IBM System z10 Enterprise Class Technical Introduction

Describes new server hardware and capabilities

Discusses key functional elements and enhancements

Reviews software support

Marian Gasparovic
Per Fremstad
Parwez Hamid
Fernando Nogal
Karl-Erik Stenfors
Bill White

Redbooks

ibm.com/redbooks

International Technical Support Organization

IBM System z10 Enterprise Class Technical Introduction

May 2008

SG24-7515-00

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

First Edition (May 2008)

This edition applies to the IBM System z10 Enterprise Class server.

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:

Redbooks (logo) ®
AIX®
CICS®
DB2 Connect™
DB2®
Domino®
DRDA®
DS8000™
e-business on demand®
ESCON®
FICON®
HiperSockets™
IBM®
IMS™
Language Environment®
Lotus®
MQSeries®
MVS™
OMEGAMON®
OS/390®
Parallel Sysplex®
Passport Advantage®
POWER™
PR/SM™
Processor Resource/Systems Manager™
RACF®
Rational Rose®
Rational®
Redbooks®
Resource Link™
RMF™
S/360™
S/390®
Sysplex Timer®
System i™
System p™
System Storage™
System z™
System z10™
System z9®
Tivoli®
TotalStorage®
WebSphere®
z/Architecture®
z/OS®
z/VM®
z/VSE™
z10™
z10 EC™
z9™
zSeries®

The following terms are trademarks of other companies:

SAP, and SAP logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries.

Oracle, JD Edwards, PeopleSoft, Siebel, and TopLink are registered trademarks of Oracle Corporation and/or its affiliates.

InfiniBand, and the InfiniBand design marks are trademarks and/or service marks of the InfiniBand Trade Association.

IPX, J2EE, Java, JVM, OpenSolaris, Solaris, Sun, and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

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

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

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

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

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

Preface

This IBM® Redbooks® publication introduces the IBM System z10™ Enterprise Class server, which is based on z/Architecture®. It builds on the inherent strengths of the System z™ platform, delivering new technologies and virtualization that are designed to offer improvements in price and performance for key workloads, as well as enabling a new range of hybrid solutions. The z10 EC™ further extends System z's leadership in key capabilities with the delivery of expanded scalability for growth and large-scale consolidation, availability to help reduce risk and improve flexibility to respond to changing business requirements, and improved security. The z10 EC is at the core of the enhanced System z platform that is designed to deliver technologies that businesses need today along with a foundation to drive future business growth.

This document provides basic information about z10 EC capabilities, hardware functions and features, and its associated software support. It is intended for systems engineers, architects, consultants, and anyone else needing to understand the new elements of the z10 EC.

This book is intended as an introduction to the z10 EC mainframe. Readers are not expected to be generally familiar with current IBM System z technology and terminology.

The team that 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.

Marian Gasparovic is an IT Specialist working for IBM Server and Technology Group in IBM Slovakia. He worked as an Administrator for z/OS® at Business Partner for six years. He joined IBM in 2004 as storage specialist. Now he works as Field Technical Sales Support for System z in the CEMAAS region as a member of a new workload team. He also works part-time for the ITSO in Poughkeepsie.

Per Fremstad is an IBM Certified Senior IT Specialist from the IBM Systems and Technology Group in IBM Norway. He has worked for IBM since 1982 and has extensive experience with mainframes and z/OS. Lately, Per also has worked a lot with Linux® on System z and z/VM®. During the past 25 years he has worked in various roles within IBM and with a large number of customers. He frequently teaches about z/OS and z/Architecture subjects and has been actively teaching at Oslo University College for the last four years. Per holds a BSc from the University of Oslo, Norway.

Parwez Hamid is a Senior Certified Consulting IT Specialist working in the UK. During the past 33 years he has worked in various roles within IBM. Parwez has co-authored a number of IBM Redbooks publications and technical material for the world-wide announcement of System z servers. Parwez is also the technical lead for the IBM System z zChampions team, which is made up of IBM key System z customer-facing technical professionals from STG and SWG. He is a Technical Staff member of the UK Technical Council and a member of the zAtlas team. zAtlas is a small group of senior technical staff who work closely with System z product development and provide input and feedback for product plans up to six years out.

Fernando Nogal is an IBM Certified Consulting IT Specialist working as an STG Technical Consultant for the Spain, Portugal, Greece, Israel, and Turkey IMT. He specializes in infrastructures and architectures for on demand. In his 25 years with IBM he has held a variety of technical positions, mainly providing support for mainframe customers. Previously, he was on assignment to the Europe Middle East and Africa (EMEA) zSeries® Technical Support group, working full time on complex solutions for e-business on zSeries. His job included, and still does, presenting and consulting in architectures and infrastructures, providing strategic guidance to zSeries customers regarding the establishment and enablement of e-business technologies on zSeries, including the z/OS, z/VM, and Linux environments. He is a zChampion and a core member of the zSeries Business Leaders Council. Fernando was very happy to return to the ITSO Poughkeepsie, where he has previously been on assignment from 1991–1993, working on mainframe virtualization technologies. An accomplished writer, he has authored and co-authored ten Redbooks and several technical papers. Other activities include chairing a virtual team of IBMers interested in e-business on System z and serving as an University Ambassador, engaged with two Portuguese Universities where he has done several seminars. He travels extensively, on direct customer engagements and as a speaker at IBM and customer events, and trade shows.

Karl-Erik Stenfors is a Senior IT Specialist in the Product and Solutions Support Centre (PSSC) in Montpellier, France. He has more than 38 years of experience in the large systems field, 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, including z/VM, z/OS, and Linux. He teaches at numerous IBM user group and IBM internal conferences. He is currently working with the System z lab in Poughkeepsie, providing customer requirement input to create an IBM System vision for the future—the zAtlas and zChampions workgroups.

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

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

Connie Beuselinck, Ellen Carbarnes, William Clark, Darelle Gent, Michael Gerhart, Gary King, Jeff Kubala, Scott Langenthal, Kenneth Oakes, Patrick Rausch, Charles Webb, Frank Wisnewski
IBM Poughkeepsie

Les Geer, Reed Mullen, Brian Valentine
IBM Endicott

Harv Emery, Greg Hutchison
IBM Gaithersburg

Monika Zimmermann, Uwe Zumpe
IBM Boeblingen, Germany

Become a published author

Join us for a two- to six-week residency program! Help write a book dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You will have the opportunity to team with IBM technical professionals, Business Partners, and Clients.

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

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

  ibm.com/redbooks

- Send your comments in an e-mail to:

  redbooks@us.ibm.com

- Mail your comments to:

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

1

Introducing the System z10 Enterprise Class

The announcement of the System z10 Enterprise Class (z10 EC) is arguably the most exciting piece of news on the mainframe scene in recent years. Equipped with the new Enterprise quad-core processor, it is the fastest in the industry at 4.4 GHz, up to 1.5 TB of memory, and with new connectivity options that enhance its open characteristics. It delivers, in a single footprint, unprecedented performance and capacity growth while drawing upon the rich heritage of previous z/Architecture servers. The z10 EC is a well-balanced general purpose server that is equally at ease on compute-intensive workloads (such as business intelligence) as it is with I/O-intensive workloads (such as transaction and batch processing).

The z10 EC offers an extensive software portfolio that spans from support for IBM WebSphere® software, full support for SOA, Web services, J2EE™, Linux, and open standards, to the more traditional ones such as CICS® and IMS™. For instance, in the Linux on System z environment alone, more than 1,100 applications are offered by over 400 independent software vendors (ISVs).

IBM mainframes traditionally provide an advanced combination of reliability, availability, security, scalability, and virtualization. The z10 EC has been designed to extend these capabilities and is optimized for today's business needs. The z10 EC is intended to be a platform of choice for integration of a new generation of applications with existing applications and data.

The z10 EC meets the requirements of large enterprises having large-scale, mission-critical transaction and data processing requirements while also delivering the scalability and granularity to meet the needs of medium-sized enterprises.

The z10 EC server introduces just-in-time deployment of additional resources, known as Capacity on Demand (CoD). CoD provides flexibility, granularity, and responsiveness by allowing the user to dynamically change capacity when business requirements change. With the proper contracts, up to four temporary capacity offerings can be installed on the server (only one was possible with the System z9®). Additional capacity resources can be dynamically activated, either partially or in totality, using granular activation controls directly from the management console of the z10 EC, without the need to interact with IBM Support.

All this scalability and flexibility means that the z10 EC can adapt and grow in response to customers' changing business requirements while continuing to drive complexity and cost out of the business. At the same time, the z10 EC enhances the mainframe's leadership in security, virtualization, and resiliency, providing businesses with an extremely dependable platform.

The System z10 Enterprise Class uses approximately 15% less energy per unit of capacity than its predecessor, the z9 EC. The energy consumption per unit of capacity in System z servers has decreased by a factor of sixteen since 1995. Over the same time period, the floor-space occupancy per unit of capacity has decreased by a factor of seventy-seven in the same period. In addition, the z10 EC provides monitoring of energy consumption and tools for building energy consumption trends that can be utilized for in-depth analysis.

There is a fast-growing number of enterprises that are reaching the limits of available space and power at their data centers. Due to its server virtualization, consolidation capabilities, and green footprint, the System z10 Enterprise Class server is well suited to help solve these problems and avoid building a new and expensive data center.

Infrastructure simplification and server consolidation

The z 10 EC supports the highest levels of consolidation of the industry. Up to 60 logical partitions (LPARs) can be used. Each one can run any of the supported operating systems: z/OS, z/VM, z/VSE™, z/TPF, and Linux on System z. And, as experience demonstrates, these can be run at up to 100% sustained utilization levels, although most customers prefer to leave a bit of white space and run at 90% or slightly under.

The Processor Resource/Systems Management (PR/SM™) function, responsible for hardware virtualization of the server, is always active and has been enhanced to provide additional performance benefits. PR/SM technology has received Common Criteria EAL5[1] security certification. Each logical partition is as secure as an isolated server.

z/VM's virtualized z/Architecture servers, known as Virtual Machines, support all operating systems and other software supported on a logical partition. In fact, a z/VM Virtual Machine is the functional equivalent of a real server.

z/VM's extreme virtualization capabilities, which have been perfected since its start in 1967, make it possible to virtualize thousands of distributed servers on a single z10 EC server. IBM itself is conducting a very large consolidation project named Project Big Green, which aims to consolidate approximately 3,900 distributed servers into approximately 30 mainframes, using z/VM and Linux on System z. It expects to achieve reductions of over 80% in the use of space and energy. Similar results have been publicly presented by some customers, and these reductions directly translate to significant monetary savings.

Consider also the gains in software licensing, since the pricing model for many distributed software products is linked to the number of processors or processor cores. Consolidating under z/VM and exploiting the specialized Internal Facility for Linux (IFL) processors may achieve a large reduction in the number of used cores. For a discussion of available processor types see "PU characterization" on page 23.

In the same footprint, you can configure all the way from a 1-way to a 64-way Symmetrical Multiprocessor (SMP), unlike the z9 EC, which had a maximum configuration of a 54-way.

The first 12 processors can be configured at three different sub-capacity levels, giving a total of one hundred distinct capacity settings in the system, providing for a range of 1:140 in

[1] Evaluation Assurance Level with specific Target of Evaluation, Certificate for System z9 published September 4th 2006.

processing power. The z10 EC continues to offer all the specialty engines available with System z9. Most upgrades are concurrent to the hardware.

As will be described later, the z10 EC reaches new availability levels by eliminating some preplanning needs and other disruptive operations.

Summing up these characteristics leads to an interesting result, as shown in Example 1-1.

Example 1-1 Result of characteristics

```
   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
=  A very wide scope of applications that can be considered for server
   consolidation and application integration.
```

This presents a significant opportunity for most enterprises to simplify their IT infrastructures. The mainframe's inherent reliability, security, and availability, as well as its operational model, can now benefit other, up to now distributed, applications.

Further simplification is possible by exploiting the z10 EC HiperSockets™[2] and z/VM's Virtual Switch functions. These may be used, at no additional cost, to replace physical routers and switches and, consequently, cables and security exposures, while simplifying configuration and administration tasks.

Other uses of virtualization include:

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

A z/VM production environment can achieve additional savings by:

- Allowing backup virtual servers to be dormant and use no resources until and if they are required. This may help reduce hardware, software, and maintenance costs.
- Pooling resources such as processor, I/O facilities, and disk space. Virtual servers are provisioned out of these pools, and when their useful life ends the resources are returned to the pools and recycled.
- Offering very fast virtual server provisioning. A complete server can be readied for use in just a few minutes, using resources from the pool and image cloning.
- Eliminating the need to re-certify servers for specific purposes. Environments are certified to the virtual server. This needs to be done only once, even if the server requires scaling up, because the underlying hardware and architecture does not change. Significant reductions and time and manpower can be achieved.

Enterprise hub

Many customers 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 customers are making *on demand* business a

[2] For a description of HiperSockets see "HiperSockets" on page 11. z/VM Virtual Switch is a z/VM system function that uses memory to emulate switching hardware.

reality. Yet other customers consider the mainframe based on the combined price/performance improvements of software and hardware.

As these customers continue to grow their traditional and new applications and leverage more of the z10 EC characteristics, they will more fully utilize the richness of current and future operating system enabled functions. Also, as enhancements are delivered via business continuity out of the Tivoli® product set, the z10 EC continues to grow its role as the resiliency hub for the enterprise, including disaster recovery. Furthermore, recent advances in IBM DS8000™ disk technology give customers the opportunity to take advantage of its increased function and value, especially in the area of secure data encryption.

The z10 EC continues the evolution of the mainframe, building upon the z/Architecture definitions. The z10 EC extends and integrates key platform characteristics: dynamic and flexible partitioning, resource management of mixed and unpredictable workload environments, availability, scalability, clustering, and security and systems management with emerging e-business on demand® application technologies, such as WebSphere, Java™, and Linux.

All these technologies and improvements come into play when the z10 EC is at the heart of the service-oriented architecture (SOA) solutions for an enterprise. There is no conflict between application modular design and integrated application deployment. In fact, the availability, as a deployment choice, of an environment with shared resources, rich operational characteristics, and time tested capabilities is highly desirable. Also, the extreme virtualization capabilities of the System z10 Enterprise Class make it possible to create dense and simplified infrastructures that are highly secure and can lower operational costs.

For a more in-depth discussion of what System z and the z/Architecture can provide, see *IBM System z Strengths and Values*, SG24-7333.

1.1 z10 EC comparison

The System z10 Enterprise Class is a follow-on to the System z9 Enterprise Class (z9 EC), which was announced in July of 2005. It employs leading-edge chip Silicon on Insulator (CMOS 11S-SOI) and other technologies, such as InfiniBand®, to provide advantages such as very high frequency chips, additional granularity options, improved availability, and enhanced on demand options. In addition, it supports the latest offerings for data encryption.

Five models of the z10 EC are offered. These are named E12, E26, E40, E56, and E64. The names represent the maximum number of configurable processors in the model.

The z10 EC system architecture ensures continuity and upgradability from the z9 EC design. Upgrading from IBM zSeries Model 990 servers is also possible.

Figure 1-1 provides a comparison of System z10 Enterprise Class with previous System z servers along four major attributes: single engine processing capacity, number of engines, memory, and I/O bandwidth.

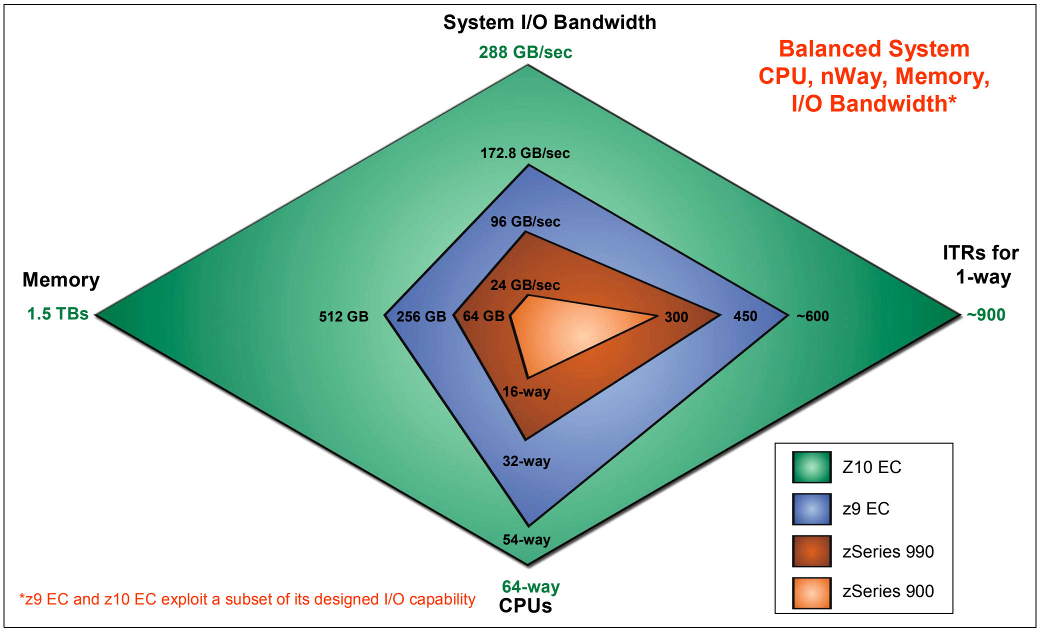

Figure 1-1 System z design comparison

1.2 z10 EC server enclosure

In this section we briefly review the most significant characteristics of the System z10 Enterprise Class. Further details can be found in Chapter 2, “Hardware overview” on page 15.

The z10 EC server is a twin-frame system. It has machine type designation of 2097. Figure 1-2 shows an external view of the server. The frames are known as the A frame and the Z frame.

Figure 1-2 System z10 Enterprise Class external view

The A frame contains:

- The Central Electronic Complex (CEC)
- Modular cooling units
- One I/O cage
- Power supplies
- An optional internal battery feature (IBF)

The Z frame contains:

- Two system support elements (SEs)
- Zero, one, or two additional I/O cages
- Power supplies
- An optional IBF

The two redundant SEs are used to configure and manage the z10 EC server (for example, defining the I/O configuration and configuring the logical partitions).

1.2.1 Central Electronic Complex

The CEC is housed in its own cage. The cage houses from one to four processor books that are fully interconnected. Each book contains a Multi-Chip-Module (MCM), memory and I/O cage connectors, and (optionally) coupling link connectors.

The z10 EC is built on the proven superscalar microprocessor architecture already deployed on the z9 EC. However, the new processor unit (PU) chip has several distinctive innovations, notably in error checking and correcting, and new specialized circuitry, for instance, to support decimal floating point operations. And, of course, it has a 4.4 GHz high-speed quad-core design.

There is one MCM that has five PU chips and two storage control (SC) chips in each book. Each PU chip has either three or four cores. The MCM continues to be cooled by modular refrigeration units (MRUs) with air cooling backup.

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

Memory has been increased, as compared with the z9 EC. In each book up to 384 GB can be installed, but because 16 GB are part of the base and reserved for the hardware system area (HSA), the maximum amount of customer purchasable memory is 1520 GB, just shy of 1.5 TB.

PU characterization

At server 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 may be characterized as follows:

CP Central processor. The standard z10 EC processors. For use with any supported operating system and user applications.

ICF Internal coupling facility. Used for z/OS clustering. ICFs are dedicated for this purpose and exclusively run Coupling Facility Control Code (CFCC).

IFL Integrated Facility for Linux. Exploited by Linux and for z/VM processing in support of Linux. z/VM is often used to host multiple Linux virtual machines (called guests). It is not possible to IPL operating systems other than z/VM or Linux on an IFL.

SAP® System assist processor. Offloads and manages I/O operations. Several are standard with the z10 EC. More may be configured if additional I/O processing capacity is needed.

zAAP[3] System z10 application assist processor. Exploited under z/OS for designated workloads, which include the IBM JVM™ and some XML System Services functions.

zIIP[3] System z10 integrated information processor. Exploited under z/OS for designated workloads, which include some XML System Services, IPSec off-load, part of DB2® DRDA®, star schema, and utilities.

Important: Work dispatched on zAAP and zIIP processors does not incur any IBM software charges.

CP Assist for Cryptographic Function (CPACF)

The z10 EC continues to use the Cryptographic Assist Architecture first implemented on z990. Further enhancements have been made to the z10 EC CPACF.

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

- Data Encryption Standard (DES)
- Triple Data Encryption Standard (TDES)
- Secure Hash Algorithm (SHA-1) 160 bit
- SHA-2 256, 384 and 512 bit

[3] z/VM 5.3 support zIIP and zAAP processors for z/OS guest workloads

- Advanced Encryption Standard (AES) for 128-bit, 192-bit, and 256-bit keys
- Pseudo Random Number Generation (PRNG)

  Note that PRNG is also a standard function supported on the Crypto Express2 feature.

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

1.2.2 I/O subsystem

Like its predecessors, the z10 EC server has a dedicated subsystem to manage all input/output operations. Known as the channel subsystem, it is composed of:

SAP System assist processor. A specialized processor that uses the installed PU cores[4]. Its role is to offload I/O operations, managing channels and the I/O operations queues. It relieves the other PUs of all I/O tasks, allowing them to be dedicated to application logic. An adequate number of SAP processors is automatically defined, depending on the number of installed books. These are part of the base configuration of the server.

HSA Hardware system area. A reserved part of the system memory, it contains the I/O configuration and is used by SAPs. On the z10 EC a fixed amount of 16 GB is reserved, which is not part of the customer-purchased memory. This provides for greater configuration flexibility and higher availability, by eliminating some planned and pre-planned disruptive situations.

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

Channel path A channel path is the means that the channel subsystem uses to communicate with the I/O devices. Due to I/O virtualization, multiple independent channel paths can be established on a single channel, allowing the channel to be shared[5] between multiple logical partitions and each partition to have its unique channel path.

Subchannels A subchannel appears to a program as a logical device, and contains the information required to perform an I/O operation. There is one subchannel for each I/O device addressable by the channel subsystem.

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

The z10 EC I/O subsystem direction is evolutionary, drawing on development from the z990 and z9 EC. It continues to be based on I/O cages, I/O cards, and I/O buses. The I/O subsystem is supported by a new I/O bus, and includes the InfiniBand infrastructure (replacing self-timed interconnect found in the prior System z servers). This new infrastructure is designed to reduce overhead and latency, and provides increased data throughput.

[4] Each z10 EC PU can be characterized as one of six different configurations. For more information see "PU characterization" on page 23.

[5] The function that allows sharing I/O paths across logical partitions is known as Multiple Image Facility (MIF).

InfiniBand

InfiniBand is an input/output (I/O) industry-standard specification that results from the merging, in 1999, of two different and competing I/O standards: Future I/O (proposed by Compaq, IBM, and Hewlett-Packard) and Next Generation I/O (proposed by Intel®, Microsoft®, and Sun™).

The InfiniBand specification includes an I/O architecture for interconnecting communications infrastructure equipment and IT equipment including storage, servers, and embedded systems. The architecture employs point-to-point channels in a switched connection, thus constituting a fabric. Connections can be either external, using both optical fiber and copper cabling, or internal, as in chassis backplanes.

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

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

InfiniBand is being exploited by the z10 EC server. Internally, in the server, the new cables from the CEC cage to the I/O cages now carry the InfiniBand protocol. For external usage, Parallel Sysplex® InfiniBand (PSIFB) links are introduced to keep up with server speed. They are used to interconnect System z servers in a Parallel Sysplex.

1.2.3 I/O connectivity

The z10 EC generation of the I/O platform, particularly through the exploitation of InfiniBand, is intended to provide significant performance improvements over the previous I/O platform used for FICON® Express4, OSA-Express2, OSA-Express3, and Crypto Express2.

I/O cage

The z10 EC has a CEC cage and, as a minimum, one I/O cage in the A frame. The Z frame can accommodate two additional I/O cages, bringing to three the total for the system. I/O cages can accommodate the following feature types:

- ESCON®
- FICON Express4, FICON Express2, or FICON Express
- OSA-Express3 and OSA-Express2
- Crypto Express2
- Coupling links

It is possible to populate the 28 I/O slots in one I/O cage with any mix of the above-mentioned cards.

ESCON channels

The Enterprise Systems Connection (ESCON) channels support connectivity to ESCON disks, tapes, and printer devices. Historically, they represent the first use of optical I/O technology on the mainframe. They are much slower than FICON channels, which are becoming the preferred technology. There are no changes in the ESCON support as compared to z9 EC.

FICON channels

Fiber Connection (FICON) channels follow the Fibre Channel (FC) standard and match data storage and access requirements as well as the latest FC technology in storage and access devices. FICON channels support the Native FICON Protocol and Fibre Channel Protocol (FCP):

- Native FICON Protocol is used to support FICON Channel-to-Channel connectivity and traffic supporting connectivity to FICON devices such as disks, tapes, and printers. These are supported by the z/OS, z/VM, z/VSE, z/TPF, and Linux on System z environments.
- Fibre Channel Protocol traffic supports connectivity to open connectivity disks and tapes through Fibre Channel switches and directors in z/VM, z/VSE, and Linux on System z environments. FCP mode is available both in long wavelength (LX) and short wavelength (SX) operation. The FCP channel can connect to FCP SAN fabrics and access FCP/SCSI devices.

It is possible to choose any combination of the supported FICON Express4, FICON Express2, and FICON Express. Depending on the feature, auto- negotiated link date rates of 1, 2, or 4 Gbps are supported. FICON Express4 provides significant improvements in start I/Os and data troughput.

Open Systems Adapter (OSA) channels

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

The z10 EC can have up to 24 features of the Open Systems Adapter family of local area network adapters, giving a maximum of 48 ports of LAN connectivity. It is possible to choose any combination of the supported OSA-Express2 and OSA-Express3 Ethernet features.

Several protocols, including fast Ethernet, Gigabit Ethernet, and 10 Gigabit Ethernet, are supported depending on the feature. There is a new OSA-Express3 feature that supports 10 Gigabit Ethernet.

Crypto Express2

The Crypto Express2 feature provides for tamperproof, high-performance cryptographic operations. Each feature has two PCI-X adapters. Each of the PCI-X adapters can be configured as either a coprocessor or an accelerator.

- Crypto Express2 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 Express2 Accelerator - for Secure Sockets Layer (SSL) acceleration
  - Designed to support clear key RSA operations
  - Offloads compute-intensive RSA public-key and private-key cryptographic operations employed in the SSL protocol

The configurable Crypto Express2 feature is supported by z/OS, z/VM, z/VSE, and Linux on System z.

Coupling links

Coupling links are used in Parallel Sysplex cluster configurations of System z servers. The links provide high-speed bidirectional communication between members of the sysplex. The z10 EC offers new links that exploit the InfiniBand technology.

HiperSockets

The HiperSockets function is an integrated function of the z10 EC that provides users with attachments to up to sixteen high-speed *virtual* local area networks with minimal system and network overhead.

HiperSockets is a function of the virtualization Licensed Internal Code (LIC) and executes memory-to-memory data transfers in a totally secure way. HiperSockets eliminates the need to utilize I/O subsystem operations and the need to traverse an external network connection to communicate between logical partitions in the same z10 EC server. Therefore, HiperSockets offers significant value in server consolidation connecting many virtual servers.

1.3 Performance

The z10 EC Model E64 is designed to offer approximately 1.7 times more capacity than the z9 EC Model S54 system. Uniprocessor performance has also increased significantly. A z10 EC Model 701 offers performance improvements of up to 1.62 times the z9 EC Model 701. However, variations on the observed performance increase are dependent upon the workload type.

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

http://www.ibm.com/servers/eserver/zseries/lspr/

The MSU ratings are available on the Web:

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

LSPR workload suite

The LSPR workload suite comprises the following workloads:

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

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

The LSPR provides performance ratios for individual workloads as well as for the default mixed workload, which is composed of equal amounts of the five workloads described above. The z10 EC LSPR continues to rate all z/Architecture processors running in LPAR mode and 64-bit mode. The single-number-metrics 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-metrics, the z10 EC LSPR contains information pertaining to Linux and z/VM environments.

Capacity ratio estimates

The graph in Figure 1-3 shows the estimated capacity ratios for z10 EC and z9 EC. The capacity estimate is based on the LSPR workload suite described above.

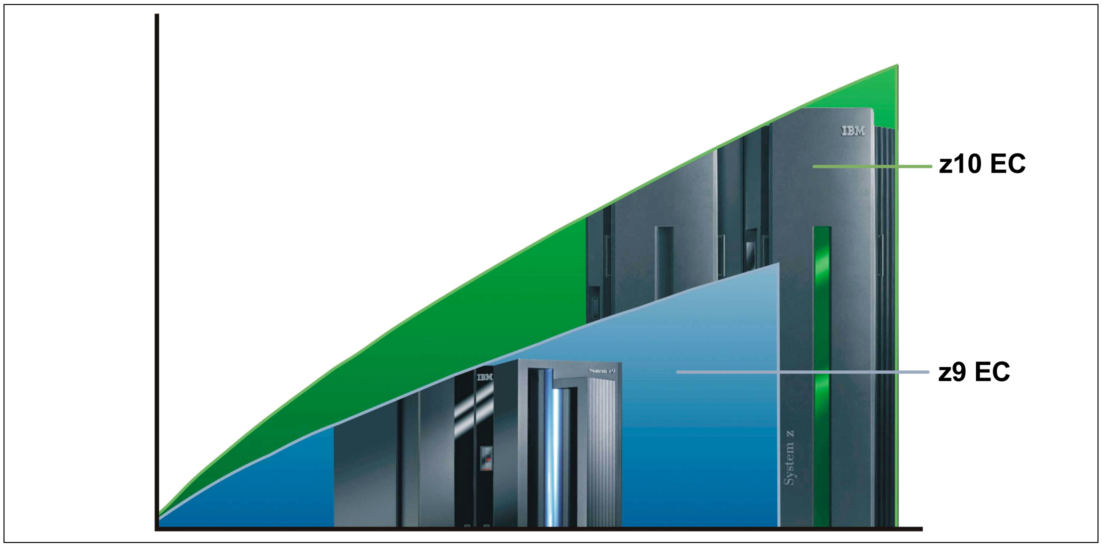

Figure 1-3 z10 EC to z9 EC performance comparison

The LSPR contains the internal throughput rate ratios (ITRRs) for the z10 EC and the previous generations of processors based upon measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user may experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, and the workload processed.

Workload performance variation

Due to the nature of the z10 EC multi-book system and resource management across those books, we expect there to be high performance variability similar to that seen on the z990 and z9 EC. This variability may be observed in several ways. The range of performance ratings across the individual workloads is likely to have a large spread. The customer impact of this increased variability will be seen as increased deviations of workloads from single-number metric-based factors such as MIPS, MSUs, and CPU time chargeback algorithms.

HiperDispatch

HiperDispatch is a new function exclusive to the z10 EC server. This function greatly enhances the processor affinity support to further improve the usable capacity of the server. In addition to the PR/SM enhancements, it requires support by z/OS. Now PR/SM and z/OS cooperate on affinity assignments, which are done at the task level rather than just at the partition level—thus a great step forward.

1.4 Capacity On Demand

New with the System z10 Enterprise Class is the possibility of doing just-in-time deployment of capacity resources. This function is designed to provide more flexibility to dynamically change capacity when business requirements change. No customer interaction is required with IBM at the time of activation. Customers now:

- Define one or more flexible configurations that can be used to solve multiple temporary situations. Previously, only one configuration was possible.
- Have multiple configurations active at once, and the configurations themselves have flexible selective activation of only what is needed.
- Add permanent capacity to the server when temporary changes are active.

1.5 Software

The z10 EC is supported by a large set of software, including ISV applications. This section only lists the supported operating systems. Further information is contained in Chapter 4, "Software support" on page 65.

Operating system support of the System z10 EC requires one of:

- z/OS Version 1 Release 7 or later. Note that z/OS.e is not supported. Also, exploitation of some features may require the latest releases.
- z/VM Version 5 Release 2 or later
- z/VSE Version 3 Release1 or later
- TPF Version 4 Release 1 and z/TPF Version 1 Release 1
- Linux on System z distributions:
  - Novell SUSE SLES 9 or SLES 10
  - Red Hat RHEL 4 or RHEL 5

2

Hardware overview

The System z10 Enterprise Class server 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 System z10 Enterprise Class server provided in Chapter 1, “Introducing the System z10 Enterprise Class” on page 1, and compares them with the System z9 EC server where relevant.

The following topics are discussed:

2.1 z10 EC highlights

The major System z10 Enterprise Class improvements over its predecessors include:

- Improved total system capacity in a 64-way server and additional subcapacity settings, offering increased levels of performance and scalability to help enable new business growth
- Quad-core 4.4 GHz processor chips that can help improve the execution of CPU-intensive workloads
- Up to 1.5 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
- A new 16 GB fixed hardware system area (HSA) that is managed separately from customer-purchased memory
- Exploitation of InfiniBand technology
- Improvements to the I/O subsystem and new features
- Additional security options for the CP Assist for Cryptographic Function (CPACF)
- A new HiperDispatch function for improved efficiencies in hardware and z/OS software
- Hardware decimal floating point on each core on the processor unit (PU)

In all, these enhancements provide customers with options for continued growth, continuity, and upgradability.

For an in-depth discussion of the IBM System z10 Enterprise Class functions and features see the *IBM System z10 Enterprise Class Technical Guide,* SG24-7516.

2.2 Models and model upgrades

The System z10 Enterprise Class has been assigned a machine type (M/T) of 2097, which uniquely identifies the server. The server is offered in five different models. These models are named E12, E26, E40, E56, and E64. The model determines the maximum number of processor units (PUs) available. PUs are delivered in single-engine increments. The first four models utilize a 17 PU Multi-Chip Module (MCM), of which 12 to 14 PUs are available for characterization. The fifth model, E64, utilizes one 17-PU MCMs and three 20-PU MCMs to provide up to 64 configurable PUs.

As with the System z9 EC, spare PUs and system assist processors (SAPs) are integral to the server. Refer to Table 2-1 on page 17 for a model summary including SAPs and spare PUs for the different models. For an explanation of PU characterization see “PU characterization” on page 23.

The five z10 EC server orderable models are:

- The z10 EC Model E12 has one book with 17 PUs, of which 12 can be characterized. The five remaining PUs are three SAPs and two spares.
- The z10 EC Model E26 has two books with 17 PUs in each book for a total of 34 PUs, of which 26 can be characterized. The eight remaining PUs are six SAPs, three in each book, and two spares, one in each book.

- The z10 EC Model E40 has three books with 17 PUs in each book for a total of 51 PUs, of which 40 can be characterized. The eleven remaining PUs are nine SAPs, three in each book, and two spares, one in book 0 and one in book 1.
- The z10 EC Model E56 has four books with 17 PUs in each book for a total of 68 PUs, of which 56 can be characterized. The 12 remaining PUs are 10 SAPs, three each in books 0, 1, and 2, and one in book 3, and two spares in book 3.
- The z10 EC Model E64 has four books with 17 PUs in book 0 and 20 PUs in books 1, 2, and 3 for a total of 77 PUs, of which 64 can be characterized. The 13 remaining PUs are eleven SAPs, one in book 0, three each in books 1 and 2, and four in book 3, and two spares, one in book 0 and one in book 1.

Table 2-1 Model summary

Model	Books/PUs	CPs	Std SAPs	Spares
E12	1/17	0–12	3	2
E26	2/34	0–26	6	2
E40	3/51	0–40	9	2
E56	4/68	0–56	10	2
E64	4/77	0–64	11	2

The z10 EC offers one hundred different capacity levels, which span a range of approximately one to 140. This is discussed in 3.2.2, “Granular capacity and capacity settings” on page 46.

Figure 2-1 summarizes the upgrade paths to z10 EC.

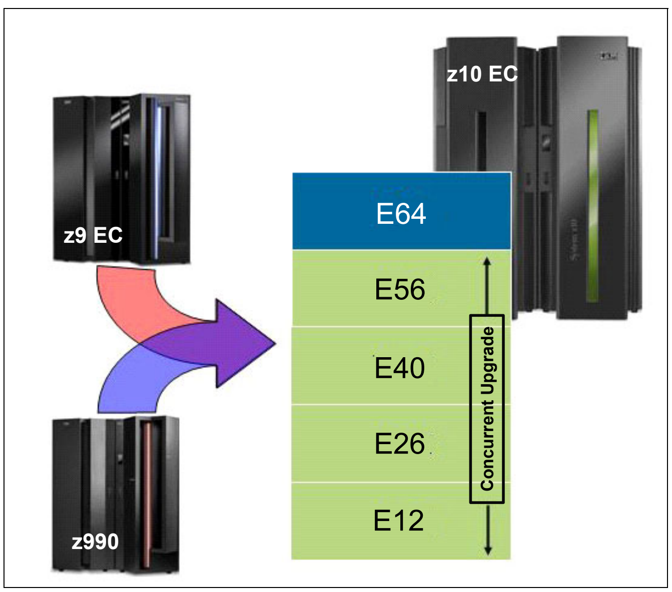

Figure 2-1 System z10 Enterprise Class upgrade paths

z10 EC upgrades

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

All z10 EC to z10 EC model upgrades are concurrent except when the target is the model E64. This is a non-concurrent upgrade because the model E64 uses a different set of MCMs.

Table 2-2 z10 EC to z10 EC upgrade paths

	E12	**E26**	**E40**	**E56**	**E64**
E12	-	Yes	Yes	Yes	Yes
E26	-	-	Yes	Yes	Yes
E40	-	-	-	Yes	Yes
E56	-	-	-	-	Yes

Upgrades to z10 EC from z9 EC

Upgrades are also available from the currently installed z9 EC servers. The five model numbers for the z9 EC servers are S08, S18, S28, S38, and S54. These upgrades are disruptive. See Table 2-3 for the upgrade paths from z9 EC to z10 EC.

Table 2-3 Upgrades from z9 EC to z10 EC

	E12	**E26**	**E40**	**E56**	**E64**
S08	Yes	Yes	Yes	Yes	Yes
S18	Yes	Yes	Yes	Yes	Yes
S28	Yes	Yes	Yes	Yes	Yes
S38	Yes	Yes	Yes	Yes	Yes
S54	Yes	Yes	Yes	Yes	Yes

Upgrades to z10 EC from z990

Similarly, the upgrades from currently installed z990s to z10 EC servers are also offered for all the models of the z990. The four model numbers for the z990 servers are A08, B16, C24, and D32. These upgrades are disruptive. See Table 2-4 for the upgrade paths from z990 to z10 EC.

Table 2-4 Upgrades from z990 to z10 EC

	E12	**E26**	**E40**	**E56**	**E64**
A08	Yes	Yes	Yes	Yes	Yes
B16	Yes	Yes	Yes	Yes	Yes
C24	Yes	Yes	Yes	Yes	Yes
D32	Yes	Yes	Yes	Yes	Yes

The System z10 Enterprise Class server is always a two-frame system. The frames are called the A frame and the Z frame. Refer to Figure 2-2 for an internal front view of the two frames. Some of the hardware elements pointed out are described later in this chapter.

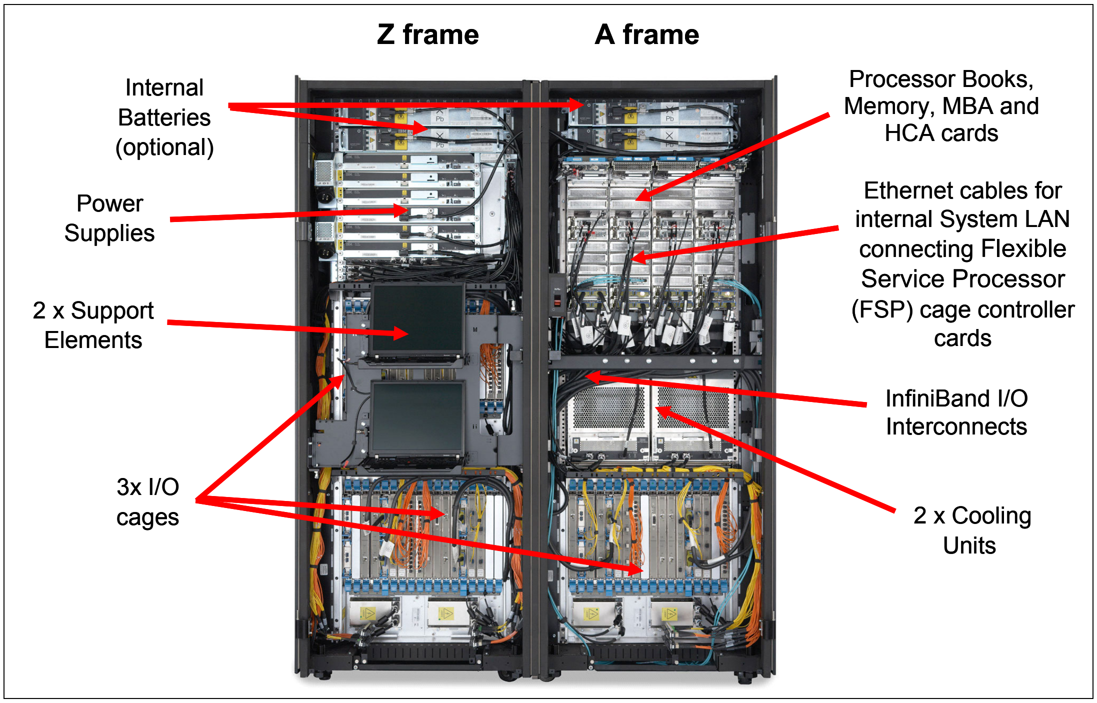

Figure 2-2 z10 EC internal front view

The z10 EC is slightly bigger than the z9 EC. Refer to Table 2-5 for the physical dimensions of the system and its frames.

Table 2-5 z10 EC physical dimensions

Frame cover	Width mm (in)	Depth mm (in)	Height mm (in)
System with covers	1565 (61.60)	1854 (71.0)	2013.2 (79.26)
System with covers and reduction	1565 (61.60)	1854 (71.0)	1785.0 (70.30)
Each frame with one side cover and without packaging	780 (30.75)	1270 (50.0)	2013.2 (79.26)

2.3 CEC cage and books

The z10 EC server has a multi-book system structure similar to the z9 EC server. A book looks like a box and plugs into one of four slots in the central electronic complex (CEC) cage of the z10 EC server. The CEC cage is located in the in A frame of z10 EC server. Refer to Figure 2-2 for a pictorial view of the CEC cage and the location of the four books.

Each book contains:

- A multi-chip module (MCM). Each MCM includes five quad-core processor Unit (PU) chips and two storage control (SC) chips. MCMs are further described in 2.4, “MCM” on page 22. Refer to Table 2-1 on page 17 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 384 GB of physical memory.
- A combination of up to eight InfiniBand Host Channel Adapter (HCA2-Optical or HCA2-Copper) fanout cards and memory bus adapter (MBA) fanout cards. Each of the cards has two ports, thereby supporting up to 16 connections. HCA2-Copper connections are for links to the I/O cages in the server, and the HCA2-Optical and MBA connections are to external servers (coupling links). MBA cards are used for ICB-4 links.
- Three distributed converter assemblies (DCAs) that provide power to the book. Loss of a DCA leaves enough book power to satisfy the book's power requirements. The DCAs can be concurrently maintained.

Note: IBM has issued the following Statement of General Direction:

> *ICB-4 links to be phased out:* IBM intends to not offer Integrated Cluster Bus-4 (ICB-4) links on future servers. The System z10 Enterprise Class is the last server to support ICB-4 links.

Figure 2-3 shows a view of a z10 EC book without the containing box.

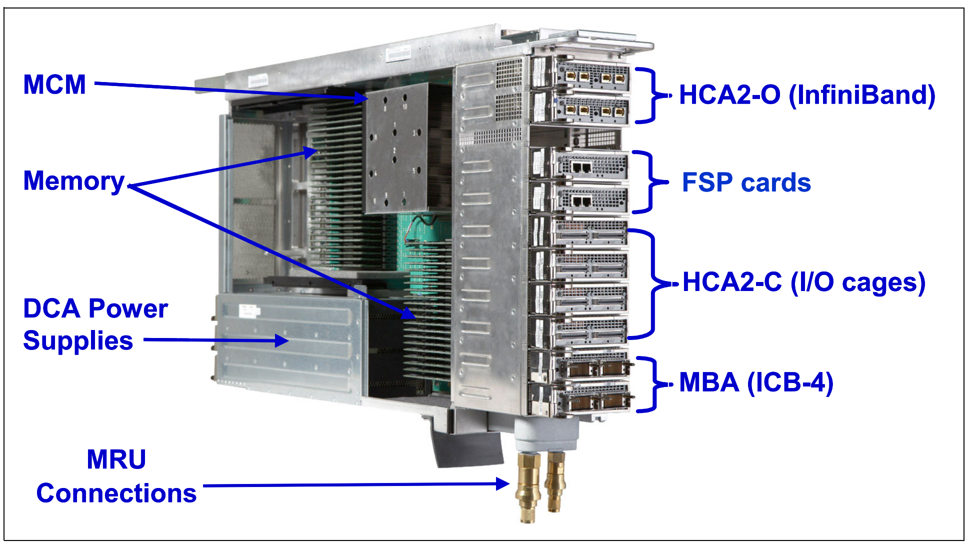

Figure 2-3 z10 EC book structure and components

The z10 EC provides a significant increase in system scalability and opportunity for server consolidation by providing a *multi-book* system structure. As shown in Figure 2-4, all books are interconnected in a star configuration with high-speed communications links via the L2 caches, which allows the system to be operated and controlled by the PR/SM facility as a symmetrical, memory-coherent multiprocessor.

The ring topology employed on the z9 EC sever is not used on the z10 EC. The point-to-point connection topology allows every book to communicate with every other book. It was designed to get the maximum benefit of the improved processor clock speed.

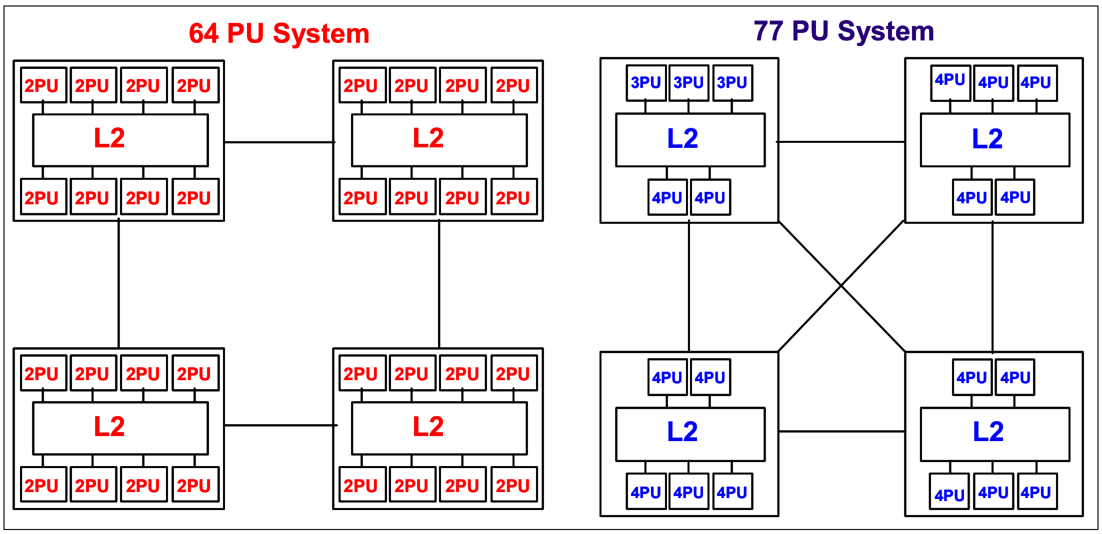

Figure 2-4 z9 EC versus z10 EC inter-book communication structure

Table 2-6 compares the book characteristics of the z9 EC to the z10 EC.

Table 2-6 z9 EC and z10 EC inter-books

	z9 EC	**z10 EC**
SMP configuration	4 books, 64 PUs	4 books, 77 PUs
Topology	Dual ring	Fully connected
Jumper books	Yes	No
Max memory	512 GB	1.5 TB
Cache levels	L1, L2	L1, L1.5, L2

2.4 MCM

The multi-chip module is a high-performance, glass-ceramic module, providing the highest level of processing integration in the industry. It is the heart of the server (Figure 2-5).

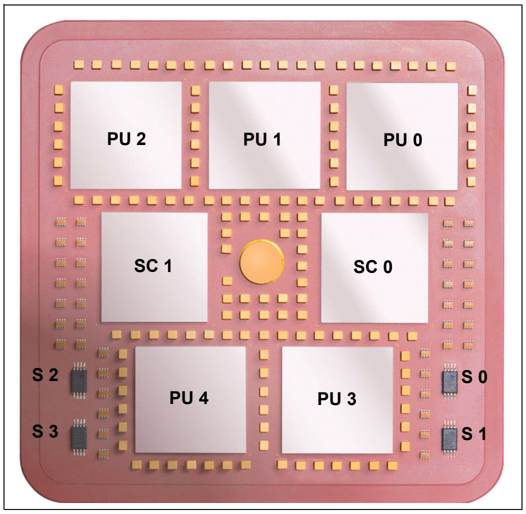

Figure 2-5 z10 EC multi-chip module

The z10 EC MCM has seven chip sites. All chip types on the MCM use CMOS[1] 11S chip technology. CMOS 11s is state-of-the-art microprocessor technology based on ten-layer Copper Interconnections and Silicon-On Insulator technologies. The chip lithography line width is 0.065 micron (65 nm). The chip contains close to 1 billion transistors in a 450 mm2 die.

There is one MCM per book and the MCM contains all of the processor chips and L2 cache of the book. The z10 EC server has five PU chips per MCM and each PU chip has up to four PUs (cores). Two MCM options are offered: with 17 or 20 processor units. All the models employ an MCM size of 17 PUs except for the model E64, which has one book with a 17 PU MCM and three books with 20 PU MCMs, for a total of 77 PUs.

The MCM also has two storage control (SC) chips. Each SC chip packs 24 MB of SRAM cache, interface logic for 20 cores, and SMP fabric logic into 450 mm2. The two SC chips are configured to provide a single 48 MB cache shared by all 17 or 20 cores on the module, yielding outstanding SMP scalability on real-world transaction processing workloads.

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

2.5 z10 EC processor chip

The z10 EC processor engine features an all-new high-frequency (4.4 GHz) four-core processor chip, a new microprocessor design, a robust cache hierarchy, and an SMP design optimized for enterprise database and transaction processing workloads, as well as for emerging workloads such as Java and Linux.

[1] CMOS - Complementary Metal Oxide of Silicon.

It leverages leading-edge technology and circuit design techniques while building on the rich heritage of mainframe system design, including industry-leading reliability, availability, and serviceability. New functional features enable increased software efficiency and scalability while maintaining full compatibility with existing software. Further detail is given in 3.2.1, “Microprocessor enhancements” on page 45.

2.6 Processor unit

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

- The basic cycle time is approximately 230 picoseconds.
- Up to two instructions may be decoded per cycle.
- Up to three instructions may be executed (finished) per cycle.
- Instructions are completed in the order in which they appeared in the instructions stream. A high-frequency, low-latency, mostly in-order pipeline, providing robust performance across a wide range of workloads, is used.
- Memory accesses might not be in the same instruction order (out-of-order operand fetching).
- Most instructions flow through a pipeline with different numbers of steps for various types of instructions. Several instructions may be in progress at any moment, subject to the maximum number of decodes and completions per cycle.

Each PU has an L1 cache divided into a 64 KB cache for instructions and a 128 KB cache for data. Each PU also has a L1.5 cache. This cache is 3 MB in size. This implementation optimizes performance of the system for high-frequency, very fast processors.

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

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

The z10 EC 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 50 new hardware instructions support more efficient code generation, particularly for Java and C++ programs.

Decimal floating point hardware fully implements the new IEEE 754r standard, helping financial institutions provide better performance and higher precision for decimal calculations.

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* processors into various types during initialization of the processor—often called a Power-On-Reset (POR) operation. It is also possible to characterize PUs dynamically without a POR. A processor that is not characterized cannot be used. Available PU types are described in “PU characterization” on page 7.

At least one CP needs to be purchased with, or before, a zAAP or zIIP can be purchased. Customers can purchase one zAAP, one zIIP, or both, for each CP (assigned or unassigned) on the server. However, a logical partition definition can contain more zAAPs or zIIPs than CPs. For example, in a server with two CPs a maximum of two zAAPs and two zIIPs can be installed. A logical partition definition for that server could contain up to two logical CPs, two logical zAAPs, and two logical zIIPs.

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

> **Notes:** Addition of ICFs, IFLs, zAAP, zIIPs, and SAPs to a server does not change the server capacity setting or its MSU rating (only CPs do).
>
> IBM does not impose any software charges on work dispatched on zAAP and zIIP processors.

2.7 Memory

Maximum physical memory sizes are directly related to the number of books in the system. Each book may contain a maximum of 384 GB of physical memory. Up to 1520 GB (~1.5 TB) of physical memory can be purchased. This is equal to 4 books x 384 GB minus 16 GB reserved for the hardware system area.

Memory can be purchased in increments of 16 GB up to a total size of 256 GB. From 256 GB the increment size doubles to 32 GB until 512 GB. From 512 GB to 944 GB the increment is 48 GB, and beyond that, up to 1520 GB, a 64 GB increment is used.

- A one-book system (z10 EC Model E12) may contain from 64 GB to 384 GB of physical memory. Memory is orderable in 16 GB and 32 GB increments for customer use up to 352 GB.
- A two-book system (z10 EC Model E26) may contain from 96 GB to 768 GB of physical memory. Memory is orderable in 16 GB, 32 GB, and 48 GB increments for customer use up to 752 GB.
- A three-book system (z10 EC Model E40) may contain from 128 GB up to a maximum of 1152 GB of physical memory. Memory is orderable in 16 GB, 32 GB, 48 GB, and 64 GB increments for customer use to up to 1136 GB.
- A four-book system (z10 EC Model E56, or z10 EC S64) may contain from 160 GB up to a maximum of 1536 GB of physical memory. Memory is orderable in 16 GB, 32 GB, 48 GB, and 64 GB increments for customer use to up to 1520 GB.

The reason that the maximum amount of orderable memory in each of the models is not equal to the maximum supported amount of physical memory is due to the fact that 16 GB of physical memory is set aside for the HSA, and after certain amounts of memory are exceeded the memory increment size changes.

Physically memory is organized as follows:

- Except on the z10 EC model E12, where a minimum of 64 GB is installed, a book always contains a minimum of 32 GB.
- A book may have more memory installed than enabled. The excess amount of memory can be installed 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 have to 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 z10 EC is still a Symmetric Multi-Processor (SMP).

Memory upgrade is concurrent when it requires no change of the physical memory cards. A memory card change is disruptive when no use is made of Enhanced Book Availability. Refer to *IBM System z10 Enterprise Class Technical Guide,* SG24-7516, 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 32 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.

Hardware system area

The Hardware System Area 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. Thus, model upgrades and also new logical partition definitions or changes required pre-planning and were sometimes disruptive, due to changes in HSA size. For further information and benefits see “Hardware system area” on page 48.

2.8 I/O Interfaces

Refer to Figure 2-6 on page 26 for the I/O system structure overview for the z9 EC server and for the z10 EC server.

The z10 EC has fanout cards residing on the front of the book package. There are three kinds of fanout cards:

- An InfiniBand HCA2-C (copper) fanout supporting ESCON, FICON, OSA, ISC-3, and Crypto Express2 cards in the I/O cages
- An InfiniBand HCA2-O (optical) supporting up to 6 GBps z10 EC to z10 EC and up to 3 GBps z10 EC to System z9 Parallel Sysplex connections
- A MBA fanout card used for ICB-4 connections only

The z10 EC supports up to eight I/O hubs (HCA-C, HCA-O, or the new MBA) for each book, with a maximum of 24 for a four-book system. Each hub comes with two ports, giving a maximum of 48 ports for I/O connectivity.

The z10 EC exploits InfiniBand (IFB) connections to I/O cages driven from the Host Channel Adapter (HCA2-C) fanout cards that are located on the front of the book. The HCA2-C fanout card is designated to connect to an I/O cage by a copper cable. The two ports on the fanout card are dedicated to I/O. This is different from the z9 EC, which uses self -interface (STI) connections driven from the Memory Bus Adapters (MBAs) to connect to the I/O cages.

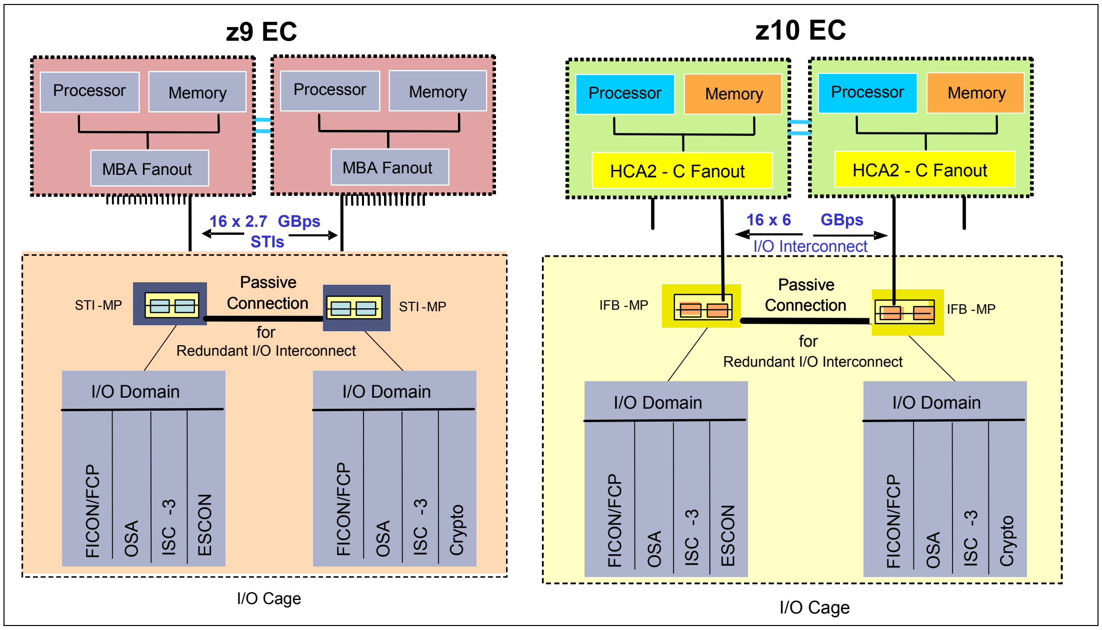

Figure 2-6 z9 EC and z10 EC system structure for I/O

For the z10 EC server there are up to eight fanout cards (numbered D1, D2, and D5 to DA) per book, each driving two IFB cables, resulting in up to 16 IFB connections per book (16 STI connections with the z9 EC server). All 16 InfiniBand (IFB) connections work with a data rate of 6 GB per second. Refer to "InfiniBand" on page 9 for a short introduction to InfiniBand.

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

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

Coupling connectivity

In addition to the HCA2-C fanout card, the z10 EC has two additional fanout cards, the HCA2-O and the MBA fanout. Both these cards are exclusively used for coupling link connectivity on a Parallel Sysplex configuration.

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

- A z10 EC and another HCA2-O card on the same or another z10 EC. This connection has a maximum link data rate of up to 6 GB per second.
- A z10 EC and a System z9 server dedicated as a coupling facility. This connection has a maximum data rate of up to 3 GB per second.

The MBA fanout card provides coupling links (ICB-4 only) to either z10 EC servers or to a z9, z990, or z890 server. This allows using existing servers in existing Parallel Sysplex environments.

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

2.9 I/O cages and features

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

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

The z10 EC server holds a minimum of one I/O cage at the bottom of the A frame and two optional I/O cages in the Z frame. Refer to Figure 2-2 on page 19, where all the three I/O cages are shown.

Each I/O cage supports up to seven I/O domains and a total of 28 I/O card slots. Each I/O domain supports four I/O features (ESCON, FICON, OSA, or ISC). See Figure 2-7 for a pictorial view of an I/O cage.

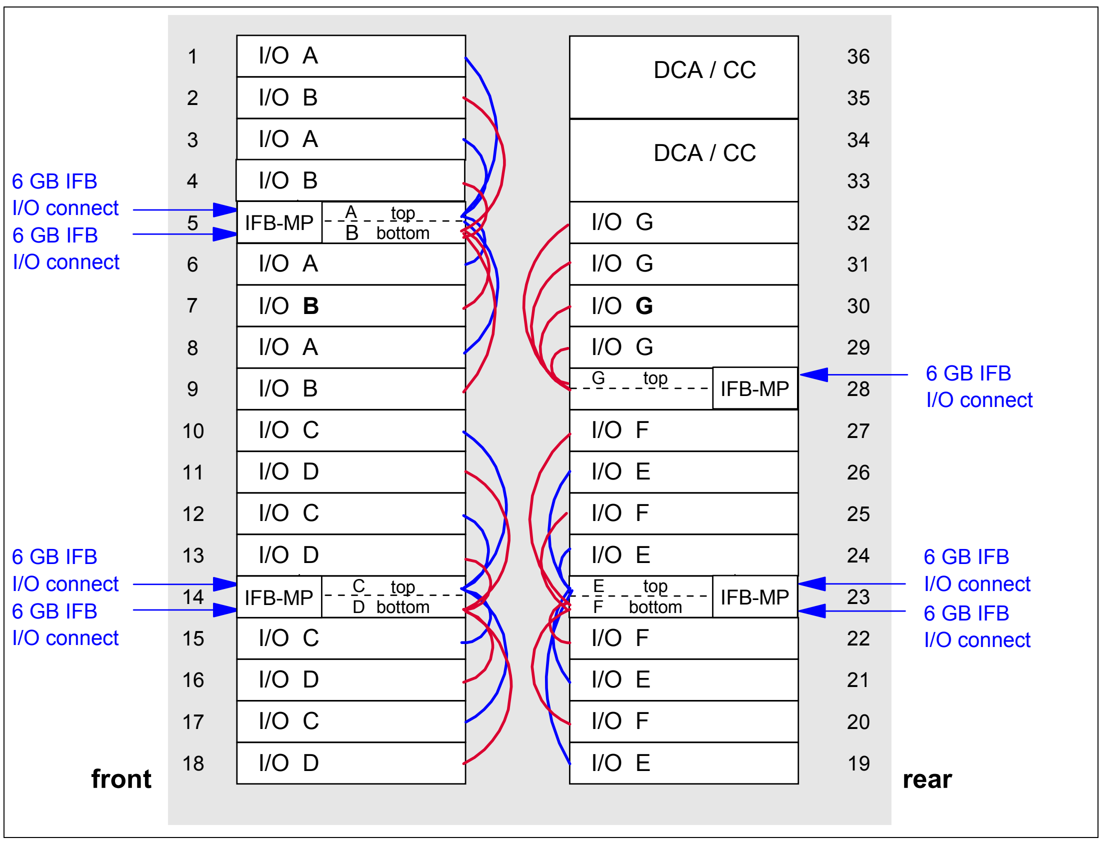

Figure 2-7 z10 EC I/O cage

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

The following I/O features can be ordered for a new z10 EC server:

- ESCON
- FICON Express4 Longwave (4 km and 10 km)
- FICON Express4 Shortwave
- OSA-Express3 10 GbE Long Reach

- OSA-Express2 Longwave
- OSA-Express2 Shortwave
- OSA-Express2 1000BASE-T Ethernet
- Crypto Express2
- ISC-3 (peer mode only)
- ICB-4 (not available on model E64)

The following features are not orderable for a z10 EC, but when present in a z9 EC or z990 server, may be *carried forward* when upgrading to a z10 EC:

- FICON Express2 (Longwave and Shortwave)
- FICON Express (Longwave and Shortwave)
- OSA-Express2 10 GbE Long Reach

The following z9 EC and z990 features are *not* supported on a z10 EC:

- FICON (pre-FICON Express)
- OSA-Express
- ICB-2
- ICB-3
- ISC-3 Links in Compatibility Mode
- PCIXCC and PCICA
- Parallel channels (use an ESCON converter)

For a list of the z10 EC supported I/O features and their characteristics refer to Appendix B, "Channel options" on page 89.

2.9.1 ESCON channels

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

16-port ESCON feature

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

The feature has 16 ports with one PCHID associated with each port, up to a maximum of 15 active ESCON channels per feature. 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 Licensed Internal Code - Control Code (LIC-CC). Each port operates at a data rate of 200 Mbps.

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

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

2.9.2 FICON channels

The FICON Express4, FICON Express2, and FICON Express LX and SX features conform to the Fibre Channel and Fiber Connection (FICON) architecture and directly attach to FICON-supported I/O devices.

FICON channels can be shared among logical partitions and can be defined as spanned. All ports on a FICON feature must be of the same type, either LX or SX.

FICON Express4

Three types of FICON Express4 channel transceivers are supported on new build z10 EC servers—two long wavelength (LX) laser versions and one short wavelength (SX) LED version:

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

All channels on a feature are of the same type, either 10 km LX, 4 km LX, or SX. The features are connected to a FICON-capable control unit, either point-to-point or switched point-to-point, through a Fibre Channel switch.

FICON Express4 10km LX feature

The FICON Express4 10 km LX feature occupies one I/O slot in the I/O cage. It has four ports, each supporting an LC duplex connector. It supports link speeds of 1 Gbps, 2 Gbps, or 4 Gbps up to an unrepeated distance of 10 km (6.2 miles). Interoperability of 10 km transceivers with 4 km transceivers is supported, provided that the unrepeated distance between the two transceivers does not exceed 4 km (2.5 miles).

Each port supports attachment to the following:

- Fibre Channel switches that support 1 Gbps, 2 Gbps, 4 Gbps
- Control units that support 1 Gbps, 2 Gbps, and 4 Gbps
- FICON channels in Fibre Channel Protocol (FCP) mode

FICON Express4 4km LX feature

The FICON Express4 4km LX feature occupies one I/O slot in the I/O cage. It has four ports, each supporting an LC duplex connector. It supports link speeds of 1 Gbps, 2 Gbps, or 4 Gbps up to an unrepeated distance of 4 km (2.5 miles). Interoperability of 10 km transceivers with 4 km transceivers is supported, provided that the unrepeated distance between the two transceivers does not exceed 4 km.

Each port supports attachment to the following:

- Fibre Channel switches that support 1 Gbps, 2 Gbps, 4 Gbps
- Control units that support 1 Gbps, 2 Gbps, and 4 Gbps
- FICON channels in Fibre Channel Protocol (FCP) mode

FICON Express4 SX feature

The FICON Express4 SX feature occupies one I/O slot in the I/O cage. It has two Peripheral Component Interconnect (PCI) cards. The PCI cards have a higher performing infrastructure, which can improve performance compared to the FICON Express2 LX feature. Each PCI card has two ports supporting an LC duplex connector, with one CHPID associated with each port, and supports link speeds of 1 Gbps, 2 Gbps, or 4 Gbps. Each port supports attachment to the following:

- Fibre Channel switches that support 1 Gbps, 2 Gbps, 4 Gbps
- Control units that support 1 Gbps, 2 Gbps, and 4 Gbps
- FICON channels in Fibre Channel Protocol (FCP) mode

FICON Express2

The FICON Express2 feature is supported on a z10 EC only if carried over on an upgrade. Two types of FICON Express2 channel transceivers are supported on z10 EC servers when carried forward on an upgrade—a long wavelength (LX) laser version and a short wavelength (SX) LED version.

- FICON Express2 LX feature
- FICON Express2 SX feature

The features are connected to a FICON-capable control unit, either point-to-point or switched point-to-point, through a Fibre Channel switch.

FICON Express2 LX feature

The FICON Express2 LX feature occupies one I/O slot in the I/O cage. It has four ports, each supporting an LC duplex connector, with one PCHID and one CHPID associated with each port. It supports link speeds of 1 Gbps or 2 Gbps.

Each port supports attachment to the following:

- Fibre Channel switches that support 1 Gbps and 2 Gbps
- Control units that support 1 Gbps and 2 Gbps
- FICON channels in Fibre Channel Protocol (FCP) mode

FICON Express2 SX feature

The FICON Express2 SX feature occupies one I/O slot in the I/O cage. It has four ports, each supporting an LC duplex connector, with one PCHID and one CHPID associated with each port. It supports link speeds of 1 Gbps or 2 Gbps.

Each port supports attachment to the following:

- Fibre Channel switches that support 1 Gbps and 2 Gbps
- Control units that support 1 Gbps and 2 Gbps
- FICON channels in Fibre Channel Protocol (FCP) mode

FICON Express

FICON Express features may be carried forward to the z10 EC when upgrading from a z9 or z990 server.

FICON Express LX feature

The FICON Express LX feature occupies one I/O slot in the I/O cage. It has two ports, each supporting an LC duplex connector, with one PCHID and one CHPID associated with each port. It supports link speeds of 1 Gbps.

Each port supports attachment to the following:

- FICON LX Bridge one-port feature of IBM 9032 ESCON Director at 1 Gbps *only*
- Fibre Channel switches that support 1 Gbps
- Control units that support 1 Gbps
- FICON channels in Fibre Channel Protocol (FCP) mode

Note: FICON Express2 and FICON Express4 features do not support FCV mode. FCV mode is available on z10 EC only if the FICON Express LX feature is carried over on upgrades.

The z10 EC is intended to be the last server to support FICON Express LX feature and CHPID type FCV.

FICON Express SX feature

The FICON Express SX feature occupies one I/O slot in the I/O cage. It has two ports, each supporting an LC duplex connector, with one PCHID and one CHPID associated with each port. It supports link speeds of 1 Gbps.

Each port supports attachment to the following:

- Fibre Channel switches that support 1 Gbps
- Control units that support 1 Gbps
- FICON channels in Fibre Channel Protocol (FCP) mode

2.9.3 OSA-Express3

This section describes the connectivity options offered by the OSA-Express3 features.

The following OSA-Express3 feature can be installed on new build z10 EC servers: OSA-Express3 10 Gigabit Ethernet (GbE) Long Range (LR).

Up to 24 OSA-Express3 features (48 ports) are supported on the z10 EC. Note that the maximum number of OSA-Express2 and OSA-Express3, in combination, is 24 features, system wide.

OSA-Express3 10 GbE LR

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

Compared to the OSA-Express2 10 GbE LR feature, the OSA-Express3 10 GbE LR feature has double port density (two ports per feature) and improved performance for standard and jumbo frames.

The OSA-Express3 10 Gbe LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only. It supports 64B/66B encoding, whereas GbE supports 8B/10B encoding. Therefore, auto-negotiation to any other speed is not possible.

When available, the OSA-Express3 10 GbE LR feature replaces the OSA-Express2 10 GbE LR feature, which will cease to be orderable.

The OSA-Express3 10 GbE LR feature is supported by z/OS, z/VM, z/VSE, z/TPF, and Linux on System z.

2.9.4 OSA-Express2

This section describes the connectivity options offered by the OSA-Express2 features.

The following OSA-Express2 features can be installed on new build z10 EC servers:

- OSA-Express2 Gigabit Ethernet (GbE) Long Wavelength (LX)
- OSA-Express2 Gigabit Ethernet (GbE) Short Wavelength (SX)
- OSA-Express2 1000BASE-T Ethernet

The following OSA-Express2 feature is available only if carried forward by an upgrade: OSA-Express2 Gigabit Ethernet 10 GbE LR.

OSA-Express features installed on previous servers are *not* supported on a z10 EC and *cannot* be carried forward on an upgrade.

A z10 EC supports up to 24 OSA-Express2 features (48 ports). The maximum number of combined OSA-Express2 and OSA-Express3 features is 24.

All supported OSA-Express2 features on the z10 EC are listed below with a short description of their respective support levels and attachment capabilities.

OSA-Express2 GbE LX

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

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

OSA-Express2 GbE SX

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

OSA-Express2 1000BASE-T Ethernet

The OSA-Express2 1000BASE-T Ethernet occupies one slot in the I/O cage and has two independent ports.

Each port supports connection to either a 1000BASE-T (1000 Mbps), 100BASE-TX (100 Mbps), or 10BASE-T (10 Mbps) Ethernet LAN. The LAN must conform either to the IEEE 802.3 (ISO/IEC 8802.3) standard or the DIX V2 specifications.

The OSA-Express2 1000BASE-T Ethernet feature supports auto-negotiation and automatically adjusts to 10 Mbps, 100 Mbps, or 1000 Mbps, depending upon the LAN.

OSA-Express2 10 GbE LR

The OSA-Express2 10 GbE LR feature occupies one slot in an I/O cage and has one port that connects to a 10 Gbps Ethernet LAN through a 9 micron single mode fiber optic cable terminated with an LC Duplex connector. The feature supports an un-repeated maximum distance of 10 km.

The OSA-Express2 10 GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

The OSA-Express 10 GbE LR feature is supported by z/OS, z/VM, z/VSE, z/TPF, and Linux on System z.

Coupling links

Coupling links are used in Parallel Sysplex cluster configurations of System z servers. The links provide high-speed bidirectional communication between members of the sysplex. The z10 EC offers new links that exploit the InfiniBand technology.

For more details refer to *Getting Started with InfiniBand on System z10 and System z9,* SG24-7539.

2.10 Cryptographic functions

The z10 EC server includes both standard cryptographic hardware and optional cryptographic features to provide flexibility and growth capability. IBM has a long history of providing hardware cryptographic solutions. Use of the cryptographic hardware function requires support by the operating system. For the z/OS operating system, the Integrated Cryptographic Service Facility (ICSF) is a base component that provides the administrative interface and a large set of application interfaces to the hardware.

Cryptographic support on the z10 EC includes:

- CP Assist for Cryptographic Function
- Crypto Express2 cryptographic adapter feature
- Trusted key entry workstation feature

2.10.1 CP Assist for Cryptographic Function

Figure 2-8 shows the layout of the z10 EC Compression and Cryptographic Accelerator (CCA). The chip contains two CCAs and each two cores on the chip share the encrypting and hashing engines of a CCA (each core has a dedicated compression engine). Every processor in the z10 EC server characterized as a CP or an IFL has access to the CP Assist for Cryptographic Function (CPACF).

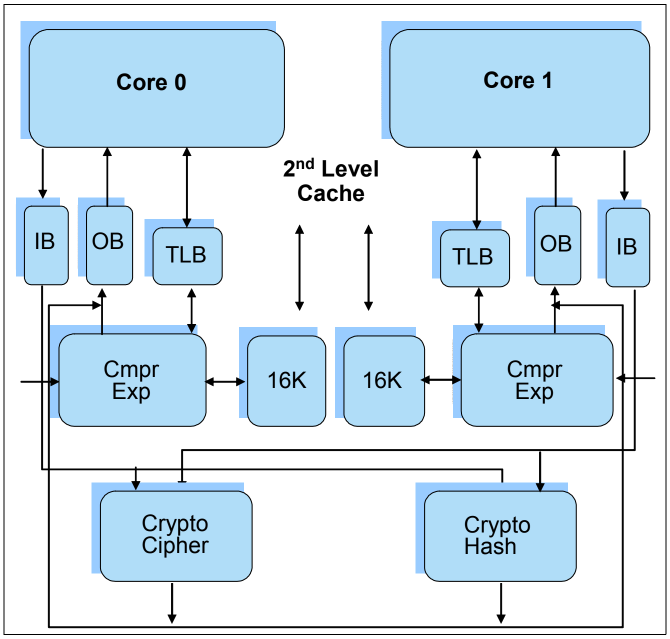

Figure 2-8 z10 EC compression and cryptography accelerator

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

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

- Data Encryption Standard (DES) data encryption and decrypting.
- Triple Data Encryption Standard (TDES) data encrypting and decrypting.
- Advanced Encryption Standard (AES) for 128-bit, 192-bit, and 256-bit keys.
- Pseudo Random Number Generation (PRNG). Note that PRNG is also a standard function supported on the Crypto Express2 feature.
- MAC message authorization.
- Secure Hash Algorithm (SHA-1, SHA-256, SHA-384, and SHA-512) hashing.

SHA-1, 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, and Linux on System z.

2.10.2 Crypto Express2 feature

The Crypto Express2 feature provides cryptographic functions on the z10 EC server. The Crypto Express2 feature has two PCI-X adapters. Each of the PCI-X adapters can be configured as either a coprocessor or an accelerator.

- Crypto Express2 Coprocessor - for secure key encrypted transactions (default)

  Designed to support security-rich cryptographic functions, use of secure encrypted key values, and user-defined extensions (UDXs). Offers functions similar to the previous PCICC or PCIXCC cryptographic feature, including the secure key functions. This mode is intended for Federal Information Processing Standard (FIPS) 140-2 Level 4 certification.

- Crypto Express2 Accelerator - for Secure Sockets Layer (SSL) acceleration
  - Designed to support clear key RSA operations
  - Offloads compute-intensive RSA public-key and private-key cryptographic operations employed in the SSL protocol

To support reliability, availability, and serviceability (RAS) requirements, the initial purchase must contain two Crypto Express2 features. Additional features may be added in increments of one.

The configurable Crypto Express2 feature is supported by z/OS, z/VM, z/VSE, and Linux on System z.

2.10.3 TKE 5.2 workstation

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

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

An optional Smart Card Reader can be attached to the TKE 5.2 workstation to allow for the use of smart cards that contain an embedded microprocessor and associated memory for data storage. Access to and the use of confidential data on the smart cards is protected by a user-defined personal identification number (PIN).

2.11 Coupling and clustering

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

Coupling connectivity on z10 EC in support of Parallel Sysplex environments gets a major boost with new Parallel Sysplex using InfiniBand (PSIFB) connections. PSIFB supports longer distances (150 m) between servers. Customers who use ISC-3 for distances under 150 m can benefit from migrating to PSIFB because of its greater bandwidth. PSIFB provides up to 3 GBps for z10 EC to z9 servers and up to 6 GBps for z10 EC to z10 EC servers.

InfiniBand technology allows you, over time, to move all of the Parallel Sysplex support to a single type of interface that provides high-speed interconnection at short distances (replacing ICB), longer distance fiber optic interconnection (replacing ISC), and interconnection between logical partitions on the same system without external cabling (replacing IC). At the time of writing this book, InfiniBand supports only short distances, up to 150 meters.

ISC-3

InterSystem Coupling Facility-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.

ICB-4

An ICB-4 connection consists of one link that attaches directly to an MBA's STI port in the system. The ICB-4 does not connect to a card in the I/O cage.

Internal Coupling (IC)

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

Parallel Sysplex InfiniBand (PSIFB) coupling

Coupling over PSIFB on z10 EC is supported by a new fanout card on the front of the book—the Host Channel Adapter2-Optical (HCA2-O). This is for coupling over fiber optic cabling exclusively. It provides:

- Twelve times the InfiniBand Double Data rate (12x IFB-DDR), 12 *lanes* (pairs of fiber) for a total link data rate of up to 6 GBps.
- One hundred and fifty meters (492 feet) point-to-point InfiniBand technology that allows System z to support a higher link data rate at greater distances in a Parallel Sysplex environment, replacing ICB, which is limited to ten meters. PSIFB is also a possible replacement for ISC-3 in situations when it is used in the data center to satisfy requirements for greater than seven-meter distances between servers.

Coupling over InfiniBand (CHPID type CIB) is designed to complement and enhance the support provided by traditional coupling link technology.

Bandwidth

PSIFB offers up to 6 GBps of bandwidth between z10 EC servers and up to 3 GBps of bandwidth between z10 EC and System z9 servers.

Time source for STP traffic

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

System-Managed CF Structure Duplexing

System-Managed Coupling Facility 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. See the Parallel Sysplex Web site at:

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

2.12 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.

2.12.1 External time reference (ETR)

Two external time reference cards are a standard feature of the z10 EC server. The ETR cards contain the ETR ports for Sysplex Timer® connection and provide a dual-path interface to the IBM Sysplex Timers, which may be used for timing synchronization between systems.

The two ETR cards are located in the processor cage of the z10 EC server.

2.12.2 Server Time Protocol (STP)

Server Time Protocol (STP) is a server-wide facility that is implemented in the Licensed Internal Code of z10 EC (and also in the z9 EC, z9 BC, z990, and z890 servers and Coupling Facilities (CFs)). STP presents a single view of time to PR/SM and provides the capability for multiple servers and CFs to maintain time synchronization with each other. A z10 EC, z9 EC, z9 BC, z990, z890, or CF may be enabled for STP by installing the STP feature.

The Server Time Protocol feature is intended to be the supported method for maintaining time synchronization between z10 EC, z9 EC, z9 BC, z990, and z890 servers and CFs. See c3.3.2, “Server Time Protocol (STP)” on page 54.

Network Time Protocol (NTP) client support

NTP client support is available on the z10 EC server and has recently (October 2007) been added to the STP code on System z9. This implementation answers the need for a single time source across the heterogeneous platforms in the enterprise. With this implementation the z10 EC server and the System z9 servers support the use of NTP servers as time sources.

2.13 HMC and SE

The Hardware Management Console (HMC) and the Support Element (SE) functions of the z10 EC server are similar to the z9 EC server, but do offer enhanced functionality.

The new HMC Version 2.10.0 will support the systems and SE versions shown in Table 2-7.

Table 2-7 HMC downward system support

Server family	Server type	Driver	SE version
z10 EC	2097	73	2.10.0
z9 BC	2096	67	2.9.2
z9 EC	2094	67	2.9.2
z890	2086	55	1.8.2
z990	2084	55	1.8.2
z800	2066	3G	1.7.3
z900	2064	3G	1.7.3
9672 G6	9672/9674	26	1.6.2
9672 G5	9672/9674	26	1.6.2

HMCs and SEs can now communicate using TCP/IP Version 4 (IPv4), TCP/IP Version 6 (IPv6), or both. It is no longer necessary to assign a static IP address to an SE if it only needs to communicate with HMCs on the same subnet because an HMC and an SE can use IPv6 link-local addresses to communicate with each other. Every IPv6 network interface is assigned a link-local IP address. This link-local address is for use on a single link (subnet) and is never routed. Two IPv6-capable hosts on a subnet can communicate using link-local addresses, without having any other IP addresses assigned.

The HMC user interface is implemented through a standard Web browser. Remote users may connect to the HMC using a browser, and have almost complete functionality. If an IPv6 address is assigned to the HMC, you can browse to it using that address. Surround the IPv6 address in the URL with brackets. Using link-local addresses is not supported by most browsers.

2.14 Power and cooling

The z10 EC server footprint is slightly bigger than the z9 EC footprint. The width of the frames are identical but the depth of the z10 EC server frames is 71.0 in (1803 MM), as compared to the 62.1 in (1577 mm) of the z9 EC frame.

The power service specifications are the same, but the power consumed by the z10 EC server can be greater. A fully loaded z10 EC server maximum consumption is 31.7 KW. Refer to Table 2-8 for the electrical service requirements for the different configurations.

Table 2-8 Electrical service requirements

Model	1 I/O cage	2 I/O cages	3 I/O cages
E12	2x60A	2x60A	2x60A
E26	2x60A	4x60A	4x60A
E40	4x60A	4x60A	4x60A
E56	4x60A	4x60A	4x60A

Model	1 I/O cage	2 I/O cages	3 I/O cages
E64	4x60A	4x60A	4x60A

2.14.1 Hybrid cooling system

The z10 EC server has a hybrid cooling system that is designed to lower power consumption. It is an air-cooled system, assisted by refrigeration. Refrigeration is provided by a closed-loop liquid cooling subsystem. The entire cooling subsystem has a modular construction. Its components and functions are found throughout the cages.

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

2.14.2 Internal Battery Feature

The Internal Battery Feature (IBF) is an optional feature on the z10 EC server. Refer to Figure 2-2 on page 19 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 over a *brownout* or for orderly shutdown, in case of a longer outage. The IBF provides up to 10 minutes of full power, depending on the I/O configuration.

3

Key functions and capabilities

The System z10 Enterprise Class is the follow-on to the System z9 EC (z9 EC) server. As its predecessor, it also offers five hardware models, but has a more powerful uniprocessor, more processor units, and new functions and features.

The z10 EC represents a new level of microprocessor technology, made possible through advances in the design and manufacturing processes.

Virtualization has been improved, additional security options are provided, and the I/O subsystem and some I/O features have been enhanced.

Based on customer requirements and changes in the market demands, the z10 EC offers some important enhancements and additional flexibility to the Capacity on Demand functions.

The z10 EC improves the reliability, availability, and serviceability of the server compared to previous System z servers by introducing solutions to minimize not only unplanned outages, but also to decrease the need for planned outages.

This chapter builds upon the information presented in Chapter 1, "Introducing the System z10 Enterprise Class" on page 1, and Chapter 2, "Hardware overview" on page 15. It discusses the following topics:

3.1 Virtualization

Virtualization is a key strength of the z10 EC server. Virtualization is embedded in the z/Architecture, which includes a precise and model-independent definition of the hardware-to-software interface. It is deeply built into the server implementation, which supports virtualization through both hardware and software.

Virtualization creates the appearance of multiple concurrent servers by sharing the existing hardware. Its major goal is to fully utilize the server resources, thus lowering the total amount of resources needed and their cost. Virtualization can be seen as an application with very demanding performance and security requirements. The z10 EC is able to handle tens, hundreds, even thousands, of virtual servers, so a very high context switching rate is to be expected, and accesses to the memory, caches, and virtual I/O devices have to be kept completely isolated. The z/Architecture, the z10 EC, and their predecessors have been *designed* to meet those requirements with very low overhead and the highest security certification in the industry: common criteria EAL5 with specific target of evaluation (logical partitions). This design has been proved in many customer installations during the last decades.

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 System z has both. In System z the hardware hypervisor is implemented in firmware and is called Processor Resource/Systems Manager™ (PR/SM). PR/SM is part of the base server and does not require any software to run. The z/VM operating system implements the software hypervisor. z/VM requires some PR/SM functions. IBM publishes a System Integrity Statement for z/OS and z/VM. Refer to the following:

- For z/OS:

  `http://www-03.ibm.com/servers/eserver/zseries/zos/racf/zos_integrity_statement.html`

- For z/VM:

  `http://www.ibm.com/security/zvminteg.html`

3.1.1 Hardware virtualization

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

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

Logical processors

Logical processors are perceived by the operating systems as real ones. They assume the following types: CPs, zAAPs, zIIPs, IFLs, and ICFs (SAPs are never part of an LPAR configuration).

PR/SM is responsible for honoring the request for logical processor work by dispatching it on the physical processors of the same type. Under certain circumstances logical zAAPs and zIIPs can be dispatched on physical CPs. Physical processors can be shared across LPARs to about 1% of their capacity, but can also be dedicated to an LPAR. However, an LPAR must have its logical processors either all shared or all dedicated.

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

Memory

To ensure security and data integrity, memory cannot be shared by active LPARs. In fact, a strict isolation is maintained. When an LPAR is activated, its defined memory is allocated in contiguous blocks, which must be a multiple of a given value. This value depends on the total allocation and varies between 64 MB and 1 GB. Thus, memory can be serially reused.

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

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

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

I/O virtualization

The z10 EC supports four channel subsystems with 256 channels each, for a total of 1024 channels. In addition to dedicated use of channels and I/O devices to an LPAR, I/O virtualization allows concurrent sharing of channels, and the I/O devices accessed through these channels, by several active LPARs. The function is known as Multiple Image Facility (MIF). The sharing channels may belong to different channel subsystems, in which case they are known as spanned channels.

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

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

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

For large installations, which usually have a large number of devices, the total number of device addresses can be very high. Thus, the concept of *channel sets* was introduced with System z9. Each channel can address two sets of 64 K device addresses, allowing the base addresses to be defined on *set 0* (IBM reserves 256 subchannels on set 0) and the aliases on *set 1*. In total, 130,816 subchannel addresses are available *per channel.*

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 z10 EC greatly eases the pre-planing requirements and enhances the flexibility and availability of these reconfigurations.

3.1.2 Software virtualization

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

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

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

Under z/VM, virtual processors, virtual central and expanded storages, and all the virtual I/O devices of the VMs are dynamically definable (provisionable). Although memory redefinition is disruptive, the other changes are concurrent. To render these concurrent definitions also nondisruptive requires support by the operating system running in the VM, which is also the case when running in an LPAR.

Although z/VM imposes no limits on the number of defined VMs, the number of active virtual machines is limited by the available resources. On a large server such as the z10 EC thousands of VMs can be activated.

IBM Statements of Direction:

- z/VM LPAR enhancements

  IBM intends to further enhance z/VM in a future release to exploit the System z10 EC support for a new logical partition (LPAR) mode "z/VM", exclusively for running z/VM. This new LPAR mode allows z/VM to utilize a wider variety of specialty processors in a single LPAR. For instance, in a z/VM mode LPAR, z/VM can manage Linux on System z guests running on IFL processors while also managing z/VSE and z/OS on central processors (CPs), and to offload z/OS system software overhead, such as DB2 workloads, on System z9 and System z10 Integrated Information Processors (zIIPs) and System z9 and System z10 Application Assist Processors (zAAPs).

- Additional support for managing z/VM systems

  IBM intends to further enhance z/VM in a future release to exploit the new Hardware Management Console (HMC) interface that allows the installation of Linux on System z into a z/VM virtual machine. Additionally, future support is planned for z/VM and the HMC to provide z/VM hypervisor-configuration tasks.

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

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

3.2 Technology improvements

The technology improvements for the z10 EC fall into five categories: microprocessor, capacity, memory, connectivity, and cryptography. They are intended to provide a more scalable, flexible, manageable, and secure consolidation and integration platform contributing to a lower total cost of ownership.

3.2.1 Microprocessor enhancements

The System z10 Enterprise Class has a newly developed microprocessor chip and a newly developed infrastructure chip. Both of those chips represent a major step forward in technology utilization for the System z products using CMOS 11 technology, resulting in increased packaging density.

Like for the z9, the microprocessor chip and the infrastructure chip for the z10 EC are packaged together on a new Multi-Chip Module (MCM). The MCM contains five microprocessor chips and two infrastructure chips, while the z9 MCMs included 16 chips in total. Each microprocessor chip contains four cores. The MCM is installed inside a book, and the z10 EC can contain from one to four books. The book also contains the memory arrays, I/O connectivity infrastructure, and various other mechanical and power controls.

The book is connected to the I/O cages through one or more cables. As new standards are making their way on to the z10 EC, these cables are now using the standard InfiniBand protocol to transfer large volumes of data between the memory and the I/O cards located in I/O cages.

z10 EC processor chip

The z10 EC chip provides more functions per chip—four cores on a single chip—thanks to technology improvements that allow designing and manufacturing more transistors per square inch. This translates into using fewer chips to implement the needed functions, which helps enhance system availability.

Both chips were developed in close cooperation with the System p™ development organization that has designed the new Power6 chip. It could be said that the new z10 EC microprocessor chip and the Power6 chip share a lot of DNA—they are siblings, but not identical twins. They have common characteristics but differ in many ways and will continue to do so.

The microprocessor chip has a significant new design when compared to the z9. 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 has now been stretched to its maximum, so a fundamental change was necessary.

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

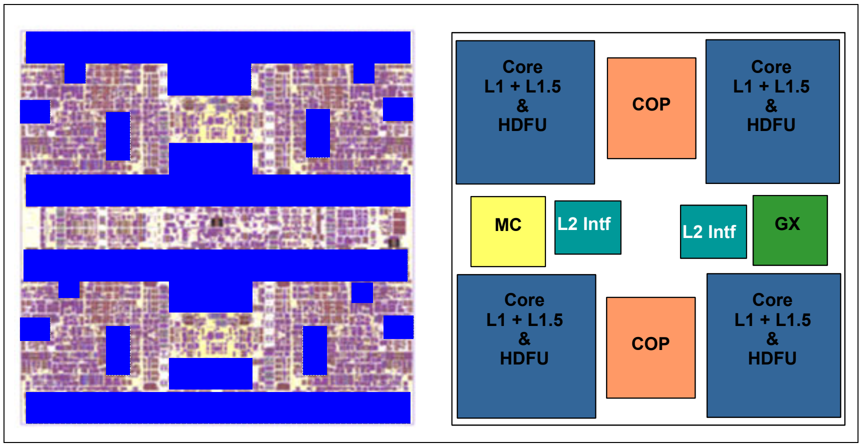

Figure 3-1 z10 EC Enterprise Quad-Core microprocessor chip

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

Hardware decimal floating point function

The z10 EC microprocessor implements a hardware decimal floating point function, 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.

New machine instructions

The z/Architecture offers a rich CISC Instruction Set Architecture (ISA). The z10 EC implements 894 instructions, of which 668 are implemented entirely in hardware. Multiple arithmetic formats are supported.

The z10 EC architectural extensions include over 50 new instructions, the bulk of which was designed in collaboration with software developers to improve compiled code efficiency. These should particularly benefit Java-based, WebSphere-based, or Linux-based workloads. New instructions are grouped under the following categories:

- Storage immediate operations
- Combined comparison and conditional branch, based on the comparison result
- Combined comparison and trap to exception handlers, based on the comparison result
- Operations on storage operands defined relative to the current instruction address
- Combined rotate and logical bit operations
- Extensions to existing instructions
- Instructions for enablement of software/hardware cache optimization
- Support of large page frames

3.2.2 Granular capacity and capacity settings

The z10 EC expands the offer on sub-capacity settings. Finer granularity in capacity levels allows the growth of installed capacity to more closely follow the enterprise growth, for a smother, pay-as-you-go investment profile. The many performance and monitoring tools available on System z environments, coupled with the flexibility of the capacity on demand options (see 3.4, “Capacity on Demand enhancements” on page 55) provide for managed growth with capacity being available when needed.

The z10 EC offers four distinct capacity levels for CPs. A processor characterized as anything other than a CP is always full capacity. This is because purchase and maintenance prices, as well as software licensing costs for non-CP processors, follow a different model.

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 z10 EC server, 64 CPs can be configured as CP7. Besides full capacity CPs, three sub-capacity levels (CP6, CP5, and CP4), each for up to twelve CPs, are offered. The four capacity levels appear in hardware descriptions as feature codes on the CPs. These feature codes (FC) are:

- CP7 is FC 6810
- CP6 is FC 6809
- CP5 is FC 6808
- CP4 is FC 6807

Granular capacity adds 36 sub-capacity settings to the 64 capacity settings that are available with full capacity CPs (CP7). Each of the 36 sub-capacity settings only apply to up to twelve CPs, independent of the z10 EC model installed. Figure 3-2 shows the relative capacity of subcapacity models.

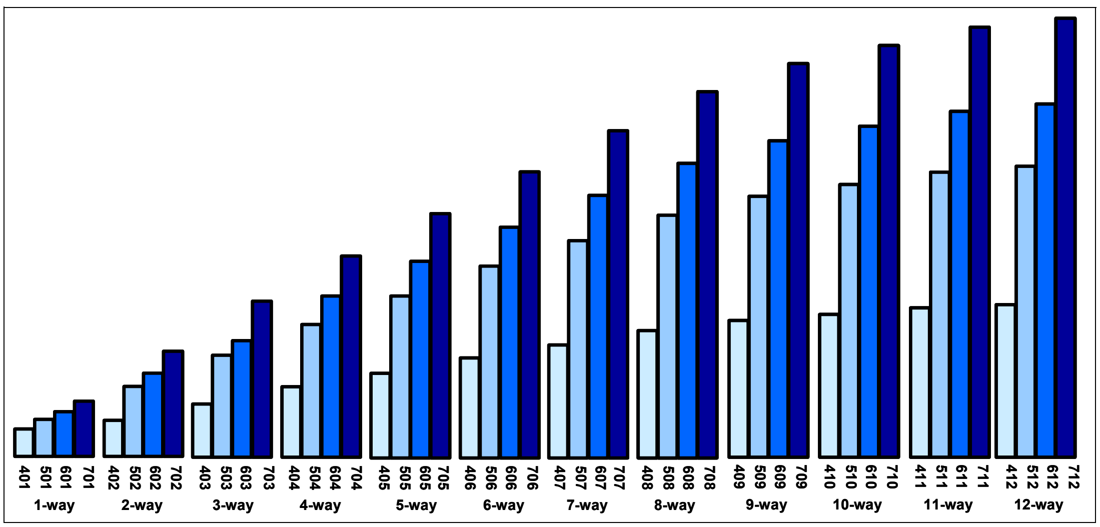

Figure 3-2 z10 EC granular capacity for up to 12 CPs

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

- 701 to 764, for capacity level CP7
- 601 to 612, for capacity level CP6
- 501 to 512, for capacity level CP5
- 401 to 412, for capacity level CP4

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

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

3.2.3 Memory enhancements

The z10 EC has greatly increased the available memory capacity over previous servers. The system can now have up to 1,520 GB of usable memory installed. The logical partitions can now be configured with 1 TB of memory, and z/OS will support the new memory size from Version 1 Release 8 and onward. In fact, z/OS 1.8 and later support up to 4 TB of main storage. For the first time, the hardware systems area (HSA) is fixed in size (16 GB) and not included in the memory that the customer orders and pays for.

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

Hardware system area

The z10 EC has a new, fixed size, hardware system area. This is intended to improve the server availability. Because the HSA is now big enough to accommodate all possible configurations for all the logical partitions, several operations that were disruptive due to HSA size change are now concurrent. Also, some pre-planning needs were eliminated.

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

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

- New logical partition to new or existing channel subsystem (CSS)
- New LCSS (Up to four can be defined.)
- New subchannel set (Up to two can be defined.)
- Dynamic I/O enabled as a default
- Add logical processors by type
- Add a cryptographic processor

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 the memory that becomes unavailable when replacing a book. On z10 EC, the additional resources required for the flexible memory configurations are provided through the purchase of pre-planned memory features along with the purchase of your memory entitlement. Flexible memory configurations are available on multi-book models E26, E40, E56, and E64 and range from 32 GB to 1136 GB, model dependent.

Contact your IBM representative to help you determine the appropriate configuration.

Large page support

The size of pages and page frames has been 4 KB for a long time. The z10 EC server introduces the capability of having large pages with the size of 1 MB. This is a performance item addressing particular workloads, and relates to large main storage usage. Large page support is exclusive to z10 EC and to z/OS V1R9 and later.

Large pages cause that the translation lookaside buffer (TLB) to better represent the working set and suffer fewer misses by allowing a single TLB entry to cover more address translations. Exploiters of large pages are better represented in the TLB and are expected to perform better.

This support is primarily of benefit for long-running applications that are memory access intensive. Large page is not recommended for general use. Short-lived processes with small working sets are normally not good candidates for large pages and would see none or little 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.

Large pages are treated as fixed pages and are never paged out. They are only available for 64-bit virtual private storage such as virtual memory located above 2 GB.

3.2.4 Connectivity enhancements

The I/O cages are similar to the ones used in the z9, and the I/O cards are the same as used in the z9, with some exceptions. See Table B-1 on page 89 for more details. There is a new OSA Express3 10 Gb Ethernet card and other connectivity changes introduced on the z10 EC.

The physical connection between the processor and memory and the I/O cages uses a new cable based on the InfiniBand technology. Up to the z9, the cables used were so-called STI cables specifically developed for this function. With new InfiniBand cables the bandwidth per cable increases from 2.7 GB per second to 6 GB per second.

Standard InfiniBand cables and protocol can be used for Parallel Sysplex coupling links for servers that are up to 150 meter apart. InfiniBand offers a longer distance than the existing ICB cable and increased bandwidth, similar to the bandwidth obtained with the cable used internally in the server. The System z9 servers can be upgraded to use the new coupling link and can participate in a Parallel Sysplex using this new technology.

Advantages of InfiniBand

InfiniBand addresses the challenges IT infrastructures face as more demands are placed on the interconnect with ever-increasing requirements for computing and storage resources. Specifically, InfiniBand has the following advantages:

- Superior performance. InfiniBand provides superior latency performance and products supporting 20 Gbs node-to-node and 60 Gbs switch-to-switch connections. In addition, InfiniBand has a defined road map to 120 Gbs—the fastest support specification of any industry-standard interconnect.
- Reduced complexity. InfiniBand allows for the consolidation of multiple I/Os on a single cable or backplane interconnect. InfiniBand also consolidates the transmission of clustering, communications, storage, and management data types over a single connection.
- Highest interconnect efficiency. InfiniBand was developed to provide efficient scalability of multiple systems. InfiniBand provides communication-processing functions in hardware—relieving the CPU of this task—and enables the full resource utilization of each node added to the cluster. In addition, InfiniBand incorporates Remote Direct Memory Access, which is an optimized data transfer protocol that further enables the server processor to focus on application processing.
- Reliable and stable connections. InfiniBand provides reliable end-to-end data connections and defines this capability to be implemented in hardware. In addition, InfiniBand facilitates the deployment of virtualization solutions, which allow multiple applications to run on the same interconnect with dedicated application partitions.

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

Statements of Direction:

- 1x IB-DDR to be introduced to support extended distance for coupling using InfiniBand

  IBM intends to offer 1x InfiniBand at double data rate (1x IB-DDR), which is designed to complement and/or replace InterSystem Channel-3 (ISC-3) in a Parallel Sysplex environment. When 1x IB-DDR is introduced it will be designed to support one “lane” (one pair of fiber - a transmit and a receive) with a 5 gigabits per second (Gbps) link data rate at an unrepeated distance of 10 km using 9 micron single-mode fiber optic cabling with LC Duplex connectors.

  The common fiber optic cabling environment is designed to facilitate a migration from ISC-3 (2 Gbps link data rate) to 1x IB-DDR (5 Gbps link data rate) in a Parallel Sysplex environment on System z10 EC.

- System z9 to support 12x IB-SDR on z9 EC and z9 BC general-purpose servers

  Support of 12x InfiniBand at single data rate (12x IB-SDR) - a link data rate of 3 gigabytes per second (3 GBps) - is planned to be made available on z9 EC and z9 BC servers whether the Processor Units (PUs) are defined as CPs, IFLs, zAAPs or zIIPs.

  This expands the support for 12x IB-SDR on z9 EC and z9 BC beyond stand-alone Coupling Facilities (ICFs only).

FICON enhancements

The z10 EC server offers some exclusive enhancements to FICON Express4 and FICON Express2 features.

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

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

Extended Distance FICON is transparent to operating systems and applies to all the FICON Express2 and FICON Express4 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 IBM System Storage™ DS8000 series with an appropriate level of Licensed Machine Code (LMC).

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. A laboratory test using one FICON

Express4 channel configured as CHPID type FCP with no other processing occurring demonstrated up to 57,000 I/O operations per second (all reads, all writes, or a mix of reads and writes), an 80% increase compared to System z9. Actual field results may differ.

A significant increase in I/O operations per second for small block sizes can also be expected with FICON Express2.

This FCP performance improvement is transparent to operating systems and applies to all the FICON Express4 and FICON Express2 features when configured as CHPID type FCP, communicating with SCSI devices.

SCSI IPL base function

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

N_Port ID Virtualization (NPIV)

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

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

Fiber Quick Connect for FICON LX

Fiber Quick Connect (FQC), an optional feature on z10 EC, is now being offered for all FICON LX (single-mode fiber) channels, in addition to the current support for ESCON (62.5 micron multimode fiber) channels. FQC is designed to significantly reduce the amount of time required for on-site installation and setup of fiber optic cabling.

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

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

Open Systems Adapter enhancements

The z10 EC offers the new OSA-Express3 feature. Other capabilities are mentioned that can help consolidate or simplify the data center environment.

Open Systems Adapter for NCP

The OSA-Express2 Gigabit Ethernet and 1000BASE-T Ethernet features have 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.

OSA-Express2 OSN support is available on the OSA-Express2 Gigabit Ethernet SX, Gigabit Ethernet LX, and 1000BASE-T Ethernet features.

VLAN support

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

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

VMAC support

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

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

OSA-Express3 10 GbE LR

The new OSA-Express3 10 Gigabit Ethernet (GbE), planned to be available in the second quarter of 2008, has been designed to increase the throughput for standard and jumbo frames, as compared to OSA-Express2 10 GbE. The performance increase results from an enhancement to the architecture that supports direct host memory access by using a data router, eliminating *store and forward* delays. When available, OSA-Express3 10 GbE LR will replace OSA-Express2 10 GbE LR, which will become unavailable for ordering.

The OSA-Express3 10 GbE has been designed with two PCI adapters, each with one port. Doubling the port density on a single feature helps to reduce the number of I/O slots required for high-speed connectivity to the LAN.

OSA-Express3 10 GbE LR is exclusive to z10 EC and is supported by z/OS, z/VM, z/VSE, z/TPF, and Linux on System z.

HiperSockets enhancements

HiperSockets has been called the *network in a box*. z10 EC provides enhancements in two areas.

HiperSockets Layer 2 support

With this support the HiperSockets internal networks on System z10 EC can support two transport modes: Layer 2 (link layer) as well as the current Layer 3 (network or IP layer). Traffic can be Internet Protocol (IP) Version 4 or Version 6 (IPv4, IPv6) or non-IP (AppleTalk, DECnet, IPX™, NetBIOS, or SNA). HiperSockets devices are now protocol-independent and Layer 3 independent. Each HiperSockets device has its own Layer 2 Media Access Control (MAC) address, which is designed to allow the use of applications that depend on the existence of Layer 2 addresses such as Dynamic Host Configuration Protocol (DHCP) servers and firewalls.

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

HiperSockets Layer 2 support is exclusive to System z10 EC, supported by Linux on System z, and by z/VM for guest exploitation.

HiperSockets Multiple Write Facility

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

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

For a complete description of the z10 EC connectivity capabilities refer to *IBM System z Connectivity Handbook*, SG24-5444.

3.2.5 Cryptography enhancements

The z10 EC delivers cryptographic facilities similar to the z9. The implementation of the processor part of the cryptographic technology (CPACF) packaging is changed, but there is no impact to the functions provided and the way to use them from the applications. The same crypto features can be used on the z10 EC as the ones used on the z9.

The cryptographic functions include improvements to encryption key lengths, in response to demand due to changing open standards emerging in the marketplace. Specifically:

- Advanced Encryption Standard (AES) has been improved with support for 192-bit and 256-bit keys.
- Secure Hash Algorithm (SHA) now offers stronger hash algorithms with 384-bit and 512-bit key support for message digest.

The enhancements to CPACF are exclusive to the System z10 EC and supported by z/OS, z/VM, z/VSE, and Linux on System z.

3.3 Common time functions

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

3.3.1 Sysplex Timer

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

The z10 EC and the other System z servers can attach to the IBM Sysplex Timer units. More information can be found in the Redbooks publication *S/390® Timer Management and IBM 9037 Sysplex Timer*, SG24-2070.

ETR attachment

New with the z10 EC server is the shipment as a standard feature of two External Time Reference (ETR) cards. Each card contains an ETR port for Sysplex Timer connection.

Thus, a redundant dual-path interface to IBM Sysplex Timers can be created, which may be used for timing synchronization between systems. This redundant design allows continued operation even if a single ETR card fails, and also allows concurrent maintenance. The two z10 EC server ETR cards are located in the processor cage of the z10 EC server.

3.3.2 Server Time Protocol (STP)

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

The Server Time Protocol feature is intended to be the supported method for maintaining time synchronization between z10 EC, z9 EC, z9 BC, z990, and z890 servers and CFs. Note that servers previous to z990 and z890 are *not* supported.

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

STP is implemented in Licensed Internal Code (LIC) as a server-wide facility of z10 EC (and also the z9 EC, z9 BC, z990, and z890 servers and Coupling Facilities). STP presents a single view of time to PR/SM and provides the capability for multiple servers and CFs to maintain time synchronization with each other. A z10 EC, z9 EC, z9 BC, z990, or z890 server or CF may be enabled for STP by installing the STP feature.

STP provides the following additional value over the Sysplex Timer:

- STP supports a multi-site timing network of up to 100 km (62 miles) over fiber optic cabling, without requiring an intermediate site. This allows a Parallel Sysplex to span these distances and reduces the cross-site connectivity required for a multi-site Parallel Sysplex.
- The STP design allows more stringent synchronization between servers and CFs using short communication links, such as PSIFB or ICB-4 links, compared to servers and CFs using long ISC-3 links across sites. With the z10 EC server, STP will support coupling links over InfiniBand.
- STP helps eliminate infrastructure requirements, such as power and space, needed to support the Sysplex Timers.
- STP helps eliminate maintenance costs associated with the Sysplex Timers.
- STP may reduce the fiber optic infrastructure requirements in a multi-site configuration. Dedicated links may not be required to transmit timing information.

The CTN concept is used in order to help meet two key goals of z10 EC and System z customers:

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

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

STP is available as a chargeable feature on the z10 EC, z9 EC, z9 BC, z990, and z890, and is supported by z/OS V1.7 and later. (PTFs are required to enable STP support.)

Network Time Protocol (NTP) client support

The use of Network Time Protocol servers as an External Time Source (ETS) is a requirement for customers who demand a time source or those who need to use a common time reference across heterogeneous platforms. These customers, in most cases, purchase an NTP server that will obtain the exact time, via satellite.

NTP client support is available on z10 EC servers and has been available in System z9 servers since October 2007. With this implementation the z10 EC server and the System z9 servers support the use of NTP servers as time sources.

NTP client support is added to the Support Element (SE) code of the z10 EC server and the System z9. The code interfaces with the NTP servers. This allows an NTP server to become the single time source for the z10 EC server and the System z9, as well as other servers that have NTP clients (UNIX®, NT, and so on). NTP can only be used for an STP-only CTN.

This support satisfies the following 2007 IBM Statement of Direction related to STP:

> IBM intends to enhance the accuracy of initializing and maintaining Coordinated Server Time to an international time standard such as Coordinated Universal Time (UTC). The then current server is planned to have the capability of attaching to an external time source, such as a Global Positioning System (GPS) receiver.

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

3.4 Capacity on Demand enhancements

Based on customer demand and changes in the market requirements, the z10 EC introduces a number of enhancements to the on demand offerings. These changes provide more flexibility and control to the customer, in order to ease the administrative burden in the handling of the offerings, and to give the customer finer control over resources needed to meet the resource requirements in various situations.

The z10 EC has the capability of *concurrent* upgrades, providing additional capacity with no *server* outage. In most cases, with prior planning and operating system support, a concurrent upgrade can also be *nondisruptive* to the operating system.

It is important to note that these upgrades are based on the enablement of resources already physically present in the z10 EC server.

Capacity upgrades cover both permanent and temporary changes to the installed capacity. This 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. Consult the IBM Resource Link site at:

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

With CoD, the z10 EC introduces the possibility of having more than one temporary capacity upgrade active at any point in time. Four different temporary upgrades can be active at the same time, with one of them being an On/Off Capacity on Demand (On/Off CoD) upgrade. The other three can be a combination of other upgrades. Furthermore, upgrades can be performed concurrently, and they can be replenished even when active. It is also possible to do permanent upgrades while temporary upgrades are active.

The content of the on demand upgrade records can be used in such a way that subsets of the capacity can be activated, and additional resources in the upgrade can be added or taken away without having to go back to the base configuration. This removes the need for a customer to have several On/Off CoD upgrade records installed and to have to switch between them, potentially impacting availability, to meet varying workload demands.

3.4.1 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 books on an existing z10 EC server, can be performed by the customer through the IBM On-line Permanent Upgrade offering, using the CIU facility. These permanent upgrades require a special contract, between the customer and IBM, through which the terms and conditions of the offering are accepted.

3.4.2 Temporary upgrades

Temporary upgrades of a System z10 server can be done by On/Off Capacity on Demand, Capacity Back Up (CBU), or Capacity for Planned Event (CPE), using the CIU facility. These temporary upgrades require a special contract, between the customer and IBM, through which the terms and conditions of the offering are accepted.

On/Off Capacity on Demand

On/Off CoD is a function available on z10 EC that enables *concurrent* and *temporary* capacity growth of the server. On/Off CoD *can* be used for customer peak workload requirements, for any length of time, and has a daily hardware charge. Customer's charges for software may vary according to the license agreement for the individual products. The IBM Software Group representative should be contacted for exact details of an impact to charges for IBM Program Products.

Using On/Off CoD, the customer can concurrently add processors (CPs, IFLs, ICFs, zAAPs, zIIPs, and SAPs), increase the CP capacity level, or both, up to the lesser of a) the limit of the installed books of the server, and b) twice the currently purchased capacity.

Capacity BackUp

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 load management of customer workload. A CBU upgrade's activation can last up to 90 days when a disaster recovery situation occurs.

CBU features are optional and require unused capacity to be available on installed books of the backup server, either as unused PUs or as a possibility to increase the Model-Capacity Identifier on a sub-capacity server, or both. A CBU contract must be in place before the special code that enables this capability can be loaded on the server. The standard CBU contract provides for five 10-day tests and one 90-day disaster activation over a five-year period.

Capacity for Planned Events

Capacity for Planned Events (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. The capacity of a CPE upgrade is used to replace temporarily lost capacity within a customer's enterprise for planned downtime events, for example, for data center changes or system maintenance. CPE cannot be used for peak load management of customer workload, and is available for customer's use up to a maximum of three days.

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

Table 3-1 summarizes the Capacity on Demand offerings that are available for the z10 EC.

Table 3-1 Capacity on Demand summary

		Upgrades	**Process**
Permanent	**Online Permanent Upgrade**	CPs, IFLs, ICFs, zAAPs, zIIPs, SAPs, and memory	Performed through the CIU facility
Temporary	**On/Off CoD**	CPs, IFLs, ICFs, zAAPs, zIIPs, and SAPs	Performed through the CIU facility
	CBU	CPs, IFLs, ICFs, zAAPs, zIIP, and SAPs	Performed through the CIU facility
	CPE	CPs, IFLs, ICFs, zAAPs, zIIPs, and SAPs	Performed through the CIU facility

3.4.3 z/OS Capacity Provisioning

Capacity Provisioning helps customers manage the CP, zAAP, and zIIP capacity of z10 EC servers that are running one or more instances of the z/OS operating system. Based on On/Off CoD, temporary capacity may be activated and deactivated under control of a defined policy. Combined with functions in z/OS, the z10 EC provisioning capability gives the customer a new, flexible, automated process to control the configuration and activation of On/Off CoD offerings.

Provisioning architecture

The provisioning architecture enhances the already rich on demand environment by opening up interfaces to the z/OS operating system. The z/OS operating system can interrogate the on demand environment and query which resources are in the On/Off CoD offerings and the status of the resources.

z/OS Capacity Provisioning simplifies the monitoring of critical workloads, and its automation features can help activate additional resources faster than manual operation. When using Capacity Provisioning you can select different levels of automation to provide you with an appropriate level of control. It is possible to:

- Activate and deactivate temporary capacity through operator commands (manual mode).
- Activate and deactivate temporary capacity based on a defined schedule, without considering workload performance.
- Have the Provisioning Manager suggest changes to the capacity of the z10 EC server based on the observation of defined workloads. In this case the operator will have to confirm the suggested changes.
- Have the Provisioning Manager automatically implement changes to the capacity of the z10 EC server based on the observation of defined workloads.

It is also possible to run Capacity Provisioning in analysis mode. In this mode the operator will be informed when an action would have occurred according to the defined rules. However, no action will be taken unless the operator manually enters the necessary commands.

The Capacity Provisioning function included in the z/OS operating system is part of the z/OS MVS™ Base Control Program (BCP). It includes the following:

- Capacity Provisioning Manager - the server program
- Capacity Provisioning Control Center - the workstation code
- Sample datasets and files

The Capacity Provisioning Manager monitors the workload on a set of z/OS instances and organizes the allocation of additional capacity when required. The systems to be observed are defined in a domain configuration file. Details of additional capacity and the rules for its allocation are held in a policy file. These two files are created and maintained using the Capacity Provisioning Control Center.

The Capacity Provisioning Control Center is installed on a workstation and is the graphical user interface to Capacity Provisioning. Through this interface the administrators work with provisioning policies and domain configurations, and can transfer these to the Capacity Provisioning Manager.

The Capacity Provisioning component includes several samples to simplify customization and help the definition of your provisioning policies.

z/OS can then apply its workload management disciplines to the on demand environment using the same mechanisms and controls already used inside z/OS. The provisioning of resources can be placed under control of automation functions using well-known parameters in the z/OS Workload Manager (WLM). Automation processes can activate resources in the OOCoD offerings as dictated by automation policies, and return the resources when they are no longer needed according to the automation policies. Manual and supervised modes (where changes need authorization) are also available.

For more information regarding the Capacity on Demand offerings, refer to the *IBM System z10 Enterprise Class Technical Guide,* SG24-7516, and *IBM System z10 Enterprise Class Capacity on Demand,* SG24-7504.

3.5 Throughput optimization enhancements

The z990 was the first server to use the concept of books. Despite the memory being distributed through the books, and books having Level 2 caches, all processors have access to all the Level 2 and memory. Thus, the server is managed as a coherent symmetric multi-processor (SMP).

Processors within the z10 EC book structure have different *distance to memory* attributes. As described on 2.3, "CEC cage and books" on page 19, books are connected in a star configuration, which helps to minimize the distance.

Other non-negligible effects result from data latency when grouping and dispatching work on a set of available logical processors. In order to minimize latency, one can the aim to dispatch and later re-dispatch work to a group of physical CPUs that share the same Level 2 (L2) cache.

PR/SM manages the utilization of physical processors by logical partitions by dispatching the logical processors on the physical processors. But PR/SM is not aware of which workloads are being dispatched by the operating system in which logical processors. The Workload Manager (WLM) component of z/OS has the information at the task level, but is unaware of physical processors. This disconnect is solved by new enhancements on z10 EC that allow PR/SM and WLM to work more closely together. They can cooperate on creating an affinity between task and processor rather than between logical partition and processor. This is known as HiperDispatch.

HiperDispatch

HiperDispatch, a z10 EC exclusive, combines two functional enhancements, one in the z/OS dispatcher and one in PR/SM. This is intended to improve efficiency both in the hardware and in z/OS.

In general, the PR/SM dispatcher assigns work to a minimum number of logical processors needed for the priority (weight) of the LPAR. The end result is to reduce the multi-processor effects and lower the interference among multiple partitions.

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

In order to use the correct logical processors, the z/OS dispatcher obtains the necessary information from PR/SM via new interfaces implemented on the z10 EC. The entire z10 EC stack (hardware, firmware, and software) now tightly collaborate to obtain the hardware's full potential.

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

3.6 Reliability, availability, and serviceability improvements

The System z10 EC server presents numerous enhancements in the reliability, availability, and serviceability areas. In the availability area focus was given to reduce the pre-planing requirements, while continuing to improve on the elimination of planned, scheduled, and unscheduled outages.

The design and packaging of the z10 EC represents a reduction in the number of chips needed to implement the processor, cache, and infrastructure functions. While the z9 was implemented using 16 chips on the MCM, the z10 EC uses only seven chips on the MCM. This reduction helps to improve the availability characteristics of the new server.

Availability enhancements include single processor core checkstop and sparing, point-to-point fabric for SMP, and fixed size HSA.

The ICB-4 connections used for Parallel Sysplex connectivity can now be reconfigured concurrently, and no longer need a server outage.

If an additional System Assist Processor (SAP) is needed on a z10 EC server (for example, as a result of a disaster recovery situation), the SAPs can be concurrently added to the server configuration.

It is now possible to concurrently add CPs, zAAPs, zIIPs, IFLs, and ICFs processors to an LPAR. This is supported by z/VM V5 R3 with appropriate PTFs.

z10 EC supports dynamically adding Crypto Express2 features to a logical partition, by providing the ability to change the cryptographic information in the image profiles without outage to the logical partition. Users can also dynamically delete or move Crypto Express2 features. Pre-planning is no longer required. This enhancement is supported by z/OS, z/VM for guest exploitation, and Linux on System z.

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

These and additional features are further described *IBM System z10 Enterprise Class Technical Guide,* SG24-7516.

3.7 Parallel Sysplex technology

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

The hardware components comprise:

- Coupling Facility (CF). This is the cluster center. It can be implemented either as a stand-alone server or as a Logical Partition of a System z server. 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 licence charges. Two CFs are recommended for availability.
- Coupling Facility Control Code (CFCC). This IBM Licensed Internal Code is the *operating system* that executes in the CF. No other code executes in the CF.[1]
- 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

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

connections between each physical server and the CF should exist. When all of the system images belong to the same physical server, internal coupling links are used.

On the software side, the z/OS operating system exploits the hardware components to create a Parallel Sysplex[2]. Normally, two or more z/OS images are clustered to create a Parallel Sysplex, although it is possible to have a configuration setting with a single image, called a *monoplex*. Multiple clusters can span several System z servers although a specific image (logical partition) can belong to only one Parallel Sysplex.

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

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

The major characteristics of a Parallel Sysplex are:

- Data sharing with integrity: The CF is key to the implementation of a share-all access to data. Every z/OS system image has access to all the data. Subsystems in z/OS declare resources to the CF. The CF accepts and manages lock and unlock requests on those resources, guaranteeing data integrity. A duplicate CF further enhances the availability. Key exploiters of this capability are DB2, WebSphere MQ, WebSphere ESB, IMS, and CICS.
- Continuous (application) availability: You can introduce changes such as software upgrades one image at a time, while the remaining images continue to process work. For additional details see the manual *Parallel Sysplex Application Considerations,* SG24-6523.
- High capacity: scales from two to 32 images. Remember that each image can have from 1 to 64 processor units. CF scalability is near linear. Contrast to 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 Coupling Facility shared storage. This exploitation enables sharing of physical resources with significant improvements in cost, performance, and simplified systems management.
- Single system image: The collection of system images in the Parallel Sysplex appears as a single entity to the operator, the user, the database administrator, and so on. A single system image ensures reduced complexity from both operational and definition perspectives.

---

[2] TPF and z/TPF also exploit the CF hardware components. However, the term Parallel Sysplex exclusively applies to z/OS exploitation of CF.

Figure 3-3 illustrates the components of a Parallel Sysplex as implemented within the System z architecture. The figure is intended only as an example. It shows one of many possible Parallel Sysplex configurations. Many other possibilities exist.

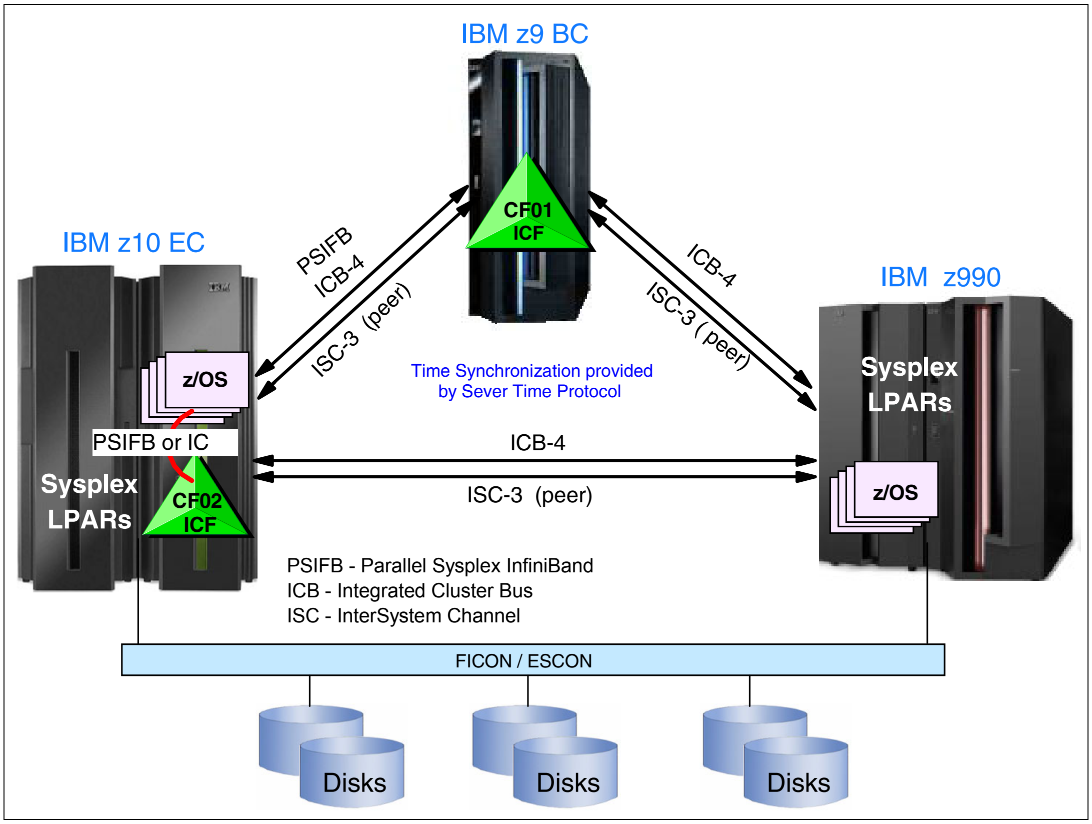

Figure 3-3 Sysplex hardware overview

Figure 3-3 shows a z10 EC server containing multiple z/OS sysplex partitions and an internal coupling facility (CF02), a z9 Business Class server containing a stand-alone CF (CF01), and a z990 containing multiple z/OS sysplex partitions. STP over coupling links provides time synchronization to all servers. Appropriate CF link technology (PSIFB, ICB-4, ISC-3) selection depends on server configuration.

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.

3.8 Summary

Multiple forces are driving a transformation of the data center, such as demands to IT improve cost and service delivery, manage escalating complexity, and better secure the enterprise. Aligning IT more closely with the business has become a primary goal.

The IBM vision for the future of the data center—the new enterprise data center—is designed to optimize service delivery and provide exceptional efficiency. The z10 EC can be the cornerstone of the new enterprise data center, helping to simplify infrastructure and provide powerful shared virtual resources while lowering cost of ownership.

The z10 EC is a big step forward in the mainframe evolution. It is, at the same time, a revolution evidenced in its new microprocessor design, use of high frequency, and additional exploitation of open technologies such as InfiniBand.

The z10 EC extreme virtualization capabilities, very large capacity range (1:140), memory growth (16 GB to 1.5 TB), and I/O bandwidth are provided *in the same footprint*. All this distinctly shows the server's readiness for IT infrastructure simplification through image consolidation and application integration.

When considering the complementarity of supported environments, with emphasis on z/OS and Linux on System z, and the wide scope of technologies exploited by the thousands of supported applications, the z10 EC is the place for both new applications and traditional ones. Furthermore, the investment value locked in traditional applications can be exploited through novel interfaces such as Web services.

z10 EC advanced security features, including key generation and management, make it the choice for an enterprise security hub. Its high availability (up to 99.999% at the *application* level with Parallel Sysplex) and resilience naturally recommend it as the platform to run the mission-critical applications and host, manage, and care for the enterprise data.

The granular capacity increments and capacity on demand options allow the server to grow with the enterprise. In fact, when also considering the advantages of server consolidations, z10 EC may well help grow the enterprise.

4

Software support

This chapter focuses on operating system requirements and support considerations for the z10 EC and its features.

The following topics are included:

Support of the System z10 EC functions is dependent on the operating system version and release. This information is subject to change. Therefore, for the most current information, refer to the Preventive Service Planning (PSP) bucket for 2097DEVICE.

4.1 Software support summary

The software portfolio for the System z10 EC server includes a large variety of operating systems and middleware that support the most recent and significant technologies. Continuing the mainframe-rich tradition, five major operating systems are supported on the z10 EC. These include z/OS, z/VM, z/VSE, z/TPF, and Linux on System z.

Operating systems summary

System z Architecture, also referred to as *z/Architecture,* defines three memory-addressing modes: 64 bit, 31 bit, and 24 bit. Changing between modes is under operating system control.

Table 4-1 summarizes the current and minimum operating system levels required, noting when software service needs to be applied, to support the z10 EC.

Table 4-1 z10 EC operating systems requirements

Operating systems	ESA/390 (31-bit mode)	z/Architecture (64-bit mode)	End of service	Notes
z/OS V1 R9	No	Yes	September 2010[a]	
z/OS V1 R8	No	Yes	September 2009[a]	Service required
z/OS V1 R7	No	Yes	September 2008	Service required
z/VM V5 R3	No	Yes	September 2010[a]	Service required
z/VM V5 R2	No	Yes	April 2009[a]	Service required
Linux on System z	Table 4-5	Table 4-5	Table 4-5	Novell SUSE SLES 9 Novell SUSE SLES 10 Red Hat RHEL 4 Red Hat RHEL 5
z/VSE V4 R1	No[b]	Yes[c]	Not announced	Service required
z/VSE V3 R1	Yes	No	July 2009[a]	Service required
TPF V4 R1	Yes	No	December 2010	Service required
z/TPF V1 R1	Yes	Yes	Not announced	Service required

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. ESA/390 is not supported. However, 31-bit mode is supported.
c. z/VSE 4.1 supports 64-bit real addressing only. It does not support 64-bit addressing for user, system, or vendor applications.

Note: Refer to the z/OS, z/VM, z/VSE, and z/TPF subsets of the 2097DEVICE Preventative Service Planning (PSP) bucket prior to installing an IBM System z10 EC.

Exploitation of some features depends on a particular operating system. In all cases, PTFs may be needed with the operating system level indicated. PSP buckets are continuously updated, and should be reviewed regularly when planning for installation of a new server. They contain the latest information about maintenance.

PSP buckets contain installation information, hardware and software service levels, service recommendations, and cross-product dependencies.

Middleware

Middleware offerings for the z10 EC 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 9 for Linux
- DB2 Connect™

Operations

The Tivoli brand has a large product set that includes:

- Tivoli Information Management for z/OS
- Tivoli Workload Scheduler
- Tivoli OMEGAMON® XE
- Tivoli System Automation

Security

A secure System z environment can be implemented at various levels using the following products:

- Security Server feature of z/OS (includes Resource Access Control Facility (RACF®) and LDAP server)
- Tivoli Access Manager
- Tivoli Federated Identity Manager
- z/OS Communications Server and Policy Agent (for policy-based network security)

Application development and languages

Many languages are available for the z10 EC environments. Since the Linux environment is similar to other servers we focus on the z/OS environment.

In additional 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 with testing and deployment on z/OS. AD tools, many of which have components based on the Eclipse platform, include:

- Rational® Application Developer for WebSphere
- Rational Developer for System z
- WebSphere developer for System z

- Rational Rose® product line
- Rational Software Architect and Software Modeler

The following Web site is organized by category and has an extensive set of links to information about software for System z:

http://www-306.ibm.com/software/sw-bycategory/systemz

4.2 Support by operating system

In this section we list the support of new System z10 EC functions by the current operating systems. See the companion manual *z10 ECIBM System z10 Enterprise Class Technical Guide*, SG24-7516, for a detailed description of z10 EC and its features. For an in-depth description of all I/O features refer to the *IBM System z Connectivity Handbook*, SG24-5444.

z/OS

z/OS Version 1 Release 7 is the earliest service release supporting the z10 EC. Service support for z/OS Version 1 Release 6 ended on September 30, 2007. Also note that z/OS.e is not supported on z10 EC and that the last release of z/OS.e was z/OS.e Version 1 Release 8. Table 4-2 summarizes the z10 EC function' support requirements for the currently supported z/OS releases.

Table 4-2 z/OS support summary

Function	z/OS V1 R9	z/OS V1 R8	z/OS V1 R7
z10 EC	Supported	Supported	Supported
Greater than 54 PUs single system image	Supported	Not supported	Not supported
Large memory > 128 GB	Supported	Supported	Not supported
Large page support	Supported	Not supported	Not supported
Hardware decimal floating point	Supported[a]	Supported[a]	Supported[a]
Enhanced CPACF	Supported[b]	Supported[b]	Supported[b]
Capacity Provisioning Manager	Supported	Not supported	Not supported
HiperDispatch	Supported	Supported	Supported[c]
HiperSockets multiple write facility	Supported[d]	Not supported	Not supported
OSA Express3 10 Gigabit Ethernet LR CHPID type OSD	Supported	Not supported	Not supported
Coupling over InfiniBand CHPID type CIB	Supported	Supported	Supported
CFLEVEL 15	Supported	Supported	Supported

a. The level of decimal floating-point exploitation varies with the z/OS release and PTF level.
b. FMIDs shipped in a Web deliverable.
c. Requires Web deliverable support for zIIP.
d. With PTFs.

z/VM

At general availability, z/VM V5 R2 and z/VM V5 R3 provide compatibility support only. Table 4-3 lists the z10 EC functions currently supported z/VM releases.

Table 4-3 z/VM support summary

Function	z/VM V5 R3	z/VM V5 R2
z10 EC	Supported[a]	Supported[a]
Greater than 54 PUs for single system image	Not supported[b]	Not supported[b]
Large memory > 128 GB	Supported[c]	Not supported[c]
Large page support	Not supported[d]	Not supported[d]
Hardware decimal floating point	Supported[e]	Not supported[d]
Enhanced CPACF	Supported[e]	Supported[e]
Execute relative guest exploitation	Supported[e]	Supported[e]
Capacity provisioning	Not supported[d]	Not supported[d]
HiperDispatch	Not supported[d]	Not supported[d]
Restore subchannel facility	Supported	Supported
HiperSockets multiple write facility	Not supported	Not supported
OSA Express3 10 Gigabit Ethernet LR CHPID type OSD	Supported	Supported
Dynamic I/O support for InfiniBand CHPIDs	Supported	Not supported
CFLEVEL 15	Supported[e]	Supported[e]

a. Compatibility support only: z/VM and guests will be supported at the System z9 functionality level. No exploitation of new hardware unless otherwise noted.
b. A maximum of 32 PUs are supported by z/VM 5.3 per system, while z/VM 5.2 supports up to 24 per system image. Guests can be configured with up to 64 virtual PUs.
c. 256 GB of central memory are supported by z/VM V5 R3 and 128 GB of central memory are supported by z/VM V5 R2. z/VM V5 R3 is designed to support more than 1 TB of virtual memory in use for guests.
d. Not available to guests.
e. Supported for guest use only.

Notes: We recommend that the capacity of z/VM logical partitions as well as of any guests, in terms of the number of IFLs and CPs, real or virtual, be adjusted in face of the PU capacity of the z10 EC.

The FCP LUN Access Control feature is not available on z10 EC. It was previewed on z990 and z890 in October 7, 2004. It was replaced by N_Port ID Virtualization (NPIV), which was announced July 27, 2005. At the time IBM stated that “NPIV is the industry-standard solution which supersedes the previewed FCP LUN Access Control.” NPIV is applicable to all FICON features on z10 EC.

z/VSE

Table 4-4 lists z10 EC support requirements for the currently supported z/VSE releases.

Table 4-4 z/VSE support summary

Function	z/VSE V4 R1	z/VSE V3 R1
z10 EC	Supported[a]	Supported[b]
Enhanced CPACF	Supported	Not supported
OSA Express3 10 Gigabit Ethernet LR CHPID type OSD	Supported	Supported

a. z/VSE V4 is designed to exploit z/Architecture, specifically 64-bit real-memory addressing, but does not support 64-bit virtual memory addressing.

b. z/VSE V3 is 31-bit mode only. It does not implement z/Architecture, and specifically does not implement 64-bit mode capabilities. z/VSE V3 is designed to exploit selected features of z10 EC and other System z servers.

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 (SLES 9, SLES 10, RHEL 4, RHEL 5) require z10 toleration support.

Table 4-5 lists the most recent service levels of the current SUSE and Red Hat releases at the time of writing.

Table 4-5 Current Linux on System z distributions as of May 2008, by z/Architecture mode

Linux distribution	ESA/390 (31-bit mode)	z/Architecture (64-bit mode)
Novell SUSE SLES 9 SP4	Yes	Yes
Novell SUSE SLES 10 SP1	No	Yes
Red Hat RHEL 4.6	Yes	Yes
Red Hat RHEL 5.1	No	Yes

For information about support-availability of Linux on System z distributions, see:

- Novell SUSE

  http://support.novell.com/lifecycle/lcSearchResults.jsp?st=Linux+Enterprise+Server&x=32&y=11&sl=-1&sg=-1&pid=1000

- Red Hat

  http://www.redhat.com/security/updates/errata/

IBM is working with its Linux distribution partners so that exploitation of selected z10 EC functions will be provided in future Linux on System z distribution releases.

We recommend that:

- SUSE SLES 10 or Red Hat RHEL 5 be used in any new projects for the z10 EC
- Any Linux distributions be updated to their latest service level before migration to z10 EC
- The capacity of any z/VM and Linux logical partitions, as well as any z/VM guests, in terms of the number of IFLs and CPs, real or virtual, be adjusted in face of the PU capacity of the z10 EC

TPF and z/TPF

Table 4-6 lists the z10 EC function' support requirements for the currently supported TPF and z/TPF releases.

Table 4-6 TPF and z/TPF support summary

Function	z/TPF V1 R1	TPF V4 R1
z10 EC	Supported	Supported
Greater than 54 PUs for single system image	Supported	Not supported
Large memory > 128 GB (4TB)	Supported	Not supported
Enhanced CPACF	Not supported	Not supported
HiperDispatch	Not supported	Not supported
OSA Express3 10 Gigabit Ethernet LR CHPID type OSD	Supported	Supported
Coupling over InfiniBand CHPID type CIB	Supported[a]	Supported[a]
CFLEVEL 15	Supported	Supported

a. Compatibility support

4.3 Support for selected functions

In this section we review the operating system support of a very small set of functions introduced by the z10 EC. They were selected due to their high importance.

Single system image

A single system image can control several processing units (PUs) such as CPs, zIIPs, zAAPs, or IFLs, as appropriate. Table 4-7 shows the maximum number of PUs supported for each operating system image.

Table 4-7 Single system image software support

Operating system	Maximum number of CPs+zIIPs+zAAPs[a] or IFLs per system image
z/OS V1 R9	64
z/OS V1 R8	32
z/OS V1 R7	32
z/VM V5 R3	32[b]
z/VM V5 R2	24[b]

Operating system	Maximum number of CPs+zIIPs+zAAPs[a] or IFLs per system image
z/VSE V4	z/VSE Turbo Dispatcher can exploit up to 4 CPs and tolerates up to 10-way LPs
z/VSE V3 R1	z/VSE Turbo Dispatcher can exploit up to 4 CPs and tolerates up to 10-way LPs
Linux on System z	Novell SUSE SLES 9: 64 CPs or IFLs Novell SUSE SLES 10: 64 CPs or IFLs Red Hat RHEL 4: 8 CPs or IFLs Red Hat RHEL 5: 64 CPs or IFLs
z/TPF V1 R1	64 CPs
TPF V4 R1	16 CPs

a. The number of purchased zAAPs and the number of purchased zIIPs cannot each exceed the number of purchased CPs. A logical partition can be defined with any number of the available zAAPs and zIIPs. The total refers to the sum of these PU characterizations.
b. z/VM guests can be configured with up to 64 virtual PUs.

Note: IBM intends to further enhance z/VM in a future release to exploit the System z10 EC support for a new logical partition (LPAR) mode z/VM, exclusively for running z/VM. This new LPAR mode allows z/VM to utilize a wider variety of specialty processors in a single LPAR. For instance, in a z/VM mode LPAR, z/VM can manage Linux on System z guests running on IFL processors while also managing z/VSE and z/OS on central processors (CPs), and can offload z/OS system software overhead, such as DB2 workloads, on IBM System z9 and IBM System z10 Integrated Information Processors (zIIPs) and IBM System z9 and IBM System z10 Application Assist Processors (zAAPs).

Large memory

Table 4-8 lists the maximum amount of main storage supported by current operating systems. Expanded storage, although part of the z/Architecture, is currently exploited only by z/VM. On z10 EC a maximum of 1 TB of main storage can be defined to a logical partition.

Table 4-8 Maximum memory supported by operating system

Operating system	Maximum supported main storage
z/OS	z/OS V1 R9 supports 4 TB and up to 1.5 TB per server z/OS V1 R8 supports 4 TB and up to 1.5 TB per server z/OS V1 R7 supports 128 GB
z/VM	z/VM V5 R3 supports 256 GB z/VM V5 R2 supports 128 GB
z/VSE	z/VSE V4 R1 supports 8 GB z/VSE V3 R1 runs in ESA/390 mode and supports 2 GB
Linux on System z (64-bit)	Novell SUSE SLES 9 supports 4 TB Novell SUSE SLES 10 supports 4 TB Red Hat RHEL 4 supports 64 GB Red Hat RHEL 5 supports 64 GB
TPF and z/TPF	z/TPF supports 4 TB TPF runs in ESA/390 mode and supports 2 GB

Hardware decimal floating point

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 some key software vendors such as Microsoft and SAP.

Hardware decimal floating point support was introduced with the z9 EC. The z10 EC, however, has a new decimal floating point accelerator feature, described in the *z10 ECIBM System z10 Enterprise Class Technical Guide*, SG24-7516. Hence, we list in Example 4-1 the operating system support for decimal floating point. See also “Decimal floating point (z/OS XL C/C++ considerations)” on page 77.

Example 4-1 Minimum support requirements for decimal floating point

Operating system	Support requirements
z/OS	z/OS V1 R9 - support includes XL, C/C++, HLASM, Language Environment®, DBX, and CDA RTLE z/OS V1 R8 - support includes HL ASM, Language Environment, DBX, and CDA RTLE z/OS V1 R7 - support of the High Level Assembler (HLASM) only
z/VM	z/VM V5 R3 - supported for guest use

MIDAW facility

The Modified Indirect Data Address Word (MIDAW) facility is a system architecture and software exploitation designed to improve FICON performance. This facility is only available on System z9 and System z10 EC servers and is exploited by the media manager in z/OS. Table 4-9 lists the minimum support requirements for the MIDAW facility.

Table 4-9 Minimum support requirements for MIDAW

Operating system	Support requirements
z/OS	z/OS V1 R7
z/VM	z/VM V5 R3[a]

a. Supported for guest exploitation

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

- MIDAW can significantly improve FICON performance for extended format (EF) data sets. Non-extended data sets can also benefit from MIDAW.
- MIDAW can improve channel utilization and can significantly improve I/O response time. This reduces FICON channel connect time, director ports, and control unit overhead.

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

For additional information relating to FICON and MIDAW, consult the following Web site for material on FICON channel performance:

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

There is also an excellent paper called “How does the MIDAW facility improve the performance of FICON channels using DB2 and other workloads?”, available at:

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

See also the manual *IBM TotalStorage DS8000 Series: Performance Monitoring and Tuning,* SG24-7146.

4.3.1 Cryptographic support

z10 EC provides two major groups of cryptographic functions. From an application program point of view there are both synchronous and asynchronous cryptographic functions:

- Synchronous cryptographic functions are provided by the CP Assist for Cryptographic Function (CPACF).
- Asynchronous cryptographic functions are provided by the Cryto Express2 feature.

The minimum software support levels are listed in the following sections. The latest PSP buckets should be obtained and reviewed to ensure that the latest support levels are known and included as part of the implementation plan.

CPACF

In z10 EC, the CP Assist for Cryptographic Function was extended to support the full standard for AES (symmetric encryption) and SHA (hashing). For a full description refer to the *z10 ECIBM System z10 Enterprise Class Technical Guide*, SG24-7516. Support for this function is provided through Web-delivered code. Table 4-10 lists the support requirements for CPACF enhancements.

Table 4-10 Minimum support requirements for CPACF enhancements

Operating system	Support requirements
z/OS[a]	z/OS V1 R7 - function varies by release
z/VM	z/VM V5 R2 - supported for guest use
z/VSE	z/VSE V4 R1
Linux on System z	Novell SUSE SLES 9 and SLES10 Red Hat RHEL 4.3 and RHEL 5.0

a. CPACF is also exploited by several IBM software product offerings for z/OS, such as IBM WebSphere Application Server for z/OS.

Crypto Express2

Table 4-11 lists the minimum software requirements for the Crypto Express2 feature when configured as a coprocessor or an accelerator, and support for the base and enhanced functions.

Table 4-11 Crypto Express2 support on z10 EC

Operating system	Support
z/OS V1 R9	Included in base product
z/OS V1 R8	Included in base product
z/OS V1 R7	Web deliverable
z/VM V5 R3	Any guest that can exploit the feature
z/VM V5 R2	Any guest that can exploit the feature; PTFs
z/VSE V4 R1	Included in the base product

Operating system	Support
z/VSE V3 R1	PTFs
Linux on System z	Novell SUSE SLES 9 SP3 and SLES 10 Red Hat RHEL 4.4 and RHEL 5.1
TPV V4 R1	Not supported
z/TPF V1 R1	PTFs (accelerator mode only)

Web deliverables

For z/OS, see the z/OS Web site at:

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

4.3.2 z/OS ICSF

Integrated Cryptographic Service Facility (ICSF) is a base component of z/OS, and is designed to transparently use the available cryptographic functions, whether CPACF or Crypto Express2, to balance the workload and help address the bandwidth requirements of your applications.

Specify support is available as a Web download for z/OS V1 R7 (FMID HCR7730), and for z/OS V1 R8 (FMID HCR7740) in support of the PCI-X cryptographic coprocessor and accelerator functions as well as the CPACF AES, PRNG, and SHA support. z/OS V1 R9 has the support (FMID HCR7740) integrated in the base, so no download is necessary.

For support of the new SHA-384 and SHA-512 function on z/OS V1 R7, V1 R8, and V1 R9, download and installation of FMID HCR7750 is required.

ICSF migration considerations

With the new version of Web-delivered ICSF code, the following points should be considered:

- Increased size of the PKDS file. This is required in order to allow 4096-bit RSA keys to be stored. If you are using the PKDS for asymmetric keys you must copy your PKDS to a larger VSAM data set before using the new version of ICSF. The ICSF options file must be updated with the name of the new data set. ICSF can then be started.

  A toleration PTF must be installed on any system that is sharing the PKDS with a system running HCR7750 ICSF. The PTF will allow the PKDS to be larger and prevent any service from accessing 4096-bit keys stored in a HCR7750 PKDS.
- Reduced support for retained private keys. Applications that make use of the retained private key capability for key management will no longer be able to store the private key in the crypto coprocessor card. The applications will continue to be able to list the retained keys and to delete them from the crypto coprocessor cards.

4.4 z/OS migration considerations

z10 EC base processor support is required in z/OS. With that exception, software changes do not require the new z10 EC functions and, equally, the new functions do not require functional software. The approach has been to, where applicable, automatically decide to enable or disable a function based on, respectively, the presence or absence of the required hardware and software.

General recommendations

The new System z10 EC introduces the latest System z technology. Although support is provided by z/OS starting with z/OS V1 R7, exploitation of z10 EC is dependent on the z/OS release. z/OS.e is *not* supported on z10 EC.

In general, we recommend that you:

- Do not migrate software releases and hardware at the same time.
- Keep members of the sysplex at the same software level other than during brief migration periods.
- Review z10 EC restrictions and migration considerations prior to creating an upgrade plan.

HCD

When using HCD on z/OS V1 R6 to create a definition for z10 EC, *all* subchannel sets must be defined or the VALIDATE will fail. On z/OS V1 R7, HCD or HCM will assist in the definitions.

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.

Large page support

The large page support function will not be enabled without the required software support. Without the large page support, page frames will be allocated at the current 4 K size.

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

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

This parameter *cannot* be changed dynamically.

HiperDispatch

There is a new HIPERDISPATCH=YES/NO parameter in the IEAOPTxx member of SYS1.PARMLIB and on the SET OPT=xx command to control whether HiperDispatch is enabled or disabled for a z/OS image. It can be changed dynamically (without an IPL or any outage).

To effectively exploit HiperDispatch, adjustment of defined WLM goals and policies may be required. We recommend that you review WLM policies and goals and update them as necessary. Keep in mind the need to turn off and on HiperDispatch while adjusting WLM goals.

A health check is provided to verify whether HiperDispatch is enabled on a z10 EC system.

Capacity provisioning

Installation of the capacity provision function on z/OS requires:

- Setting up and customizing z/OS RMF™, including the Distributed Data Server (DDS)
- Setting up the z/OS CIM Server (a z/OS base element with z/OS V1 R9)
- Performing capacity provisioning customization as described in the *z/OS MVS Capacity Provisioning User's Guide*, SA33-8299

Exploitation of the capacity provisioning function requires:

- TCP/IP connectivity to observed systems
- TCP/IP connectivity from the observing system to the HMC of observed systems
- RMF Distributed Data Server must be active
- CIM Server must be active
- Security and CIM customization
- Capacity Provisioning Manager customization

In addition, the Capacity Provisioning Control Center has to be downloaded from the host and installed on a PC workstation. This application is only used to define policies. It is not required to manage operations.

Customization of the capacity provisioning function is required on the operating system that will observe other z/OS systems in one or multiple sysplexes. For a description of capacity provisioning domain refer to the *z/OS MVS Capacity Provisioning User's Guide*, SA33-8299. There is also an IBM Redbooks publication *IBM System z10 Enterprise Class Capacity on Demand,* SG24-7504, that covers capacity provisioning in more detail.

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

There are two new options for the C/C++ compiler: ARCHITECTURE and TUNE. They require z/OS V1 R9.

The ARCHITECTURE C/C++ compiler option selects the minimum level of machine architecture on which your program will run. Note that certain features provided by the compiler require a minimum architecture level. ARCH(8) exploits instructions available on the z10 EC.

The TUNE compiler option allows optimization of the application for a specific machine architecture within the constraints imposed by the ARCHITECTURE option. The TUNE level must not be lower than the setting in the ARCHITECTURE option.

For more information about the ARCHITECTURE and TUNE compiler options refer to the *z/OS V1R9.0 XL C/C++ User's Guide*, SC09-4767.

Note: A C/C++ program compiled with the ARCH() or TUNE() options can only run on z10 EC servers, or an operation exception will result. This is a consideration for programs that may need to run on different level servers during development, test, production, and fallback or DR.

4.5 Coupling Facility and CFCC considerations

Coupling Facility connectivity to a z10 EC server is supported on the z9 EC, z9 BC, z990, z890, or another z10 EC server. The logical partition running the Coupling Facility Control Code (CFCC) can reside on any of the supported servers previously listed.

Because coupling link connectivity to z800 and z900 is *not* supported, this may impact the introduction of z10 EC into existing installations, and require additional planning. For more information refer to the *IBM System z10 Enterprise Class Technical Guide*, SG24-7516System z servers is CFCC level 15. To support migration from one CFCC level to the next, different levels of CFCC can be run concurrently as long as the Coupling Facility logical partitions are running on different servers (CF logical partitions running on the same server share the same CFCC level).

For additional details on CFCC code levels, see the Parallel Sysplex Web site at:

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

4.6 IOCP

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 z10 EC is V2 R1 L0 (IOCP 2.1.0). See the *Input/Output Configuration Program User's Guide,* SB10-7037, for details.

4.7 ICKDSF

ICKDSF Release 17 is required on all systems that share disk subsystems with a z10 EC server.

ICKDSF supports a modified format of the CPU information field, which contains a two-digit logical partition identifier. ICKDSF uses the CPU information field instead of CCW reserve/release for concurrent media maintenance. It prevents multiple systems from running ICKDSF on the same volume, and at the same time allows user applications to run while ICKDSF is processing. In order to prevent any possible data corruption, ICKDSF must be able to determine all sharing systems that may potentially run ICKDSF. Therefore, this support is required for z10 EC.

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.

4.8 Software licensing considerations

The System z10 EC mainframe software portfolio includes operating system software (that is, z/OS, z/VM, z/VSE, and z/TPF) and middleware that runs on these operating systems. It also includes middleware for Linux on System z environments.

Two major metrics for software licensing are available from IBM, depending on the software product: Monthly Licence Charges (MLC) or International Program License Agreement (IPLA).

Monthly Licence Charges pricing metrics have a recurring charge that applies each month. In addition to the right to use the product, the charge includes access to IBM product support during the support period. MLC metrics, in turn, include a variety of offerings. Those applicable to the System z10 EC are:

- Workload License Charges (WLC)
- System z New Application License Charges (zNALC)
- Parallel Sysplex License Charges (PSLC)
- Midrange Workload License Charges (MWLC)

International Program License Agreement (IPLA) metrics have a single, up-front charge for an entitlement to use the product. Optionally, a separate annual charge called *subscription and support* entitles customers to receive future releases and versions at no additional charge and also access to IBM product support during the support period.

For details, consult the *IBM System z Software Pricing Reference Guide*, which is available at:

http://www-03.ibm.com/servers/eserver/zseries/library/refguides/sw_pricing.html

4.8.1 Workload License Charges

Workload License Charges requires z/OS or z/TPF operating systems in 64-bit mode. Any mix of z/OS, z/VM, Linux, z/VSE, TPF, and z/TPF images is allowed.

There are two WLC license types:

- Flat WLC (FWLC): Software products licensed under FWLC are charged at the same flat rate, no matter what is the capacity of (MSUs) the server.
- Variable WLC (VWLC) applies to products such as z/OS, DB2, IMS, CICS, MQSeries®, and Lotus® Domino®. VWLC software products can be charged in two different ways:
  - Full-capacity: The server's total number of MSUs is used for charging. Full-capacity is applicable when the server is not eligible for Sub-capacity.
  - Sub-capacity: Software charges are based on the logical partition's utilization where the product is running.

WLC Sub-capacity allows software charges based on logical partition utilizations instead of the server's total number of MSUs. Sub-capacity removes the dependency between software charges and server (hardware) installed capacity.

Sub-capacity is based on the logical partition's rolling four-hour average utilization. It is *not* based on the utilization of each product,[1] but on the utilization of the logical partition or partitions where it runs. The VWLC licensed products running on a logical partition will be charged by the maximum value of this partition's rolling four-hour average utilization within a month.

The logical partition's rolling four-hour average utilization can be limited by a *defined capacity* definition on the partition's image profiles. This activates the *soft capping* function of PR/SM, limiting four-hour average partition utilizations above the defined capacity value. Soft capping controls the maximum rolling four-hour average utilization (the *last* four-hour average value at every five-minute interval), but does *not* control the maximum *instantaneous* partition utilization.

There is also a LPAR group capacity limit, which allows you to set soft capping of PR/SM for a group of logical partitions running z/OS.

[1] With the exception of products licensed using the SALC pricing metric

Even using the soft capping option, the partition's utilization can reach up to its maximum share based on the number of logical processors and weights in the image profile. Only the rolling four-hour average utilization is tracked, allowing utilization peaks above the defined capacity value.

As with the Parallel Sysplex License Charges (PSLC) software license charge type, the aggregation of servers' capacities within the same Parallel Sysplex is also possible in WLC, following the same prerequisites.

Entry Workload License Charges (EWLC) is not offered for IBM System z10 Enterprise Class.

For further information about WLC and details on how to combine logical partitions utilization, see the publication *z/OS Planning for Workload License Charges*, SA22-7506.

4.8.2 System z New Application License Charges

System z New Application License Charges (zNALC) 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 at:

http://www-03.ibm.com/servers/eserver/zseries/swprice/znalc.html

4.8.3 Select Application License Charges (SALC)

Select Application License Charges applies to WebSphere MQ for System z only. It is designed to allow a WLC customer to licence MQ under product utilization rather than the sub-capacity pricing provided under WLC.

WebSphere MQ is typically a low-usage product that runs pervasively throughout the environment. Clients who run WebSphere MQ at a very low usage may benefit from SALC. Alternatively, you can still choose to license WebSphere MQ under WLC.

A reporting function, which IBM provides in the operating system IBM Software Usage Report Program, is used to calculate the daily MSU number. The rules to determine the billable SALC MSUs for WebSphere MQ use the following algorithm:

1. Determine the highest daily usage of a program[2] family, which is the highest of 24 hourly measurements recorded each day.
2. Determine the monthly usage of a program[2] 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 MWLC Web page at:

http://www.ibm.com/servers/eserver/zseries/swprice/other.html

[2] Program refers to all active versions of MQ

4.8.4 Midrange Workload Licence Charges

Midrange Workload Licence Charges (MWLC) applies to z/VSE V4 when running on System z servers (z9 EC, z9 BC, and z10 EC). The exception is the IBM System z9 BC 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 V5R2 or later is required.

To report utilization, the sub-capacity report tool is used. One SCRT report per server is required.

For additional information see the MWLC Web page at:

http://www.ibm.com/servers/eserver/zseries/swprice/mwlc.html

4.8.5 System z international licensing agreement

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 4 and 5.

Generally, there are three pricing metrics that apply to IPLA products for System z:

VU Value unit pricing, which applies to the IPLA products that run on z/OS. Value unit pricing is typically based upon 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 and OS/390®

EBVU Engine-based value unit pricing enables a lower cost of incremental growth with additional engine-based licenses purchased. Examples of eligible products include z/VM V5 and certain z/VM middleware, which are priced based on the number of engines.

PVU Processor value units. Here the number of engines is converted into processor value units under the Passport Advantage® terms and conditions. Most Linux middleware is also priced based on the number of engines.

For additional information see the System z IPLA Web page at:

http://www.ibm.com/servers/eserver/zseries/swprice/zipla/

4.9 References

For current planning information for each operating system, refer to the support Web page for each operating system:

- z/OS

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

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

  http://www.ibm.com/software/htp/tpf/pages/maint.htm
- z/VSE

  http://www.ibm.com/servers/eserver/zseries/zvse/
- Linux on System z

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

A

Frequently asked questions

Q: What is System z?
A: IBM System z is a brand name for IBM mainframe computers. It is the line of computers that started in 1964 with S/360™ and evolved over the decades. It still preserves backward compatibility with previous systems while bringing new features and technologies.

Q: Does z10 EC use the Power processor?
A: The processor in z10 EC was developed in cooperation with the Power6 team. Both processors share a lot of components: IBM 65nm SOI technology, design building blocks, large portions of execution units, floating point units, EI3 interface technology, core pipeline design style, high frequency, low latency, mostly in order instruction execution. But they have different personalities, different cache hierarchy, SMP topology and protocol, and different chip organization. The z10 EC implements the Complex Instruction Set (CISC) of z/Architecture and POWER™ implements a RISC architecture. They are siblings, not identical twins.

Q: Processor clock speed has more than doubled. Why does the capacity in MIPS not show double growth?
A: The published processor capacity indexes (not MIPS) are comparisons of the capacity of a processor running a standard workload mix with a defined processor taken as the base. Clock speed is and has always been only one of many factors influencing the performance characteristics of a microprocessor in a given operating environment. Penalties to be paid by increasing clock speeds are partially offset by architecture, hardware, and operating system design improvements. The combination of these has led to the performance figures published for specific workloads. Variability of workload characteristics will show similar variability in performance characteristics.

Q: It is well known that there are many problems associated with high-frequency processors. Why do you still use them?
A: The classic scaling of transistor switching speeds as defined by Moore's law has slowed down because at ultra-high frequencies power leakage must be overcome. In the coming years slower clock speed increases must be expected while the industry as a whole is searching for and implementing techniques based on complex chip, processor, and system designs, while software development must concentrate on introducing more parallelism in their designs.

Q: What is new about PU sparing?
A: Sparing works the same way as in System z9 servers. There are two spare PUs per server, regardless of number of books. Sparing of a single core is now possible.

Q: What is HiperDispatch?
A: HiperDispatch is a name for several improvements in interaction between PR/SM and z/OS. It is a mechanism that recognizes the physical processor where the work was started and then dispatches subsequent work to the same physical processor. This helps to reduce the movement of cache and data and improves overall system throughput. HiperDispatch is available only with z10 EC PR/SM and z/OS functions.

Q: What are the consequences when I switch HiperDispatch off in z/OS?
A: For some workloads that are dispatched on many processors it may decrease the performance because processor caches have to be reloaded more often.

Q: Can I run Linux on z10 EC?
A: Yes. Linux on System z distributions are available from Novell SUSE and Red Hat. IBM is working with these Linux distribution partners to provide Linux with appropriate functionality on all of its hardware platforms.

Q: Can I run AIX® on z10 EC?
A: No, there is no AIX version available that would run on z10 EC. AIX is designed for IBM System p and System i™ servers.

Q: Can I run MS Windows® on z10 EC?
A: No, there is no Windows operating system available for z10 EC.

Q: Can I run Sun Solaris™ on z10 EC?
A: No. In a joint Sun and IBM press call in August 2007 it was mentioned that IBM would begin to collaborate with the Open Source community to investigate the feasibility of bringing OpenSolaris™ to System z. This is a project to understand the technical feasibility of technology, environments, and so on. There is no Solaris on z10 EC offering.

Q: Can any z990 or z9 EC model be upgraded to a z10 EC?
A: Yes.

Q: Can I do all upgrades concurrently?
A: It depends on whether the upgrade requires a hardware change. For example, if memory is already installed in the book, enabling it with Licensed Internal Code (LIC) is a concurrent action. But if the memory has to be installed physically first, the action may not be concurrent. It also depends on the number of available books and their configuration.

Q: What is the difference between concurrent and nondisruptive upgrade?
A: In general, concurrency addresses the continuity of operations of just the hardware part of an upgrade, for instance, whether a server (as a box) is required to be switched off during the upgrade. Disruptive versus nondisruptive refers to whether the running software or operating system has to be restarted for the upgrade to take an effect. Thus, even concurrent upgrades can be disruptive to those operating systems or programs that do not support them while at the same time being nondisruptive to others.

Q: What is the benefit of 1 MB page size? Should I switch from 4 k to 1 MB?
A: Large pages cause the Translation Lookaside Buffer (TLB) to better represent the working set and suffer less misuses by allowing a single TLB entry to cover more address translations. Exploiters of large pages are better represented in the TLB and are expected to perform better. Long-running memory access-intensive applications especially benefit. Short processes with small working sets see none or little improvement. The use of large pages

must be decided based on knowledge obtained from measurement of memory usage and page translation overhead for a specific workload.

Under z/OS, large pages are treated as fixed pages and are never paged out. They are only available for 64-bit virtual private storage such as virtual memory located above 2 GB.

IBM is working with its Linux distribution partners to include support in future Linux on System z distribution releases, where the benefits will be similar as described above.

Q: What is CPE?
A: Capacity for Planned Events is a Capacity on Demand offering. It delivers a replacement capacity for planned event-like data center planned outage, server move, and so on. When it is activated, all available capacity (both PUs and memory) can be set online and used for up to three days. Upon activation, it does not incur any additional charges from IBM.

Q: Can I use NTP instead of STP?
A: No. However, z10 EC server can be synchronized against an external time source through NTP.

Q: Mainframes are known for EBCDIC code pages. How can Linux, which is ASCII, run on z10 EC?
A: There is no requirement for EBCDIC defined in the z/Architecture. For example, z/OS supports EBCDIC, ASCII, and Unicode. Linux on System z uses ASCII and Unicode.

Q: How many new machine instructions were added to z10 EC?
A: More than 50 instructions were added.

Q: What are they used for?
A: The instructions added to z/Architecture are the result of active collaboration between hardware and software designers, specifically with the compiler teams. Hardware and software are being co-optimized, while maintaining full upward compatibility. Most of these instructions are specifically targeted to use by compilers to improve the efficiency of generated code. Examples are combining two simple functions into a single instruction or reducing the number of active general registers needed for a sequence. Another group of new instructions improves software/hardware synergy by enabling software to give hints to the hardware on caching of specific blocks of memory and by communicating the effective SMP topology so that processes can be kept close to their cached data. The remaining instructions provide minor extensions to existing functions.

Q: When using a sub-capacity model, is it possible to mix different CP feature codes?
A: No, all CP feature codes must be the same. In other words, all CPs must be on the same capacity level.

Q: Are there subcapacity versions of specialty engines available?
A: No, specialty engines are always on maximum capacity.

Q: Is it possible to have more than one book although purchased processors would fit into fewer books?
A: Yes, there is no restriction that prevents this.

Q: Are there any new specialty engines added to z10 EC?
A: No.

Q: I cannot have more zAAPs or zIIPs than CPs. Can I have more zAAPs and zIIPS than CPs?
A: Yes. For each CP you can have one zAAP, one zIIP, or one zAAP and one zIIP.

Q: Can I mix dedicated and shared processors in one logical partition?
A: No. Dedicated and shared processors cannot be mixed in one logical partition, regardless of their type.

Q: Is it possible to order an IFL-only server?
A: Yes.

Q: Why can I not use zAAP for Java workload in Linux?
A: zAAP is designed to offload Java workload from CPs in z/OS to keep MSU values lower while providing more processing capacity for Java workload. Because there is no MSU measurement for Linux, it makes no sense to use zAAP there.

Q: What is the Capacity Provisioning Manager?
A: The Capacity Provisioning Manager is a software delivered with the z/OS BCP feature. It is a component that allows you to switch OOCoD records on and off automatically according to defined policies. It monitors RMF metrics to decide when to activate OOCoD and when to deactivate it.

Q: What is the Watts per square foot ratio for z10 EC?
A: It is approximately 1 163. For a comparison, it is about 709 for System z9 EC.

Q: How much memory should I plan for HSA?
A: Each z10 EC server by default contains a 16 GB memory that is fixed and used for HSA. This memory is *not* part of the memory purchased by customer. HSA never occupies additional memory outside of its 16 GB.

Q: Can I run out of HSA space?
A: No. The HSA size is big enough to hold all possible definitions.

Q: What are the increments for ordering the memory?
A: Up to 256 GB, the increment is 16 GB. From 256 GB to 512 GB it is 32 GB. From 512 GB to 944 GB it is 48 GB. From 944 GB it is 64 GB.

Q: Can I connect z10 EC to a SAN?
A: Yes. FICON cards support both FICON and FC protocol. If the operating system supports FC protocol, it can participate in the same environment as any other operating system supporting FC protocol.

Q: Can I use the 2-port FICON Express4 adapter with a z10 EC server?
A: No. It is unique to the z9 BC.

Q: Can I carry forward my older OSA-Express adapters?
A: In general, yes. Check Table B-1 on page 89 for details.

Q: There are several FICON choices. Is there a preferred version?
A: The FICON Express4 4-port 4km LW version might be considered the standard version.

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

Q: Why was InfiniBand implemented on the z10 and what on the z10 takes advantage of it?
A: The goal of System z was to introduce an industry-standard, high-speed host bus physical interface to replace the self-timed interconnect (STI) proprietary host bus interface. At the

same time, System z was looking for an industry-standard protocol to provide enhanced performance for Parallel Sysplex coupling links. The InfiniBand host bus physical interface supports 12x Double Data Rate (12x IB-DDR) with a link speed of up to 6 GBps when attached to a z10 EC and supports 12x Single Data Rate (12x IB-SDR) with a link speed of up to 3 Gbps when a z10 EC is attached to a z9 EC or z9 BC.

Q: How can I use InfiniBand on z10 EC?
A: The z10 EC takes advantage of InfiniBand to implement:

- A new I/O bus that includes the InfiniBand Double Data Rate (IB-DDR) infrastructure and replaces the self-timed interconnect features found in prior System z servers.
- Parallel Sysplex coupling over InfiniBand (PSIFB). This link has a bandwidth of up to 6 GBps between two z10 EC servers and of up to 3 GBps between z10 EC and System z9 servers.

Q: Why is the InfiniBand connection speed to z9 only up to 3 GB per second while it is up to 6 GB per second to z10 EC?
A: The System z9 internal bus cannot handle more than 3 GBps. The connection speed automatically adjusts to the slower server.

Q: Why can I not use a non-dedicated CF on a System z9 in a PSIFB environment?
A: Such a configuration is not supported yet. IBM has announced a Statement of Direction to support PSIFB links on general-purpose System z9 EC and BC servers. When that support is made available, System z9 servers with general-purpose processors, IFLs, zAAPs, or zIIPs will be able to utilize 12x IB-SDR links. Currently, 12x IB-SDR links are available on System z9 servers that have only ICF processors.

Q: What CFCC level is supplied with the z10 EC server?
A: CFCC 15 is the initial level in z10 EC.

B

Channel options

Table B-1 lists the attributes of the channel options supported on z10 EC servers with the required connector and cable types, the maximum unrepeated distance, and the bit rate.

The z10 EC Model E12 supports up to 64 I/O slots (960 CHPIDs in three I/O cages), while models E26, E40, E56, and E64 support up to 84 I/O slots (1024 CHPIDs in three I/O cages). At least one ESCON, FICON, ICB, or ISC feature is required.

Table B-1 System z10 EC channel feature support

Channel feature	Feature codes	Bit rate	Connector	Cable type	Maximum unrepeated distance[a]
Cryptography					
Crypto Express2	0863	N/A	N/A	N/A	N/A
Enterprise Systems CONnection (ESCON)					
16-port ESCON	2323	200 Mbps	MT-RJ	MM 62.5 µm	3 km (800)
FIber Connection (FICON)					
FICON Express[b] LX	2319	2 Gbps	LC Duplex	SM 9 µm	10 km/20 km
FICON Express[b] SX	2320	2 Gbps	LC Duplex	MM 62.5 µm MM 50 µm	120 m (200) 300 m (500)
FICON Express2[b] LX	3319	2 Gbps	LC Duplex	SM 9 µm	10 km/20 km
FICON Express2[b] SX	3320	2 Gbps	LC Duplex	MM 62.5 µm MM 50 µm	120 m (200) 300 m (500)

Channel feature	Feature codes	Bit rate	Connector	Cable type	Maximum unrepeated distance[a]
FICON Express4 SX	3322	4 Gbps	LC Duplex	MM 62.5 µm MM 50 µm	55 m (160) 70 m (200) 150 m (500) 270 m (2000)
		2 Gbps	LC Duplex	MM 62.5 µm MM 50 µm	120 m (160) 150 m (200) 300 m (500) 500 m (2000)
		1 Gbps	LC Duplex	MM 62.5 µm MM 50 µm	250 m (160) 300 m (200) 500 m (500) 860 m (2000)
FICON Express4 4KM LX	3324	4 Gbps	LC Duplex	SM 9 µm	4 km
FICON Express4 10KM LX	3321	4 Gbps	LC Duplex	SM 9 µm	10 km/20 km[c]
Open Systems Adapter (OSA)					
OSA-Express2 GbE LX	3364	1 Gbps	LC Duplex	SM 9 µm	5 km
				MCP	550 m (500)
OSA Express2 GbE SX	3365	1 Gbps	LC Duplex	MM 62.5 µm	220 m (166) 275 m (200)
				MM 50 µm	550 m (500)
OSA Express2 1000BASE-T Ethernet	3366	10/100/ 1000	RJ45	UTP Cat5	100 m
OSA Express2 10 GbE LR	3368	10 Gbps	SC Duplex	SM 9 µm	10 km
OSA Express3 10 GbE LR	3370	10 Gbps	LC Duplex	SM 9 µm	10 km
Parallel Sysplex					
IC	n/a		N/A	N/A	N/A
ICB-4	3393	2 GBps	0229/0230[c]		10 m
ISC-3 (peer mode)	0217 0218 0219	2 Gbps	LC Duplex	SM 9 µm MCP 50 µm	10 km/20 km[c] 550 m (400)
ISC-3 (RPQ 8P2197 Peer mode at 1 Gbps)[d]		1 Gbps		SM 9 µm	20 km[c]
IFB	0163	6 GBps	MPO	OM3 MM 50 µm	150 m
ETR[e]	n/a	8 Mbps	MT-RJ	MM 62.5 µm MM 50 µm	3 km (26 km) 2 km (24 km)

a. Minimum fiber bandwidth in MHz/km for multi-mode fiber optic links are included in parentheses were applicable.

b. Feature is only available if carried forward by an upgrade from a previous server.

c. FC 0229 cable: z10 EC to z9, z990, z890.
FC 0230 cable: z10 EC to z10 EC.

d. RPQ 8P2197 enables the ordering of a different 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 have to be activated.

e. The External Time Reference (ETR) replaces the ETR feature available in previous servers. Two ETR cards are a standard feature in the z10 EC servers.

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 Redbooks" on page 94. Note that some of the documents referenced here may be available in softcopy only.

- *IBM System z10 Enterprise Class Technical Guide,* SG24-7516
- *IBM System z Strengths and Values*, SG24-7333
- *Getting Started with InfiniBand on System z10 and System z9,* SG24-7539
- *IBM System z Connectivity Handbook*, SG24-5444
- *Server Time Protocol Planning Guide*, SG24-7280
- *Server Time Protocol Implementation Guide*, SG24-7281
- *IBM System z9 Enterprise Class Configuration Setup,* SG24-7571
- *IBM System z10 Enterprise Class Capacity On Demand,* SG24-7504
- *IBM TotalStorage DS8000 Series: Performance Monitoring and Tuning,* SG24-7146
- *How does the MIDAW Facility Improve the Performance of FICON Channels Using DB2 and other workloads?*, REDP-4201
- *Introduction to the New Mainframe: z/VM Basics*, SG24-7316

Online resources

These Web sites are also relevant as further information sources:

- ResourceLink Web site

  `http://www-01.ibm.com/servers/resourcelink/svc03100.nsf?OpenDatabase`
- 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 2.10.0,* SC28-6867
- *Support Element Operations Guide V2.10.0,* SC28-6868
- *IOCP User's Guide,* SB10-7037

- *Stand-Alone Input/Output Configuration Program User's Guide,* SB10-7152
- *Planning for Fiber Optic Links,* GA23-0367
- *System z10 Enterprise Class Capacity on Demand User's Guide,* SC28-6871
- *CHPID Mapping Tool User's Guide,* GC28-6825
- *Common Information Model (CIM) Management Interfaces,* SB10-7154
- *System z10 Enterprise Class Installation Manual,* GC28-6865
- *System z10 Enterprise Class Installation Manual for Physical Planning,* GC28-6864
- *System z10 Enterprise Class Processor Resource/Systems Manager Planning Guide,* SB10-7153
- *System z10 Enterprise Class System Overview,* SA22-1084
- *System z10 Enterprise Class Service Guide,* GC28-6866
- *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 Application Programmer's Guide,* SA22-7522
- *z/OS Cryptographic Services Integrated Cryptographic Service Facility Messages,* SA22-7523
- *z/OS Cryptographic Services Integrated Cryptographic Service Facility Overview,* SA22-7519
- *z/OS Cryptographic Services Integrated Cryptographic Service Facility System Programmer's Guide,* SA22-7520

How to get Redbooks

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

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

IBM Redbooks

IBM System z10 Enterprise Class Technical Introduction

IBM System z10 Enterprise Class Technical Introduction

Describes new server hardware and capabilities

Discusses key functional elements and enhancements

Reviews software support

This IBM Redbooks publication introduces the IBM System z10 Enterprise Class server, which is based on z/Architecture. It builds on the inherent strengths of the System z platform, delivering new technologies and virtualization that are designed to offer improvements in price and performance for key workloads, as well as enabling a new range of hybrid solutions. The z10 EC further extends System z's leadership in key capabilities with the delivery of expanded scalability for growth and large-scale consolidation, availability to help reduce risk and improve flexibility to respond to changing business requirements, and improved security. The z10 EC is at 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 z10 EC capabilities, hardware functions and features, and its associated software support. This book is intended for systems engineers, architects, consultants, and anyone else needing to understand the new elements of the z10 EC.

This book is intended as an introduction to the z10 EC mainframe. Readers are not expected to be generally 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

SG24-7515-00 ISBN 0738480571