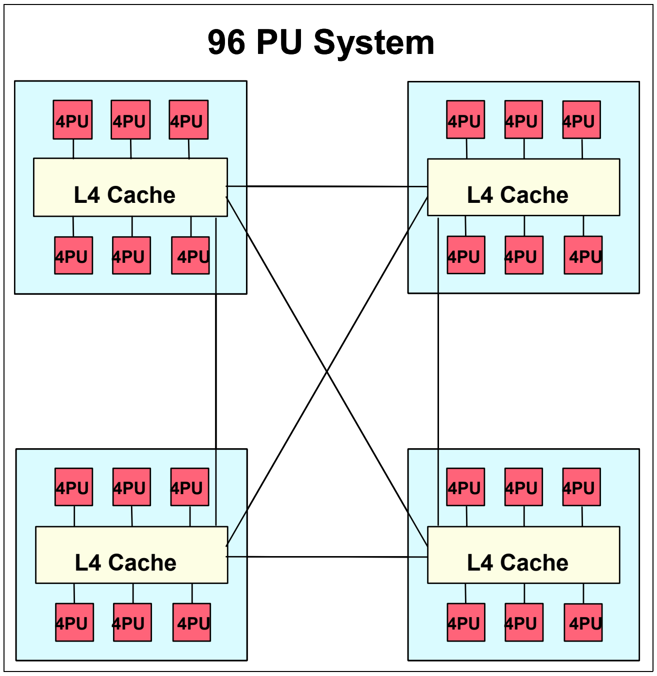

Figure 3-8 z196 inter-book communication structure

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

z196 MCM

The MCM (Figure 3-9) is a high-performance, glass-ceramic module, providing the highest level of processing integration in the industry. It is the heart of the system.

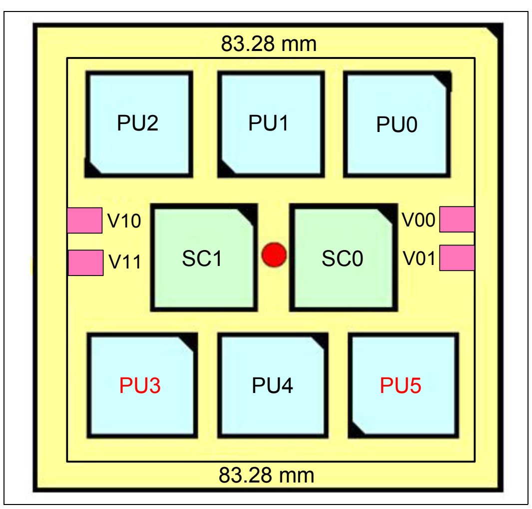

Figure 3-9 z196 multi-chip module

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 has six PU chips per MCM and each PU chip has up to four active PUs (cores) (Figure 3-10). 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.

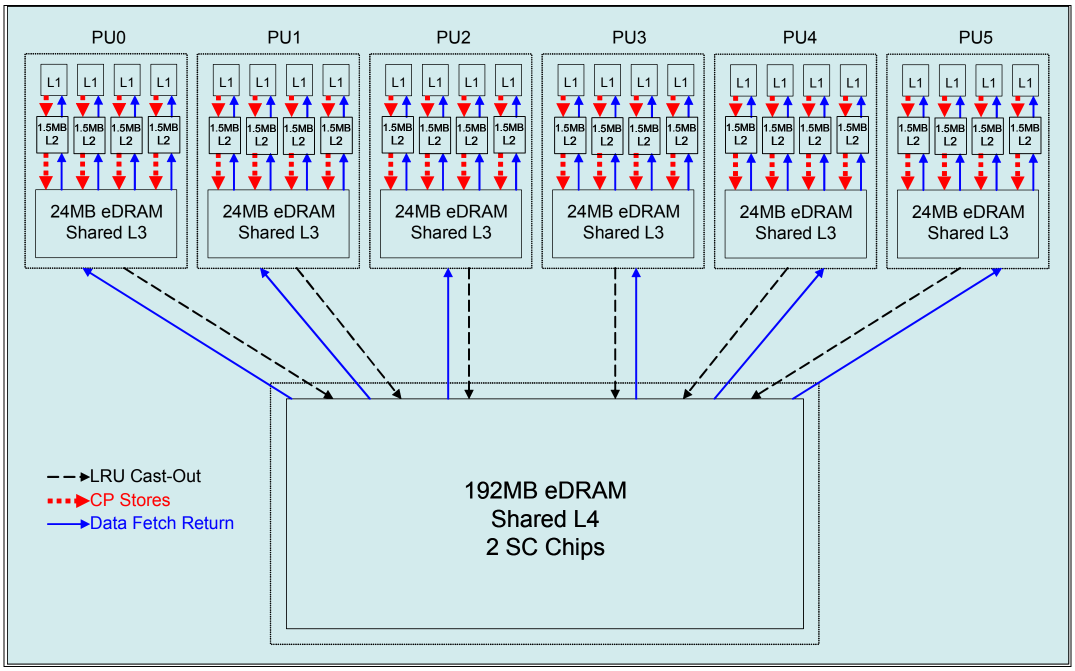

Figure 3-10 MCM chips and cache structure

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 mm2. The two SC chips are configured to provide a single 192 MB L4 cache shared by all 20 or 24 cores on the module, yielding outstanding SMP scalability on real-world 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.

3.4 z114 processor drawer and SCM

The z114 central processor complex (CPC) uses a packaging concept for its processors based on processor drawers. A processor drawer contains single chip modules (SCMs), memory, and connectors to I/O drawers, PCIe I/O drawers, and other systems. The z114 M05 has one processor drawer installed and z114 M10 has two processor drawers installed. Figure 3-11 shows a processor drawer and its components.

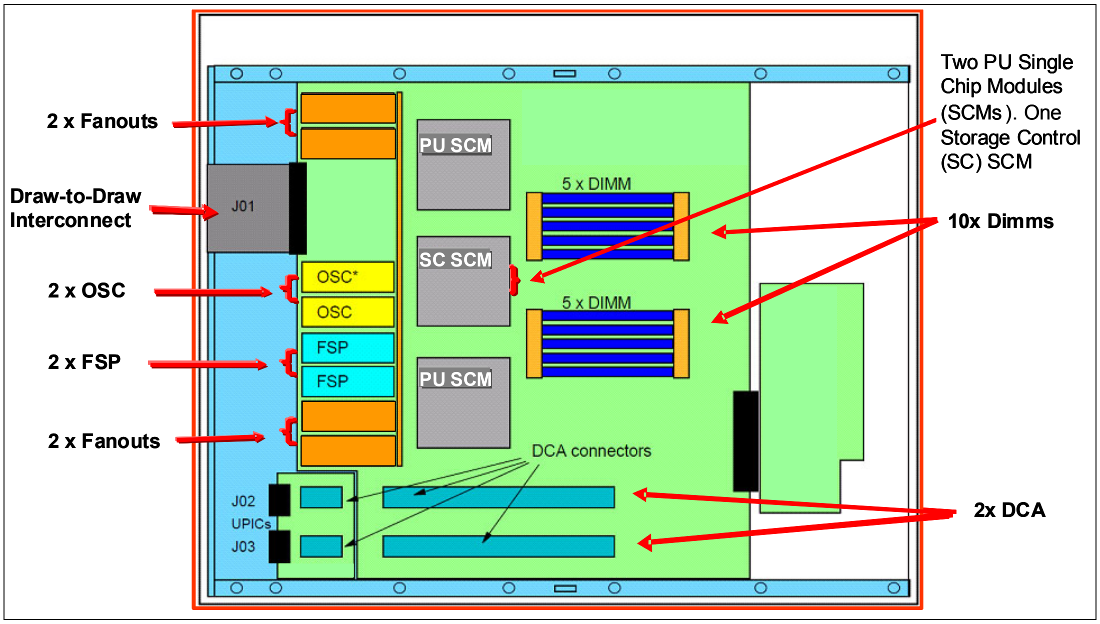

Figure 3-11 z114 processor drawer structure and components

Each processor drawer contains:

- One SC single chip module with 96 MB L4 cache.
- Two PU SCMs. One PU SCM has three active cores, while the other PU SCM has four active cores.
- Memory DIMMs plugged into 10 available slots, providing up to 160 GB of physical memory installed in a processor drawer.
- A combination of up to four fanouts. PCIe fanout connections are for links to the PCIe I/O drawers in the CPC, HCA2-Copper connections are for links to the I/O drawers in the CPC, and HCA2-Optical and HCA3-Optical connections are to external systems (coupling links).
- Two distributed converter assemblies (DCAs) that provide power to the processor drawer. Loss of a DCA leaves enough power to satisfy the processor drawer’s power requirements (N+1 redundancy). The DCAs can be concurrently maintained.
- Two flexible service processor (FSP) cards for system control.
- Two oscillator (OSC) cards with Pulse Per Second (PPS) connections

On the z114 M10, fabric book connectivity (FBC) provides the point-to-point connectivity between the two processor drawers of the CPC.

Both PU and storage control (SC) chips on the SCM use CMOS 12s chip technology. CMOS 12s is 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). On the SCM, two serial electrically erasable programmable ROM (SEEPROM) chips are rewriteable memory chips that hold data without power, use the same technology, and are used for retaining product data for the SCM and relevant engineering information.

Each PU SCM is 50 x 50 mm in size and the SC SCM size is 61 x 61 mm. Figure 3-12 shows the cache structure on the z114.

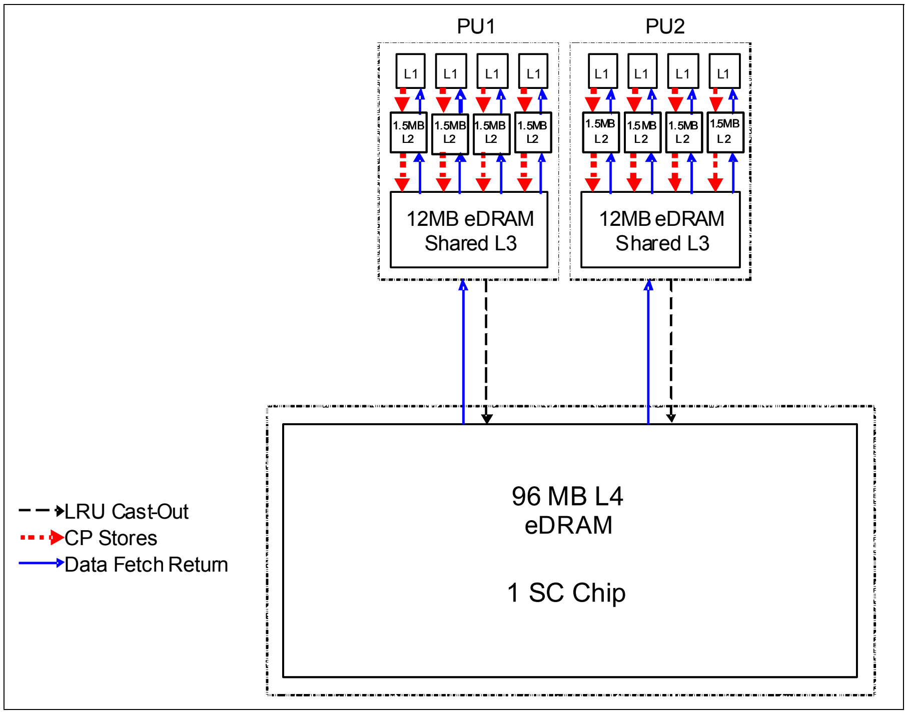

Figure 3-12 z114 processor drawer cache levels structure

3.5 Processor chip

The zEnterprise CPC features a high-frequency quad-core processor chip, an advanced microprocessor design, a robust cache hierarchy, and an SMP design optimized for enterprise database and transaction processing workloads and for workloads such as Java and Linux. It makes use of 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 details are given in 4.2.1, "Microprocessor" on page 88.

3.6 Processor unit

A PU is the generic term for the z/Architecture processor on the multi-chip module (MCM) or single chip module (SCM). 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 varying numbers of steps for various types of instructions. Several instructions can 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. In addition, each PU chip contains a L3 cache, shared by all four PUs on the chip. The shared L3 cache uses eDRAM. The z196 has a 24 MB L3 cache and z114 has a 12 MB L3 cache. The cache structures of z196 and z114 are shown, respectively, in Figure 3-10 on page 56 and Figure 3-12 on page 58. This implementation optimizes performance of the system for high-frequency, 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. This provides the reliability and availability required for true mainframe quality.

The zEnterprise 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 each processor unit into one of 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, on a server with two CPs, a maximum of two zAAPs and two zIIPs can be installed. A logical partition definition for that server can 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 CPC does not change the system capacity setting or its MSU rating (only CPs do).
>
> IBM does not impose any software charges on work dispatched on zAAP and zIIP processors.

3.7 Memory

This section discusses memory for the z196 and z114.

z196 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 in up to 768 GB of available memory per book and up to 3072 GB (3 TB) per system. As the HSA memory has a fixed amount of 16 GB and is managed separately from customer memory, you can order up to 752 GB on the one-book model and up to 3056 GB on the four-book server models.

Table 3-4 lists the maximum and minimum memory sizes for each z196 model.

Table 3-4 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. 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 systems, the memory granularity varies from 32 GB, for customer memory sizes from 32 to 256 GB, up to 256 GB, for systems having from 1776 GB to 3056 GB of customer memory. Table 3-5 shows the memory granularity depending on installed customer memory.

Table 3-5 z196 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 physically installed memory.
- A book can 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. See *IBM zEnterprise 196 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.

z114 memory

A z114 CPC has more physical memory installed than ordered. Part of the physical installed memory is used to implement the redundant array of independent memory (RAIM) design, resulting in up to 128 GB of available memory for M05 and up to 256 GB for M10. As the HSA memory has a fixed amount of 8 GB and is managed separately from customer memory, you can order up to 120 GB for M05 and up to 248 GB for M10.

Table 3-6 shows the maximum and minimum memory sizes that customer can order for each z114 model and the increments.

Table 3-6 z114 servers memory sizes

Model	Number of processor drawer	Increment (GB)	Customer memory (GB)
M05	1	8	8 - 120
M10	2	8	16 - 120
M10	2	32	152 - 248

On z114 CPCs, the memory granularity is 8 GB for customer memory sizes from 8 to 120 GB and 32 GB for systems with 152 GB to 248 GB of customer memory. Physically, memory is organized as follows:

- A processor drawer always contains a minimum of 40 GB of physically installed memory.
- A processor drawer can 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 processor drawer with additional memory is necessary.

When activated, a logical partition can use memory resources located in any processor drawer. No matter in which processor drawer the memory resides, a logical partition has access to that memory if so allocated.

A memory upgrade is concurrent when it requires no change of the physical memory cards. A memory card change is disruptive.

3.7.1 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 zEnterprise CPCs 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.

3.7.2 Redundant array of independent memory

zEnterprise CPCs introduce the redundant array of independent memory (RAIM) to System z, 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.

3.7.3 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. Starting with z10, HSA is a fixed size, large enough to accommodate any LPAR definitions or changes, thus eliminating those outage possibilities. For further information and benefits, see 4.2.3, “Memory” on page 92.

3.8 I/O system structure

The z196 and z114 support two internal I/O infrastructures:

- InfiniBand-based infrastructure for I/O cages (z196 only) and I/O drawers
- PCIe-based infrastructure for PCIe I/O drawers, which use a different form factor drawer, and I/O features

The InfiniBand I/O infrastructure consists of:

- InfiniBand fanouts supporting the current 6 GBps InfiniBand I/O interconnect
- InfiniBand I/O card domain multiplexers with Redundant I/O interconnect in:
  - The 14U, 28-slot, 7-domain I/O cage (z196 only)
  - The 5U, 8-slot, 2-domain I/O drawer (z114 and z196)
- I/O features

The PCIe I/O infrastructure consists of:

- PCIe fanouts for processor book or processor drawer connectivity to the PCIe I/O drawer
- 7U PCIe I/O drawer with 32 slots (eight slots per I/O domain) for PCIe I/O features
- I/O features

Note that I/O cages (z196), I/O drawers, or PCIe I/O drawers cannot be directly ordered. Ordering of I/O feature types will determine the appropriate mix of I/O cages, I/O drawers, and PCIe I/O drawers.

Figure 3-13 shows a high-level view of the I/O system structure for the zEnterprise CPCs.

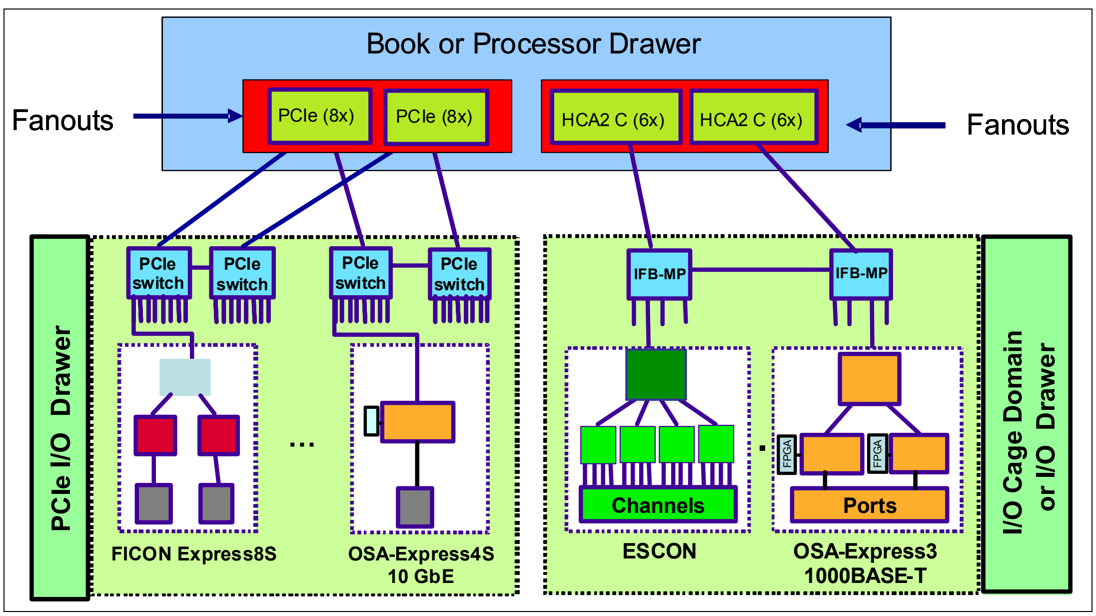

Figure 3-13 zEnterprise CPC I/O system structure

The zEnterprise CPCs have six fanout types, which reside in the front of the book for the z196 or in the front of the processor drawer of the z114. The fanout types are:

- PCIe
- HCA2-C
- HCA2-O
- HCA2-O LR
- HCA3-O
- HCA3-O LR

The HCA3-O LR fanout comes with four ports, and each of the other fanouts come with two ports.

The types of internal I/O connectivity that support the I/O cage, I/O drawer, and PCIe I/O drawer are:

- InfiniBand (IFB) connections to the I/O cages and I/O drawers are from the Host Channel Adapter (HCA2-C) fanouts via copper cables. The two ports in the fanout are dedicated to ESCON, FICON, OSA, ISC-3, and Crypto features in the I/O drawer, or in the I/O cage.
- PCIe connections to the PCIe I/O drawers are from the PCIe fanouts via copper cables. The two ports on the fanouts support both FICON Express8S and OSA-Express4S features in the PCIe I/O drawers.

For coupling link connectivity in a Parallel Sysplex configuration, the zEnterprise CPCs support the following fanouts:

- HCA2-O
- HCA2-O LR
- HCA3-O
- HCA3-O LR

The z196 can have up to eight fanouts (numbered D1, D2, and D5 to DA) per book. The z114 is equipped with up to two processor drawers. Each drawer can have up to four fanouts (D1, D2, D7, and D8). In a system configured for maximum availability, alternate paths maintain access to critical I/O devices, such as disks, networks, and so on.

Each I/O cage supports up to seven I/O domains and a total of 28 I/O feature slots. Each I/O domain supports four I/O features (ESCON, FICON, OSA, or ISC).

The I/O domains (A, B, C, D, E, F, and G) and the InfiniBand MultiPlexer (IFB-MP), which connects to the CPC cage and 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. Note that I/O cages are not supported on the z114.

Each I/O drawer supports two I/O domains (A and B) for a total of eight I/O 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 features in the I/O drawer are installed horizontally. The two Distributed Converter Assemblies (DCAs) distribute power to the I/O drawer.

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

The PCIe I/O drawer is a two-sided drawer (I/O features on both sides) that is 7U high (one half of I/O cage) and fits into a 24-inch System z frame. The drawer contains 32 slots, four switch cards (two in front, two in rear), two DCAs to provide the redundant power, and two air moving devices (AMDs) for redundant cooling.

3.9 I/O features

As mentioned, the z196 and z114 support two internal I/O infrastructures:

- InfiniBand-based infrastructure for I/O cages and I/O drawers, supporting the following I/O features:
  - Crypto Express3
  - ESCON
  - FICON Express8
  - FICON Express4
  - ISC-3
  - OSA-Express3
- PCIe-based infrastructure for PCIe I/O drawers, supporting the following I/O features:
  - FICON Express8S
  - OSA-Express4S

All new system builds, migration offerings, and exchange programs will offer FICON Express8S and OSA-Express4S features. Crypto Express3, ESCON, ISC-3, OSA-Express3 1000BASE-T, and PSC features cannot be used in the PCIe I/O drawer.

The following features are no longer orderable for zEnterprise CPCs:

- FICON Express2 (all features)
- OSA-Express2 10 GbE LR feature
- Crypto Express2

In addition, ICB-4 links are not supported. See Table C-1 on page 153 for more details.

3.9.1 ESCON

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) optical 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 features is fully allocated (by seven ESCON port groups, using 28 ports), single features are used for additional ESCON port groups.

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

Statement of Direction: The IBM zEnterprise 196 and IBM zEnterprise 114 are the last System z servers to support ESCON channels.

IBM plans to not offer ESCON channels as an orderable feature on System z servers that follow the z196 (machine type 2817) and z114 (machine type 2818). In addition, ESCON channels cannot be carried forward on an upgrade to such follow-on servers.

System z customers should continue to eliminate ESCON channels from the mainframe wherever possible. Alternate solutions are available for connectivity to ESCON devices.

For more information see the following web page:

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 the following web page:

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.

3.9.2 FICON Express8S

Two types of FICON Express8S transceivers are supported on new build zEnterprise CPCs, one long wavelength (LX) laser version and one short wavelength (SX) laser version:

- FICON Express8S 10KM LX feature
- FICON Express8S SX feature

Each port supports attachment to the following elements:

- 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 have been withdrawn from marketing.

When upgrading to a zEnterprise CPC, replace your FICON Express, FICON Express2, and FICON Express4 features with FICON Express8S features. The FICON Express8S features offer better performance and increased bandwidth.

FICON Express8S 10KM LX feature

The FICON Express8S 10KM LX feature occupies one I/O slot in the PCIe I/O drawer. It has two 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.

FICON Express8S SX feature

The FICON Express8S SX feature occupies one I/O slot in the PCIe I/O drawer. It has two 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.

3.9.3 FICON Express8

Two types of FICON Express8 transceivers are supported on zEnterprise CPC:

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

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.

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.

3.9.4 FICON Express4

Three types of FICON Express4 transceivers are supported on z196 and z114 servers:

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

The following two types of FICON Express4 are supported only on the z114 server:

- FICON Express4-2C 4KM LX feature
- FICON Express4-2C SX feature

Statement of Direction: The IBM zEnterprise 196 and the zEnterprise 114 are the last System z servers to support FICON Express4 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

Note: The FICON Express4 feature is the last feature to negotiate down to 1 Gbps. Review the usage of your installed FICON Express4 channels and, where possible, migrate to FICON Express8 or FICON Express8S channels.

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.

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.

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.

FICON Express4-2C 4KM LX feature

The FICON Express4 4KM LX feature is only available on the z114 and occupies one I/O slot in the I/O drawer. It has two ports, each supporting an LC duplex connector and link speeds of 1 Gbps, 2 bps, or 4 Gbps up to an unrepeated maximum distance of 4 km.

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 Express4-2C SX feature

The FICON Express4 SX feature is only available on z114 and occupies one I/O slot in the I/O drawer. It has two 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.

3.9.5 OSA-Express4S

This section describes the connectivity options offered by the OSA-Express4S features. The following OSA-Express4S features can be installed on zEnterprise CPCs:

- OSA-Express4S 10 Gigabit Ethernet (GbE) Long Reach (LR)
- OSA-Express4S 10 Gigabit Ethernet Short Reach (SR)
- OSA-Express4S Gigabit Ethernet long wavelength (GbE LX)
- OSA-Express4S Gigabit Ethernet short wavelength (GbE SX)

OSA-Express4S 10 GbE LR feature

The OSA-Express4S 10 GbE LR feature occupies one slot in a PCIe I/O drawer. It has one port that connects to a 10 Gbps Ethernet LAN through a 9 µm single-mode fiber optic cable terminated with an LC Duplex connector. The feature supports an unrepeated maximum distance of 10 km.

OSA-Express4S 10 GbE SR feature

The OSA-Express4S 10 GbE SR feature occupies one slot in the PCIe I/O drawer. It has one port that connects to a 10 Gbps Ethernet LAN through a 62.5 µm or 50 µm multi-mode fiber optic cable terminated with an LC Duplex connector. The maximum supported unrepeated distance is 33 m on a 62.5 µm multi-mode fiber optic cable, and 300 m on a 50 µm multi-mode fiber optic cable.

OSA-Express4S GbE LX feature

The OSA-Express4S GbE LX occupies one slot in the PCIe I/O drawer. It has two ports representing one CHPID 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. 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.

OSA-Express4S GbE SX feature

OSA-Express4S GbE SX occupies one slot in the PCIe I/O drawer. It has two ports representing one CHPID 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).

3.9.6 OSA-Express3

This section describes the connectivity options offered by the OSA-Express3 features. The following OSA-Express3 feature are supported on the zEnterprise CPCs:

- OSA-Express3 1000BASE-T Ethernet
- 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)

The following OSA-Express3 features can be installed only on a z114:

- OSA-Express3-2P Gigabit Ethernet short wavelength (GbE SX)
- OSA-Express3-2P 1000BASE-T Ethernet

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, with two ports per CHPID, 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 or Cat6 cabling, which supports a maximum distance of 100 meters.

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 µm single-mode fiber optic cable terminated with an LC Duplex connector. The feature supports an unrepeated maximum distance of 10 km.

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 µm or 50 µm multi-mode fiber optic cable terminated with an LC Duplex connector. The maximum supported unrepeated distance is 33 m on a 62.5 µm multi-mode fiber optic cable, and 300 m on a 50 µm 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, with two ports per CHPID, 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. 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.

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, with two ports per CHPID, 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-Express3-2P GbE SX feature

OSA-Express3 GbE SX is only available on the z114 and occupies one slot in the I/O drawer. It has two ports that belong to one CHPID and 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-Express3-2P 1000BASE-T Ethernet feature

OSA-Express3 1000BASE-T is only available on the z114 and occupies one slot in the I/O drawer. It has two ports that belong to one CHPID and connect to a 1000 Mbps (1 Gbps), 100 Mbps, or 10 Mbps Ethernet LAN. Each port has an RJ-45 receptacle for UTP Cat5 or Cat6 cabling, which supports a maximum distance of 100 meters.

3.9.7 OSA-Express2

This section describes the connectivity options offered by the OSA-Express2 features. The following OSA-Express2 features are supported on zEnterprise CPC:

- 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 System z servers are not supported on a zEnterprise CPC 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 uses a long wavelength laser as the optical transceiver.

A multimode (62.5 or 50 µm) fiber cable can be used with the OSA-Express2 GbE LX feature. The use of these multimode cable types requires a 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 uses a short wavelength laser as the optical transceiver.

OSA-Express2 1000BASE-T Ethernet feature

The OSA-Express2 1000BASE-T Ethernet feature 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 or Cat6 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.

3.9.8 Coupling links

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

- Internal Coupling (IC) links, 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, using 9 µm single mode fiber optic cables, and repeated distances up to 100 km using System z-qualified DWDM equipment. ISC-3s are supported exclusively in peer mode.

  > **Statement of Direction:** z196 and z114 are the last systems to offer ordering of ISC-3. Customers who use ISC-3 connections in their current environment are encouraged to migrate to InfiniBand coupling links.

- InfiniBand (HCA2-O) coupling links (12x IFB) are used for z/OS-to-CF communication, CF-to-CF traffic, or STP messaging at distances up to 150 meters (492 feet) using industry standard OM3 50 µm fiber optic cables.
  - 12x InfiniBand coupling links support single data rate (SDR) at 3 GBps when z196, z114, or z10 is connected to z9.
  - 12x InfiniBand coupling links support double data rate (DDR) at 6 GBps for a z196, z114, or z10 to z196, z114, or z10 connection.
- InfiniBand (HCA3-O) coupling links (12x IFB) are used for z/OS-to-CF communication at distances up to 150 meters using industry standard OM3 50 µm fiber optic cables.

  When no more than four CHPIDs are defined per port, and an HCA3-O to HCA3-O connection is set up, the IFB3 protocol is used. When using the IFB3 protocol, synchronous service times are designed to be 40% faster than when using the IFB protocol.

  An HCA3-O to HCA2-O connection is supported, but then the standard IFB protocol is used. HCA3-O cannot connect to HCA1-O on a System z9®.

  12x InfiniBand coupling links support double data rate (DDR) at 6 GBps for a z196, z114, or z10 to z196, z114, or z10 connection.
- InfiniBand (HCA2-O LR[1]) coupling links (1x IFB) 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. Connectivity to HCA3-O LR is supported.
- InfiniBand (HCA3-O LR) coupling links (1x IFB) 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. Connectivity to HCA2-O LR is supported.

  The HCA3-O LR has four ports, as opposed to all other HCA-O fanouts, which have two ports.

HCA3-O LR and HCA2-O LR now support up to 32 subchannels (devices) per CHPID, versus the current seven devices per CHPID.

1 HCA2-O LR is only available on z114 or z196 when carried forward during an upgrade.

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 IFB) might be higher than ICB-4 and InfiniBand (1x IFB) 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-4 links or ISC-3 links. ICB-4 is not supported on z196 and z114.

See the *Coupling Facility Configuration Options* white paper for a more specific explanation regarding the use of ICB-4 or ISC-3 technology versus migrating to InfiniBand coupling links. The white paper is available at the following web page:

http://www.ibm.com/systems/z/advantages/pso/whitepaper.html

For details about all InfiniBand features see the *IBM System z Connectivity Handbook*, SG24-5444, or *Infiniband Coupling Links on System z,* SG24-7539.

3.10 Cryptographic functions

The zEnterprise CPCs include 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 functions 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 zEnterprise CPC includes the following elements:

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

3.10.1 CP Assist for Cryptographic Function

The CP Assist for Cryptographic Function (CPACF) 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.

Figure 3-14 shows the layout of the zEnterprise compression and cryptographic coprocessor (CoP). Each chip contains two coprocessors for data compression and encryption functions. Each CoP is shared by two cores. Every processor in the zEnterprise CPC characterized as a CP, zAAP, zIIP, or IFL has access to the CPACF.

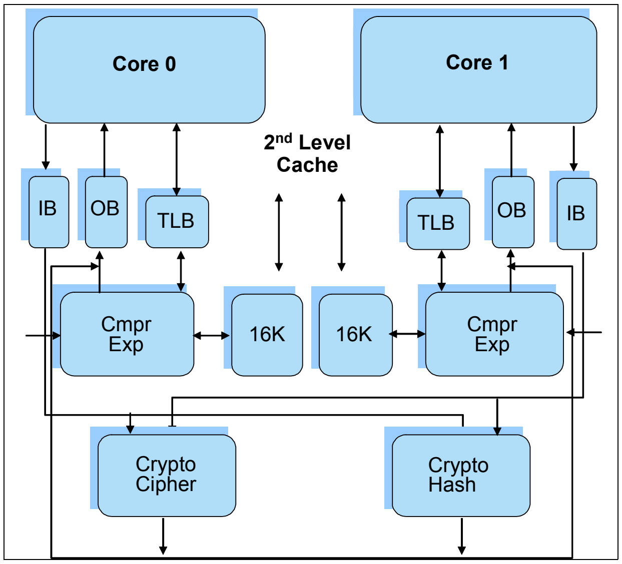

Figure 3-14 zEnterprise compression and cryptographic coprocessor

CPACF supports the full standard in SHA algorithm for hashing and AES and DES algorithms for symmetric encryption. CPACF is designed to provide significant throughput improvements for encryption of large volumes of data and 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 exploiting the CPACF protected key encryption.

The CPACF protected key helps to ensure that keys are not visible to applications and operating systems when used for encryption.

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

3.10.3 Trusted key entry workstation

The Trusted Key Entry (TKE) workstation is an optional feature that provides a secure, remote and local, flexible method of providing master key and operational key entry and managing PCIe Cryptographic Coprocessors. The TKE workstation is a combination of hardware and software. Included with the system unit are a mouse, keyboard, flat panel display, PCIe adapter, and writable USB media to install TKE Licensed Internal Code (LIC). 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 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 V7.1, which is required to support the zEnterprise CPCs.

3.11 Coupling and clustering

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

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

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).

A new generation of host channel adapters (HCA3s) for coupling, the HCA3-O fanout for 12x InfiniBand (12x IFB), is available for z114 and z196. Designed for improved service times, using the 12x InfiniBand3 (12x IFB3) protocol and the HCA3-O LR fanout for 1x InfiniBand (1x IFB) now provides four ports of connectivity and optional additional subchannels for extended-distance solutions.

3.11.1 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: The IBM zEnterprise 196 and the zEnterprise 114 are the last System z servers to offer ordering of ISC-3.

Enterprises should begin migrating from ISC-3 to IFB coupling links.

3.11.2 Internal Coupling (IC)

The IC channel emulates in LIC the coupling facility connection between images within a single system. It operates at memory speed, and no hardware is required.

3.11.3 InfiniBand coupling links

IFB coupling links are high-speed links that are available on zEnterprise System, System z10, and System z9 servers. There are four types of Host Channel Adapter (HCA) fanouts used for IFB coupling links on the z196 and z114:

- HCA2-O fanout supporting 12x InfiniBand (12x IFB)
- HCA2-O Long Reach (LR) fanout supporting 1x InfiniBand (1x IFB)
- HCA3-O fanout supporting InfiniBand 12x InfiniBand (12x IFB)
- HCA3-O Long Reach (LR) fanout supporting 1x InfiniBand (1x IFB)

The HCA3-O delivers improved service times with the IFB3 protocol, and the HCA3-O LR delivers four ports of connectivity and optional additional subchannels for extended-distance solutions.

Time source for STP traffic

IFB links 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 *Infiniband Coupling Links on System z,* SG24-7539.

3.11.4 Coupling Facility Control Code (CFCC) Level 17

CFCC Level 17 is available on the z196 and z114 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 CF link throughput can increase.
- CFCC Level 17 supports up to 2047 structures, while up to 1023 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 more structures than ever before.

CF structure sizing changes are expected when going from CFCC Level 16 to CFCC Level 17. CFCC code now requires 512 MB of memory. We suggest reviewing the CF LPAR size by using the CFSizer tool available at the following web page:

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

3.12 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.

3.12.1 Pulse Per Second

Two external clock facility (ECF) cards are a standard feature of the z196. The ECF cards are located in the processor cage of the z196 and provide a dual-path interface for Pulse Per Second (PPS) support. The z114 provides PPS connectivity on the two Master Oscillator Cards located in the first processor drawer. A cable connection from the PPS port to the PPS output of the NTP server is required when the z196 or z114 is using STP and configured in an STP-only CTN using NTP with PPS as the external time source.

3.12.2 Server Time Protocol (STP)

STP 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 CPCs and CFs to maintain time synchronization with each other.

A System z or CF can 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, see *Server Time Protocol Planning Guide,* SG247280, and *Server Time Protocol Implementation Guide*, SG24-7281.

Statement of Direction: The IBM zEnterprise 196 and the zEnterprise 114 are the last System z servers to support dial-up modems for use with the Remote Support Facility (RSF) and the External Time Source (ETS) option of Server Time Protocol (STP).

The currently available Network Time Protocol (NTP) server option for ETS and internet time services available using broadband connections can be used to provide the same degree of accuracy as dial-up time services.

Enterprises should begin migrating from dial-up modems to Broadband for RSF connections.

3.12.3 Network Time Protocol (NTP) support

NTP support is available on z196, z114, and z10 EC servers and has been added to the STP code on z9. This implementation answers the need for a single time source across the heterogeneous platforms in the enterprise. With this implementation, the z196, z114, z10, and z9 systems support the use of NTP as time sources.

3.12.4 Time coordination for zBX components

Network Time Protocol (NTP) clients, running on blades in the zBX, can synchronize their time to the NTP server provided by the Support Element (SE) every hour. Therefore, it is important for the SE's clock to maintain time accuracy.

An enhancement has been made to improve the time accuracy of the SE's Battery Operated Clock (BOC) by synchronizing the SE's BOC to the server's Time-of-Day (TOD) clock every hour, instead of the previous synchronization, which took place every 24 hours. This provides the capability for the components in the zBX to maintain an approximate time accuracy of 100 milliseconds to an NTP server if they synchronize to the SE's NTP server at least once an hour. This enhancement is exclusive to the z196 and z114.

3.13 HMC and SE

The HMC and SE are appliances that together provide hardware platform management for System z. Hardware platform management covers a complex set of configuration, operation, monitoring, service management tasks, and services that are essential to the use of the hardware platform product. See 4.3, “Hardware Management Console functions” on page 103, for more information about HMC functions and capabilities.

The HMC is attached to a LAN, as is the server's 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 HMC, 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:

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

An HMC consists of the following elements:

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

The zEnterprise CPC is supplied with a pair of integrated ThinkPad SEs. One is always active while the other is an alternate. Power for the SEs is supplied by the CPC's power supply, and there are no additional power requirements. The SEs connect to the bulk power hub (BPH) in the Z frame of the zEnterprise CPC. There is an additional connection from the hub to the HMC using the VLAN capability of the BPH.

3.14 Power and cooling

The power service specifications for the zEnterprise CPCs are the same as their particular predecessors, but the power consumption is more efficient. A fully loaded z196 CPC maximum consumption is 27.4 kW, the same as a z10 EC, but with a maximum performance ratio of 1.64, which is a much higher exploitation on the same footprint.

Power considerations

The zEnterprise CPCs operate with two completely redundant power supplies. Each of the power supplies have its individual line-cords or pair of line-cords for the z196 or z114, depending on the configuration. The number of line cords required depends on the system configuration. Line cords attach 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. There is a Balanced Power Plan Ahead feature available for future growth, also assuring adequate and balanced power for all possible configurations. With this feature, downtimes for upgrading a server will be eliminated by including the maximum power requirements in terms of Bulk Power Regulators (BPR) and line cords to your installation.

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

For the z114 M05 with a maximum of two I/O drawers or an I/O drawer plus a PCIe I/O drawer, or the z114 M10 with a maximum of one I/O drawer, an optional dual-power cable 1-phase connection is also available. The Balanced Power Plan Ahead feature is not available in conjunction with the 1 phase power option.

The power requirements depend on the cooling facility installed and on number of books, or respectively processor drawer in the z114 and the number and kind of I/O units installed. Maximum power consumption tables for the various configurations and environments can be found in *Installation Manual - Physical Planning (IMPP),* GC28-6897, for the z196 and in *Installation Manual - Physical Planning (IMPP),* GC28-6907, for the z114.

The z114 can operate in raised-floor and non-raised-floor environments. For both kinds of installation a new overhead power cable option for top exit of the cables is available. In the case of a non-raised-floor environment, the top exit option must be chosen for the top exit power cables the I/O cables.

Statement of Direction: The IBM zEnterprise 196 and the zEnterprise 114 are the last System z servers to support the Power Sequence Controller (PSC) feature.

IBM intends to not offer support for the PSC on future System z servers after the z196 (2817) and z114 (2818). PSC features cannot be ordered and cannot be carried forward on upgrade to such a follow-on server

3.15 zEnterprise BladeCenter Extension

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

The zBX Model 002 (2458-002) connects to the zEnterprise CPCs 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 chassis[2] with a combination of up to 112 blades[3]
- Redundant infrastructure for fault tolerance and higher availability
- Management support through the zEnterprise System HMC and SE

The IBM Smart Analytics Optimizer solution is also offered with the zBX. See 4.9.2, “IBM Smart Analytics Optimizer solution” on page 115 for more information.

The first rack (rack B) in the zBX is the primary rack where one or two BladeCenter chassis reside. Two pairs of top of rack (TOR) switches are included in rack B, one pair for intranode management network (INMN) and another pair for 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 (zCPC and zBX). The OSA-Express for Unified Resource Manager (OSM) CHPID type supports the connectivity from the zEnterprise CPC to the Bulk Power Hubs (BPHs). The BPHs are also connected to the INMN TOR switches in the zBX.

[2] IBM Smart Analytics Optimizer solution can be placed in a maximum of four BladeCenter chassis.
[3] Maximum number of blades supported: 56 for the IBM Smart Analytics Optimizer solution, 28 for DataPower XI50z, 28 for System x XH5.

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 zEnterprise CPC to the zBX.

Figure 3-15 shows the ensemble node connections through the OSA-Express3 1000BASE-T features (CHPID type OSM) and OSA-Express3 10 GbE or OSA-Express4S 10 GbE features (CHPID type OSX) in the zEnterprise CPC.

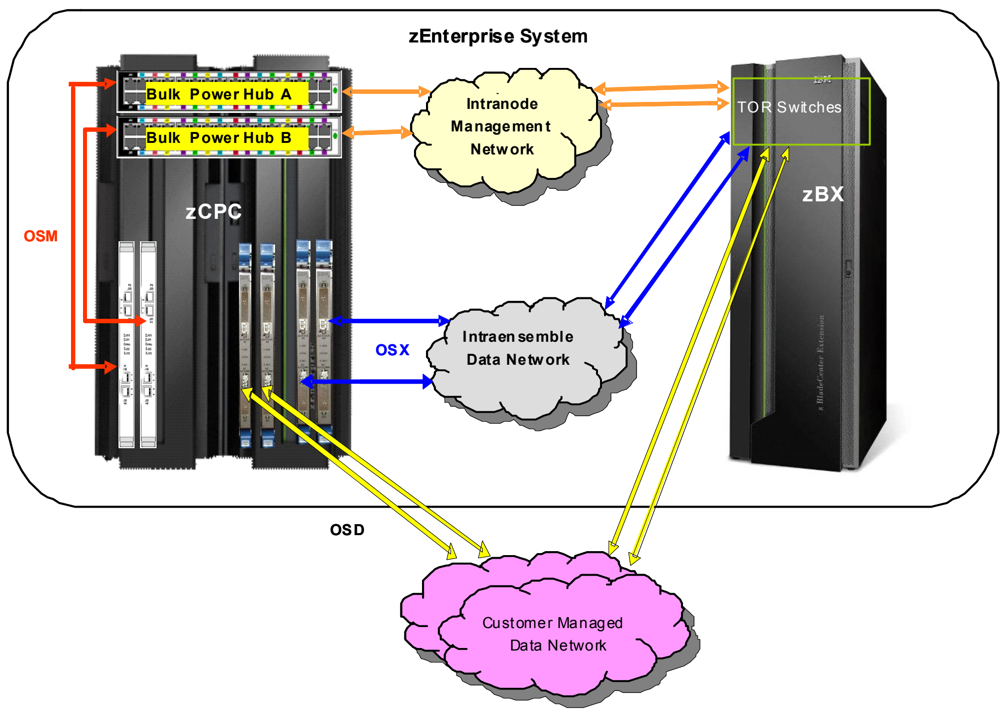

Figure 3-15 INMN, IEDN, and customer-managed data networks in an ensemble

Optionally, as part of the ensemble, any OSA-Express2[4], OSA-Express3 or OSA-Express4S features (with CHPID type OSD) in the zEnterprise CPC can connect to the customer-managed data network. 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 through a SAN switch. 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 Restrictions might apply. OSA-Express2 features can only be carried forward on CPC upgrades from either z9 or z10.

Figure 3-16 shows a rear view of a two-rack zBX configuration.

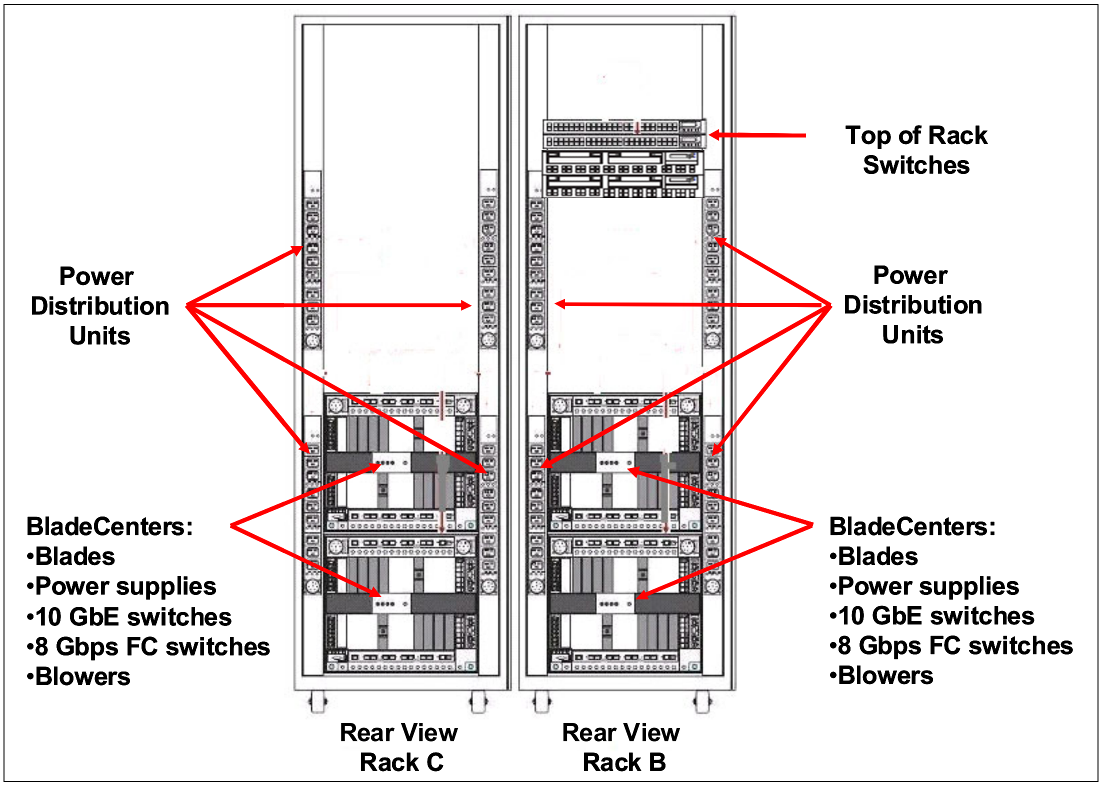

Figure 3-16 zBX rear view: two-rack configuration

The zBX racks include:

- Two TOR 1000BASE-T switches (rack B only) for the intranode management network (INMN)
- Two TOR 10 GbE 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 blade slots
  – Two Advanced Management Modules (AMM)
  – Two Ethernet Switch Modules (ESM)
  – Two 10 GbE high speed switch (HSS) modules
  – Two 8 Gbps Fibre Channel switch modules
  – Two blower modules
- Power Distribution Units (PDUs)

The following blade types are supported in zBX.

- IBM POWER7 PS701 Express blades
- IBM System x blades (HX5 7873 dual-socket 16-core)
- IBM WebSphere DataPower XI50 for zEnterprise blades (double-width)
- Blades for the IBM Smart Analytics Optimizer solution offering

PowerVM is the hypervisor on the POWER7 blades, and the supported operating system is AIX. Linux on System x is the supported operating system for select System x blades, utilizing the zBX integrated hypervisor for IBM System x blades (using Kernel-based virtual machine). Note that both hypervisors are shipped, serviced, and deployed as System z LIC; booted automatically at power-on reset; and isolated on the internal platform management network.

Client-supplied external disks are required with the zBX. Supported FC disk types and vendors with IBM blades are listed on the following web page:

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

Refer to *IBM zEnterprise 196 Technical Guide,* SG24-7833, and *IBM zEnterprise 114 Technical Guide,* SG24-7954, for more information about the number of blades permitted and configuration options.

4

Key functions and capabilities of the zEnterprise System

Compared to previous System z servers, the zEnterprise System has additional security options, improved reliability, availability, and serviceability, and enhanced virtualization capabilities. The zEnterprise System is an integrated platform that brings mainframe and distributed technologies together. The zEnterprise System is made up of the zEnterprise central processor complex (CPC), the zEnterprise BladeCenter Extension, and the zEnterprise Unified Resource Manager.

The z196 and z114 are, respectively, the follow-ons to the System z10 EC and System z10 BC. Like its predecessor, the z196 offers five hardware models, but has a more powerful uniprocessor, more processor units, and new functions and features. The z114 offers two hardware models and employs the same technologies as the z196.

The zEnterprise BladeCenter Extension (zBX) Model 002 with select IBM blades, the IBM WebSphere DataPower XI50 for zEnterprise appliance, and the IBM Smart Analytics Optimizer solution offer specific services and components. The consistent size and shape (or form factor) of each blade allows it to fit in a BladeCenter chassis.

The zEnterprise Unified Resource Manager has an integrated System z management facility to unify resource management, thereby extending System z qualities of service across the infrastructure.

In this chapter we discuss the following functions and capabilities:

4.1 Virtualization

The zEnterprise System is fully virtualized, with the goal of maximizing the utilization of its resources, thereby 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.

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 the zEnterprise System has both. The hardware hypervisor for the zEnterprise CPCs is known as Processor Resource/Systems Manager™ (PR/SM). PR/SM is implemented in firmware as part of the base system, fully virtualizes the system resources, 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 Enterprise Edition is the hypervisor that offers a virtualization solution for any Power workload. It allows use of the POWER7 blade CPU cores and other physical resources, providing better scalability and reduction in resource costs. IBM System x blades have an integrated hypervisor with identical objectives.

PowerVM is EAL4+ certified and is isolated on the internal network of the zEnterprise System, providing intrusion prevention, integrity, and secure virtual switches with integrated consolidation. PowerVM, as well as the integrated hypervisor for the System x blades, is managed by the zEnterprise Unified Resource Manager, so 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 2, “Achieving a (r)evolutionary IT infrastructure” on page 25.

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

4.1.1 zEnterprise CPCs 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. Certain 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 on a z196 (30 on a z114). 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 zEnterprise CPCs and its predecessors) has been designed to meet those stringent requirements with low overhead and the highest security certification in the industry: common criteria EAL5[1] with a specific target of evaluation (logical partitions). This design has been proven in many customer installations over several decades.

For example, the z196 and z114 can handle tens, hundreds, or even (in the case of z196) thousands of virtual servers, so a high context switching rate is to be expected, and accesses to the memory, caches, and virtual I/O devices must be kept completely isolated.

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

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 and are not virtualized.

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 inside the machine are strictly enforced at all times.

Logical processors can be dynamically added to and removed from LPARs. Operating system support is required to take advantage of this capability. Starting with z/OS V1R10, z/VM V5R4, and z/VSE V4R3, 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 available to the operating systems and, in the case of z/VM, to its guests.

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 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 use a wider variety of specialty processors in a single LPAR. The following processor types can be configured to a z/VM-mode partition:

- CPs
- IFLs
- zIIPs
- zAAPs
- ICFs

Memory

To ensure security and data integrity, memory cannot be concurrently shared by active LPARs. In fact, a strict isolation is maintained. When an LPAR is activated, its defined memory is allocated in blocks of 256 MB.

Using the plan-ahead capability, memory can be physically installed with holding enabled. It can subsequently be enabled when it is necessary. z/OS and z/VM support dynamically increasing the memory size of the LPAR.

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.

z/OS is able to take advantage of this support by nondisruptively acquiring and releasing memory from the reserved area. z/VM V5R4 and later versions 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.

LPAR memory is said to be virtualized in the sense that, in each LPAR, memory addresses are contiguous and start at zero. LPARs' memory addresses are different from the machine's absolute memory addresses, which are contiguous and have a single "zero" byte. Do not confuse this with the operating system virtualizing its LPAR memory, which is done through the creation and management of multiple address spaces.

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 1023 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 a program's right to read or write in each page frame by comparing the page key with the key of the program requesting access. Definition and management of the address spaces is under operating system control. This mechanism has been in use since the System 370. Memory keys were part of, and used by, the original System 360 operating systems. Three addressing modes, 24-bit, 31-bit, and 64-bit, are simultaneously supported.

Operating systems can 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.

I/O virtualization

The zEnterprise CPCs support multiple channel subsystems with 256 channels each, four in the case of the z196 for a total of 1024 channels, and two in the case of thez114 for a total of 512 channels. In addition to the dedicated use of channels and I/O devices by an LPAR, I/O virtualization allows concurrent sharing of channels, and the I/O devices accessed through these channels, by several active LPARs. This function is known as *Multiple Image Facility (MIF)*. The shared channels can 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 permits accessing a device by several addresses (normally one base address and three aliases), thereby 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 (and the equivalent function in z/VM's Control Program) to create PAV structures dynamically, depending on the current I/O demand in the system, 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 high. Thus, the concept of *channel sets* was introduced with the IBM System z9. On the z196, up to three sets of 64 K device addresses are available (up to two on the z114), 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 (z196 only). In total, 196,352 subchannel addresses are available per channel subsystem (130,816 on the z114). 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, thereby providing more connectivity through channel set 0.

To reduce the complexity of managing large I/O configurations further, System z introduces *Extended Address Volumes (EAV)*. EAV is designed to build large disk volumes using virtualization technology. By extending the disk volume size, potentially fewer volumes can be required to hold the amount of data, making systems and data management less complex.

The health checker function in z/OS V1R10 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 zEnterprise CPCs greatly eases the planning requirements and enhances the flexibility and availability of these reconfigurations.

4.1.2 zEnterprise CPCs software virtualization

Software virtualization is provided by the z/VM product. Strictly speaking, it is a function of the Control Program (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 can emulate I/O peripherals (for instance, printers) by using spooling and other techniques, and LAN switches and disks by exploiting memory.

z/VM allows fine-grained allocation of resources. As an example, in the case of processor sharing, the minimum is approximately 1/10,000 of a processor. As 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 can support the dynamic addition of memory. All other changes are concurrent. To render these concurrent definitions nondisruptive requires support by the operating system running in the guest.

Although z/VM imposes no limits on the number of defined VMs, the number of active VMs is limited by the available resources. On a large server, such as the z196, thousands of VMs 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, available from the following web page:

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

4.2 zEnterprise CPC technology improvements

The technology used with the zEnterprise CPCs falls into five categories:

- Microprocessor
- Capacity settings
- Memory
- I/O capabilities
- Cryptography

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

4.2.1 Microprocessor

The zEnterprise CPCs have 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 use for the System z products, resulting in increased packaging density.

As with the z10 EC, the microprocessor chip and the storage control chip for the z196 are packaged together on a multi-chip module (MCM). The MCM contains six microprocessor chips and two storage control chips, while the z10 MCMs included seven chips in total. 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 PCIe I/O drawers, I/O drawers, and I/O cages through one or more cables.

The z114 uses the same chips as the z196, but housed in single-chip modules (SCM). Each microprocessor chip contains four cores with either three or four active cores. The z114 SCM is installed in a processor drawer. The processor drawer also contains the memory arrays, I/O connectivity infrastructure, and various other mechanical and power controls.

The processor drawer also uses cables to connect to the PCIe I/O drawers and I/O drawers.

As new standards are making their way on to the zEnterprise CPCs, these cables are now using the standard PCIe and InfiniBand protocols to transfer large volumes of data between the memory and the I/O cards located in the PCIe I/O drawers, I/O drawers, and I/O cages.

zEnterprise CPC processor chip

The zEnterprise CPC chip provides more functions per chip (four cores on a single chip) because of 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 zEnterprise CPC microprocessor chip has an improved design when compared with the System 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. Design improvements include out-of-order instruction execution, redesigned in-core Level 1 caches and in-chip Level 2 and Level 3 caches, redesigned hardware decimal floating point unit, and over 110 new instructions.

The processor chip (Figure 4-1) 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.

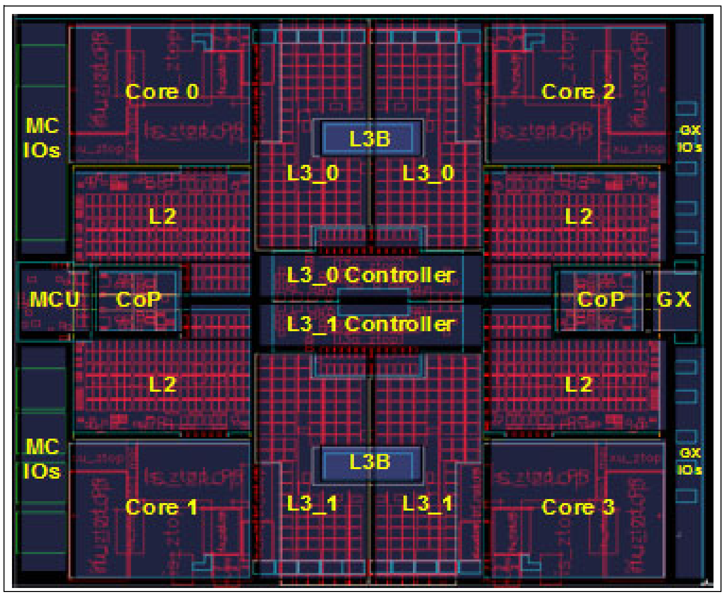

Figure 4-1 zEnterprise CPC Quad-Core microprocessor chip

On-chip cryptographic hardware includes extended key and hash sizes for the DES, AES, and SHA algorithms, as well as new modes and protected key.

Hardware decimal floating point (HDFP) function

HDFP support was introduced with the z9 EC. The zEnterprise CPC, 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 *IBM zEnterprise 196 Technical Guide,* SG24-7833, or *IBM zEnterprise 114*

Technical Guide, SG24-7954. See also “Decimal floating point (z/OS XL C/C++ considerations)” on page 145.

New machine instructions

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

The zEnterprise CPC architectural extensions include over 110 new instructions, the bulk of which was designed in collaboration with software developers to improve compiled code efficiency. These are 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

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 14 for a description.

Table 4-1 lists the maximum number of PUs supported for each operating system image. The physical limits of the hardware determine the usable number of PUs.

Table 4-1 Single system image software support

Operating system	Maximum number of (CPs+zIIPs+zAAPs)[a] or IFLs per system image
z/OS V1R10 and later	80
z/OS V1R9	64
z/OS V1R8	32
z/OS V1R7	32
z/VM V5R4 and later	32[b c]
z/VSE V4 and later	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 6: 80 CPs or IFLs Red Hat RHEL 5: 80 CPs or IFLs
z/TPF V1R1	64 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.

4.2.2 Capacity settings

The z196 and the z114 expand the offer on sub-capacity settings. Finer granularity in capacity levels allows the growth of installed capacity to follow the enterprise growth more closely, 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 4.6, “zEnterprise CPC Capacity on Demand (CoD)” on page 109) provide for managed growth with capacity being available when needed.

z196

The z196 offers four distinct capacity levels for the first 15 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 separate pricing model for non-CP processors regarding purchase and maintenance prices, and 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) are offered for up to 15 CPs, independently of the z196 model installed. The four capacity levels appear in hardware descriptions as feature codes on the CPs.

If more than 15 CPs are configured in the system, they all must be full capacity because all CPs must be on the same capacity level. Granular capacity adds 45 sub-capacity settings to the 80 capacity settings that are available with full capacity CPs (CP7).

z114

Unlike the z196, the z114 offers a larger range of 130 distinct capacity levels for CPs (full capacity and 25 sub-capacities). A processor characterized as anything other than a CP is always set at full capacity. There is, correspondingly, a pricing model for non-CP processors regarding purchase and maintenance prices, and various offerings for software licensing costs. Granularity levels are similar to z10 BC to facilitate upgrades and incremental growth.

A capacity level is a setting of each CP to a sub-capacity of the full CP capacity. Full capacity CPs are identified as CP-Z. On the z114 server, five CPs can be configured as CP-Z. Besides full-capacity CPs, 25 sub-capacity levels (CP-A to CP-Y), each for the five CPs, are offered. The 26 capacity levels appear in hardware descriptions as feature codes on the CPs.

Capacity setting A00 is for ICF or IFL systems only. For either of these systems, its possible to have up to 10 ICFs or IFLs.

Model M10 provides specialty engine scale out capabilities.

Note: The actual throughput that a user experiences varies depending on considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, or the workload being processed.

To help size a System z server to fit your requirements, IBM provides a no-cost tool that reflects the latest IBM LSPR measurements, called the *IBM Processor Capacity Reference (zPCR)*. The tool can be downloaded from the following web page:

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

4.2.3 Memory

The z196 has greatly increased the available memory capacity over previous systems. The system can 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 system area (HSA) is fixed in size (16 GB) and is not included in the memory which the customer orders.

Up to 120 GB of usable memory can be ordered on z114 M05 and up to 248 GB of usable memory can be ordered on z114 M10. In addition, the hardware system area (HSA) is fixed in size (8 GB) and not included in the memory which the customer orders.

> **Note:** 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 later support up to 4 TB.
- z/VM V5R4 and later support up to 256 GB.
- z/VSE V4R2 and later support up to 32 GB.
- z/TPF V1R1 supports up to 4 TB.
- Novell SUSE SLES 11 supports 4 TB and Red Hat RHEL 6 supports 3 TB.

Hardware system area

The zEnterprise CPC has a fixed-size hardware system area. Because the HSA is big enough to accommodate all possible configurations for all logical partitions, several operations that were disruptive on previous servers due to HSA size are now concurrent, improving availability. In addition, planning needs are eliminated. The HSA has a fixed size of 16 GB on the z196 and 8GB on the z114, and resides in a reserved area of memory, separate from customer-purchased memory.

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

- New logical partition to new or existing channel subsystem (CSS)
- New CSS (up to four can be defined on z196 and up to two can be defined on z114)
- New subchannel set (up to three can be defined on z196 and up to two can be defined on z114)
- Maximum number of devices in each subchannel set
- Dynamic I/O enabled as a default
- Logical processors by type
- Cryptographic adapters

Plan-ahead memory

Planning for future memory requirements and installing dormant memory in the CPC allows 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 is activated and the rest remains 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.

Do not confuse plan-ahead memory 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 replace temporarily 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.

The flexible memory option is not available on z114.

Contact your IBM representative to help determine the appropriate configuration for your business.

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 benefits 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 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.

4.2.4 I/O capabilities

The z196 and z114 have many I/O capabilities for supporting high-speed connectivity to resources inside and outside the system. The connectivity of the z196 and z114 is designed to maximize application performance and satisfy clustering, security, storage area network (SAN), and local area network (LAN) requirements.

Multiple subchannel sets (MSS)

MSS is designed to provide greater I/O device configuration capabilities for large enterprises. Up to two subchannel sets for z114 and up to three subchannel sets for z196 can be defined to each channel subsystem.

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

The first subchannel set (SS0) allows the definition 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 Metro Mirror secondary devices only. The third subchannel set applies to ESCON, FICON, and zHPF protocols and is supported by z/OS and Linux on System z.

IPL from an alternate subchannel set

z196 supports IPL from subchannel set 1 (SS1) or subchannel set 2 (SS2), and z114 supports IPL from subchannel set 1 (SS1) and subchannel set 0. Devices used early during IPL processing can now be accessed using subchannel set 1 or subchannel set 2. This allows the users of Metro Mirror (PPRC) secondary devices defined using the same device number and a new device type in an alternate subchannel set to be used for IPL, IODF, and stand-alone dump volumes, when needed.

I/O infrastructure

The FICON features in the zEnterprise CPCs can provide connectivity to servers, FC switches, and a variety of devices (control units, disk, tape, printers) in a SAN environment, while delivering improved throughput, reliability, availability, and serviceability.

High Performance FICON for System z (zHPF)

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 online 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 64 Kb), and z196 removes the 64 kb 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, z114, and System z10. The FICON Express8S, FICON Express8, and FICON Express4 (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 page:

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

Modified Indirect Data Address Word (MIDAW) facility

The MIDAW facility is a system architecture and software exploitation designed to improve FICON performance. This facility was introduced with 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 use 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 IBM Redpaper™ publication called *How does the MIDAW Facility Improve the Performance of FICON Channels Using DB2 and other workloads?*, REDP-4201, provides valuable information. It is available at the following web page:

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

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

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.

Improved IU pacing can optimize the use of the link (for example, help keep a 4 Gbps link fully used 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 can 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 can 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, FICON Express8, and FICON Express8S 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).

z/OS discovery and autoconfiguration

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

Users 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 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 incorporates 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 simplify I/O configuration on zEnterprise CPCs running z/OS and reduce complexity and setup time. zDAC applies to all FICON features supported on zEnterprise CPCs when configured as CHPID type FC.

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 performance of 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 and z114 exclusive function is transparent to operating systems and applicable to all FICON Express8S, 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 Express8S, 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 page:

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

SCSI IPL base function

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

FCP channels to support T10-DIF for enhanced reliability

Recognizing that high reliability is important to maintaining the availability of business-critical applications, the System z Fibre Channel Protocol (FCP) has implemented support of the American National Standards Institute's (ANSI) T10 Data Integrity Field (DIF) standard. Data integrity protection fields are generated by the operating system and propagated through the storage area network (SAN). System z helps to provide added end-to-end data protection between the operating system and the storage device.

An extension to the standard, Data Integrity Extensions (DIX), provides checksum protection from the application layer through the host bus adapter (HBA), where cyclical redundancy checking (CRC) protection is implemented.

T10-DIF support by the FICON Express8S and FICON Express8 features, when defined as CHPID type FCP, is exclusive to z196 and z114. Exploitation of the T10-DIF standard requires support by the operating system and the storage device.

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 VMs. 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.

WWPN tool

A part of the installation of your zEnterprise 196 server is the planning of the SAN environment. IBM has made available a stand-alone tool to assist with this planning prior to the installation. The tool, known as the WWPN 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 using N_Port Identifier Virtualization (NPIV). Thus, the SAN can be set up in advance, allowing operations to proceed much faster after the server is installed.

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

`http://www.ibm.com/servers/resourcelink/`

Fiber Quick Connect for FICON LX

Fiber Quick Connect (FQC), an optional feature on z196 and z114, is now 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 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 can 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.

LAN connectivity

The z196 and z114 offer a wide range of functions that can help consolidate or simplify the LAN environment with the supported OSA-Express features, while also satisfying the demand for more throughput. Improved throughput (mixed inbound/outbound) is achieved via the data router function introduced in the OSA-Express3 features.

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).

QDIO optimized latency mode (OLM)

QDIO 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 or OSA-Express4S without requiring additional program controlled interrupts (PCIs).
- For outbound processing, the OSA-Express3 or OSA-Express4S 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 (IWQ)

IWQ has been designed to reduce overhead and latency for inbound z/OS network data traffic and implement an efficient way for initiating parallel processing. This is achieved by using an OSA-Express3 or OSA-Express4S 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 improves performance, because IWQ can exploit the SMP architecture of the zEnterprise CPC.

zEnterprise ensemble connectivity

With zEnterprise CPC, two CHPID types are introduced to support the zEnterprise ensemble:

- OSM for the intranode management network
- OSX for the intraensemble data network

The INMN is one of the ensemble's two private and secure internal networks. INMN is used by the Unified Resource Manager functions in the primary HMC.

The z196 introduced the OSA-Express for Unified Resource Manager (OSM) CHPID type. The OSM connections are through the Bulk Power Hubs (BPHs) in the zEnterprise CPC. 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.

The IEDN is the ensemble's other private and secure internal network. IEDN is used for communications across the virtualized images (LPARs and virtual machines).

The z196 introduced the OSA-Express for zBX (OSX) CHPID type. The OSX connection is from the zEnterprise CPC to the IEDN TOR switches in zBX. The IEDN requires two OSA Express3 10 GbE or OSA-Express4S 10 GbE ports from separate features.

Virtual local area network (VLAN) support

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.

Virtual MAC (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 requirement 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 the following elements:

- 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 port. Sharing occurs as it does today, but the ability to communicate between sharing QDIO data connections can be restricted through 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 and OSA-Express4S features (CHPID type OSD) on zEnterprise CPC. z/OS supports a similar capability.

QDIO interface isolation for z/OS

Certain environments require strict controls for routing data traffic between severs 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 discards 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 later and for all OSA-Express3 and OSA-Express4S features on zEnterprise CPC.

Open Systems Adapter for NCP (OSN)

The OSN support 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.

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 and OSA-Express4S have the capability for the operating system to query and display the current OSA configuration information (similar to OSA/SF) directly. z/OS exploits this OSA capability with a TCP/IP operator command Display OSAINFO. Display OSAINFO allows the operator to monitor and verify the current OSA configuration, which improves the overall management, serviceability, and usability of OSA-Express3 and OSA-Express4S features.

Display OSAINFO is exclusive to OSA-Express4S and 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." HiperSockets simulates LANs entirely in the hardware. The data transfer is from LPAR memory to LPAR memory, mediated by microcode. zEnterprise CPCs support up to 32 HiperSockets. One HiperSocket can be shared by up to 60 LPARs on a z196 and up to 30 LPARs on a z114. Up to 4096 communication paths support a total of 12,288 IP addresses across all 32 HiperSockets.

HiperSockets Layer 2 support

The HiperSockets internal networks can support two transport modes:

- Layer 2 (link layer)
- 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 way as they do for 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 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 use 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 use 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 unblocks the sending application.

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

HiperSockets Network Traffic Analyzer (HS NTA)

HS NTA is a function available in the zEnterprise CPCs 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.

Statements of Direction for HiperSockets

- HiperSockets Completion Queue

  IBM plans to support transferring HiperSockets messages asynchronously, in addition to the current synchronous manner, on z196 and z114. This could be especially helpful in burst situations. The Completion Queue function is designed to allow HiperSockets to transfer data synchronously if possible and asynchronously if necessary, thus combining ultra-low latency with more tolerance for traffic peaks. HiperSockets Completion Queue is planned to be supported in the z/VM and z/VSE environments.

- HiperSockets integration with the IEDN

  Within a zEnterprise environment, it is planned for HiperSockets to be integrated with the intraensemble data network (IEDN), extending the reach of the HiperSockets network outside of the CPC to the entire ensemble, appearing as a single Layer 2 network. HiperSockets integration with the IEDN is planned to be supported in z/OS V1R13 and z/VM in a future deliverable.

Cryptography

zEnterprise CPCs provide cryptographic functions that, from an application program perspective, can be grouped into 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 and new modes.

CPACF protected key is an enhancement designed to facilitate the continued privacy of cryptographic key material while keeping the desired high performance. CPACF ensures that key material is not visible to applications or operating systems during encryption operations.

CPACF protected key provides substantial throughput improvements for large-volume data encryption and low latency for encryption of small blocks of data.

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

- 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)
- Message Authentication Code (MAC)
- 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, z/VSE, z/TPF, and Linux on System z.

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 functions being exploited. Several functions require software support, which can be downloaded from the web. Also see "ICSF" on page 139 for additional information.

The Crypto Express3 features, residing in the I/O drawer or I/O cage of the zEnterprise CPC, continue 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 zEnterprise are supported:

- Elliptic Curve Cryptography (ECC) Digital Signature Algorithm
- Elliptic Curve Diffie-Hellman (ECDH) algorithm
- Secure key HMAC (Keyed-Hash Message Authentication Code)
- PIN block decimalization table protection
- ANSI X9.8 PIN security
- Enhanced CCA key wrapping to comply with ANSI X9.24-1 key bundling requirements
- PKA RSA OAEP with SHA-256 algorithm
- Enhanced ANSI TR-31 interoperable secure key exchange
- Expanded key support for AES algorithm
- Concurrent Driver Upgrade (CDU) and Concurrent Path Apply (CPA)

Additional key features of Crypto Express3 are as follows:

- 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 Channel Subsystems (CSSs) having access to 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.

Web deliverables

For z/OS downloads, see the z/OS web page:

http://www.ibm.com/systems/z/os/zos/downloads/

4.3 Hardware Management Console functions

HMC/SE Version 2.11.1 is the current version available for the zEnterprise System CPCs.

The HMC and SE are appliances that 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 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, *Create Ensemble*, allows 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 issue commands to their CPCs and zBXs.

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

- Ensemble Membership Flag
- Manage Firmware Suite
- Automate/Advanced Management Firmware Suite (optional)

Additional features may be needed, depending on the selection of Blade options for the zBX.

See 2.2, “Unified Resource Manager” on page 28, for more information about these HMC functions and capabilities.

4.3.1 HMC enhancements

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

- The Monitors Dashboard supersedes the System Activity Display (SAD), 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 zCPC on the HMC. This task provides similar data along with a historical summary of processor and channel use. 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.
- 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 through a web browser.
- Partition capping values can be scheduled and are specified on the Change LPAR Controls scheduled operation support. Viewing details about an existing Change LPAR Controls schedule operation is available on the SE.
- A removable writable media was added as an alternate to the HMC DVD-RAM. This media is the USB Flash Memory Drive (UFD).

For more information about the HMC, refer to:

- *IBM zEnterprise 196 Technical Guide,* SG24-7833
- *IBM zEnterprise 114 Technical Guide,* SG24-7954

4.3.2 Considerations for multiple HMCs

Often multiple HMC instances to manage an overlapping collection of systems are deployed. Today, all 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 2.2, “Unified Resource Manager” on page 28, 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 (Figure 3-15 on page 79).

4.4 zEnterprise CPC 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 a Sysplex, coordinating the time across all the systems in the Sysplex is critical to its operation.

The z196 and z114 support the Server Time Protocol and can participate in a common time network.

4.4.1 Server time protocol (STP)

STP 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 or z114 and CFs in Sysplex environments.

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

STP is implemented in LIC as a system-wide facility of z196 or z114 (and other System z servers and CFs). STP presents a single view of time to PR/SM and provides the capability for multiple CPCs and CFs to maintain time synchronization with each other. A System z server or CF is 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 CPCs and CFs using short communication links, such as IFB links, compared with CPCs and CFs using long ISC-3 links across sites. With the z196 and z114, STP supports 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 can reduce the fiber optic infrastructure requirements in a multi-site configuration. Dedicated links cannot be required to transmit timing information.

The CTN concept is used to meet two key goals of z196, z114, and System z clients:

- Concurrent migration from an existing External Time Reference (ETR) network to a timing network using STP.
- Capability of CPCs 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 recovery enhancement

When HCA3-O or HCA3-O LR coupling links are used, an unambiguous “going away signal” will be sent when the server on which the HCA3 is running is about to enter a failed state. When the going away signal sent by the Current Time Server (CTS) in an STP-only CTN is received by the Backup Time Server (BTS), the BTS can safely take over as the CTS without relying on the previous recovery methods of Offline Signal (OLS) in a two-server CTN or the Arbiter in a CTN with three or more servers.

Note that the already available STP recovery design is still available for the cases when a going away signal is not received or for other failures besides a server failure.

4.4.2 Network time protocol (NTP) client support

The use of NTP 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 available in the support element (SE) code of the z196 and z114. The code interfaces with the NTP servers. This allows an NTP server to become the single time source for z196 and z114 and for other servers that have NTP clients. NTP can be used only for an STP-only CTN environment.

Pulse per second (PPS) support

Certain 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 PPS signal. The redundant design allows continuous operation, in case of failure of one card, and concurrent card maintenance.

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

For z114 the PPS connectivity is provided through two Master Oscillator Cards in the first processor drawer.

NTP server on HMC

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

Statement of Direction: The zEnterprise 196 and the zEnterprise 114 are the last systems to support dial-up modems for use with the ETS option of Server Time Protocol (STP).

The currently available Network Time Protocol (NTP) server option for ETS as well as Internet time services available using broadband connections can be used to provide the same degree of accuracy as dial-up time services.

Enterprises should begin migrating from dial-up modems to Broadband for RSF connections.

For a more in-depth discussion of STP, see the *Server Time Protocol Planning Guide,* SG247280, and the *Server Time Protocol Implementation Guide*, SG24-7281.

4.5 zEnterprise CPC Power functions

As environmental concerns raise the focus on energy consumption, zEnterprise CPCs offer a holistic focus on the environment. New efficiencies and functions like power saving and power capping enable a dramatic reduction of energy usage and floor space when consolidating workloads from distributed servers.

4.5.1 High voltage DC power

In data centers today, many businesses are paying increasing electric bills and are also running out of power. The zEnterprise CPC 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 CPCs to directly use the high voltage DC distribution in 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 zEnterprise CPC'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. Because HV DC is a new technology, there are multiple proposed standards. The zEnterprise CPC 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 zEnterprise CPC 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.

4.5.2 Internal Battery Feature (IBF)

IBF is an optional feature on the zEnterprise CPCs. See Figure 3-3 on page 51 for the z196 or Figure 3-6 on page 53 for the z114 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 at the z196 and up to 25 minutes at the z114 of full power, depending on the I/O configuration.

4.5.3 Power capping and power saving

zEnterprise Systems support power capping, which gives the ability to control the maximum power consumption and reduce cooling requirements (especially with zBX). To use power capping, the Automate Firmware Suite feature must be ordered. This feature is used to enable the Automate suite of functionality associated with the Unified Resource Manager. The Automate suite includes representation of resources in a workload context, goal-oriented monitoring and management of resources, and energy management. A static power-saving mode is also available for the z196 when the Automate Firmware Suite feature is installed. It uses frequency and voltage reduction to reduce energy consumption and can be set up ad hoc or as a scheduled operation. It means, for example, in periods of low utilization or on CBU systems, that customers can set the system in a static power saving mode. Power Saving functions are also provided for the blades in the zBX.

4.5.4 Power estimation tool

The power estimator tool for the zEnterprise CPC allows you to enter your precise server configuration to produce an estimate of power consumption. Log in via the Resource Link and navigate to **Planning** → **Tools** → **Power Estimation Tools**. Specify the quantity for the features that are installed in your machine. This tool estimates the power consumption for the specified configuration. The tool does not verify that the specified configuration can be physically built.

Note: The exact power consumption for your machine will vary. The objective of the tool is to produce an estimation of the power requirements to aid you in planning for your machine installation. Actual power consumption after installation can be confirmed on the HMC monitoring tools.

4.5.5 z196 Hybrid cooling system

The z196 CPC 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 machine refrigeration unit (MRU) technology, 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.

4.5.6 z196 Water cooling

The z196 introduces water cooling unit (WCU) technology, which provides the ability to cool systems with user-chilled water. 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 MRU to WCU require a frame roll and is not done in the field. The water cooling option is not available on z114 upgrades to the z196.

4.5.7 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 vendor 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, AIX and Linux on System p, and Linux on System z. See 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 must be taken that limiting power consumption does not affect performance.
  - Sensors and alerts can warn the user if limiting power to this server can affect performance.
- Certain data can be gathered from the SNMP API on 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

4.6 zEnterprise CPC Capacity on Demand (CoD)

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

The zEnterprise CPCs can perform 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 zEnterprise CPCs.

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 web page:

http://www.ibm.com/servers/resourcelink

For more information regarding the CoD offerings, see the *IBM zEnterprise 196 Technical Guide,* SG24-7833, or the *IBM zEnterprise 114 Technical Guide,* SG24-7954.

4.6.1 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 processor capacity on an existing zEnterprise CPC, 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.

4.6.2 Temporary upgrades

Temporary upgrades of a zEnterprise CPC 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 CoD function

On/Off CoD is a function available on the zEnterprise CPC that enables concurrent and temporary capacity growth of the CPC. On/Off CoD can be used for customer peak workload requirements, for any length of time, and has a daily hardware charge and can have an associated SW charge. On/Off CoD offerings can be pre-paid or post-paid. Capacity tokens are available on zEnterprise CPCs. 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.

CBU function

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 (or processor drawers) of the backup system, either as unused PUs or as a possibility to increase the CP capacity level on a sub-capacity system, or both. A CBU contract must be in place before the LIC-CC code that enables this capability can be loaded on the system. 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.

CPE function

CPE 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 can be active for a maximum of three days.

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

4.6.3 z/OS capacity provisioning

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

4.7 Throughput optimization with zEnterprise CPC

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 in 3.3, “z196 processor cage, books, and MCM” on page 53, books are connected in a star configuration to minimize the distance. The z114 (as with the z10 BC) uses the processor drawer concept, which is analogous to, but not the same as, the book concept in the high-end System z servers. Other non-negligible effects result from data latency when grouping and dispatching work on a set of available logical processors. 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 use 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 zEnterprise CPC 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 zEnterprise CPCs, 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 the minimum number of logical processors needed for the priority (weight) of the LPAR. PR/SM attempts to group the logical processors into the same book (on a z196) and, if possible, the same chip. The result is to reduce the multi-processor effects, maximize use of shared cache, and lower the interference among multiple partitions.

The z/OS dispatcher has been 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 can be dispatched to a small subset of logical processors, which PR/SM ties 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 zEnterprise CPC. The entire zEnterprise CPC 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.

4.8 zEnterprise CPC performance

The System z microprocessor chip of the zEnterprise CPC has a high-frequency design that leverages IBM leadership technology and offers more cache than other chips. In addition, an enhanced instruction execution sequence delivers world-class per-thread performance. There are more than 110 new instructions that will help to deliver CPU-centric performance. For CPU-intensive workloads, additional gains can be achieved via multiple compiler-level improvements. Improved performance of the zEnterprise CPC is a result of the enhancements that we covered in Chapter 3, “zEnterprise System hardware overview” on page 45.

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.

For small and medium businesses, the z114 system offers up to 14 microprocessors running at 3.8 GHz and provides up to 18% uniprocessor performance improvement and up to 12% improvement in total system capacity for z/OS, z/VM, and Linux on System z workloads, as compared to the z10 BC.

On average, the z114 offers 12% improvements in performance over its predecessor, the IBM System z10 Business Class.

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 the following web page:

`https://www-304.ibm.com/servers/resourcelink/lib03060.nsf/pages/lsprindex?`

The MSU ratings are available at the following web page:

`http://www-03.ibm.com/systems/z/resources/swprice/reference/exhibits/`

LSPR workload suite: zEnterprise CPC changes

Historically, LSPR capacity tables, including pure workloads and mixes, have been identified with application names or a software characteristic. Examples are CICS, IMS, OLTP-T[2], CB-L[3], LoIO-mix[4], and TI-mix[5]. However, capacity performance is more closely associated with how a workload uses and interacts with a particular processor *hardware design*. Workload capacity performance is sensitive to three major factors:

- Instruction path length
- Instruction complexity
- Memory hierarchy

With the availability of CPU measurement facility (MF) data, the ability to gain insight into the interaction of workload and *hardware design* in production workloads has arrived. CPU MF data helps LSPR to adjust workload capacity curves based on the underlying hardware sensitivities, in particular the processor access to caches and memory. This is known as *nest*

2 Traditional online transaction processing workload (formerly known as IMS)
3 Commercial batch with long-running jobs
4 Low I/O Content Mix Workload
5 Transaction Intensive Mix Workload

activity intensity. With this, the LSPR introduces three new workload capacity categories that replace all prior primitives and mixes:

- LOW (relative nest intensity)

  A workload category representing light use of the memory hierarchy. This is similar to past high scaling primitives.
- AVERAGE (relative nest intensity)

  A workload category representing average use of the memory hierarchy. This is similar to the past LoIO-mix workload and is expected to represent the majority of production workloads.
- HIGH (relative nest intensity)

  A workload category representing heavy use of the memory hierarchy. This is similar to the past TI-mix workload.

These 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. CPU MF data can be collected by z/OS System Measurement Facility on SMF 113 records.

Guidance in converting LSPR previous categories to the new ones is provided, and built-in support has been added to the zPCR tool.

In addition to low, average, and high categories, the latest zPCR provides the low-average and average-high categories, which allow better granularity for workload characterization.

The 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 zCPC LSPR tables contain information pertaining to Linux and z/VM environments.

The LSPR contains the internal throughput rate ratios (ITRRs) for the zEnterprise CPC 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.

Experience demonstrates that System z servers can be run at up to 100% use levels, sustained, although most clients prefer to leave a bit of white space and run at 90% or slightly under. For any capacity comparison, using "one number" like MIPS or MSU metric is not a valid method. That is why, while doing capacity planning, we suggest using zPCR and involving IBM technical support. For more information about zEnterprise CPC performance, refer to *IBM zEnterprise 196 Technical Guide,* SG24-7833, and *IBM zEnterprise 114 Technical Guide,* SG24-7954.

4.9 zEnterprise BladeCenter Extension

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

zEnterprise System characteristics make it especially valuable for mission-critical workloads. 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 and 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 that 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 includes the System Assist Processor for I/O handling, which was implemented the 370-XA architecture, the CF, the Cryptographic processors, and of course all the I/O cards, which are nothing less than specialized dedicated hardware with sophisticated software that offloads processing from the System z processor units (PUs).

The common theme with these specialized hardware components is their seamless integration within the mainframe.

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 as well as Linux workloads on IBM System x blades[6].

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 are not System z PUs, the zBX is handled by System z firmware called *zEnterprise Unified Resource Manager*. The zBX hardware features are part of the mainframe, not add-ons.

zBX also supports appliances such as the IBM WebSphere DataPower XI50 for zEnterprise and *IBM Smart Analytics Optimizer,* a cost-optimized solution for running Data Warehouse and business intelligence queries against DB2 for z/OS.

4.9.1 IBM blades

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

PowerVM handles all the access to the hardware resources, providing a Virtual I/O Server (VIOS) function and the ability to create logical partitions. The logical partitions can be either dedicated processor LPARs, which require a minimum of one core per partition, or shared processor LPARs (micro-partitions), which in turn can be as small as 0.1 core per partition.

A select set of IBM System x blades is available for the zBX. These blades have an integrated hypervisor, and their virtual machines run Linux on System x[6].

Also available on the zBX is the IBM WebSphere DataPower XI50 for zEnterprise appliance.

See 2.1, “zEnterprise ensembles and virtualization” on page 27, for an overview of virtualized resources in the zBX.

[6] Statement of General Direction: In the future, IBM intends to offer select IBM System x blades running Microsoft Windows on the zEnterprise BladeCenter Extension Model 002.

4.9.2 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 BI solution on zEnterprise.

The offering is made up 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 the following benefits:

- 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 AQTs) to be queried. After completing a workload assessment as part of the solution assurance process, the required number of blades can be ordered in quantities of 7, 14, 28, 42, or 56.

About AQTs: Eligible queries for the IBM Smart Analytics Optimizer solutions are 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 tables on which the MQT definition is based. See the article at the following web page:

http://www.ibm.com/developerworks/data/library/techarticle/dm-0509melnyk

More information about Smart Analytics Optimizer and the solution assurance process can be found at the following web page:

http://www-01.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 zEnterprise CPCs, but is controlled from a single zEnterprise CPC server, called the *owning server*. A DB2 data-sharing group spanning the z/OS LPARs of the up to eight zEnterprise CPCs can exploit a single IBM Smart Analytics Optimizer (on the zBX). Other data-sharing groups, running in the same servers, can exploit a different IBM Smart Analytics Optimizer (on different zBXs).

This is an integrated solution offering a centralized environment that extends System z legendary availability and security to heterogeneous BI and DW workloads. These benefits come without change to current applications, because 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 about the benefits and usage of the IBM Smart Analytics Optimizer solution, see *Using IBM System z As the Foundation for Your Information Management Architecture*, REDP-4606.

4.10 Reliability, availability, and serviceability (RAS)

The zEnterprise systems present numerous enhancements in the RAS 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 zEnterprise CPC can concurrently activate a new LIC Driver level. Concurrent activation of the select new LIC Driver level is supported at specifically released synchronization points. However, there are certain LIC updates where a concurrent update or upgrade is not possible.

Availability enhancements include single processor core checkstop and sparing, enhanced cooling system including a water-cooled option for the z196, humidity and altimeter sensors, point-to-point fabric for SMP, and fixed size HSA.

The zEnterprise CPCs introduce a way to increase memory availability, called Redundant Array of Independent Memory (RAIM), where a fully redundant memory system can identify and correct memory errors without stopping. This new concept is unique to the zEnterprise CPCs. 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 zEnterprise CPC (for example, as a result of a disaster recovery situation), the SAPs can be concurrently added to the CPC 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, and also (with appropriate PTFs) by z/OS and z/VSE V4R3 and later. Previously, proper planning was required to add CPs, zAAPs, and zIIPs to a z/OS LPAR concurrently. Adding memory to an LPAR concurrently is possible and is supported by z/OS and z/VM.

zEnterprise CPCs support 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 SAD screens (on the HMC) now include energy efficiency displays.

4.10.1 RAS capability for the HMC

The HMC for the zEnterprise is where the Unified Resource Manager routines are executed.

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 ensure 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 if the primary fails. Although the primary HMC can perform the classic HMC activities in addition to the Unified Resource Manager activities, the backup HMC can only be the backup. No additional tasks or activities can be performed at the backup HMC.

4.10.2 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 through the zEnterprise hardware management structure. The HMC/SE functions of the zEnterprise CPC 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 two top of rack (TOR) switches for each network (INMN and IEDN). These switches provide N + 1 connectivity for the private networks between the zEnterprise CPC 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 zEnterprise CPC. The same zEnterprise CPC processes and controls are used. Any fixes to the zBX machine are downloaded to the owning zEnterprise CPC'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 zEnterprise CPC's HMC.

These and additional features are further described in *IBM zEnterprise 196 Technical Guide,* SG24-7833 and *IBM zEnterprise 114 Technical Guide,* SG24-7954.

4.11 High availability technology for zEnterprise

System z, renowned for the designed reliability, availability, and serviceability capabilities, now provides extended availability technology with PowerHA™ for Power blades in the zBX. First we discuss the System z Parallel Sysplex technology and then the PowerHA technology:

- 4.11.1, "High availability for zEnterprise CPC with Parallel Sysplex" on page 117
- 4.11.2, "PowerHA in zBX environment" on page 120

4.11.1 High availability for zEnterprise CPC with Parallel Sysplex

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 are made up of the following elements:

- 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.

- System-managed CF structure duplexing

  System-managed 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 page:

http://www.ibm.com/systems/z/pso/index.html

- 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.

  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.

- 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 must 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 Sysplex.

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

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 shared-all access to data. This is facilitated by System z I/O virtualization capabilities such as MIF. MIF allows several logical partitions to share I/O paths in a totally secure way, maximizing use 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*. Rather than a quick recovery of a failure, the Parallel Sysplex design objective is *zero failure.*

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 *Parallel Sysplex Application Considerations*, SG24-6523.

- High capacity

  Scales from two to 32 images. Remember that each image can have from one to 80 processor units. CF scalability is near-linear. This contrasts with other forms of 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 CF 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, user, database administrator, and so on. A single system image ensures reduced complexity from both operational and definition perspectives.

Figure 4-2 illustrates the components of a Parallel Sysplex as implemented within the System z architecture. It shows one of many possible Parallel Sysplex configurations.

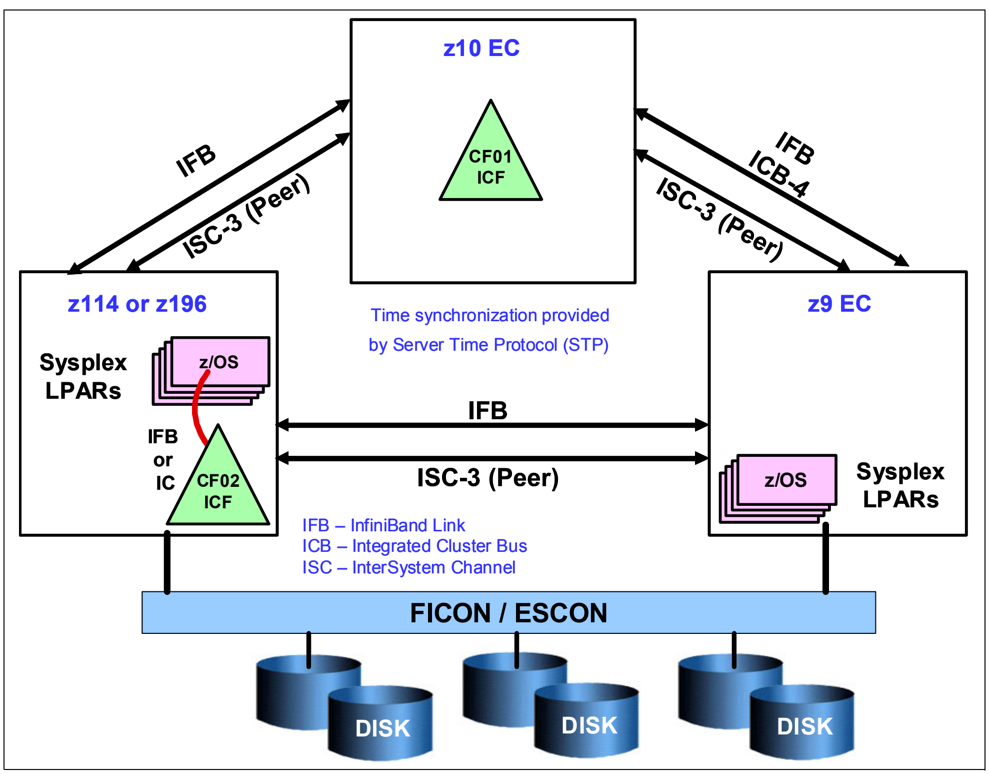

Figure 4-2 Sysplex hardware overview

Figure 4-2 on page 119 shows a z196 or z114 system containing multiple z/OS sysplex partitions and an internal coupling facility (CF02), a z10 EC server containing a stand-alone CF (CF01), and a z9 EC containing multiple z/OS sysplex partitions. STP over coupling links provides time synchronization to all servers. Appropriate CF link technology (IFB or ISC-3) selection depends on server configuration. The ICB-4 coupling link is not supported on zEnterprise CPCs.

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.

4.11.2 PowerHA in zBX environment

An application running on AIX can be fitted with high availability by the use of the PowerHA System Mirror for AIX (formerly known as HACMP™[7]). PowerHA is easy to configure (menu driven) and provides high availability for applications running on AIX.

PowerHA helps define and manage resources (required by applications) running on AIX by providing service/application continuity through platform resources and application monitoring, and automated actions (start/manage/monitor/restart/move/stop).

> **Terminology:** Resource movement and application restart on the second server is known as *failover*.

Automating the failover process speeds up recovery and allows for unattended operations, thus providing improved application availability.

A PowerHA configuration or *cluster* consists of two or more servers[8] (up to 32) that have their resources managed by PowerHA cluster services to provide automated service recovery for the applications managed. Servers can have physical or virtual I/O resources, or a combination of both.

PowerHA performs the following functions at the cluster level:

- Manage/monitor OS and HW resources.
- Manage/monitor application processes.
- Manage/monitor network resources.
- Automate application (start/stop/restart/move).

The virtual servers defined and managed in zBX use only virtual I/O resources. PowerHA can manage both physical and virtual I/O resources (virtual storage and virtual network interface cards).

PowerHA can be configured to perform automated service recovery for the applications running in virtual servers deployed in zBX. PowerHA automates application failover from one virtual server in a POWER blade to another virtual server in a different POWER blade that has a similar configuration.

Failover protects service (masks service interruption) in case of unplanned or planned (scheduled) service interruption. During failover, users might experience a short service interruption while resources are configured by PowerHA on the new virtual server.

[7] High Availability Cluster Multi-Processing
[8] Servers can be also virtual servers. One server equals one instance of the AIX Operating System.

The PowerHA configuration for the zBX environment is similar to standard POWER environments, with the particularity that it uses only virtual I/O resources. Currently, PowerHA for zBX support is limited to failover inside the same zBX.

Figure 4-3 shows a typical PowerHA cluster.

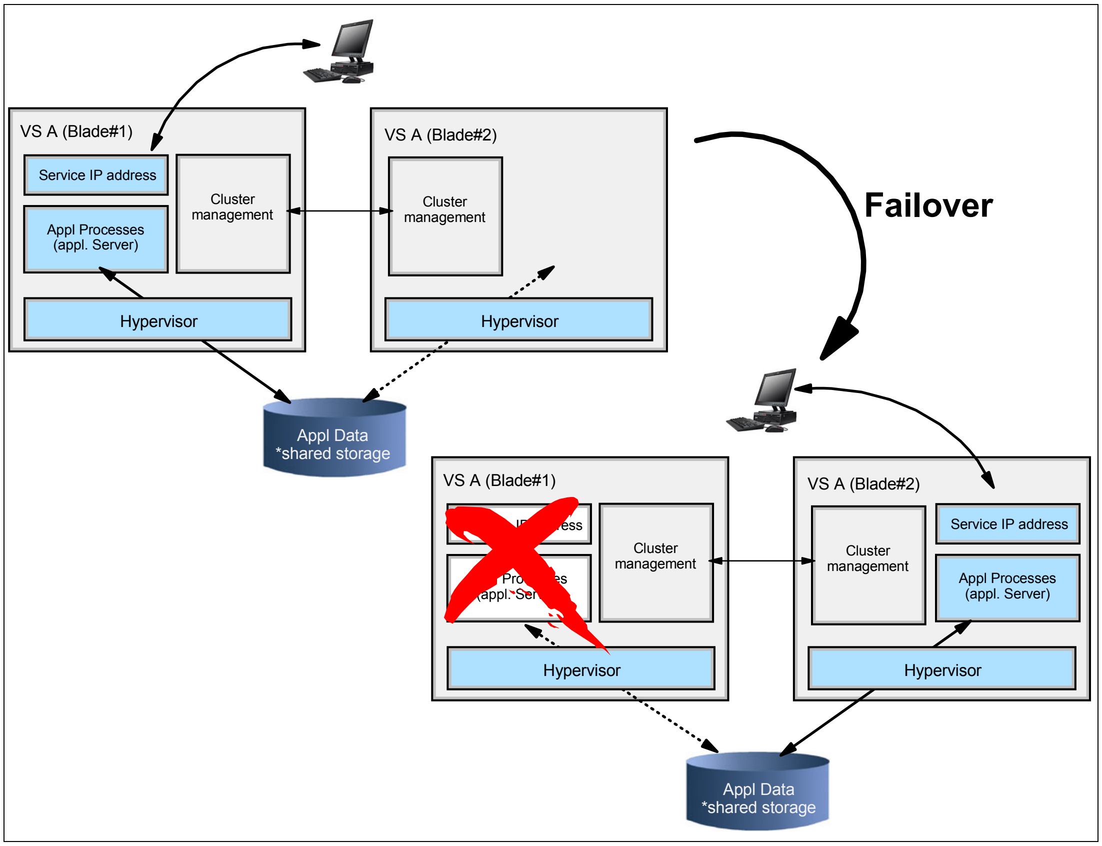

Figure 4-3 Typical PowerHA cluster diagram

A PowerHA configuration must cover the following:

- Network planning (VLAN and IP configuration definition and for server connectivity)
- Storage planning (Shared storage must be accessible to all blades that provide resources for a PowerHA cluster.)
- Application planning (start/stop/monitoring scripts and OS, CPU, and memory resources).
- PowerHA SW installation and cluster configuration
- Application integration (integrating storage, networking, and application scripts)
- PowerHA cluster testing and documentation

A

Operating systems support and considerations

This appendix contains operating system requirements and support considerations for the zEnterprise CPCs and their features.

This chapter discusses the following topics:

Support of the zEnterprise CPCs functions is dependent on the operating system version and release. This information is subject to change. Therefore, for the most current information, see the Preventive Service Planning (PSP) bucket for 2817DEVICE and 2818DEVICE.

Software support summary

The software portfolio for the zEnterprise CPCs 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:

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

For zBX software support see "Software support for zBX" on page 137.

Operating systems summary

Table A-1 summarizes the current and minimum operating system levels required to support the z196 and the z114. Operating system levels that are no longer in service are not covered in this publication. These older levels can provide support for certain 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 V1R13	No	Yes	Not announced[a]	Availability planned for September 2011
z/OS V1R12	No	Yes	September 2013[a]	See the z/OS, z/VM, z/VSE, and z/TPF subsets of the 2817DEVICE and 2818DEVICE Preventative Service Planning (PSP) buckets prior to installing the IBM z196 and IBM z114, respectively.
z/OS V1R11	No	Yes	September 2012[a]	
z/OS V1R10	No	Yes	September 2011[b]	
z/OS V1R9[c]	No	Yes	September 2010[c]	
z/OS V1R8[d]	No	Yes	September 2009[d]	
z/VM V6R1[e]	No[f]	Yes	April 2013	
z/VM V5R4	No[f]	Yes	September 2013[a]	
z/VSE V5R1[g]	No	Yes[h]	Not announced	
z/VSE V4R3	No[i]	Yes[j]	Not announced	
z/VSE V4R2	No[i]	Yes[j]	Not announced	
z/TPF V1R1	Yes	Yes	Not announced	
Linux on System z	See Table A-6 on page 135.	See Table A-6 on page 135.	See footnote [k]	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 V1.10, fee-based corrective service can be ordered for up to two years after the withdrawal of service for z/OS V1R10.
c. 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.
d. 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.

e. z/VM V6R1 requires an architectural level set exclusive to z10 and successors.
f. z/VM supports both ESA/390 mode and z/Architecture mode virtual machines.
g. z/VSE V5R1 has been previewed. It requires an architectural level set exclusive to z9 and successors.
h. z/VSE V5R1 has been previewed. It is designed to support virtual 64-bit addressing both for applications and ISV products.
i. ESA/390 is not supported. However, 31-bit mode is supported.
j. z/VSE V4R2 and V4R1 support 64-bit real addressing only. They do not support 64-bit addressing for user, system, or vendor applications.
k. For information about support-availability of Linux on System z distributions, see:
Novell SUSE:
http://support.novell.com/lifecycle
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 are reviewed regularly when planning for installation of a new system. 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 zEnterprise CPCs 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:

- The Security Server feature of z/OS and z/VM (includes Resource Access Control Facility (RACF®))
- IBM Tivoli Directory Server for z/OS
- z/OS Communications Server and Policy Agent (for policy-based network security)

Application development and languages

Many languages are available for the zEnterprise CPCs 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, are:

- 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 page 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
- 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 makes use of 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, makes use of 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 from their workstation.

Support by operating system

In this section we list the support by in-service operating systems of selected functions of the zEnterprise CPCs. For a detailed description of zEnterprise CPCs and their features see the companion manuals *IBM zEnterprise 196 Technical Guide,* SG24-7833, and *IBM zEnterprise 114 Technical Guide,* SG24-7954. For an in-depth description of all I/O features see the *IBM System z Connectivity Handbook*, SG24-5444.

z/OS

z/OS Version 1 Release 10 is the earliest in-service release supporting the zEnterprise CPCs. The IBM Lifecycle Extension program provides a fee-based extension for defect support for up to two years after end of service. This is currently available for V1.9 and V1.8 and will be required by V1.10 after withdrawal of service on September 30, 2011.

Also note that z/OS.e is not supported on zEnterprise CPCs and that the last release of z/OS.e was z/OS.e Version 1 Release 8.

Table A-2 summarizes the support requirements of selected zEnterprise CPCs function 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	V1 R13	V1 R12	V1 R11	V1 R10[a]	V1 R9[a]	V1 R8[a]
z196 and z114 [b]	Y	Y	Y	Y	Y	Y
Greater than 54 PUs single system image	Y	Y	Y	Y	Y	N
Dynamic add of logical CPs	Y	Y	Y	Y	N	N
zAAP on zIIP	Y	Y	Y	Y[d]	Y[d]	N
Large memory > 128 GB	Y	Y	Y	Y	Y	Y
Large page support	Y	Y	Y	Y	Y	N
Hardware decimal floating point [c]	Y	Y	Y	Y	Y[d]	Y[d]
Capacity Provisioning Manager	Y	Y	Y	Y	Y[d]	N
HiperDispatch	Y	Y	Y	Y	Y	Y
CPACF	Y	Y	Y	Y	Y	Y
CPACF AES-128, AES-192, and AES-256	Y	Y	Y	Y	Y	Y
CPACF SHA-1, SHA-224, SHA-256, SHA-384, SHA-512	Y	Y	Y	Y	Y	Y
CPACF protected key [d]	Y	Y	Y[e]	Y[e]	Y[e]	N
Crypto Express3 [d]	Y	Y	Y[e]	Y[e]	Y[e]	Y[ef]
Crypto Express3-1P [d g]	Y	Y	Y[e]	Y[e]	Y[e]	Y[ef]
Elliptic Curve Cryptography (ECC)	Y	Y[e]	Y[e]	Y[e]	N	N
HiperSockets multiple write facility	Y	Y	Y	Y	Y[d]	N
High Performance FICON	Y	Y	Y	Y[d]	Y[d]	Y[d]
FICON Express8S [h]	Y	Y[d]	Y[d]	Y[d]	Y[d]	Y[d]
FICON Express8 [h]	Y	Y	Y[d]	Y[d]	Y[d]	Y[d]
FICON Express4	Y	Y	Y	Y	Y	Y
FICON Express4-2C [g]	Y	Y	Y	Y	Y	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y	Y	Y	Y	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y	Y[d]	Y[d]	Y[d]	N	N
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	Y	Y	Y	Y	Y[d]	Y[d]
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	Y	Y	Y	Y	Y	Y

Function	V1 R13	V1 R12	V1 R11	V1 R10[a]	V1 R9[a]	V1 R8[a]
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y	Y	Y	Y	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSX	Y	Y[d]	Y[d]	Y[d]	N	N
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN [i] (using four ports)	Y	Y	Y	Y	Y[d]	Y[d]
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN [i] (using two ports)	Y	Y	Y	Y	Y	Y
OSA-Express3-2P [g] Gigabit Ethernet SX CHPID type OSD	Y	Y	Y	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID types OSC and OSD (using four ports)	Y	Y	Y	Y	Y[d]	Y[d]
OSA-Express3 1000BASE-T Ethernet CHPID types OSE and OSN [i] (using four ports)	Y	Y	Y	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet using two ports CHPID types OSC, OSD, OSE, and OSN [i]	Y	Y	Y	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID type OSM [j] (using two ports)	Y	Y[d]	Y[d]	Y[d]	N	N
OSA-Express3-2P [g] 1000BASE-T Ethernet CHPID types OSC, OSD, OSE, and OSN [i]	Y	Y	Y	Y	Y	Y
OSA-Express3-2P [g] 1000BASE-T Ethernet CHPID type OSM [j]	Y	Y	Y	Y	N	N
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
InfiniBand coupling links (1x IB-SDR or 1x IB-DDR) at an unrepeated distance of 10 km	Y	Y	Y	Y[d]	Y[d]	Y[d]
CFCC Level 17	Y	Y[d]	Y[d]	Y[d]	N	N
CFCC Level 16	Y	Y	Y	Y[d]	Y[d]	Y[d]
Assembler instruction mnemonics	Y	Y	Y	Y	Y[d]	Y[d]
C/C++ exploitation of hardware instructions	Y	Y	Y	Y	Y[d]	Y[d]
Layer 3 VMAC	Y	Y	Y	Y	Y	Y[d]
Large dumps	Y	Y	Y	Y	Y[d]	Y[d]
CPU measurement facility	Y	Y	Y	Y	Y[d]	Y[d]

a. With the announcement of IBM Lifecycle Extension for z/OS V1.10, fee-based corrective service can be ordered for up to two years after the withdrawal of service for z/OS V1R10. A similar offering is available for z/OS V1.9 and z/OS v1.8.
b. PTFs are required for support of z196 latest level and z114.
c. The level of decimal floating-point exploitation varies with z/OS release and PTF level.
d. PTFs are required.
e. FMIDs are shipped in a web deliverable.

f. Toleration support only.
g. Available on the z114. Not available on the z196.
h. Support varies with operating system and level.
i. CHPID type OSN does not use ports. LPAR-to-LPAR communication is used.
j. One port is configured for OSM. The second port of the pair is unavailable.

z/VM

At general availability, z/VM V6R1 provides compatibility exploitation support of features, and z/VM V5R4 provides compatibility support only. Table A-3 summarizes the support requirements of selected zEnterprise CPCs function for the currently supported 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 and z114	Y	Y[f]
Greater than 32 PUs for single system image [a]	N	N
Dynamic add of logical CPs	Y	Y
zAAP on zIIP [b e]	Y	Y
Large memory > 128 GB [c]	Y	Y
Large page support [d]	N	N
Hardware decimal floating point [e]	Y	Y
Capacity provisioning [d]	N	N
CPU measurement facility counter	Y[ef]	Y[ef]
HiperDispatch [d]	N	N
CPACF[e]	Y	Y
CPACF AES-128, AES-192, and AES-256[e]	Y	Y
CPACF SHA-1, SHA-224, SHA-256, SHA-384, SHA-512[e]	Y	Y
CPACF protected key [ef]	Y	Y
Crypto Express3 [ef]	Y	Y
Crypto Express3-1P [efh]	Y	Y
Elliptic Curve Cryptography (ECC) [ef]	Y	Y
Execute relative guest exploitation [e]	Y	Y
Restore subchannel facility	Y	Y
HiperSockets multiple write facility [d]	Y[e]	Y[e]
High Performance FICON	N[d]	N[d]
FICON Express8S	Y	Y
FICON Express8 [g]	Y[f]	Y[f]

Function	V6R1	V5R4
FICON Express4	Y[f]	Y[f]
FICON Express4-2C [h]	Y	Y
OSA-Express QDIO data connection isolation for z/VM environments	Y	Y[f]
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y[f]
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y[f]	Y[f]
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	Y	Y[f]
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	Y	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSX	Y[f]	Y[f]
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN [j] (using four ports)	Y[f]	Y[f]
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN [j] (using two ports)	Y	Y
OSA-Express3-2P [h] Gigabit Ethernet SX CHPID type OSD	Y	Y
OSA-Express3 1000BASE-T Ethernet using four ports CHPID types OSC, OSD[i], OSE, and OSN[j] (using four ports)	Y[f]	Y[f]
OSA-Express3 1000BASE-T Ethernet CHPID types OSC, OSD, OSE, and OSN[j] (using two ports)	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID type OSM [f k] (using two ports)	Y	Y
OSA-Express3-2P [h] 1000BASE-T Ethernet CHPID types OSC, OSD, OSE, and OSN [j]	Y	Y
OSA-Express3-2P [h] 1000BASE-T Ethernet CHPID type OSM [f k]	Y	Y
Dynamic I/O support for InfiniBand CHPIDs	Y[l]	Y[l]
InfiniBand coupling links (12x IB-SDR or 12x IB-DDR) at a distance of 150 m	N	N
InfiniBand coupling links (1x IB-SDR or 1x IB-DDR) at an unrepeated distance of 10 km	N	N
CFCC Level 17	Y[e f]	Y[e f]
CFCC Level 16	Y[e]	Y[e]

a. z/VM supports up to 32 real PUs and up to 64 logical PUs per guest.
b. Available for z/OS on virtual machines without virtual zAAPs defined when the z/VM LPAR does not have zAAPs defined to it.
c. 256 GB of central memory are supported by z/VM V5R4 and later. z/VM V5R4 and later support more than 1 TB of virtual memory in use for guests.
d. Not available to guests.

e. Supported for guest use only.
f. Service is required.
g. Support varies with operating system and level.
h. Available on the z114. Not available on the z196.
i. PTFs are required for CHPID type OSD.
j. CHPID type OSN does not use ports, it uses LPAR-to-LPAR communication.
k. One port is configured for OSM. The second port of the pair is unavailable.
l. Support is for dynamic I/O configuration only.

Notes: We suggest that the capacity of z/VM logical partitions and any guests, in terms of the number of IFLs and CPs, real or virtual, be reviewed and adjusted in face of the PU capacity of the z196 and z114.

z/VSE

Table A-4 summarizes the support requirements of selected zEnterprise CPCs function 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	V5R1[a]	V4R3[b]	V4R2[b]
z196 and z114	Y[c]	Y[c]	Y[c]
Dynamic add of logical CPs	Y	Y	N
Large page support for data spaces	Y	Y	N
CPACF	Y	Y	Y
CPACF AES-128, AES-192, and AES-256	Y	Y	Y
CPACF SHA-1, SHA-224, SHA-256, SHA-384, SHA-512	Y	Y	Y
CPACF protected key	N	N	N
Crypto Express3 [d]	Y	Y	Y[e]
Crypto Express3-1P [df]	Y	Y	Y[e]
Elliptic Curve Cryptography (ECC)	N	N	N
FICON Express8S [g]	Y	Y	Y
FICON Express8 [g]	Y	Y	Y
FICON Express4	Y	Y	Y
FICON Express4-2C [f]	Y	Y	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y	N	N
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	Y	Y	Y

Function	V5R1[a]	V4R3[b]	V4R2[b]
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using one port)	Y	Y	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSX	Y	N	N
OSA-Express3 Gigabit Ethernet LX and SX CHPID type OSD and OSN [i] (using four ports [h])	Y	Y	Y
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN [i] (using two ports)	Y	Y	Y
OSA-Express3-2P [f] Gigabit Ethernet SX CHPID type OSD	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID types OSC, OSD, OSE, and OSN [i] (using four ports [h])	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID types OSC, OSD, OSE, and OSN [i] (using two ports)	Y	Y	Y
OSA-Express3-2P [f] 1000BASE-T Ethernet CHPID types OSC, OSD, OSE, and OSN [i]	Y	Y	Y

a. z/VSE V5 is designed to exploit z/Architecture, specifically 64-bit real and virtual-memory addressing. z/VSE V5R1 requires an architectural level set available with IBM System z9 or later.
b. z/VSE V4 is designed to exploit z/Architecture, specifically 64-bit real-memory addressing, but does not support 64-bit virtual-memory addressing.
c. PTFS are required for z114 support.
d. IBM TCP/IP for VSE/ESA V1.5.0 with PTFs is required. Some functions are not supported.
e. Service is required.
f. Available on the z114. Not available on the z196.
g. Support varies with operating system and level.
h. Exploitation of two ports per CPHID type OSD requires a minimum of z/VSE V4R1 with PTFs.
i. CHPID type OSN does not use ports. All communication is LPAR to LPAR.

z/TPF

Table A-5 summarizes the support requirements of selected zEnterprise CPCs function 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-5 TPF and z/TPF support summary

Function	z/TPF V1R1
z196 and z114	Y
Greater than 54 PUs for single system image	Y
Large memory > 128 GB (4 TB)	Y
CPACF[a]	Y
CPACF AES-128, AES-192, and AES-256[a]	Y[b]

Function	z/TPF V1R1
CPACF SHA-1, SHA-224, SHA-256, SHA-384, SHA-512[a]	Y[c]
CPACF protected key	N
Crypto Express3[a]	Y[d]
Crypto Express3-1P[a]	Y[d]
Elliptic Curve Cryptography (ECC)	N
FICON Express8S	Y
FICON Express8	Y
FICON Express4	Y
FICON Express4-2C	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y[e]
OSA-Express4S Gigabit Ethernet LX and SX CHPID types OSD and OSN [f] (using four ports)	Y[e]
OSA-Express4S Gigabit Ethernet LX and SX CHPID types OSD and OSN [f] (using two ports)	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSD	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSX	Y
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN[f] (using four ports)	Y[e]
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN[f] (using two ports)	Y
OSA-Express3-2P [g] Gigabit Ethernet SX CHPID type OSD	Y
OSA-Express3 1000BASE-T Ethernet using four ports CHPID types OSC, OSD, and OSN [f]	Y
OSA-Express3 1000BASE-T Ethernet using two ports CHPID types OSC, OSD, and OSN [f]	Y
OSA-Express3-2P [g] 1000BASE-T Ethernet CHPID types OSC, OSD, and OSN [f]	Y
Coupling over InfiniBand CHPID type CIB	Y[h]
CFCC Level 17	Y
CFCC Level 16	Y

a. Service is required.
b. Supports only AES-128 and AES-256.
c. Supports only SHA-1 and SHA-256
d. Supports only running in accelerator mode (CEX3A)

e. Requires PUT 4 with PTFS.
f. CHPID type OSN does not use ports. All communication is LPAR to LPAR.
g. Available on the z114. Not available on the z196.
h. Compatibility is supported.

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 and z114 toleration support, so any release of these distributions can run on zEnterprise CPCs.

Table A-6 lists the most recent service levels of the current SUSE and Red Hat releases at the time of writing.

Table A-6 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 SP4	No	Yes
Red Hat RHEL 6.1	No	Yes

Table A-7 lists selected zEnterprise CPCs features, showing the minimum level of Novell SUSE and Red Hat distributions that support each feature.

Table A-7 Linux on System z support summary

Function	Novell SUSE	Red Hat
z196 and z114	SLES 10	RHEL 5
Large page support	SLES 10 SP2	RHEL 5.3
Hardware decimal floating point	SLES 11	N[d]
CPACF	SLES 9	RHEL 4.4
CPACF AES-128, AES-192, and AES-256	SLES 10 SP2	RHEL 5.2
CPACF SHA-1, SHA-224, SHA-256, SHA-384, SHA-512	SLES 10 SP2	RHEL 5.2
CPACF protected key	N	N
Crypto Express3	SLES 10 SP3[a] SLES 11 SP1[b]	RHEL 5.4[a] RHEL 6.1[b]
Crypto Express3-1P	SLES 10 SP3[a] SLES 11 SP1[b]	RHEL 5.4[a] RHEL 6.1[b]
Elliptic Curve Cryptography (ECC)	N	N
HiperSockets Layer 2 support	SLES 10 SP2	RHEL 5.3
High Performance FICON	Note[c]	Note[c]

Function	Novell SUSE	Red Hat
FICON Express8S CHPID type FC	SLES 11[c] SLES 10	RHEL 6[c] RHEL 5
FICON Express8, FICON Express4 [d] CHPID types FC and FCP	SLES 10	RHEL 5
FICON Express4-2C [e] CHPID types FC and FCP	SLES 10	RHEL 5
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSD	SLES 10	RHEL 5
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSX	SLES 10 SP4	RHEL 5.6
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	SLES 10 SP2	RHEL 5.2
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	SLES 10	RHEL 5
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSD	SLES 10	RHEL 5
OSA-Expres3 Gigabit Ethernet LX and SX CHPID type OSD and OSN [f](using four ports)	SLES 10 SP2	RHEL 5.2
OSA-Expres3 Gigabit Ethernet LX and SX CHPID types OSD and OSN [f] (using two ports)	SLES 10	RHEL 5
OSA-Express3-2P [e] Gigabit Ethernet SX CHPID OSD	SLES 10	RHEL 5
OSA-Express3 1000BASE-T Ethernet using four ports CHPID types OSD and OSN [f]	SLES 10 SP2	RHEL 5.2
OSA-Express3 1000BASE-T Ethernet using two ports CHPID types OSD, and OSN [f]	SLES 10	RHEL 5
OSA-Express3-2P [e] 1000BASE-T Ethernet CHPID types OSD and OSN [f]	SLES 10	RHEL 5
OSA-Express3 1000BASE-T Ethernet CHPID type OSM [g] (using two ports)	SLES 10 SP4	RHEL 5.6
OSA-Express3-2P 1000BASE-T Ethernet CHPID type OSM	SLES 10 SP4	RHEL 5.6

a. Crypto Express3 toleration mode.
b. Crypto Express3 exploitation.
c. For support of zHPF single track operations.
d. FICON Express4 10KM LX, 4KM LX, and SX features are withdrawn from marketing. All FICON Express2 and FICON features are withdrawn from marketing.
e. Available on the z114. Not available on the z196.
f. CHPID type OSN does not use ports. All communication is LPAR-to-LPAR.
g. One port is configured for OSM. The second port of the pair is unavailable.

IBM is working with its Linux distribution partners so that exploitation of further zEnterprise CPCs functions are provided in future Linux on System z distribution releases. We suggest the following implementations:

- SUSE SLES 11 or Red Hat RHEL 6 be used in any new projects for the zEnterprise CPCs
- Any Linux distributions be updated to their latest service level before migration to zEnterprise CPCs
- The capacity of any z/VM and Linux logical partitions, and 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 and z114

Software support for zBX

The zEnterprise BladeCenter Extension offers the following types of application environments:

- Special purpose, dedicated environments such as the IBM WebSphere DataPower XI50 for zEnterprise and 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.
- System x XH5 based blades. The integrated hypervisor provides a virtualized environment. Linux on System x can be run in this environment[1].

References

Planning information for each operating system is available on the following support web pages:

- z/OS

  http://www.ibm.com/systems/z/os/zos
- z/VM

  http://www.ibm.com/systems/z/os/zvm
- z/TPF

  http://www.ibm.com/systems/z/os/tpf
- z/VSE

  http://www.ibm.com/systems/z/os/zvse
- Linux on System z

  http://www.ibm.com/systems/z/os/linux

z/OS considerations

z196 and z114 base processor support is required in z/OS. With that exception, software changes do not require the new z196 or the new z114 functions and, equally, the new

[1] IBM has issued a Statement of General Direction regarding support of the Windows operating system on System x blades. See 1.5.6, "zEnterprise BladeCenter Extension" on page 20.

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. The z114 employs the same technologies, and with it new technologies were introduced. A notable case is the PCIe I/O support, also made available in the z196. Although support is provided by z/OS starting with z/OS V1R8, exploitation of zEnterprise CPCs' functions is dependent on the z/OS release. z/OS.e is not supported on zEnterprise CPCs.

In general, we suggest the following approaches:

- 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.
- Migrate to STP or Mixed-CTN network prior to introducing a z196 or a z114 into a Sysplex.
- Review z196 and z114 restrictions and considerations prior to creating an upgrade plan.

zAAP on zIIP capability

This capability, first made available on z9 servers, enables, under defined circumstances, workloads eligible to run on Application Assist Processors (zAAPs) to run on Integrated Information Processors (zIIP). This 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 the following situations:

- When there are zIIPs but no zAAPs installed in the system.
- 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 system can have zAAPs installed. Because z/VM can dispatch both virtual zAAPs and virtual zIIPs on real CPs, the z/VM partition does not require any real zIIPs defined to it, although we suggest the use of real zIIPs due to software licensing reasons.

> **Note:** 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.

Large page support

Memory reserved for large page support can be defined with the following parameter in the IEASYSxx member of SYS1.PARMLIB:

```
LFAREA=xx%|xxxxxxM|xxxxxxG
```

This parameter cannot be changed dynamically.

HiperDispatch

The IEAOPTxx member of SYS1.PARMLIB and the **`SET OPT=xx`** command have a HIPERDISPATCH=YES/NO parameter 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 exploit HiperDispatch effectively, adjustment of defined WLM goals and policies might be required. We suggest that you review WLM policies and goals and update them as necessary. You might want to operate with the new policies and HiperDispatch active for a period, turn it off and use the older WLM policies while analyzing the results of using

HiperDispatch, re-adjust the new policies, and repeat the cycle as needed. To change WLM policies, turning HiperDispatch off then on is not necessary.

A health check is provided to verify whether HiperDispatch is enabled on z196 and z114 systems.

Capacity provisioning

Installation of the capacity provision function on z/OS requires the following tasks to be completed:

- 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 the following elements:

- 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 observes other z/OS systems in one or multiple sysplexes. For a description of the capacity provisioning domain see 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 use the available cryptographic functions transparently, whether CPACF or Crypto Express features, to balance the workload and help address the bandwidth requirements of the your applications.

Despite being a z/OS base component, ICSF new functions are generally made available through a web deliverable support a couple of months after a new z/OS release is launched. Due to this fact, new functions must be related to an ICSF FMID instead of a z/OS version.

Table 4-2 lists the ICSF FMIDs and web-deliverable codes for z/OS V1R7 through V1R13. Later FMIDs include the functions of previous ones.

Table 4-2 z/OS ICSF FMIDs

z/OS	ICSF FMID	Web deliverable name	Supported function
V1R7	HCR7720	Included as a z/OS base element	► Crypto Express2 support ► Support for 64-bit callers ► Support for clear DES/TDES key tokens ► 2048-bit key RSA operations ► 19-digit Personal Account Numbers (PANs)
	HCR7730	Cryptographic Support for z/OS V1R6/R7 and z/OS.e V1R6/R7[a]	► Crypto Express2 Accelerator ► CPACF AES-128 and SHA-256 support ► Support for clear AES key tokens ► CKDS Sysplex-wide consistency
	HCR7731	Enhancements to Cryptographic Support for z/OS and z/OS.e V1R6/R7[a]	► ATM Remote Key Loading ► PKDS Key Management ► ISO 16609 CBC Mode TDES MAC ► Enhanced PIN Security
	HCR7750	Cryptographic support for z/OS V1R7-V1R9 and z/OS.e V1R7-V1R8[b]	► CPACF AES-192 and AES-256 ► CPACF SHA-224, SHA-384 and SHA-512 ► 4096-bit RSA keys ► ISO-3 PIN block format
V1R8	HCR7731	Included as a z/OS base element	► ATM Remote Key Loading ► PKDS Key Management ► ISO 16609 CBC Mode TDES MAC ► Enhanced PIN Security
	HCR7750	Cryptographic support for z/OS V1R7-V1R9 and z/OS.e V1R7-V1R8[b]	► CPACF AES-192 and AES-256 ► CPACF SHA-224, SHA-384 and SHA-512 ► 4096-bit RSA keys ► ISO-3 PIN block format
	HCR7751	Cryptographic Support for z/OS V1R8-V1R10 and z/OS.e V1R8[a]	► IBM System z10 BC support ► Secure key AES ► Key store policy ► PKDS Sysplex-wide consistency ► In-storage copy of the PKDS ► 13-digit through 19-digit personal account numbers (PANs) ► Crypto Query service ► Enhanced SAF checking

z/OS	ICSF FMID	Web deliverable name	Supported function
V1R9	HCR7740	Included as a z/OS base element	► PKCS#11 support ► CFB and PKCS#7 padding
	HCR7750	Cryptographic support for z/OS V1R7-V1R9 and z/OS.e V1R7-V1R8[b]	► CPACF AES-192 and AES-256 ► CPACF SHA-224, SHA-384 and SHA-512 ► 4096-bit RSA keys ► ISO-3 PIN block format
	HCR7751	Cryptographic Support for z/OS V1R8-V1R10 and z/OS.e V1R8[a]	► IBM System z10 BC support ► Secure key AES ► Key store policy ► PKDS Sysplex-wide consistency ► In-storage copy of the PKDS ► 13-digit through 19-digit PANs ► Crypto Query service ► Enhanced SAF checking
	HCR7770	Cryptographic support for z/OS V1R9-V1R11	► Crypto Express3 and Crypto Express3-1P support ► PKA Key Management Extensions ► CPACF Protected Key ► Extended PKCS#11 ► ICSF Restructure (Performance, RAS, ICSF-CICS Attach Facility)

z/OS	ICSF FMID	Web deliverable name	Supported function
V1R10	HCR7750	Included as a z/OS base element	► CPACF AES-192 and AES-256 ► CPACF SHA-224, SHA-384 and SHA-512 ► 4096-bit RSA keys ► ISO-3 PIN block format
	HCR7751	Cryptographic Support for z/OS V1R8-V1R10 and z/OS.e V1R8[a]	► IBM System z10 BC support ► Secure key AES ► Key store policy ► PKDS Sysplex-wide consistency ► In-storage copy of the PKDS ► 13-digit through 19-digit PANs ► Crypto Query service ► Enhanced SAF checking
	HCR7770	Cryptographic support for z/OS V1R9-V1R11	► Crypto Express3 and Crypto Express3-1P support ► PKA Key Management Extensions ► CPACF Protected Key ► Extended PKCS#11 ► ICSF Restructure (Performance, RAS, ICSF-CICS Attach Facility)
	HCR7780	Cryptographic support for z/OS V1R10-V1R12	► IBM zEnterprise 196 support ► Elliptic Curve Cryptography ► Message-Security-Assist-4 ► HMAC Support ► ANSI X9.8 Pin ► ANSI X9.24 (CBC Key Wrapping) ► CKDS constraint relief ► PCI Audit ► All callable services AMODE(64) ► PKA RSA OAEP with SHA-256 algorithm[c]

z/OS	ICSF FMID	Web deliverable name	Supported function
V1R11	HCR7751	Included as a z/OS base element	► IBM System z10 BC support ► Secure key AES ► Key store policy ► PKDS Sysplex-wide consistency ► In-storage copy of the PKDS ► 13-digit through 19-digit PANs ► Crypto Query service ► Enhanced SAF checking
	HCR7770	Cryptographic support for z/OS V1R9-V1R11	► Crypto Express3 and Crypto Express3-1P support ► PKA Key Management Extensions ► CPACF Protected Key ► Extended PKCS#11 ► ICSF Restructure (Performance, RAS, ICSF-CICS Attach Facility)
	HCR7780	Cryptographic support for z/OS V1R10-V1R12	► IBM zEnterprise 196 support ► Elliptic Curve Cryptography ► Message-Security-Assist-4 ► HMAC Support ► ANSI X9.8 Pin ► ANSI X9.24 (CBC Key Wrapping) ► CKDS constraint relief ► PCI Audit ► All callable services AMODE(64) ► PKA RSA OAEP with SHA-256 algorithm[c]
	HCR7790	Cryptographic Support for z/OS V1R11-V1R13[d]	► Expanded key support for AES algorithm ► Enhanced ANSI TR-31 ► PIN block decimalization table protection ► Elliptic Curve Diffie-Hellman (ECDH) algorithm ► RSA in the Modulus Exponent (ME) and Chinese Remainder Theorem (CRT) formats.

z/OS	ICSF FMID	Web deliverable name	Supported function
V1R12	HCR7770	Included as a z/OS base element	► Crypto Express3 and Crypto Express3-1P support ► PKA Key Management Extensions ► CPACF Protected Key ► Extended PKCS#11 ► ICSF Restructure (Performance, RAS, ICSF-CICS Attach Facility)
	HCR7780	Cryptographic support for z/OS V1R10-V1R12	► IBM zEnterprise 196 support ► Elliptic Curve Cryptography ► Message-Security-Assist-4 ► HMAC Support ► ANSI X9.8 Pin ► ANSI X9.24 (CBC Key Wrapping) ► CKDS constraint relief ► PCI Audit ► All callable services AMODE(64) ► PKA RSA OAEP with SHA-256 algorithm[c]
	HCR7790	Cryptographic Support for z/OS V1R11-V1R13[d]	► Expanded key support for AES algorithm ► Enhanced ANSI TR-31 ► PIN block decimalization table protection ► Elliptic Curve Diffie-Hellman (ECDH) algorithm ► RSA in the ME and CRT formats.
V1R13	HCR7780	Included as a z/OS base element	► IBM zEnterprise 196 support ► Elliptic Curve Cryptography ► Message-Security-Assist-4 ► HMAC Support ► ANSI X9.8 Pin ► ANSI X9.24 (CBC Key Wrapping) ► CKDS constraint relief ► PCI Audit ► All callable services AMODE(64) ► PKA RSA OAEP with SHA-256 algorithm[c]
	HCR7790	Cryptographic Support for z/OS V1R11-V1R13[d]	► Expanded key support for AES algorithm ► Enhanced ANSI TR-31 ► PIN block decimalization table protection ► Elliptic Curve Diffie-Hellman (ECDH) algorithm ► RSA in the ME and CRT formats.

a. Download is no longer available and has been replaced by the Cryptographic Support for z/OS V1R10-V1R12 web deliverable.
b. This download is installable on V1R6 and V1R7, but it is not supported. Customers running z/OS V1R9 or later should install the Cryptographic Support for z/OS V1R10-V1R12 web deliverable.
c. Service is required.
d. Download is planned to be available during 2H2011.

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 suggest that you consult the *Coupling Facility Configuration Options* white paper when planning to exploit the InfiniBand coupling technology. The white paper is available at the following web page:

http://www.ibm.com/systems/z/advantages/pso/whitepaper.html

Decimal floating point (z/OS XL C/C++ considerations)

The two 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 runs. 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 see 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 runs only on zEnterprise CPCs. 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 disaster recovery.

Coupling Facility and CFCC considerations

Coupling Facility connectivity to a zEnterprise CPC is supported on the z10 BC, z9 EC, z9 BC, or another zEnterprise CPC. The logical partition running the Coupling Facility Control Code (CFCC) can reside on any of these supported systems.

Because coupling link connectivity to z990, z890, and previous System z servers is *not* supported, this might affect the introduction of z196 or z114 into existing installations and require additional planning. For more information see the companion books *IBM zEnterprise 196 Technical Guide,* SG24-7833, and *IBM zEnterprise 114 Technical Guide,* SG24-7954.

zEnterprise CPCs 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 CPCs. (CF logical partitions running on the same CPC share the same CFCC level.)

For additional details about CFCC code levels, see the Parallel Sysplex web page:

http://www.ibm.com/systems/z/pso/cftable.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 zEnterprise CPC systems.

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. To prevent any possible data corruption, ICKDSF must determine all sharing systems that can potentially run ICKDSF. Therefore, this support is required for the z196 and z114.

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

Software licensing

This appendix briefly describes the software licensing options available for zEnterprise CPCs.

Software licensing considerations

The zEnterprise CPC's IBM software portfolio includes operating system software[1] (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 metric groups 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.

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 clients to access IBM product support during the support period and also receive future releases and versions at no additional charge. For details, consult the following resources:

- The *IBM System z Software Pricing Reference Guide*, web page:

  http://public.dhe.ibm.com/common/ssi/ecm/en/zso01378usen/ZSO01378USEN.PDF

- IBM System z Software Pricing web pages, which can be reached from:

  http://www.ibm.com/systems/z/resources/swprice/mlc/index.html

MLC pricing metrics

MLC pricing applies to z/OS z/VSE or z/TPF operating systems. Any mix of z/OS, z/VM, Linux, z/VSE, and z/TPF images is allowed. Charges are based on processor capacity, which is measured in millions of service units (MSU) per hour.

There is a variety of Workload Licence Charges (WLC) pricing structures that support two charge models:

- Variable charges (several pricing metrics)

  Variable charges apply to products such as z/OS, z/VSE, z/TPF, DB2, IMS, CICS, MQSeries®, and Lotus Domino®. There are several pricing metrics employing the following charge types:

  - Full-capacity

    The CPC's total number of MSUs is used for charging. Full-capacity is applicable when the client's CPC is not eligible for sub-capacity.

  - Sub-capacity

    Software charges are based on the utilization of the logical partitions where the product is running.

- Flat charges

  Software products licensed under flat charges are not eligible for sub-capacity pricing. There is a single charge per CPC on the z196. On the z114, price is based on the CPC size.

[1] Linux on System z distributions are not IBM products.

Sub-capacity

For eligible programs, sub-capacity allows software charges based on utilization of logical partitions instead of the CPC's total number of MSUs. Sub-capacity removes the dependency between software charges and CPC (hardware) installed capacity.

Sub-capacity is based on the highest observed rolling 4-hour average utilization for the logical partitions. It is not based on the utilization of each product, but on the utilization of the logical partitions where the product runs (with the exception of products licensed using the SALC pricing metric).

The sub-capacity licensed products are charged monthly based on the highest observed rolling 4-hour average utilization of the logical partitions in which the product runs. This requires measuring the utilization and reporting it to IBM.

The logical partition's rolling 4-hour average utilization can be limited by a defined capacity value on the partition's image profile. This activates the soft capping function of PR/SM, limiting the rolling 4-hour average partition utilization to 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 by 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 utilization is tracked, allowing utilization peaks above the defined capacity value.

Some pricing metrics apply to stand-alone System z servers. Others apply to the aggregation of multiple System z servers' workloads within the same Parallel Sysplex.

For further information about WLC and details about how to combine logical partitions utilization, see the publication *z/OS Planning for Workload License Charges*, SA22-7506, available from the following web page:

http://www-03.ibm.com/systems/z/os/zos/bkserv/find_books.html

IBM zEnterprise 196

Metrics applicable to a stand-alone IBM zEnterprise 196 are:

- Advanced Workload License Charges (AWLC)
- System z New Application License Charges (zNALC)
- Parallel Sysplex License Charges (PSLC)

Metrics applicable to a IBM zEnterprise 196 in an actively coupled Parallel Sysplex are:

- Advanced Workload License Charges (AWLC), when all CPCs are z196 or z114

  Variable Workload License Charges (VWLC) are only allowed under the AWLC Transition Charges for Sysplexes when not all CPCs are z196 or z114.

- System z New Application License Charges (zNALC)

- Parallel Sysplex License Charges (PSLC)

IBM zEnterprise 114

Metrics applicable to a stand-alone IBM zEnterprise 114 are:

- Advanced Entry Workload License Charges (AEWLC)
- System z New Application License Charges (zNALC)

Metrics applicable to a IBM zEnterprise 114 in an actively coupled Parallel Sysplex are:

- Advanced Workload License Charges (AWLC), when all CPCs are z114 or z196

  Variable Workload License Charges (VWLC) are only allowed under the AWLC Transition Charges for Sysplexes when not all CPCs are z196 or z114.
- System z New Application License Charges (zNALC)
- Parallel Sysplex License Charges (PSLC)

Advanced Workload License Charges (AWLC)

Advanced Workload License Charges were introduced with the IBM zEnterprise 196. They utilize the measuring and reporting mechanisms, as well as the existing MSU tiers, from VWLC.

As compared with VWLC, the prices per tier have been lowered, and prices for tiers 4, 5, and 6 are different, allowing for lower costs for charges above 875 MSU. AWLC offers improved price performance as compared with VWLC for all customers above 3 MSUs.

Similarly to Workload Licence Charges, AWLC can be implemented in full-capacity or sub-capacity mode. AWLC applies to z/OS and z/TPF and their associated middleware products such as DB2, IMS, CICS, MQSeries, and Lotus Domino when running on a z196.

For additional information see the AWLC web page:

http://www-03.ibm.com/systems/z/resources/swprice/mlc/awlc.html

Advanced Entry Workload License Charges (AEWLC)

Advanced Entry Workload License Charges were introduced with the IBM zEnterprise 114. They utilize the measuring and reporting mechanisms, as well as the existing MSU tiers, from Entry Workload Licence Charges (EWLC) pricing metric and Midrange Workload Licence Charges (MWLC) pricing metric.

AEWLC also offers improved price performance as compared with EWLC and MWLC for most customers.

Similarly to Workload Licence Charges, AEWLC can be implemented in full-capacity or sub-capacity mode. AEWLC applies to z/OS and z/TPF and their associated middleware products such as DB2, IMS, CICS, MQSeries, and Lotus Domino when running on a z114.

For additional information see the AEWLC web page:

http://www-03.ibm.com/systems/z/resources/swprice/mlc/aewlc.html

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/systems/z/resources/swprice/mlc/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 use 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 low usage can benefit from SALC. Alternatively, you can still choose to license WebSphere MQ under the same metric as the z/OS software stack.

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. Determine the highest daily usage of a program family, which is the highest of 24 hourly measurements recorded each day. *Program* refers to all active versions of MQ.
2. Determine the monthly usage of a program family, which is the fourth highest daily measurement recorded for a month.
3. Use 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/systems/z/resources/swprice/mlc/other.html

Midrange Workload Licence Charges (MWLC)

Midrange Workload Licence Charges applies to z/VSE V4 when running on z196, System z10, and 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.

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/systems/z/resources/swprice/mlc/mwlc.html

Parallel Sysplex Licence Charges (PWLC)

PSLC applies to a large range of mainframe servers. The list can be obtained from the web page:

http://www-03.ibm.com/systems/z/resources/swprice/reference/exhibits/hardware.html

Although it can be applied to stand-alone CPCs, the metric only provides aggregation benefits when applied to group of CPCs in an actively coupled Parallel Sysplex cluster according to IBM terms and conditions.

Aggregation allows charging a product based on the total MSU value of the machines where the product executes (as opposed to all the machines in the cluster). In an uncoupled environment, software charges are based on the MSU capacity of the machine.

For additional information see the PSLC web page:

http://www.ibm.com/systems/z/resources/swprice/mlc/pslc.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:

- Value unit (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.

- Engine-based value unit (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 v6, and certain z/VM middleware, which are priced based on the number of engines.

- Processor value unit (PVU)

  Processor value units are determined from the number of engines, 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:

http://www.ibm.com/systems/z/resources/swprice/zipla/index.html

C

Channel options

Table C-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 IFB feature is required.

Statement of Direction:

- The z196 and the z114 are the last servers to offer ordering of ESCON channels. IBM intends not to support ESCON channels on future System z servers.
- The z196 and the z114 are the last System z servers to offer ordering of ISC-3 Links.

Table C-1 z196 and z114 channel feature support

Channel feature	Feature codes	Bit rate	Cable type	Maximum unrepeated distance[a]	Ordering information	Remarks
Enterprise Systems Connection (ESCON)						
16-port ESCON	2323	200 Mbps	MM 62.5 µm	3 km (800)	New build	
Fiber Connection (FICON)						
FICON Express4-2C 4KM LX	3323	1, 2, or 4 Gbps	SM 9 µm	4 km	Carry forward[b]	Only on z114
FICON Express4 4KM LX	3324				Carry forward	
FICON Express4 10KM LX	3321			10 km	Carry forward	
FICON Express8 10KM LX	3325	2, 4, or 8 Gbps	SM 9 µm	10 km	Carry forward[b]	
FICON Express8S 10KM LX	0409				New build	

Channel feature	Feature codes	Bit rate	Cable type	Maximum unrepeated distance[a]	Ordering information	Remarks
FICON Express4 SX	3322	1, 2, or 4 Gbps	OM1,OM2, OM3	See Table C-2 on page 155.	Carry forward	
FICON Express4-2C SX	3318				Carry forward[b]	Only on z114
FICON Express8 SX	3326	2, 4, or 8 Gbps			Carry forward only	
FICON Express8S SX	0410				New build	
Open Systems Adapter (OSA)						
OSA-Express2 GbE LX	3364	1 Gbps	SM 9 µm MCP 50 µm	5 km 550 m (500)	Carry forward	
OSA-Express3 GbE LX	3362				Carry forward[b]	
OSA-Express4S GbE LX	0404				New build	
OSA-Express2 GbE SX	3365	1 Gbps	MM 62.5 µm MM 50 µm	220 m (166) 275 m (200) 550 m (500)	Carry forward	
OSA-Express3 GbE SX	3363				Carry forward[b]	
OSA-Express3-2P GbE SX	3373				Carry forward[b]	Only on z114
OSA-Express4S GbE SX	0405				New build	
OSA-Express2 1000BASE-T Ethernet	3366	10, 100, or 1000 Mbps	UTP Cat5 or Cat6	100 m	Carry forward	
OSA-Express3 1000BASE-T Ethernet	3367				New build	
OSA-Express3-2P 1000BASE-T Ethernet	3369				New build	Only on z114
OSA-Express3 10 GbE LR	3370	10 Gbps	SM 9 µm	10 km	Carry forward[b]	
OSA-Express4S 10 GbE LR	0406				New build	
OSA-Express3 10 GbE SR	3371	10 Gbps	MM 62.5 µm MM 50 µm	33 m (200) 300 m (2000) 82 m (500)	Carry forward[b]	
OSA-Express4S 10 GbE SR	0407				New build	
Parallel Sysplex						
IC	N/A		N/A	N/A	N/A	

Channel feature	Feature codes	Bit rate	Cable type	Maximum unrepeated distance[a]	Ordering information	Remarks
ISC-3 (peer mode)	0217 0218 0219	2 Gbps	SM 9 µm MCP 50 µm	10 km 550 m (400)	New build	
ISC-3 (RPQ 8P2197 Peer mode at 1 Gbps)[c]		1 Gbps	SM 9 µm	20 km	New build	
HCA2-O (12x IFB)	0163	6 GBps	OM3	150 m	New build	
HCA2-O LR (1x IFB)	0168	2.5 or 5 Gbps	SM 9 µm	10 km	Carry forward	
HCA3-O (12x IFB)	0171	6 GBps	OM3	150 m	New build	
HCA3-O LR (1x IFB)	0170	2.5 or 5 Gbps	SM 9 µm	10 km	New build	
Cryptography						
Crypto Express3	0864	N/A	N/A	N/A	New build	
Crypto Express3-1P	0871	N/A	N/A	N/A	New build	Only on z114

a. Minimum fiber bandwidth distance product in MHz·km for multi-mode fiber optic links are included in parentheses were applicable.
b. Ordering of this feature is determined by the fulfillment process.
c. RPQ 8P2197 enables the ordering of a daughter card supporting 20 km unrepeated 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).

Table C-2 Maximum unrepeated distance for FICON SX features

Cable type\bit rate	1 Gbps	2 Gbps	4 Gbps	8 Gbps
OM1 (62,5 µm at 200 MHz·km)	300 meters	150 meters	70 meters	21 meters
	984 feet	492 feet	230 feet	69 feet
OM2 (50 µm at 500 MHz·km)	500 meters	300 meters	150 meters	50 meters
	1640 feet	984 feet	492 feet	164 feet
OM3 (50 µm at 2000 MHz·km)	860 meters	500 meters	380 meters	150 meters
	2822 feet	1640 feet	1247 feet	492 feet

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 158. Note that some of the documents referenced here might be available in softcopy only.

- *IBM zEnterprise 196 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*Server Time Protocol Implementation Guide*, SG24-7281
- *IBM System z10 Enterprise Class Capacity On Demand*, SG24-7504
- *DS8000 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:

- A collection of frequently asked questions about the IBM zEnterprise System

  http://public.dhe.ibm.com/common/ssi/ecm/en/zsq03051usen/ZSQ03051USEN.PDF
  http://public.dhe.ibm.com/common/ssi/ecm/en/zsq03052usen/ZSQ03052USEN.PDF

- IBM ResourceLink

  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 Guid*e, SB10-7152
- *Planning for Fiber Optic Links*, GA23-0367
- *CHPID Mapping Tool User's Guid*e, 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, and 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

T

U

V

W

IBM Redbooks

IBM zEnterprise System Technical Introduction

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 the IT industry has 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. 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, which ensures a tighter fit with IT requirements of an enterprise.

This IBM Redbooks publication introduces the revolutionary scalable IBM zEnterprise System, which consists of the IBM zEnterprise 196 (z196), the IBM zEnterprise 114 (z114), the IBM zEnterprise BladeCenter Extension (zBX), and the IBM zEnterprise Unified Resource Manager. IBM is taking a bold step by integrating heterogeneous platforms under the proven System z hardware management capabilities, while extending System z qualities of service to those platforms. The z196 and z114 are general-purpose systems that are equally at ease with compute-intensive workloads and with I/O-intensive workloads. The integration of heterogeneous platforms is based on IBM BladeCenter technology, allowing improvements in price and performance for key workloads and enabling a new range of heterogeneous platform solutions. The z196 and z114 are the core of the enhanced System z platform, which 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, z114, zBX, and Unified Resource Manager capabilities, hardware functions and features, and 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 familiar with current IBM System z technology and terminology.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-7832-01 0738435945