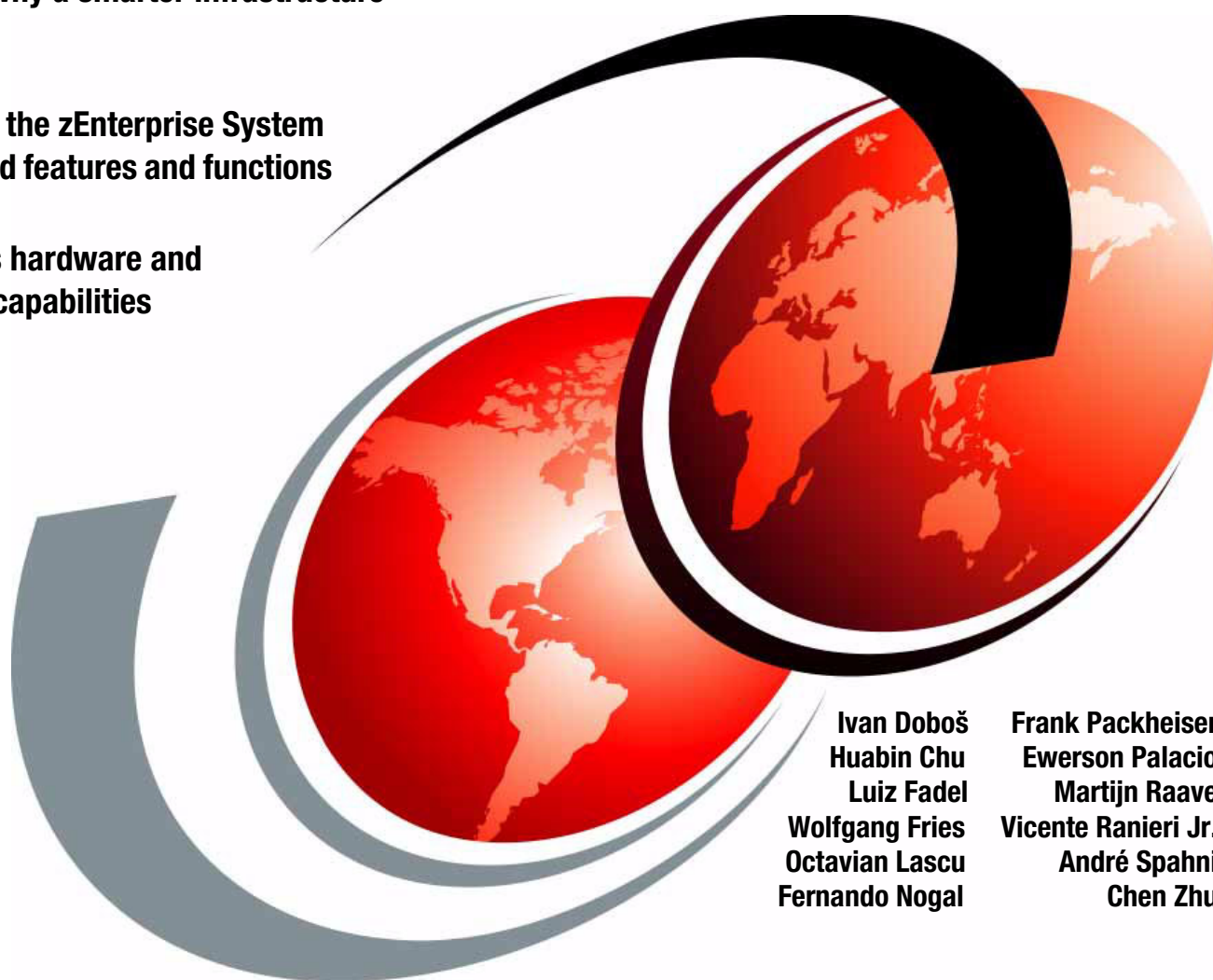


IBM zEnterprise System Technical Introduction

Explains why a smarter infrastructure
is needed

Describes the zEnterprise System
and related features and functions

Discusses hardware and
software capabilities



Ivan Doboš
Huabin Chu
Luiz Fadel
Wolfgang Fries
Octavian Lascu
Fernando Nogal

Frank Packheiser
Ewerson Palacio
Martijn Raave
Vicente Ranieri Jr.
André Spahni
Chen Zhu

Redbooks



International Technical Support Organization

IBM zEnterprise System Technical Introduction

March 2014

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

Second Edition (March 2014)

This edition applies to the IBM zEnterprise EC12, IBM zEnterprise BC12 and the IBM zEnterprise BladeCenter Extension Model 003.

© Copyright International Business Machines Corporation 2012, 2014. All rights reserved.

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

Contents

Notices	vii
Trademarks	ix
Preface	xi
Authors	xi
Now you can become a published author, too!	xiv
Comments welcome	xiv
Stay connected to IBM Redbooks	xiv
Chapter 1. Smartly essential	1
1.1 zEC12 and zBC12 technical description	4
1.1.1 IBM zEnterprise EC12	4
1.1.2 IBM zEnterprise BC12	6
1.1.3 Processor cages and drawers	8
1.1.4 Processor unit characterization	9
1.1.5 I/O subsystem	11
1.1.6 IBM Mobile Systems Remote	17
1.1.7 Hardware Management Console and Support Elements	17
1.1.8 Capacity on demand and performance	18
1.2 IBM zEnterprise BladeCenter Extension	19
1.3 Unified Resource Manager	20
1.4 Reliability, availability, and serviceability (RAS)	22
1.5 Software	22
Chapter 2. Hardware overview	25
2.1 zEC12 and zBC12 highlights, models, and upgrades	26
2.1.1 zEC12 highlights	26
2.1.2 zEC12 models	27
2.1.3 zEC12 upgrades	28
2.1.4 zBC12 highlights	29
2.1.5 zBC12 models	30
2.1.6 zBC12 upgrades	31
2.2 Frames	32
2.2.1 zEC12 frames	32
2.2.2 zBC12 frame	33
2.3 zEC12 processor cage, books, and multiple chip modules	34
2.4 zBC12 processor drawer and single chip modules	37
2.5 Processor chip	39
2.6 Processor unit	39
2.7 Memory	41
2.7.1 zEC12 memory	41
2.7.2 zBC12 memory	42
2.7.3 Concurrent memory upgrade	43
2.7.4 Redundant array of independent memory	43
2.7.5 Hardware system area	43
2.8 I/O system structure	44
2.9 I/O features	46
2.9.1 FICON Express8S	47
2.9.2 FICON Express8	47

2.9.3 FICON Express4	48
2.9.4 OSA-Express5S	48
2.9.5 OSA-Express4S	49
2.9.6 OSA-Express3	50
2.9.7 Flash Express	51
2.9.8 zEDC Express	52
2.9.9 10 Gigabit Ethernet RoCE Express	52
2.9.10 Coupling links	53
2.9.11 Cryptographic features	53
2.10 Coupling and clustering	54
2.10.1 InfiniBand coupling links	55
2.10.2 Internal Coupling (IC)	56
2.10.3 InterSystem Channel-3 (ISC-3)	56
2.10.4 Coupling Facility Control Code (CFCC)	56
2.11 Time functions	57
2.11.1 Server Time Protocol (STP)	58
2.11.2 Network Time Protocol support	58
2.11.3 Time coordination for zBX components	58
2.12 Hardware Management Console and Support Element	58
2.13 Power and cooling	59
2.14 zEnterprise BladeCenter Extension	62
 Chapter 3. Key functions and capabilities of IBM zEnterprise EC12 and IBM zEnterprise BC12	 67
3.1 Virtualization	69
3.1.1 zEC12 and zBC12 hardware virtualization	69
3.1.2 zEnterprise software virtualization	73
3.1.3 zBX virtualized environments	74
3.1.4 zEnterprise-based clouds	75
3.2 zEC12 and zBC12 technology improvements	75
3.2.1 Microprocessor	76
3.2.2 Memory	79
3.2.3 Native PCIe features and integrated firmware processor	81
3.2.4 Flash Express	82
3.2.5 zEDC Express	82
3.2.6 10GbE RoCE Express	84
3.2.7 I/O capabilities	84
3.2.8 Cryptography	93
3.3 Capacity and performance	95
3.3.1 Capacity settings	95
3.3.2 zEnterprise capacity on demand (CoD)	97
3.3.3 Permanent upgrades	98
3.3.4 Temporary upgrades	98
3.3.5 z/OS capacity provisioning	99
3.3.6 Throughput optimization with zEC12 and zBC12	99
3.3.7 zEnterprise EC12 and BC12 performance	100
3.4 zEnterprise common time functions	102
3.4.1 Server Time Protocol (STP)	102
3.4.2 Network Time Protocol (NTP) client support	103
3.5 Hardware Management Console functions	104
3.5.1 Hardware Management Console key enhancements for zEC12 and zBC12	105
3.5.2 Considerations for multiple Hardware Management Consoles	105
3.6 zEnterprise CPC power and cooling functions	106

3.6.1 High voltage DC power	106
3.6.2 Integrated battery feature (IBF)	106
3.6.3 Power capping and power saving	106
3.6.4 Power estimation tool	107
3.6.5 IBM Systems Director Active Energy Manager	107
3.6.6 Top Exit Power	108
3.7 zEnterprise BladeCenter Extension Model 003	108
3.7.1 IBM blades	108
3.8 Reliability, availability, and serviceability (RAS)	109
3.8.1 IBM System z Advanced Workload Analysis Reporter (IBM zAware)	110
3.8.2 RAS capability for the HMC	110
3.8.3 RAS capability for zBX	111
3.9 High availability technology for zEnterprise	111
3.9.1 High availability for zEnterprise with Parallel Sysplex	111
3.9.2 PowerHA in zBX environment	115
3.10 zEnterprise and emerging paradigms	116
Chapter 4. Operating system support and considerations	119
4.1 Software support summary	120
4.2 Support by operating system	123
4.2.1 z/OS	123
4.2.2 z/VM	127
4.2.3 z/VSE	130
4.2.4 z/TPF	133
4.2.5 Linux on System z	135
4.2.6 References	138
4.3 Software support for zBX Model 003	139
4.4 z/OS considerations	139
4.5 Coupling Facility and CFCC considerations	144
4.6 Input/output configuration program (IOCP) considerations	144
4.7 IBM Device Support Facilities (ICKDSF) considerations	144
Chapter 5. A smarter infrastructure	145
5.1 Integrated hybrid infrastructures	146
5.1.1 A cornerstone of a smart IT infrastructure	146
5.2 Benefiting from a smart infrastructure	147
5.3 Cloud computing	152
Appendix A. Software licensing	159
Software licensing considerations	160
Monthly License Charge pricing metrics	160
Advanced Workload License Charges	162
Advanced Entry Workload License Charges	163
System z New Application License Charges	163
Select Application License Charges	163
Midrange Workload License Charges	164
Parallel Sysplex License Charges	164
System z International Program License Agreement	165
zBX licensed software	165
zEnterprise Unified Resource Manager	165
Appendix B. Channel options	167
Related publications	171

IBM Redbooks publications	171
Other publications	171
Online resources	172
Help from IBM	172

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 grant 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.

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality.

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 websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites 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.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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:

AIX®	OMEGAMON®	RMF™
BladeCenter®	Parallel Sysplex®	Smarter Planet®
CICS®	Passport Advantage®	Sysplex Timer®
DataPower®	Power Systems™	System p®
DB2 Connect™	POWER6®	System Storage®
DB2®	POWER7®	System x®
Domino®	PowerHA®	System z10®
DRDA®	PowerVM®	System z9®
DS8000®	POWER®	System z®
ESCON®	PR/SM™	SystemMirror®
FICON®	Processor Resource/Systems Manager™	Tivoli®
GDPS®	RACF®	WebSphere®
Global Technology Services®	Rational Rose®	z/Architecture®
HACMP™	Rational®	z/OS®
HiperSockets™	Redbooks®	z/VM®
IBM Systems Director Active Energy Manager™	Redpaper™	z/VSE®
IBM®	Redbooks (logo)  ®	z10™
IMS™	Resource Link®	z9®
Lotus®	Resource Measurement Facility™	zEnterprise®
		zSeries®

The following terms are trademarks of other companies:

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

Linux is a trademark of Linus Torvalds 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.

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

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

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

Preface

In a smarter planet, information-centric processes are exploding in growth. The mainframe has always been the IT industry's leading platform for transaction processing, consolidated and secure data serving, and support for available enterprise-wide applications. IBM® has extended the mainframe platform to help large enterprises reshape their client experiences through information-centric computing and to deliver on key business initiatives.

IBM zEnterprise® is recognized as the most reliable and trusted system, and the most secure environment for core business operations. The new zEnterprise System consists of the IBM zEnterprise EC12 (zEC12) or IBM zEnterprise BC12 (zBC12), the IBM zEnterprise Unified Resource Manager, and the IBM zEnterprise IBM BladeCenter® Extension (zBX) Model 003.

This IBM Redbooks® publication describes the zEC12 and zBC12, with their improved scalability, performance, security, resiliency, availability, and virtualization. The zEnterprise System has no peer as a trusted platform that also provides the most efficient transaction processing and database management. With efficiency at scale delivering significant cost savings on core processes, resources can be freed up to focus on developing new services to drive growth.

This book provides a technical overview of the zEC12, zBC12, zBX Model 003, and Unified Resource Manager. This publication is intended for IT managers, architects, consultants, and anyone else who wants to understand the elements of the zEnterprise System. For this introduction to the zEnterprise System, readers are not expected to be familiar with current IBM System z® technology and terminology.

Authors

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

Ivan Doboš is an IBM Certified Consulting IT Specialist working as a mainframe consultant at STG Lab Services Central & Eastern Europe. He has 15 years of experience with IBM System z. He joined IBM in 2003 and worked in different sales and technical roles supporting mainframe clients, as Technical Leader for Linux on System z projects in the System z Benchmark Center, IT Optimization Consultant in the System z New Technology Center, and Mainframe Technical Sales Manager in Central & Eastern Europe. During the past ten years, he has worked with many clients and spent most of his time supporting new workloads on System z projects. Ivan has authored several Redbooks publications and Redpaper publications.

Huabin Chu is an Advisory I/T Specialist in China. He has seven years of experience with IBM Global Technology Services® and in supporting clients of large System z products. His areas of expertise include IBM z/OS®, IBM Parallel Sysplex, System z high availability solutions, and IBM GDPS®.

Luiz Fadel is an IBM Distinguished Engineer responsible for supporting System z for the Latin America region, part of the Growth Markets Unit. He joined IBM in 1969 and has supported Large Systems ever since, including working on two assignments with the International Technical Support Organization (ITSO). Luiz is a member of the Latin America

Advanced Technical Support team, which is responsible for handling Client Critical Situation and client claims within System z, Early Support Programs, new product installations, internal product announcements, and second level client support, as well as managing complex Proof of Concepts (POC). He is a member of the zChampions team and the co-author of several IBM Redbooks publications.

Wolfgang Fries is a Senior Consultant in the System z HW Support Center in Germany. He spent several years at the European Support Center in Montpellier, France, providing international support for System z servers. Wolfgang has more than 35 years of experience in supporting large System z customers. His areas of expertise include System z servers and connectivity. Wolfgang has co-authored a number of IBM Redbooks publications.

Octavian Lascu is a Senior IT Consultant for IBM Romania with over 20 years of experience. He is specialized in designing and supporting complex IT infrastructure environments (Systems, Storage, Networking), including High Availability and Disaster Recovery solutions and High performance Computing deployments. He has developed and taught over 50 workshops for technical audiences around the world. He has authored several Redbooks publications and Redpaper publications.

Fernando Nogal is an IBM Certified Consulting IT Specialist working as an STG Technical Consultant for the Spain, Portugal, Greece, and Israel IMT. He specializes in advanced infrastructures and architectures. In his 30+ years with IBM, he has held a variety of technical positions, mainly providing support for mainframe clients. Previously, he was on assignment to the Europe Middle East and Africa (EMEA) System z Technical Support group, working full-time on complex solutions for e-business. His job included, and still does, presenting and consulting in architectures and infrastructures, and providing strategic guidance to System z clients regarding the establishment and enablement of advanced technologies on System z, including the z/OS, IBM z/VM, and Linux environments. He is a zChampion and a member of the System z Business Leaders Council. An accomplished writer, he has authored and co-authored over 30 IBM Redbooks publications and several technical papers.

Frank Packheiser is a Senior z IT Specialist at the Field Technical Sales Support office in Germany. He has 20 years of experience in zEnterprise, System z, IBM zSeries®, and predecessor mainframe servers. He has worked for 10 years for the IBM education center in Germany, developing and providing professional training. He also provides professional services to System z and mainframe clients. In 2008 and 2009, he supported clients in the Middle East/North Africa (MENA) as a zIT Architect. Besides co-authoring several Redbooks publications since 1999, he has been an ITSO guest speaker on ITSO workshops for the last two years.

Ewerson Palacio is an IBM Distinguished Engineer and a Certified Consulting IT Specialist for Large Systems in Brazil. He has more than 40 years of experience in IBM Large Systems. Ewerson holds a Computer Science degree from Sao Paulo University. His areas of expertise include System z Servers Technical and Client Support, Mainframe architecture, infrastructure implementation and design. He is an International Technical Support Organization (ITSO) System z Hardware official speaker and has presented technical ITSO seminars, workshops, and private sessions to IBM clients, IBM IT Architects, IT Specialists, and Business Partners around the globe. He has also been a System z Hardware Top Gun training designer, developer, and instructor for the last generations of the IBM high-end servers. Ewerson leads the Mainframe Specialty Services Area (MF-SSA), part of GTS - Delivery, Technology, and Engineering (GTS - DT&E) and he is an IBM Academy of Technology Member.

Martijn Raave is a certified System z Client Technical Specialist for the IBM Systems and Technology Group (STG) in the Netherlands. Over a period of 15 years, his professional career has revolved around the mainframe platform. Before joining IBM through a strategic outsourcing deal in 2005, he worked for a large Dutch customer as a systems programmer with expertise in the areas of z/OS, (Globally Dispersed) Parallel Sysplex and hardware. Four years ago, he decided to explore the other aspects of the mainframe ecosystem within IBM and joined STG in his current role. As a Client Technical Specialist, he supports several Dutch System z customers, IBM Business Partners, and IBM Sales Representatives on technical topics and in sales engagements. He is also a board member of Guide Share Europe (GSE) Netherlands.

Vicente Ranieri Jr. is an Executive IT Specialist and the Lead Architect at the High End Design Center in Latin America. He has more than 30 years of experience working for IBM. Ranieri used to be a member of the Advanced Technical Skills (ATS) team in Latin America and his areas of expertise include System z security, IBM Parallel Sysplex®, System z hardware, and z/OS. Vicente has co-authored several Redbooks publications and has been an ITSO guest speaker since 2001, teaching System z security update workshops worldwide. Ranieri is certified as a Distinguished IT Specialist by the Open group and he is a member of the zChampions team, of the Technology Leadership Council – Brazil, and of the IBM Academy of Technology.

André Spahni is a Senior Support Center Representative working for IBM Global Technology Services® in Switzerland. He has 11 years of experience working with and supporting System z clients. André has been working for the Technical Support Competence Center (TSCC) Hardware FE System z for Switzerland, Germany, and Austria since 2008. His areas of expertise include System z hardware, Parallel Sysplex, and connectivity.

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

A special thanks is due to the authors of the first version of this IBM Redbooks publication for their efforts in creating the groundwork for this edition. Those authors include: Parwez Hamid, Gerard Laumay, Swee Seng Ng, Karan Singh, Esra Ufacik, Hans Wijngaard, and Zhaoxu Zhang.

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

Debbie Beatrice, James Caffrey, Ellen Carbarnes, Edward Chencinski, Doris Conti, Jim Coon, Kathleen Fadden, Leslie Geer, Susan Greenlee, Christian Jacobi, Paul Jones, Kenneth Kerr, Georgette Kurdt, Amy Lander, Patrick Rausch, Peter Relson, Robert Rogers, Donald M. Smith, James Sweeny, Elpida Tzortzatos, Romney White, and Doug Zobre
IBM US

Marcel Mitran
IBM Canada

Jia Lei Ma
IBM China

Juergen Kuhn and Horst Sinram
IBM Germany

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

ibm.com/redbooks

- Send your comments in an email to:

redbooks@us.ibm.com

- Mail your comments to:

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

Stay connected to IBM Redbooks

- Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- Follow us on Twitter:

<http://twitter.com/ibmredbooks>

- Look for us on LinkedIn:

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

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

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

- Stay current on recent Redbooks publications with RSS Feeds:

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



Smartly essential

Ever since computing technology started to be used by the public and private sectors, its use has been relentless and the pace ever increasing. Moreover, the *scope and focus* have been shifting: It all used to happen in the *Data Processing* department, remember? Now *information technology (IT)* is used by virtually everyone. One may not realize it, but heavy usage of IT is an integral part of our daily lives, from the common cash withdrawal at an ATM to receiving on our mobile devices a personalized product offering, which was shaped by the mining of the vast amounts of data generated by social applications.

One consequence is that organizations, in their pursuit of “a better IT,” have accumulated a huge diversity of equipment, networks, architectures, software, operational procedures, standards, and practices. This has often led to complex, heavy, and inflexible infrastructures, with many components and interconnections. Those infrastructures have no underlying framework and no common operational tools able to provide global and unified views and management. The diversified environments demand a vast set of skills and are labor inefficient. These IT infrastructures are reaching the point of becoming unmanageable.

Technologies, IT included, are increasingly seen as drivers of change, and CEOs¹ see harnessing the benefits of technologies as a key way to differentiate their organizations and outperform the competition. IT is being looked upon to support the CMOs’, CFOs’ and CEOs’ strategic initiatives, empowering employees, and providing the means for new collaborative modes and organizational ways.

For instance, mobility, of clients, partners and employees, is a growing interaction paradigm which requires agility in conceiving and deploying applications. Mobility, through the applications that support it, places heavy requirements upon the underlying infrastructure, from security to rapid provisioning of resources, which dovetails with another growing IT delivery model: cloud computing.

Enterprises are looking to provide innovative services and tailored offerings, derived from better insights into clients needs, which can be obtained by mining the vast amounts of data generated by and through these interactions, as well as social computing. These services are preferentially delivered through mobile end-points, as studies show that over 90% of mobile users keep their devices with them all the time and 75% of mobile shoppers take action after receiving a location based message.

¹ See the *IBM 2012 Global Chief Executive Officer Study* at <http://www.ibm.com/ceostudy>

And do not forget “the Internet of things,” now with more devices than the number of humans alive, which is also generating new services and other offerings.

All of these considerations are highly transactional in nature, and it is becoming very clear that two broad categories of systems are needed to support them: *Systems of Interaction*² and *Systems of Record*³. Both make specific requirements of their supporting infrastructure, with agility, reliability, availability, and security as common characteristics.

Management demands adaptability and flexibility because that is the best guarantee against an uncertain future; a future that no one can foresee but, all agree, will be very different from the present.

IBM meets those requirements by proposing an enterprise-wide, comprehensive view of the IT systems, which includes an optimized, smarter infrastructure, one that is flexible, scalable, and provides diversified components that are tuned to the task.

IBM zEnterprise System

Forming an essential part of this smart infrastructure, IBM zEnterprise System offers the breadth and scope to support both systems of interaction and systems of record, on a cohesively managed and highly resource optimized heterogeneous system. At its heart are the IBM zEnterprise EC12 and IBM zEnterprise BC12 systems, shown in Figure 1-1 along with the other zEnterprise components, the zEnterprise BladeCenter Extension (zBX) and the zEnterprise Unified Resource Manager.



Figure 1-1 IBM zEnterprise EC12, zEnterprise BladeCenter Extension (zBX), zEnterprise Unified Resource Manager, and IBM zEnterprise BC12

² *Systems of Interaction* provide flexibility and scale in back-office systems to keep pace with the rate of change, and connect them to all of the new endpoints of front-office engagements.

³ *Systems of Record* provide the authoritative, trustworthy source for data, guaranteeing data integrity and validity.

The zEC12 and zBC12 are heirs, respectively, to the IBM zEnterprise 196 and IBM zEnterprise 114. The zEC12 and the zBC12 are hybrid systems that IBM introduced to help overcome fundamental problems of today's IT infrastructures and, simultaneously, provide a foundation for the future. With industry-leading virtualization, the highest security, trusted availability, and massive data abilities, they are the preferred and reference platforms for secure cloud, enterprise mobility, and operational analytics.

With zEnterprise systems, IBM brought together multiple platforms and created a scalable solution that simplifies hardware, firmware management, support, and the definition and management of a network of virtualized servers. Thus, clients can start to replace individual islands of computing with a more integrated and hybrid infrastructure. This configuration can reduce complexity, improve security and manageability, and bring applications closer to the data that they need.

This integrated hybrid infrastructure has three main components:

- ▶ The *IBM zEnterprise central processor complex (CPC)* implements the System z platform environments.
- ▶ The *IBM zEnterprise BladeCenter Extension (zBX)* implements IBM POWER® and IBM System x® environments, and specialized solutions and appliances.
- ▶ The *IBM zEnterprise Unified Resource Manager* provides the overall management capabilities for the other components with advanced end-to-end management capabilities for the diverse environments within the zBX.

Ensembles

This heterogeneous infrastructure also includes the concept of an *ensemble*: a collection of highly virtualized diverse systems that can be managed as a single logical entity, and where diverse workloads can be deployed.

An ensemble is composed of up to eight members, or *nodes*, each of which is composed of a CPC and its optional zBX. So, an ensemble is a collection of up to eight CPCs and up to 896 blades that are housed in up to eight zBXs. Also included are dedicated integrated networks for ensemble management and data communication across the ensemble's members, and the Unified Resource Manager function.

The concept of an ensemble is similar to that of a cloud. An ensemble provides a perfect infrastructure to support a cloud because the real purpose of an ensemble is to provide infrastructure resources in a way that ensures that the workloads which run on it achieve their business requirement objectives. Those objectives are specified through policies, which the ensemble implements.

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

Many mission-critical workloads today have one or more components on System z, using System z environments for database, transactional systems, and other capabilities, in reality, Systems of Record. Other components, mostly found on Systems of Interaction, can exploit other zEnterprise environments, notably Linux on System z, as well as, on the zBX, AIX, Linux on System x, and Windows environments. The ability to collocate all of the workload components *under the same management platform*, and thereby benefit from uniformly high qualities of service, is appealing and provides tangible benefits and a rapid ROI.

1.1 zEC12 and zBC12 technical description

Organizations' requirements span a very broad range, so IBM has two offerings, exploiting the same technologies, that cover that diversity: the IBM zEnterprise EC12 (zEC12), for organizations with medium to very large workload requirements; and the IBM zEnterprise BC12 (zBC12) when workload requirements range from entry to medium. However, some characteristics, such as security and reliability, have the same high standard in both offerings.

This chapter introduces the IBM zEnterprise EC12 and IBM zEnterprise BC12 systems and explains how, with their innovations and traditional strengths, they can play an essential role in a smarter IT infrastructure. Chapter 2, "Hardware overview" on page 25, provides additional technical details.

Terminology: In the remainder of the book, we use the designation *CPC* to refer to the *central processor complex*. The reference *zEnterprise CPC* includes zEC12, zBC12, z196, and z114.

1.1.1 IBM zEnterprise EC12

The IBM zEnterprise EC12 exploits several leading-edge technologies, including silicon-on-insulator 32 nm (CMOS 13s-SOI), storage-class memory, InfiniBand, and Ethernet. The zEC12, when compared to its predecessor, the IBM zEnterprise 196, offers improvements in several areas such as a faster and redesigned high-frequency chip, more granularity options, better availability, and enhanced on-demand options.

The IBM zEnterprise EC12 is the first system to offer a *Transactional Execution Facility*, known in the industry as *hardware transactional memory*, which can increase transaction rates by reducing resource collision. In addition, several features are introduced, namely in the connectivity and data encryption areas, and the IBM Flash Express for Storage Class Memory, a solid-state disk-based offering. The innovative *IBM zAware*, an analytical and statistical based offering appliance, possesses sophisticated detection and diagnostic capabilities that contribute to system availability.

zEC12 is a *symmetric multiprocessor (SMP)* system with a scalable design. Five models are offered: H20, H43, H66, H89, and HA1.

The model name represents the maximum number of processors that can be configured in the model ("A1" stands for 101).

The zEC12 system architecture ensures continuity and upgradability from the z196 and IBM z10™ EC systems. The IBM commitment to the zEC12 and its sustained investment in the system and its predecessors is portrayed in Figure 1-2. The figure provides a comparison of the zEC12 with previous Enterprise Class System z systems, regarding four major attributes:

- ▶ Single engine processing capacity (based on the *Processor Capacity Index (PCI)*⁴)
- ▶ Number of engines
- ▶ Memory
- ▶ I/O bandwidth (servers use a subset of their designed I/O capability)

⁴ PCI values can be obtained from SC28-1187 *Large Systems Performance Reference*, at this website: <https://www-304.ibm.com/servers/resourceLink/lib03060.nsf/pages/1sprindex>

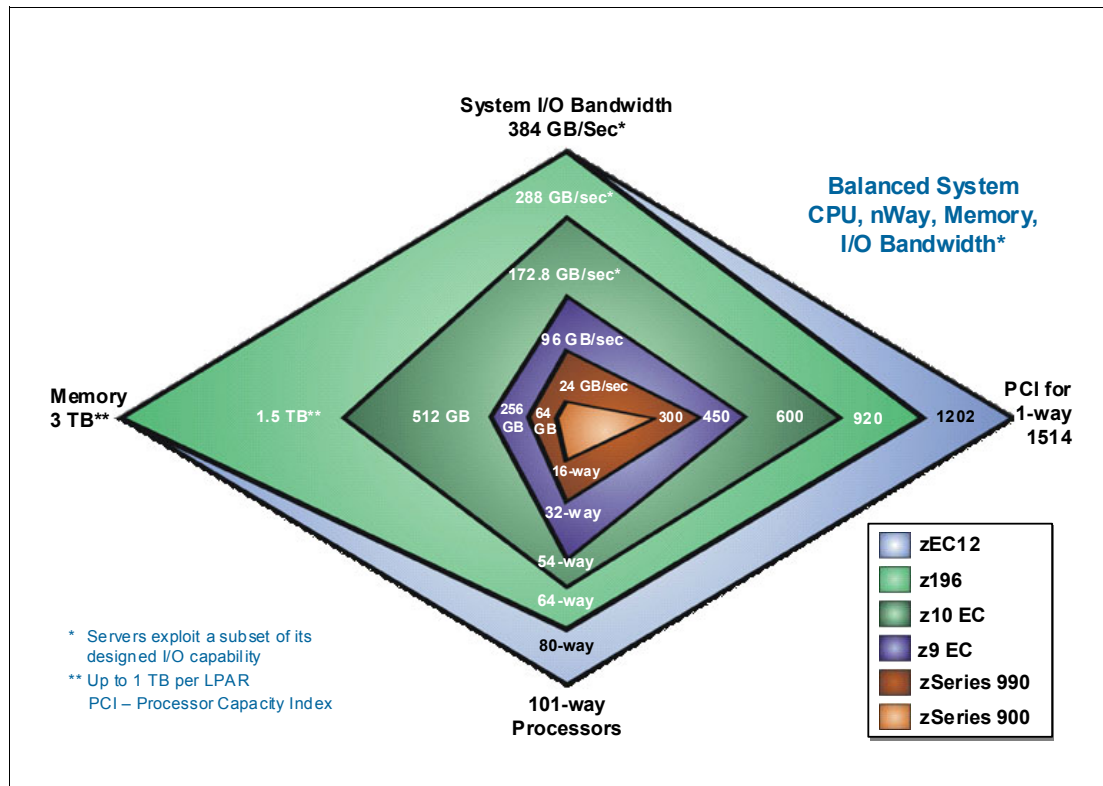


Figure 1-2 High End System z systems design comparison

The zEC12 has a machine type (M/T) designation of 2827 and is a two-frame system. The frames are known as the *A frame* and the *Z frame*.

The A frame contains the following elements:

- ▶ The processor cage, which houses the processor books
- ▶ Modular cooling units (different for water and air cooling)
- ▶ PCIe I/O drawers, I/O drawers, I/O cages, and their I/O features, available in several combinations
- ▶ An optional integrated battery feature (IBF)

The Z frame contains the following elements:

- ▶ Two redundant Support Elements (SE)
- ▶ PCIe I/O drawers, I/O drawers, and their I/O features, available in several combinations
- ▶ Power supplies
- ▶ An optional IBF

The SEs can be used to configure and manage the zEC12 system. When configured for an ensemble environment, the SE can also be used to manage the controlled zBX.

For better energy efficiency, the zEC12 offers, in addition to a radiator air-cooled option, water-cooled and high-voltage DC power options. These features mean potentially lower costs, without significantly changing the system physical footprint (the water cooling option adds a few inches of depth to the back of both system frames).

The zEC12 offers top exit cabling options for power and I/O, as an alternative to having all the cables exiting at the bottom of the CPC to under the raised floor. A non-raised floor installation of the zEC12 air-cooled systems is also possible, although water-cooled systems must be installed on a raised floor. Top exit cabling can also help to increase the air flow. These options are offered on new build and MES orders. The increased flexibility allows choosing the options that best meet the data center requirements.

1.1.2 IBM zEnterprise BC12

The zBC12 employs the same technologies as the zEC12, namely leading-edge silicon-on-insulator 32 nm (CMOS 13s-SOI) and other technologies, such as storage-class memory, InfiniBand, and Ethernet.

The IBM zEnterprise zBC12, when compared to its predecessor, the IBM zEnterprise 114, offers improvements in several areas, which include a faster and redesigned high-frequency chip, more granularity options, better availability, and enhanced on-demand options.

Exclusive to the zEC12 and zBC12 is the *Transactional Execution Facility*, known in the industry as *hardware transactional memory*, which can increase transaction rates by reducing resource collision. In addition, several features are introduced, namely in the connectivity and data encryption areas, and the IBM Flash Express for Storage Class Memory, a solid-state disk-based offering. The innovative *IBM zAware*, an analytical and statistical based appliance offering, possesses sophisticated detection and diagnostic capabilities that contribute to system availability.

zBC12 is a *symmetric multiprocessor (SMP)* system with a scalable design. Two models are offered: H06 and H13. The model name represents the maximum number of processors that can be configured in the model.

The zBC12 system architecture ensures continuity and upgradability from the z114 and IBM z10™ BC systems. The IBM commitment to the zBC12 and its sustained investment in the system and its predecessors is portrayed in Figure 1-3. The figure provides a comparison of the zBC12 with previous Business Class System z systems, regarding four major attributes:

- ▶ Single engine processing capacity (based on the *Processor Capacity Index (PCI)*⁵)
- ▶ Number of engines
- ▶ Memory
- ▶ I/O bandwidth (servers use a subset of their designed I/O capability)

⁵ PCI values can be obtained from SC28-1187 *Large Systems Performance Reference*, at this website:
<https://www-304.ibm.com/servers/resourceLink/lib03060.nsf/pages/1sprindex>

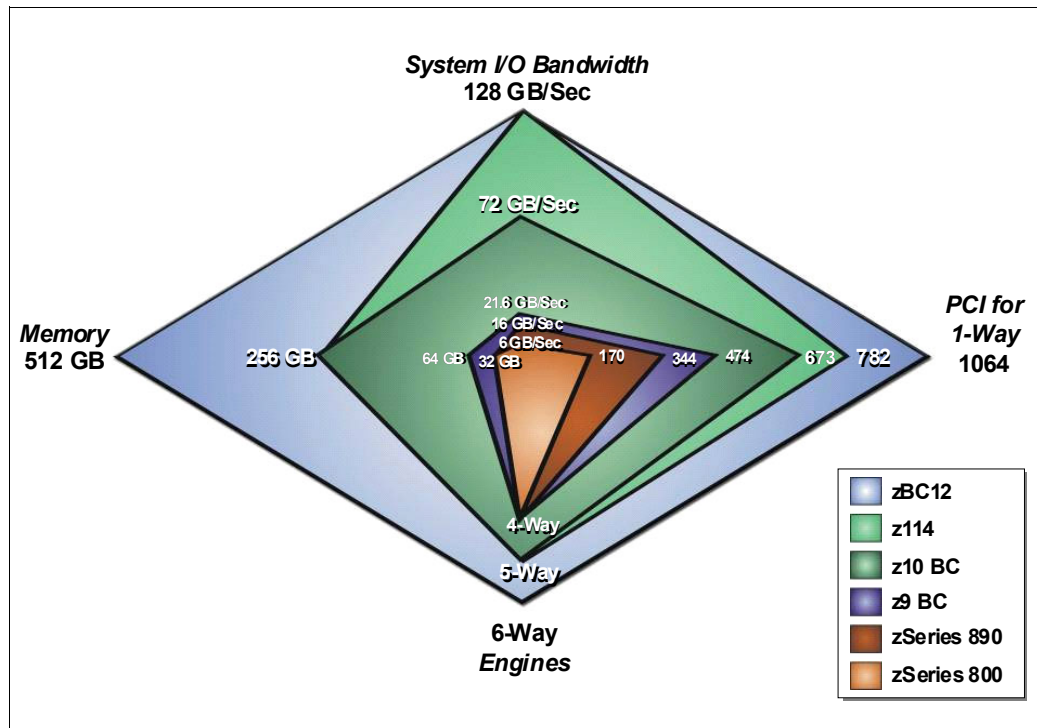


Figure 1-3 Mid-sized System z systems design comparison

The zBC12 has a machine type (M/T) designation of 2828 and is a single-frame system. The frame is known as the *A frame*.

The A frame contains the following elements:

- ▶ The processor drawers, which house the processor single-chip modules (SCM) and memory
- ▶ Modular cooling units
- ▶ PCIe I/O drawers, I/O drawers, and their I/O features, available in several combinations
- ▶ Two redundant Support Elements (SE)
- ▶ Power supplies
- ▶ An optional integrated battery feature (IBF)

The SEs can be used to configure and manage the zBC12 system. When configured for an ensemble environment, the SE can also be used to manage the controlled zBX.

The zBC12 is an air-cooled system. When planning to build a green data center, the optional high-voltage DC power provides better energy efficiency, with the potential to lower costs.

The zBC12 offers top exit cabling options for power and I/O, as an alternative to having all the cables exiting at the bottom of the CPC to under the raised floor. A non-raised floor installation of the zBC12 air-cooled systems is also possible. Top exit cabling can also help to increase the air flow. These options are offered on new build and MES orders. The increased flexibility allows choosing the options that best meet the data center requirements.

1.1.3 Processor cages and drawers

The zEC12 employs the same technologies of the z196 but also incorporates new ones. Notably, the IBM z/Architecture® processor chip is redesigned and operates at increased frequency over its predecessor, retaining industry leadership. zBC12 employs the same technologies of the zEC12 and also draws from the z114 design.

Although the zEC12 and zBC12 potentially include several hundred processor chips, only the z/Architecture processor chips are described in the section that follows.

zEC12 processor cage

On a zEC12, the processor cage houses from one to four processor books that are fully interconnected. Each book contains a *multiple chip module (MCM)*, memory, and connectors to the PCIe I/O drawers, I/O drawers, I/O cage, and coupling link connectors. Despite the multi-book design, the system is a symmetric multiprocessor, scalable up to 120 cores.

The zEC12 is the first mainframe to implement a high-speed six-core design. At 5.5 GHz, it is the fastest commercial processor chip in the industry, at the time of writing. Each core is known as a *PU (processor unit)*.

zEC12 is built on the proven superscalar microprocessor architecture that is already deployed on the z196. However, the PU chip has several distinctive innovations, notably in the out-of-order instruction execution design and on-chip caches. Improvements have been made in error checking and correcting (namely in the memory design) and specialized circuitry (for instance, to support improved out-of-order execution and decimal floating point operations). zEC12 is the first system to implement hardware transactional memory, through its Transactional Execution Facility.

In each book, the MCM houses eight chips: six PU chips and two storage control chips. Each PU chip has either four, five, or six cores enabled.

zEC12 offers two options to cool the MCMs:

- ▶ *Radiator units* with air-cooling backup, which exchange heat with an internal, closed-loop, water system
- ▶ *Water Cooling Units*, which are connected to building chilled water systems with back door heat-exchange units

One of the options must be selected at ordering time, because they are factory installed. It is not possible to convert between the options in the field.

Up to 3 TB of memory are available, with up to 1 TB configurable per logical partition, as with the z196. For enhanced availability, memory is implemented as a *Redundant Array of Independent Memory (RAIM)*. In each book, of the 960 GB that can be installed, part is redundant, so that up to 768 GB of usable memory can be configured. In addition, 32 GB are part of the base and reserved for the *hardware system area (HSA)*, making the maximum amount of purchasable memory 3040 GB, just short of 3 TB (with redundancy, a total of 3.75 GB are installed). *Plan-ahead memory*, a capability whereby memory can be installed but not enabled for use until needed, further enhances system availability for continuous operations.

zBC12 processor drawer

The zBC12 employs a processor drawer. The drawer was introduced with the z10 BC. The zBC12 Model H06 has one processor drawer and the Model H13 has two. The processor drawer is air-cooled.

Each processor drawer houses three single-chip modules (SCM), memory, I/O drawer connectors, and (optionally) coupling link connectors. The zBC12 uses the same chip technology as the zEC12.

There are two PU SCMs and the third is a storage control single-chip module (SC SCM). One PU SCM has four active PUs, while the other has five active PUs, for a total of nine, running at 4.2 GHz each.

On the H06 any unassigned PU can be used as a spare. The H13 has two PUs designated as spares. Each individually characterized PU can be transparently spared, as with the z114.

For enhanced availability, memory is implemented as a *Redundant Array of Independent Memory (RAIM)*. Up to 512 GB of memory can be installed, of which 16 GB are part of the base system and are reserved for the hardware system area (HSA), making the maximum amount of customer purchasable memory 496 GB. *Plan-ahead memory*, a capability whereby memory can be installed but not enabled for use until needed, further enhances system availability for continuous operations.

1.1.4 Processor unit characterization

At system initialization time, each installed and enabled *processor unit (PU)* is characterized as one of various types. It is also possible to characterize PUs dynamically. A PU that is not characterized cannot be used.

Some PUs are part of the base system, that is, they are *not* part of customer purchasable PUs and are characterized by default:

- ▶ *System Assist Processor (SAP)*:
SAPs offload and manage I/O operations. Several are standard with the zEC12 and two SAPs are available per system on a zBC12. Additional SAPs can be configured if increased I/O processing capacity is needed.
- ▶ *Integrated firmware processor (IFP)*:
The integrated firmware processor (IFP) is a single PU dedicated to the support of the native PCIe features (10GbE RoCE Express and zEDC Express).
- ▶ Two spare PUs (on all zEC12 models and on the zBC12 Model H13), which can transparently assume any characterization, in the case of the permanent failure of another PU.

Customer purchasable PUs can assume any of the following characterizations:

- ▶ *Central processor (CP)*:
The standard processor. For use with any supported operating system and user applications.
- ▶ *Integrated Facility for Linux (IFL)*:
Exclusively used with Linux on System z and for running the z/VM hypervisor in support of Linux. z/VM is often used to host multiple Linux virtual machines (called *guests*).
- ▶ *IBM System z Application Assist Processor (zAAP)*:
Used under z/OS⁶ for designated workloads, which include the IBM Java Virtual Machine (JVM) and XML System Services functions.

⁶ z/VM V5R4 and later support zIIP and zAAP processors for z/OS guest workloads.

Statement of Direction: zEC12 and zBC12 are planned to be the last System z servers to offer support for zAAP specialty engine processors. IBM intends to continue support for running zAAP workloads on zIIP processors (“zAAP on zIIP”).

► *IBM System z Integrated Information Processor (zIIP):*

Used under z/OS⁶ for designated workloads, which include various XML System Services, IPsec offload, certain parts of IBM DB2 DRDA®, star schema, IBM HiperSockets™ for large messages, and the IBM GBS Scalable Architecture for Financial Reporting.

zAAP and zIIP:

- Work that is dispatched on zAAP and zIIP does not incur any IBM software charges. zAAPs and zIIPs contribute to a lower cost of computing by taking some of the z/OS load that would otherwise run on CPs.
- It is possible to run a zAAP-eligible workload on zIIPs. This capability is offered to enable the optimization and maximization of investments on zIIPs.
- The z/OS system measurement facility (SMF), when properly configured, is able to project the potential utilization of zAAPs and zIIPs by the workloads running on the system, without actually requiring the purchase and installation of these processors. This estimate may help size the number of required zAAPs or zIIPs, and justify their acquisition.

► *Internal Coupling Facility (ICF):*

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

- *Additional System Assist Processor (SAP)*, to be used by the channel subsystem.

CP Assist for Cryptographic Function

The zEC12 and zBC12 continue to offer the cryptographic assist implementation, first deployed in 2003, known as *CP Assist for Cryptographic Function (CPACF)*.

CPACF is physically implemented in the six-core chip by the compression and cryptography accelerators. Each core has one dedicated coprocessor (CoP) integrating the CPACF and the compression unit. This configuration eliminates any interferences that could occur with the implementation on the z196, z114, and z10 where two cores share the coprocessor.

The CPACF offers the full complement of the *Advanced Encryption Standard (AES)* algorithm and *Secure Hash Algorithm (SHA)* along with the *Data Encryption Standard (DES)* algorithm. CPACF must be explicitly enabled, using a no-charge enablement feature, except for the SHAs, which are shipped enabled with each server.

The CP Assist for Cryptographic Function supports the following functions:

- *DES*
- *AES*
- *SHA*
- *Message authentication code (MAC)*:
- *Pseudo Random Number Generation (PRNG)* for cryptographic key generation
- Protected key capabilities

Keys: The keys can only be provided in clear form.

1.1.5 I/O subsystem

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

The z/Architecture also specifies that *peripheral devices* are managed by *control units* and are reached through *channels* from the CPC. A control unit provides controlling function for a device or set of devices, and may be physically implemented with the device or in an independent unit.

The I/O subsystem is fully virtualized and allows full sharing of its elements across multiple logical partitions.

As with its predecessors, the zEC12 and zBC12 implement the z/Architecture I/O subsystem through a dedicated subsystem, known as the *channel subsystem*, which is composed of the following elements:

- ▶ System Assist Processor:

System Assist Processor (SAP) is a specialized processor that uses the installed PUs⁷. Its role is to offload I/O operations and manage channels and the I/O operations queues. It relieves the other PUs of all I/O tasks, allowing them to be dedicated to application logic. Enough SAP processors are automatically defined, depending on the model of the machine. The SAPs are part of the base configuration of the system.

- ▶ Hardware System Area:

Hardware system area (HSA) is a reserved part of the system memory and contains the I/O configuration. It is used by SAPs. On the zEC12, a fixed amount of 32 GB is reserved, and on the zBC12 the HSA has a fixed size of 16 GB. The HSA is not part of the client-purchased memory. This amount provides for greater configuration flexibility and higher availability by eliminating planned and pre-planned outages.

- ▶ Channels:

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

- ▶ Channel path:

Channel paths are the means by which the channel subsystem communicates with the I/O devices. Because of I/O virtualization, multiple independent channel paths can be established on a single channel, allowing sharing of the channel between multiple logical partitions, with each partition having a unique channel path. The function that allows sharing I/O paths across logical partitions is known as the multiple image facility (MIF). On the zEC12 and zBC12, the Channel subsystem enhancement for I/O resilience was introduced, providing improved throughput and reduced I/O service times.

- ▶ Subchannels:

Subchannels are displayed to a program as a logical device (programs do not directly communicate with the devices) and contain the information that is required to perform an I/O operation. One subchannel exists for each I/O device addressable by the channel subsystem. The zEC12 has three subchannel sets⁸ and the zBC12 has two.

⁷ Each zEC12 and zBC12 PU can be characterized as one of seven types. For more information, see “Processor unit characterization” on page 9.

⁸ Subchannel set 0 can have up to 63.75 K devices (256 devices are reserved). Subchannels 1 and 2 can each have up to 64 K minus one.

In addition to the channel subsystem, zEC12 and zBC12 also implement a *queued direct I/O (QDIO)* infrastructure, present also on predecessor systems. QDIO is a highly efficient data transfer mechanism that is designed to dramatically reduce system overhead and improve throughput by using system memory queues and a signaling protocol. Data is directly exchanged between the I/O features and the network software.

QDIO is exploited by Open Systems Adapter (OSA) features, HiperSockets using the QDIO Accelerator function, and the FICON Channels when operating in FCP mode.

The I/O subsystem direction of the zEC12 and zBC12 is evolutionary, expanding on developments from the z196, and includes Peripheral Component Interconnect Express (PCIe), InfiniBand, enhanced cards, and protocols (High Performance FICON for System z (zHPF)). It is intended to provide significant performance improvements over the I/O platforms of previous systems both by reducing overhead and latency and providing increased data throughput.

The I/O infrastructure includes: I/O infrastructure elements, the PCIe I/O drawer, the I/O drawer⁹, and the I/O cage⁹.

Peripheral Component Interconnect Express

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

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

PCIe in the zEC12 and the zBC12 provides an internal I/O infrastructure that positions the system for continued support of the industry's direction for high-performance I/O.

The PCIe I/O bus connects the processor cage to the PCIe I/O drawers. PCIe I/O drawers house PCIe features.

InfiniBand

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

A single connection can carry 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 pervasive.

InfiniBand is exploited by the zEC12 and the zBC12. Within the system, the InfiniBand I/O bus connects the processor cage to I/O drawers or the I/O cage. I/O cages and I/O drawers house legacy I/O features. For external usage, InfiniBand (IFB) links are available to interconnect zEnterprise and z10 systems in a Parallel Sysplex (z/OS cluster). IFB links can completely replace the InterSystem Channel-3 (ISC-3) and ICB-4 offerings available on previous systems.

⁹ No new orders. Carry forward only.

I/O cage

The *I/O cage* accommodates up to 28 I/O features, in any combination. On the zEC12, a maximum of one I/O cage is supported, when carried-forward on an upgrade from a previous system. It is housed, along with the processor cage, in the A frame. zBC12 does not support I/O cages.

I/O drawer

An *I/O drawer* provides increased I/O granularity and capacity flexibility, as compared with the I/O cage. I/O drawers can be concurrently added and removed in the field, an advantage over I/O cages, which also eases pre-planning. When carried forward on an upgrade, up to two I/O drawers can be installed on a zEC12, and one I/O drawer can be installed on a zBC12. The I/O drawers can be installed, on the zEC12, both on the A frame and on the Z frame, and on the zBC12 A frame. I/O drawers were first offered with the z10 BC, and each can accommodate up to eight I/O features, in any combination.

PCIe I/O drawer

The *PCIe I/O drawer* was introduced with the z196. This drawer provides for a higher number of cards (four times as much as the I/O drawer and a 14% increase over the I/O cage) and finer port granularity. The PCIe I/O drawers can be concurrently installed and repaired in the field. Each drawer can accommodate up to 32 PCIe I/O features in any combination. Up to five PCIe I/O drawers can be installed on the zEC12 and two on the zBC12.

PCIe I/O features and I/O features

The zEC12 and zBC12 support the following PCIe features, which can be installed only in the PCIe I/O drawers:

- ▶ FICON Express8S
- ▶ OSA-Express5S
- ▶ OSA-Express4S
- ▶ RoCE Express
- ▶ Crypto Express4S
- ▶ Flash Express
- ▶ zEDC Express

When carried forward on an upgrade, the zEC12 and zBC12 also support the following features, which can be installed in the I/O drawers or in an I/O cage:

- ▶ FICON Express8
- ▶ FICON Express4 10 KM LX and SX
- ▶ OSA-Express3 10 GbE LR and SR
- ▶ OSA-Express3 GbE LX and SX
- ▶ OSA-Express3 1000BASE-T Ethernet
- ▶ OSA-Express3-2P 1000BASE-T (for zBC12 only)
- ▶ OSA-Express3-2P GbE SX (for zBC12 only)
- ▶ Crypto Express3
- ▶ Crypto Express3-1P (for zBC12 only)
- ▶ ISC-3 coupling links

In addition, IFB coupling links, which attach directly to the processor books, are supported.

For a description of each I/O feature that is supported by the zEC12 and zBC12, refer to 2.9, “I/O features” on page 46.

Native PCIe and integrated firmware processor (IFP)

All the features residing in the PCIe I/O drawers, I/O drawers, and I/O cages are supported, from a hardware point of view, from the Support Elements. This includes installing and updating the features' Licensed Internal Code and other operational tasks.

Many features have an application-specific integrated circuit (ASIC) that handles the adaptation layer functions required to present, in a uniform manner, the necessary features to the rest of the system. Thus, all the operating systems have the same interface into the I/O subsystem.

With zEC12 and zBC12, System z introduces two features, 10GbE RoCE Express and zEDC Express, with industry standard PCIe adapters, termed native PCIe adapters.

For native PCIe features, there is no adaptation layer, but the device driver is presented in the operating system. The adapter management functions (such as diagnostics and firmware updates) are provided by Resource Groups.

Resource Groups (there are two, for reliability, availability, and serviceability) use the integrated firmware processor (IFP), which is part of the system's base configuration and transparent to customers' operations.

Storage connectivity

Storage connectivity is provided on the zEC12 and zBC12, by Fibre Connection (FICON) features, which support several protocols. Enterprise Systems Connection (IBM ESCON®) features of previous servers are not supported by both the zEC12 and the zBC12¹⁰.

IBM Fibre Connection (FICON) channels

FICON channels follow the Fibre Channel (FC) standards. They support data storage and access requirements, and the latest FC technology in storage and access devices. FICON channels support the following protocols:

- ▶ **Native FICON:** An enhanced protocol (over FC) providing for communication across channels, channel-to-channel (CTC) connectivity, and with FICON devices such as disks, tapes, and printers. Includes the zHigh Performance FICON (zHPF) protocol and is used in z/OS, z/VM, IBM z/VSE® (no zHPF), z/TPF, and Linux on System z environments.
- ▶ **Fibre Channel Protocol (FCP):** A standard protocol for communicating with disk and tape devices through Fibre Channel switches and directors. The FCP channel can connect to FCP SAN fabrics and access FCP/SCSI devices. FCP is used by z/VM, z/VSE, and Linux on System z environments.

There are some restrictions on combining the FICON Express8S, FICON Express8, and FICON Express4 features.

Depending on the feature, auto-negotiated link data rates of 1, 2, 4, or 8 Gbps are supported (1, 2, and 4 for FICON Express4; 2, 4, and 8 for FICON Express8 and FICON Express8S).

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

¹⁰ IBM Facilities Cabling Services - ESCON to FICON migration services are available for managing the transition from ESCON to FICON. This transition includes the flexible migration of ESCON devices to match your expected lifecycle and investment priorities.

Networking connectivity

The zEC12 and zBC12 are fully virtualized systems, able to support many system images. Network connectivity, thus, covers not only those connections between the system and its outside, but also specialized internal connections for inter-system communication.

Open Systems Adapter

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

There are some restrictions on combining the OSA-Express4S and OSA-Express3. The OSA-Express2 10 GbE LR features are not supported, fulfilling the IBM Statement of General Direction.

HiperSockets

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

HiperSockets is a function of the virtualization Licensed Internal Code (LIC) and provides LAN connectivity across multiple system images on the same CPC by performing memory-to-memory data transfers in a secure way. The HiperSockets function eliminates having to use I/O subsystem operations and having to traverse an external network connection to communicate between logical partitions in the same CPC. Therefore, HiperSockets offers significant value in server consolidation by connecting virtual servers and simplifying the Enterprise network.

HiperSockets improved functions (also available on z196 and z114) include the ability to integrate in the *intraensemble data network (IEDN)*, and support for bridging to z/VM virtual switches.

10GbE RoCE Express

The 10 Gigabit Ethernet (10GbE) RoCE Express feature exploits Remote Direct Access Memory over Converged Ethernet and is designed to provide very fast memory-to-memory communications between two CPCs. It is transparent to applications.

Use of the 10GbE RoCE Express feature may help reduce consumption of CPU resources for applications utilizing the TCP/IP stack (such as WebSphere accessing a DB2 database), and may also help reduce network latency with memory-to-memory transfers utilizing Shared Memory Communications - Remote Direct Memory Access (SMC-R) in z/OS V2R1.

Coupling and Server Time Protocol (STP) connectivity

Clustering of System z systems under a single logical system image is possible under z/OS and exploits specialized hardware. Note that z/VM Single System Image has a different design and does not require the features described in this section.

Coupling links

Coupling links are used when clustering zEC12, zBC12, and System z systems running the z/OS operating system on the same or different CPCs. A clustered configuration is known as a *Parallel Sysplex* and can have up to 32 member nodes. The links provide high-speed bidirectional communication between members of a sysplex. The zEC12 and zBC12 support the following links:

- ▶ Internal coupling (IC) links for memory-to-memory transfers between LPARs on the same CPC
- ▶ 12x InfiniBand links for distances up to 150 meters (492 feet)
- ▶ For unrepeated distances up to 10 Km (6.2 miles):
 - 1x InfiniBand links
 - InterSystem Channel-3 (ISC-3) links, when carried forward on an upgrade

Statement of Direction: The IBM zEnterprise EC12 and IBM zEnterprise BC12 are planned to be the last System z servers to offer support of the InterSystem Channel-3 (ISC-3) for Parallel Sysplex environments at extended distances. ISC-3 will not be supported on future System z servers as carried forward on an upgrade. Enterprises should continue migrating from ISC-3 features to InfiniBand Coupling Links.

Special purpose features

This section overviews several features that, although installed in the PCIe I/O drawer or in the I/O drawer, provide specialized functions without actually performing I/O operations, that is, no data is moved between the CPC and externally attached devices.

Cryptography

The Crypto Express4S and Crypto Express3 features provide tamper-sensing and tamper-responding, high-performance cryptographic operations. Each Crypto Express4S and each Crypto Express3-1P feature has one PCI Express adapter and each Crypto Express3 feature has two PCI Express adapters. Each of the adapters can be configured in one of these modes:

- ▶ Secure IBM CCA coprocessor: For secure key encrypted transactions using CCA callable services (default).
- ▶ Accelerator: For public key and private key cryptographic operations that are used with Secure Sockets Layer/Transport Layer Security (SSL/TLS) acceleration.
- ▶ Secure IBM Enterprise PKCS #11 (EP11) coprocessor (Crypto Express4S only): Implements industry standardized set of services that adhere to the PKCS#11.

These features have specialized hardware to perform DES, TDES, AES, RSA, SHA-1, and SHA-2 cryptographic operations. The tamper-resistant hardware security module (HSM), which is contained in the Crypto Express4S and Crypto Express3 features, is designed to meet the FIPS 140-2 Level 4 security requirements for hardware security models.

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

A paper, which is written by atsec information security corporation, on Payment Card Industry compliance, recognizes the inherent qualities of the mainframe and the simplification in the infrastructure that it can provide. For more information, see this website:

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

Flash Express

Flash Express is an optional feature introduced with the zEC12 and also available with the zBC12. When used under z/OS V1R13 and above, Flash Express might help improve availability and handling of paging workload spikes. Using Flash Express can help availability by eliminating slow downs that can occur at the start of the workday. It can also help to eliminate delays that might occur when collecting diagnostic data during failures. Flash Express might, therefore, be able to help organizations meet their most demanding service level agreements.

Flash Express is easy to configure, requires no special skills, and provides rapid time to value. Additional usage of Flash Express is expected to be supported later.

More details about the Flash Express feature are provided in 3.2.4, “Flash Express” on page 82.

IBM zEnterprise Data Compression capability and zEDC Express

The growth of data that needs to be captured, transferred, and stored for large periods of time is not relenting. On the contrary, software implemented compression algorithms are costly in terms of processor resources, and storage costs are not negligible either.

The zEDC Express, an optional feature exclusive to zEC12 and zBC12, addresses that problem by providing hardware-based acceleration for data compression and decompression.

The z/OS V2R1 *IBM zEnterprise Data Compression* acceleration capability exploits the *zEDC Express* feature. It is designed to deliver an integrated solution to help reduce CPU consumption, optimize performance of compression related tasks, and enable more efficient use of storage resources, while providing a lower cost of computing and also helping to optimize the cross-platform exchange of data.

1.1.6 IBM Mobile Systems Remote

IBM Mobile Systems Remote is a mobile application developed by IBM, which is intended to help customers monitor and manage their zEnterprise environment from a mobile device.

By interfacing with the zEnterprise HMC, the application allows authorized personnel to hold on the palm of their hands almost all the information normally viewed on the HMC. Customers will be able to monitor their zEnterprise CP and, in case of an ensemble, also the IBM BladeCenters and installed blades in the zBX.

For more information on this freely downloadable application and links to the different application stores, see the IBM Mobile Systems Remote web site:

<http://ibmremote.com/>

1.1.7 Hardware Management Console and Support Elements

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

For an example of the HMC and SE connectivity, see Figure 2-13 on page 59.

Role of the Hardware Management Console in an ensemble

When part of an ensemble, the HMC and SE have additional roles, supplied through the zEnterprise Unified Resource Manager, which is installed in the Hardware Management Console (HMC) and, alongside other functionality, enables extending those tasks to an ensemble.

The HMC allows viewing and managing multinodal configurations with virtualization, I/O networks, support networks, power subsystems, cluster connectivity infrastructure, and storage subsystems. The HMC has a management responsibility for the entire ensemble, while the SE has management responsibility at the individual node level. An ensemble is managed by a primary/alternate HMC pair.

The HMC possesses a highly interactive and dynamic web-based user interface. The views, management, and monitoring tasks of the HMC user interface provide everything needed for complete management of the virtual servers' lifecycle across the zEnterprise hypervisors:

- ▶ IBM PR/SM™
- ▶ z/VM¹¹
- ▶ IBM PowerVM Enterprise Edition
- ▶ System x blades integrated hypervisor (using Kernel-based virtual machines)

Virtual servers can be managed from their inception all the way through monitoring, migration, and policy-based administration during their deployment. Functions also include instantiating an ensemble, defining virtual servers and workloads, and assigning those virtual servers to one or more workloads.

1.1.8 Capacity on demand and performance

In the same footprint, the zEC12 101-way system can deliver up to 50% more capacity than the largest 80-way z196. The zEC12 1-way system has approximately 25% more capacity than the z196 1-way. Numerous improvements in the processor chip design, including new instructions, refinements to out-of-order execution, and restructured caches contribute to the additional capacity. Exploitation of some of the functionality is available only by using the latest levels of compilers and JVMs.

Similarly, the zBC12 6-way (z/OS environment) can deliver up to 58% more capacity than the z114 5-way. In a Linux environment, a zBC12 13-way can deliver as much as 62% more capacity than z114-10 way. The zBC12 uses the same chip as the zEC12, providing similar enhancements, when compared to the z114, as the zEC12 when compared to the z196.

The zEC12 and the zBC12 continue to offer all the specialty engines available with z196 and z114. See "Processor unit characterization" on page 9.

The zEC12 and the zBC12 enhance the availability and flexibility of just-in-time deployment of more system resources, which is known as *Capacity on Demand (CoD)*.

On the zEC12 and zBC12, it is possible to perform just-in-time deployment of processor capacity resources. The Capacity on Demand (CoD) function is designed to provide more flexibility, granularity, and responsiveness to business requirements changes by allowing the user to dynamically change the available system capacity.

A similar capability is not available with the zBX. For more information, see 3.7, "zEnterprise BladeCenter Extension Model 003" on page 108.

¹¹ z/VM V6R2 only.

1.2 IBM zEnterprise BladeCenter Extension

The *IBM zEnterprise BladeCenter Extension Model 003 (zBX)* is available as an option with the zEC12 and zBC12 systems. This model provides several distributed environments (IBM AIX® on IBM POWER7®, Linux on System x, and Microsoft Windows on System x) on a blade form factor, which are connected to the zEnterprise CPCs through virtual LANs supported on a high-speed private network.

The zBX is managed through the Support Elements of its controlling zEC12 or zBC12 and by using the Unified Resource Manager functions. In bringing together multiple platforms, IBM has created a scalable solution that simplifies hardware and firmware management and support, and the definition and management of a network of virtualized servers.

The zBX Model 003 consists of the following components:

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

The INMN provides connectivity for management purposes between the SE and zBX of the CPC.

- ▶ Two TOR 10 GbE switches for the IEDN¹
- ▶ The IEDN is used for data paths between the zEnterprise CPCs and the zBX, and the other ensemble members.
- ▶ Eight Gbps Fibre Channel switch modules for connectivity to an SAN
- ▶ Power distribution units (PDU) and cooling fans

The zBX is configured with redundant components to provide qualities of service similar to that of zEC12 or zBC12, such as firmware management and the capability for concurrent upgrades and repairs.

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

GDPS/PPRC and GDPS/GM support zBX hardware components, providing workload failover for automated multi-site recovery. These capabilities can help facilitate the management of planned and unplanned outages across zEC12 or zBC12.

Statement of Direction: IBM intends to deliver new functionality with IBM Systems Director offerings to support the IBM zBX. Such planned new capabilities will be designed to provide virtual image management and enhanced energy management functions for IBM Power Systems™ and System x blades.

The zBX Model 003 can be controlled only from a zEC12 or zBC12, and the zBX Model 002 can be controlled only from a z196 or z114. However, zEC12 and zBC12 can use a zBX Model 002 controlled by a z196 or a z114, and vice-versa, by connecting the CPCs to the IEDN TOR switch or by using the zBX to zBX IEDN switch ports.

A zBX Model 002 controlled by a z196 or a z114 must be upgraded to Model 003 at the time of upgrading the controlling z196 or z114 to, respectively, a zEC12 or zBC12. During the upgrade, the virtualization and configuration data is preserved; however, the process is disruptive, and requires a planned outage.

IBM blades

IBM offers a selected set of IBM blades that can be installed and operated on the zBX Model 003. These blades were tested to ensure compatibility and manageability in the zEnterprise environment. The following blades are available:

- ▶ Select IBM POWER7 PS701 Express blades
- ▶ Select IBM System x blades (HX5 7873 dual-socket 16-core)
- ▶ IBM DataPower® XI50z blades (double-width)

The POWER7 blades offer a virtualized environment through the IBM PowerVM® Enterprise Edition hypervisor. The virtual servers run the AIX operating system. The System x blades have an integrated hypervisor using *Kernel-based virtual machines (KVM)*, which provides a virtualized environment for running the Linux and Windows operating systems.

IBM WebSphere DataPower Integration Appliance X150 for zEnterprise

The *IBM WebSphere® DataPower Integration Appliance X150 for zEnterprise (DataPower XI50z)* is integrated into the zEnterprise infrastructure. DataPower XI50z is a multifunctional appliance that can help provide multiple levels of XML optimization. It can also streamline and secure valuable service-oriented architecture (SOA) applications, and provide drop-in integration for heterogeneous environments by enabling core Enterprise Service Bus (ESB) functionality, including routing, bridging, transformation, and event handling. This appliance can help to simplify, govern, and enhance the network security for XML and web services.

For a more detailed description of the IBM DataPower X150z integration appliance, see this website:

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

1.3 Unified Resource Manager

The zEC12 and zBC12 systems perfectly fit in a smart infrastructure, continuing IBM high-end systems' leadership and being both the next step in the evolution of mainframes and a premier solution for centrally managed enterprise environments including clouds. They integrate a true hybrid computing system that is composed of virtualized heterogeneous resources that are integrated and managed as a single system by the IBM zEnterprise Unified Resource Manager.

The Unified Resource Manager is an integral part of the zEC12 and zBC12 systems. It provides end-to-end management of CPC and zBX resources, and of virtualized environments, with the ability to align those resources according to individual workload requirements.

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

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

The functions that pertain to an ensemble are provided by the Hardware Management Console (HMC) and Support Elements. For more information, see 3.5, “Hardware Management Console functions” on page 104.

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

Resource management suites

The functions that are delivered by the Unified Resource Manager are accessed through the Hardware Management Console (HMC) and provide the following capabilities:

- ▶ Integrated hardware management across all elements of the system, the CPC, the zBX, and the integrated networks.
- ▶ Fully automatic and coherent integrated resource discovery and inventory for all elements of the system without requiring user configuration, deployment of libraries or sensors, or user scheduling.
- ▶ Hypervisors for the zBX components are shipped, serviced, and deployed as Licensed Internal code (LIC). They are booted automatically at power-on reset, and managed through the isolated intranode management network (INMN).
- ▶ Virtual server lifecycle management, enabling uniform directed and dynamic virtual server provisioning across all zBX elements’ hypervisors from a single point of control.
- ▶ Representation of the physical and virtual resources that are used in the context of a deployed business function as a named workload.
- ▶ Monitoring and trend reporting of CPU energy efficiency, which can be helpful in managing the costs of deployed workloads.
- ▶ Delivery of system activity through a user interface, the Monitors Dashboard (which augments the existing System Activity Display), enabling a broader and more granular view of system resources consumption.

The Unified Resource Manager offers the ability to optimize technology deployment according to individual workload requirements. To achieve this optimization, the Unified Resource Manager is delivered in two suites of tiered functionality:

- ▶ Manage
- ▶ Automate/Advanced Management

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

- ▶ *Building an Ensemble Using IBM zEnterprise Unified Resource Manager*, SG24-7921
- ▶ *IBM zEnterprise EC12 Technical Guide*, SG24-8049
- ▶ *IBM zEnterprise BC12 Technical Guide*, SG24-8138

1.4 Reliability, availability, and serviceability (RAS)

The zEC12 and zBC12 continue to offer the high quality of service and *reliability, availability, and serviceability (RAS)* that is traditional in IBM mainframes.

The RAS strategy employs a building-block approach developed to meet the client's stringent requirements for achieving continuous reliable operation. Those building blocks are error prevention, error detection, recovery, problem determination, service structure, change management, measurement, and analysis.

Most hardware upgrades can be installed concurrently. The zEC12 and the zBC12 reach new availability levels by eliminating various pre-planning needs and other disruptive operations.

The RAS strategy is focused on a recovery design that is necessary to mask errors and make them transparent to client operations. One example is the use of *Redundant Array of Independent Memory (RAIM)*, a concept similar to RAID (disk). An extensive hardware recovery design is implemented to detect and correct array faults. In cases where total transparency cannot be achieved, the system can restart with the maximum possible capacity.

The IBM mainframe systems have gone through decades of intense engineering development. The introduction of zEC12 and zBC12 adds, once again, new, carefully engineered RAS features, providing the highest possible level of RAS.

For a more detailed description of the RAS features, see the corresponding chapter in the *IBM zEnterprise EC12 Technical Guide*, SG24-8049 or the *IBM zEnterprise BC12 Technical Guide*, SG24-8138.

1.5 Software

The IBM zEnterprise EC12 and IBM zEnterprise BC12 are supported by a large set of software, including over 7200 independent software vendor (ISV) applications. The extensive software portfolio available for the zEC12 and zBC12 spans from mobile offerings, IBM WebSphere and full support for service-oriented architecture (SOA), web services, Java Platform Enterprise Edition, Linux, and open standards, to the more traditional batch and transactional environments. Examples of these types of environments include IBM Customer Information Control System (CICS®) and IBM Information Management System (IBM IMS™).

In addition, any AIX products that run today on IBM POWER servers continue to run on the zBX POWER7 blades' virtualized AIX environment, and System x blades support Linux and Windows. There are also specialized solutions such as the DataPower XI50z appliance.

Operating systems

Use of some features might require the latest releases. The following operating systems are supported by the zEC12 and zBC12:

- ▶ z/OS Version 2 Release 1
- ▶ z/OS Version 1 Release 13 with PTFs
- ▶ z/OS Version 1 Release 12 with PTFs
- ▶ z/OS Version 1 Release 11 with the IBM Lifecycle Extension with PTFs
- ▶ z/VM Version 6 Release 3 with PTFs
- ▶ z/VM Version 6 Release 2 with PTFs
- ▶ z/VM Version 5 Release 4 with PTFs
- ▶ z/VSE Version 4 Release 3 or later, with PTFs

- ▶ z/TPF Version 1 Release 1
- ▶ Linux on System z distributions:
 - SUSE: SLES 10 and SLES 11¹²
 - Red Hat: RHEL 5¹³ and RHEL 6

The following operating systems support IBM blades on the zBX Model 003:

- ▶ For the POWER7 blades: AIX Version 5 Release 3 or later, with the PowerVM Enterprise Edition
- ▶ For the System x blades:
 - Linux on System x (64-bit only):
 - Red Hat: RHEL 5.5 and up, RHEL 6.0 and up
 - SUSE: SLES 10 (SP4) and up, SLES 11 SP1 and up
- ▶ Microsoft Windows Server 2012, Windows Server 2008 R2, and Windows Server 2008 SP2 (Datacenter Edition is recommended), 64-bit only

IBM compilers

Compilers are built with knowledge about the system's processors and cache topologies, which is used in code generation. Using the latest compilers is, thus, essential to extract the maximum benefit of a server's new capabilities. Empower your business applications with IBM compilers on the IBM zEnterprise System.

With IBM Enterprise COBOL and Enterprise PL/I, decades of IBM experience in application development can be used to integrate COBOL and PL/I with web services, XML, and Java. Such interoperability enables capitalizing on existing IT investments while smoothly incorporating new, web-based applications into your organizations' infrastructure.

z/OS XL C/C++ helps creating and maintaining critical business applications that are written in C or C++ to maximize application performance and improve developer productivity. z/OS XL C/C++ can transform C or C++ source code to fully use System z hardware, including zEnterprise. This function is possible through hardware tailored optimizations, built-in functions, performance-tuned libraries, and language constructs that simplify system programming and boost application runtime performance.

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

More information about software support can be found in Chapter 4, "Operating system support and considerations" on page 119.

¹² SLES is the abbreviation for SUSE Linux Enterprise Server.

¹³ RHEL is the abbreviation for Red Hat Enterprise Linux.



Hardware overview

The IBM zEnterprise EC12 and the IBM zEnterprise BC12 are the next step in the evolution of the mainframe family. They continue this evolution by introducing several innovations and expanding existing functions, building upon the z/Architecture.

The zEC12 and zBC12 are designed to deliver new levels of performance and capacity for large-scale consolidation and growth, and in support of cloud infrastructures, support for the next generation of digital signature security, cutting edge pattern recognition analytics for smart monitoring of system health, and enhanced environmental capabilities.

This chapter expands upon the overview of key hardware elements of the zEC12 and zBC12 provided in 1.1, “zEC12 and zBC12 technical description” on page 4, and compares them with the predecessor IBM System z systems, where relevant.

This chapter describes the following topics:

- ▶ zEC12 and zBC12 models
- ▶ Frames
- ▶ zEC12 processor cage, books, and multiple chip modules
- ▶ zBC12 processor drawer and single chip modules
- ▶ Processor chip
- ▶ Processor unit
- ▶ Memory
- ▶ I/O system structure
- ▶ I/O features
- ▶ Cryptographic features
- ▶ Coupling and clustering
- ▶ Time functions
- ▶ Hardware Management Console and Support Element
- ▶ Power and cooling
- ▶ zEnterprise BladeCenter Extension

2.1 zEC12 and zBC12 highlights, models, and upgrades

The zEC12 and zBC12 models, as well as the improvements and upgrades over their predecessors, are discussed in this section.

2.1.1 zEC12 highlights

The IBM zEnterprise EC12 major improvements over its predecessors include the following features:

- ▶ Increased total system capacity in a 120-way system (with 101 characterizable PUs) and more subcapacity settings, offering increased levels of performance and scalability to help enable new business growth.
- ▶ Hex-core 5.5 GHz processor chips that can help improve the execution of processor-intensive workloads.
- ▶ Up to 3 TB of available real memory per system (with up to 1 TB real memory per logical partition) ensuring high availability in the memory subsystem by the proven technology of *redundant array of independent memory (RAIM)*.
- ▶ A 32 GB fixed *hardware system area (HSA)* that is managed separately from client-purchased memory (double the size of the z196).
- ▶ Third-generation high frequency, second-generation out-of-order design, with numerous pipeline improvements, based on z10 and z196 designs, and the number of instructions in flight is increased by 25%.
- ▶ Cache structure improvements and larger cache sizes that can benefit most production workloads. Achieved by a new structure for second level private cache and doubling the cache size as compared to z196 on third level (48 MB) and fourth level cache (384 MB).
- ▶ Improved cryptography functionality and performance achieved by the introduction of one dedicated Co-processor per core.
- ▶ Channel subsystem enhancement for I/O resilience. The zEC12 channel subsystem has enhanced channel path selection algorithms that are designed to provide improved throughput and I/O service times when abnormal conditions occur. See “Modified Indirect Data Address Word (MIDAW) facility” on page 86 for more information.

zEC12 also introduces several features and functions:

- ▶ Flash Express, implemented on flash solid-state drives (SSD). These are mounted in Flash Express feature cards and can be used to handle paging workload spikes and improve availability.
- ▶ The Crypto Express4S feature, with enhanced support of cryptographic functions.
- ▶ I/O features, including the OSA-Express4S 1000 BASE-T, implemented with the PCI Express (PCIe) I/O infrastructure.
- ▶ I/O features, including the OSA-Express5S, implemented with the PCI Express (PCIe) I/O infrastructure and native PCIe features 10GbE RoCE Express and zEDC Express.
- ▶ The IBM System z Advanced Workload Analysis Reporter (IBM zAware) software appliance, which provides a smart solution for detecting and diagnosing anomalies in z/OS systems.

The driver 15 level of the Licensed Internal Code introduces support for the following functions:

- ▶ IBM zEnterprise Data Compression (zEDC), designed to reduce CPU consumption, optimize performance of compression related tasks, and enable more efficient use of storage resource, helping to provide a lower cost of computing.
- ▶ Remote Direct Memory Access (RDMA) over 10GbE RoCE (RDMA over Converged Enhanced Ethernet) through the use of the SMC-R (Shared Memory Communications - Remote Direct Memory Access) protocol. This feature is designed to allow high speed memory-to-memory data movement between zBC12/zEC12 Systems. It is supported by z/OS using SMC-R. It improves network latency and throughput, reducing CPU overhead, z/OS network congestion and cost related to remote off stack data movement.

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

For an in-depth discussion of the IBM zEnterprise EC12 functions and features, see the *IBM zEnterprise EC12 Technical Guide*, SG24-8049.

2.1.2 zEC12 models

The zEC12 has an assigned machine type (M/T) of 2827, which uniquely identifies the *central processor complex (CPC)*. The zEC12 is offered in five models:

- ▶ zEC12 H20. Includes one book and a maximum of 20 customizable *processor units (PU)*.
- ▶ zEC12 H43. Includes two books and a maximum of 43 customizable PUs.
- ▶ zEC12 H66. Includes three books and a maximum of 66 customizable PUs.
- ▶ zEC12 H89. Includes four books and a maximum of 89 customizable PUs.
- ▶ zEC12 HA1. Includes four books and a maximum of 101 customizable PUs.

The model determines the maximum number of processor units (PU) that are available for characterization. PUs are delivered in single-engine increments. The first four models use 27-PU multiple chip modules (MCM), the fifth model, HA1, uses 30-PU MCMs to provide a maximum of 101 configurable PUs.

Spare PUs and system assist processors (SAP) and one integrated firmware processor (IFP) are integral to the system. Table 2-1 provides a model summary that includes SAPs and spare PUs for the various models. For an explanation of PU characterization, see “Processor unit characterization” on page 40.

Table 2-1 zEC12 Model summary

Model	Books/PUs	CPs	Standard SAPs	Spares	Integrated firmware processor
H20	1/27	0–20	4	2	1
H43	2/54	0–43	8	2	1
H66	3/81	0–66	12	2	1
H89	4/108	0–89	16	2	1
HA1	4/120	0–101	16	2	1

The zEC12 offers 161 capacity levels. There are 101 capacity levels that are given by the number of physically used *central processors (CP)*, plus the possibility of 60 subcapacity models for the first 20 CPs. There is also one model for all *Integrated Facility for Linux (IFL)* or all *Internal Coupling Facility (ICF)* configurations. This topic is described in more detail in 3.2.2, “Memory” on page 79.

2.1.3 zEC12 upgrades

Figure 2-1 summarizes the upgrade paths to the zEC12.

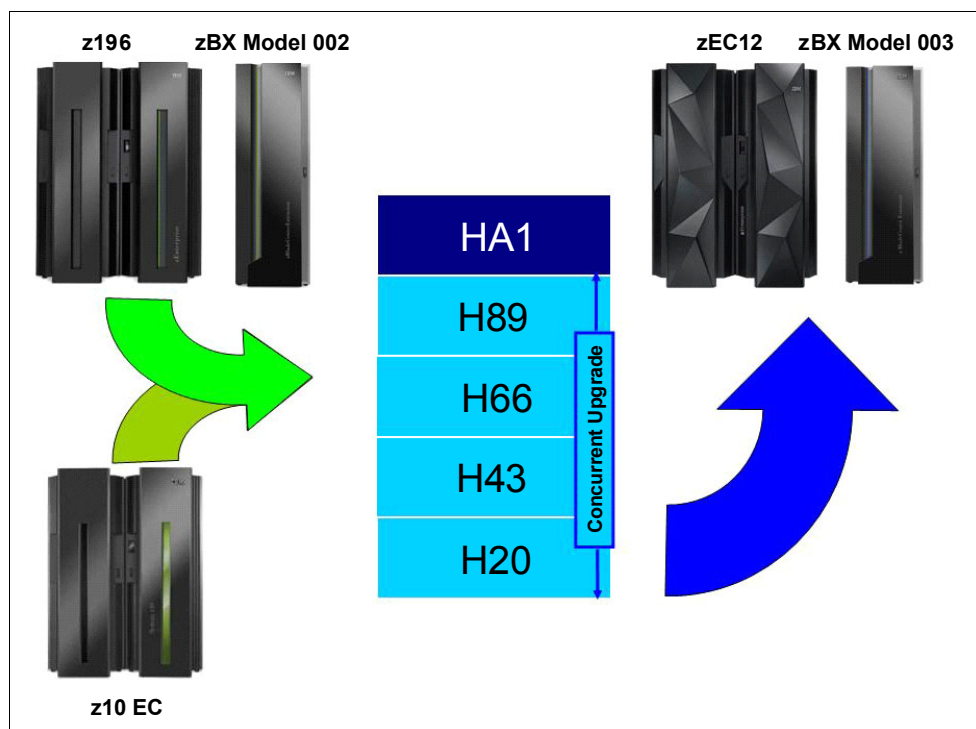


Figure 2-1 zEC12 upgrade paths

Concurrent upgrades of CPs, IFLs, ICFs, zAAPs, zIIPs, or SAPs are available for the zEC12. However, concurrent PU upgrades require that more PUs are installed (at a previous time), but not activated.

If an upgrade request cannot be accomplished within the specified configuration, a hardware upgrade is required. The upgrade enables the addition of one or more books to accommodate the wanted capacity. Additional books can be installed concurrently. Upgrades from any zEC12 (model H20, H43, H66, H89) to a model HA1, are disruptive because this upgrade requires the replacement of all installed books.

Spare PUs are used to replace defective PUs. There are always two spare PUs on a zEC12. In the rare event of a PU failure, a spare PU is concurrently and transparently activated, and assigned the characteristics of the failing PU.

When a z196 with a zBX Model 002 is upgraded to zEC12, the zBX is converted to a Model 003. The virtualization and configuration data is preserved, however the process is disruptive, and requires a planned outage.

2.1.4 zBC12 highlights

The IBM zEnterprise BC12 major improvements over its predecessors include the following features:

- ▶ Increased total system capacity in a 18-way system (with 13 characterizable PUs) and more subcapacity settings, offering increased levels of performance and scalability to help enable new business growth.
- ▶ Hex-core processor chips with a clock speed of 4.2 GHz, that can help improve the execution of processor-intensive workloads.
- ▶ Up to 512 GB of available real memory per system ensuring high availability in the memory subsystem by the proven technology of *redundant array of independent memory (RAIM)*.
- ▶ A 16 GB fixed *hardware system area (HSA)* that is managed separately from client-purchased memory (double the size of the z114).
- ▶ Third-generation high frequency, second-generation out-of-order design with numerous pipeline improvements, based on z10BC and z114 designs, and the number of instructions in flight is increased by 25%.
- ▶ Cache structure improvements and larger cache sizes that can benefit most production workloads. Achieved by a new structure for second level private cache and doubling the cache size as compared to z114 on third level (24 MB) and fourth level cache (192 MB).
- ▶ Improved cryptography functionality and performance achieved by the introduction of one dedicated co-processor per core.
- ▶ Channel subsystem enhancement for I/O resilience. The zBC12 channel subsystem has enhanced channel path selection algorithms that are designed to provide improved throughput and I/O service times when abnormal conditions occur. See “Modified Indirect Data Address Word (MIDAW) facility” on page 86 for more information.

zBC12 also introduces several features and functions when compared with the z114:

- ▶ Flash Express, implemented on flash solid-state drives (SSD). These are mounted in Flash Express feature cards and can be used to handle paging workload spikes and improve availability.
- ▶ The Crypto Express4S feature, with enhanced support of cryptographic functions.
- ▶ I/O features, including the OSA-Express5S, implemented with the PCI Express (PCIe) I/O infrastructure and native PCIe features 10GbE RoCE Express and zEDC Express.
- ▶ The IBM System z Advanced Workload Analysis Reporter (IBM zAware) software appliance, which provides a smart solution for detecting and diagnosing anomalies in z/OS systems.
- ▶ The IBM zEnterprise Data Compression (zEDC), designed to reduce CPU consumption, optimize performance of compression related tasks, and enable more efficient use of storage resources, helping to provide a lower cost of computing.
- ▶ Remote Direct Memory Access (RDMA) over 10GbE RoCE (RDMA over Converged Enhanced Ethernet) through the use of the SMC-R (Shared Memory Communications - Remote Direct Memory Access) protocol. This feature is designed to allow high speed memory-to-memory data movement between zBC12/zEC12 Systems. It is supported by z/OS using SMC-R. It improves network latency and throughput, reducing CPU overhead, z/OS network congestion and cost related to remote off stack data movement.

2.1.5 zBC12 models

The zBC12 has an assigned machine type (M/T) of 2828, which uniquely identifies the *central processor complex (CPC)*. The zBC12 is offered in two models:

- zBC12 H06. Includes one processor drawer and a maximum of 6 customizable *processor units (PU)*.
- zBC12 H13. Includes two processor drawers and a maximum of 13 customizable PUs.

The model determines the maximum number of processor units (PU) that are available for characterization. PUs are delivered in single-engine increments. The two models use 3 single chip modules (SCM) per drawer (two PUs and one System Control (SC)), of which 6 or 13 PUs are available for characterization.

Spare PUs, system assist processors (SAP) and one integrated firmware processor (IFP) are integral to the system. See Table 2-2.

Table 2-2 zBC12 Model summary

Model	Drawers/PUs	CPs	Standard SAPs	Dedicated Spares	IFP
H06	1/9	0–6	2	0	1
H13	2/18	0–6	2	2	1

The zBC12 offers 156 capacity levels. CP processors can be configured in any of 26 levels (from A to Z - 26 capacity indicators).

2.1.6 zBC12 upgrades

Figure 2-2 summarizes the upgrade paths to the zBC12.

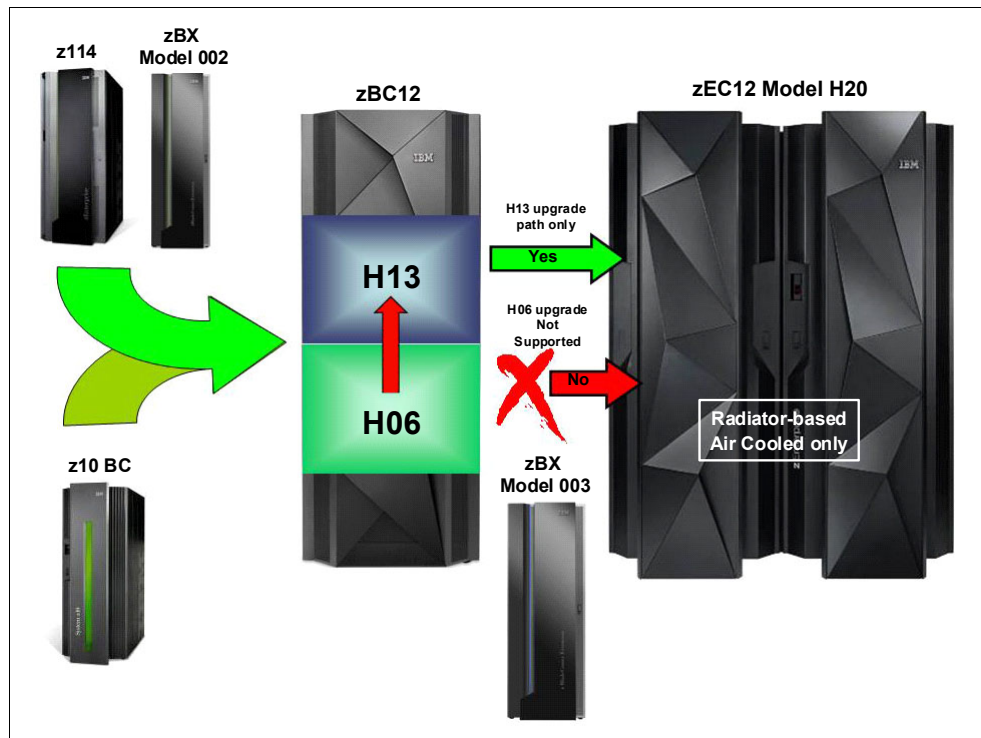


Figure 2-2 zBC12 upgrade paths

Any z10 BC or z114 can be upgraded to a zBC12. There is an upgrade path from the zBC12 H13 to a zEC12 radiator-based air cooled system.

An upgrade from zBC12 model H06 to a zBC12 model H13 is disruptive.

Concurrent upgrades of CPs, IFLs, ICFs, zAAPs, or zIIPs are available on the zBC12. However, concurrent PU upgrades require that more PUs are installed (at a previous time), but not activated.

If an upgrade request cannot be accomplished within the specified configuration, a hardware upgrade is required. The upgrade enables the addition of a second processor drawer to accommodate the wanted capacity and is disruptive.

Spare PUs are used to replace defective PUs. There is no dedicated spare PU on zBC12 H06, however the zBC12 H13 has two dedicated spare PUs. In the rare event of a PU failure, a spare or an unassigned PU is concurrently and transparently activated, and assigned the characteristics of the failing PU.

When a z114 with a zBX Model 002 is upgraded to zBC12, the zBX is converted to a Model 003. The virtualization and configuration data is preserved, however the process is disruptive, and requires a planned outage.

2.2 Frames

The frames of the zEC12 and zBC12 are described in this section.

2.2.1 zEC12 frames

The zEC12 is always a two-frame system. The frames are called the *A Frame* and the *Z Frame*. The zEC12 can be delivered as an air-cooled system or as a water-cooled system.

Figure 2-3 shows an internal front view of the two frames for an air-cooled CPC. The number and type of I/O drawers can vary and depends on the number of I/O features. For a new build system, all I/O drawers are PCIe drawers, with a maximum of five. Miscellaneous Equipment Specifications (MES) can carry forward up to one I/O cage and up to two I/O drawers (8 slot I/O drawer). This configuration provides for a maximum number of 44 non-PCIe feature cards that can be carried forward.

In Figure 2-3, the system is shown with the maximum of five PCIe I/O drawers.

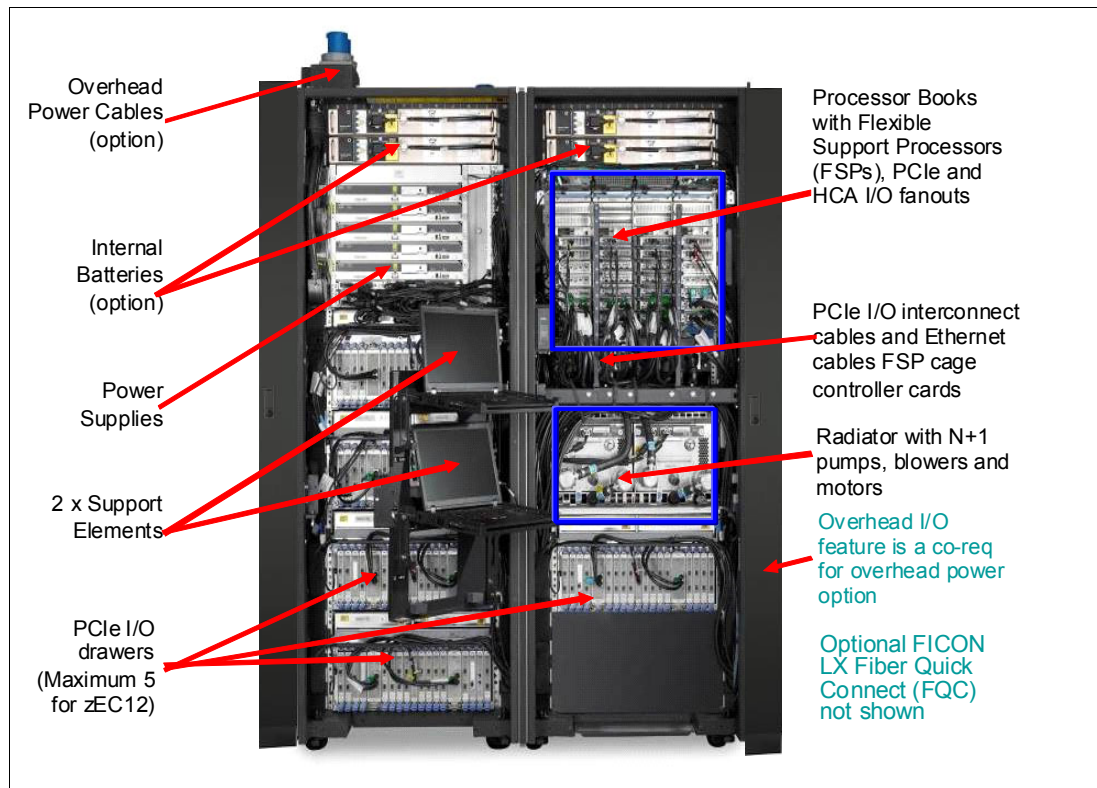


Figure 2-3 zEC12 internal front view: air-cooled CPC

Figure 2-4 shows an internal front view of the two frames of a water-cooled CPC. The I/O cage that is shown in the A Frame is carried forward and is not installed in a new build system.

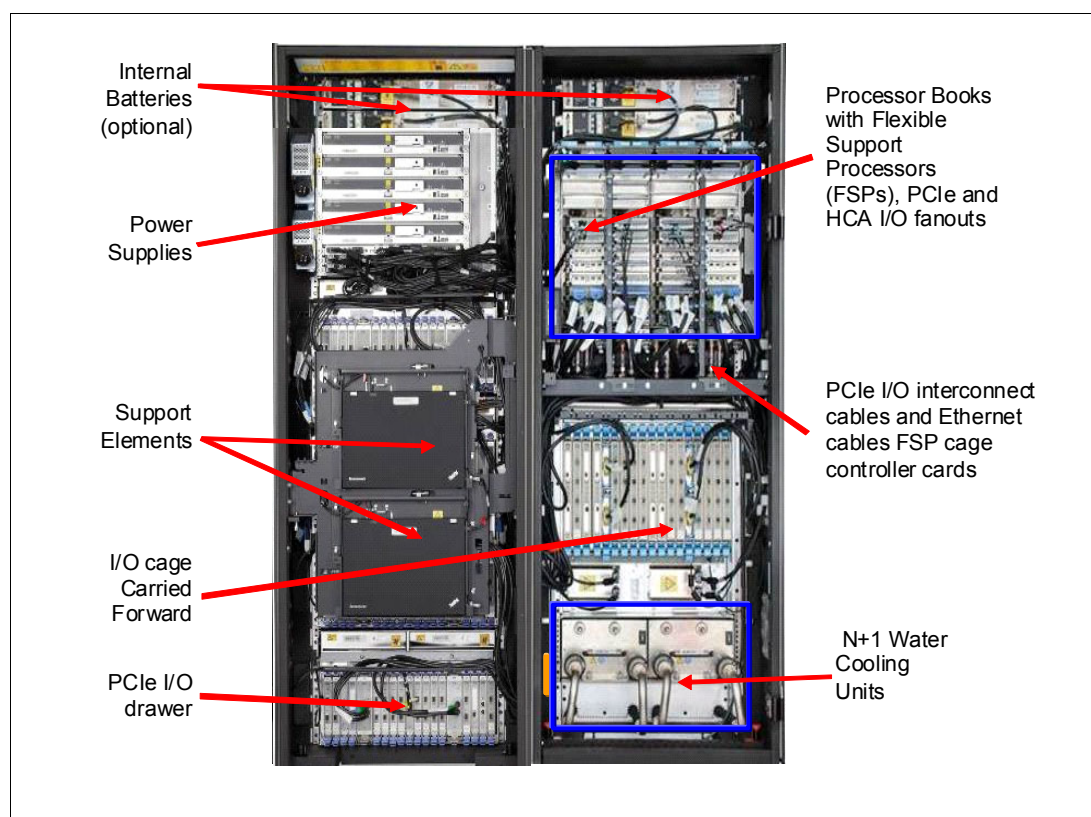


Figure 2-4 zEC12 internal front view: water-cooled CPC

Top exit I/O and power cabling

zEC12 and its predecessor, the z196, have the option of ordering the infrastructure to support the top exit of fiber optic cables (FICON, OSA, 12x InfiniBand, 1x InfiniBand, ISC-3 and RoCE) and copper cables for the 1000BASE-T Ethernet features.

On the zEC12, the top exit capability is designed to provide an additional option, the overhead power cabling option. Figure 2-3 on page 32 shows this overhead power cable feature, present on the Z Frame. Instead of all the cables exiting under the CPC to under the raised floor, there is now the flexibility to choose the options that best meet the data center requirements. A non-raised floor installation of the zEC12 air-cooled systems is also possible. Top exit cabling can also help to increase air flow. These options are offered on new build and MES orders.

2.2.2 zBC12 frame

System z frames are enclosures that are built to Electronic Industry Association (EIA) standards. The zBC12 central processor complex (CPC) has one 42U EIA frame, which is shown in Figure 2-5. The frame has locations for one or two processor drawers and a combination of I/O drawers or PCIe I/O drawers.

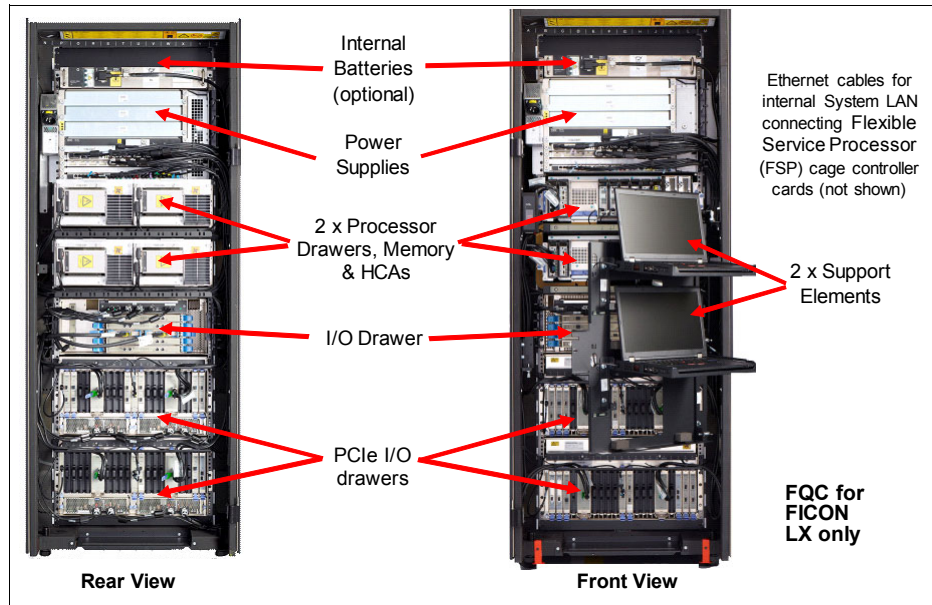


Figure 2-5 zBC12 front and rear view of processor drawers and I/O drawers

The number and type of I/O drawers can vary and depends on the number of I/O features. For a new build system, all I/O drawers are PCIe drawers, with a maximum of two. Miscellaneous Equipment Specifications (MES) upgrade can carry forward up to one I/O drawer. The maximum configuration will be two PCIe drawers plus one I/O drawer.

2.3 zEC12 processor cage, books, and multiple chip modules

The zEC12 system has a multi-book system structure similar to the z196. Up to four books can be installed on a CPC. A book looks like a box and plugs into one of the four slots of the processor cage, which resides in the A frame of the CPC. Figure 2-3 on page 32 provides a pictorial view of the processor cage and the location of the four books. The location and structure of the processor cage is the same for an air-cooled or a water-cooled system.

Each book contains the following elements:

- ▶ A multiple chip module (MCM):

Each MCM includes six hex-core PU chips and two storage control (SC) chips. MCMs are described in “zEC12 multiple chip module” on page 36. See Table 2-1 on page 27 for the model summary and the relation between the number of books and number of available PUs.

- ▶ Memory:

A minimum of 32 and a maximum of 768 GB of memory for client use. See Table 2-3 on page 41 for details.

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

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

- Three Distributed Converter Assemblies (DCA) that provide power to the book:
Loss of one of the DCAs leaves enough power to satisfy the power requirements of the book. The DCAs can be concurrently maintained.

Figure 2-6 shows a view of a zEC12 book without the containing box.

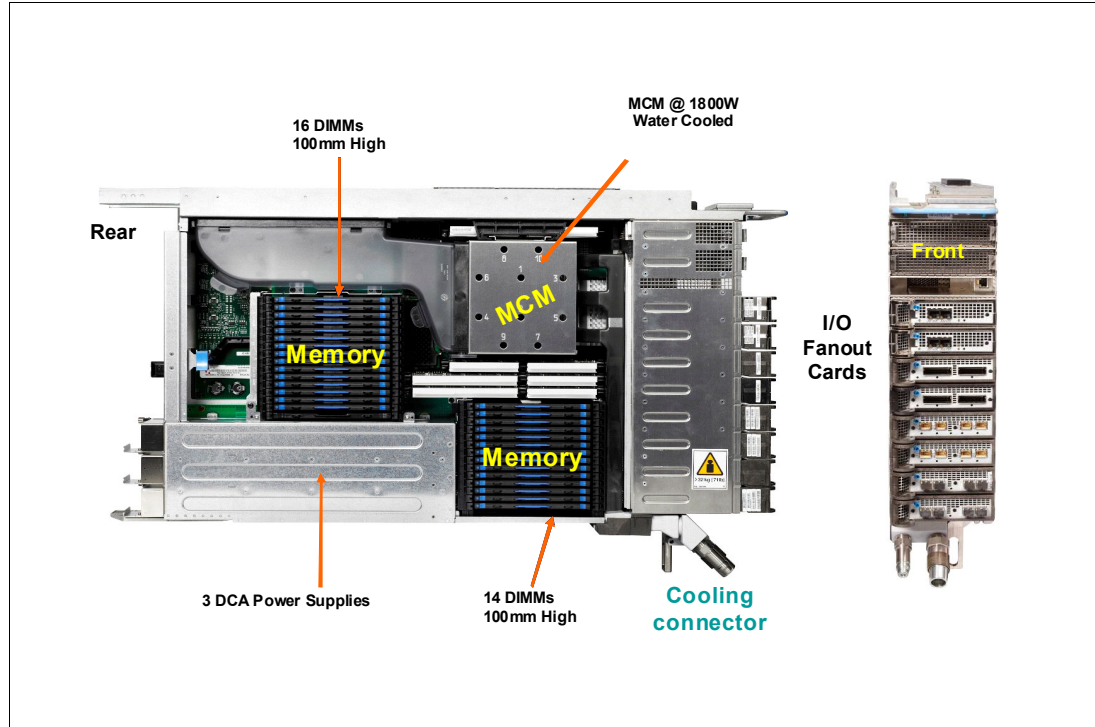


Figure 2-6 zEC12 book structure and components

The zEC12 continues the significant increase in system scalability and opportunity for server consolidation by providing a multi-book system structure. As shown in Figure 2-7, all books are interconnected in a star configuration with high-speed communications links through the L4 shared caches. The zEC12 has 384 MB of L4 cache per book, which is the double of its predecessor. This star design allows the system to be operated and controlled by the PR/SM facility as a symmetrical, memory-coherent multiprocessor.

Figure 2-7 shows the zEC12 inter-book communication structure.

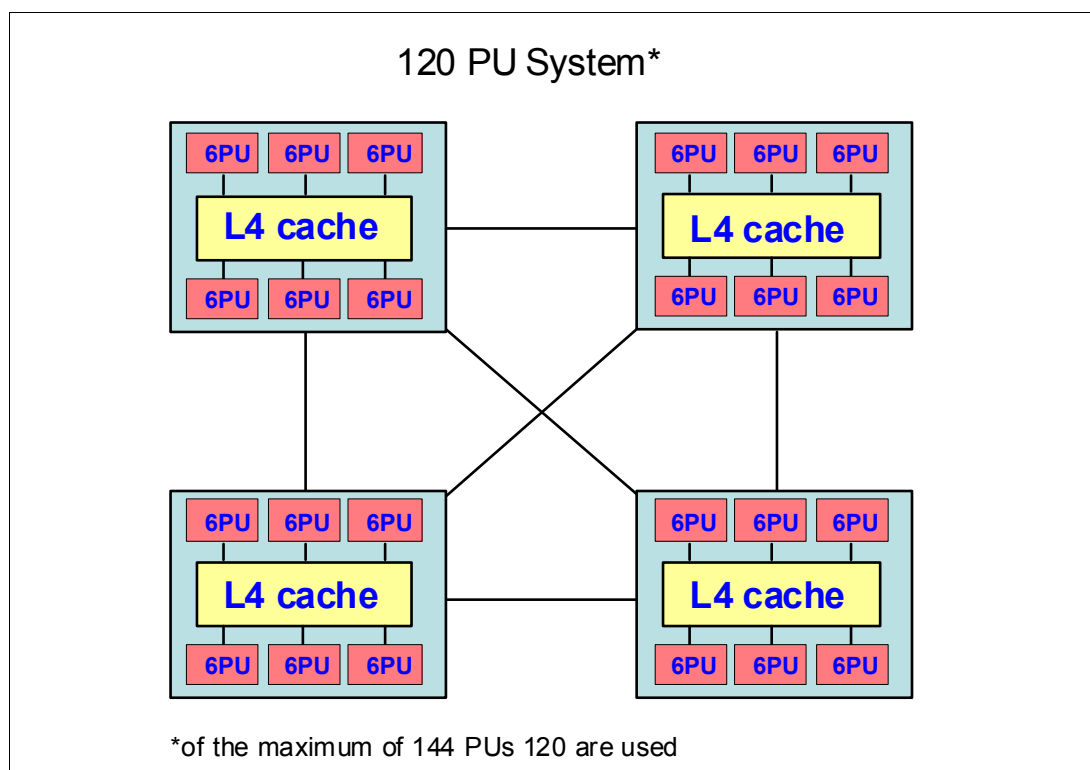


Figure 2-7 zEC12 inter-book communication structure

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

zEC12 multiple chip module

The *multiple chip module (MCM)* (Figure 2-8) is a high-performance, glass-ceramic module, providing the highest level of processing integration in the industry. It is the heart of the system.

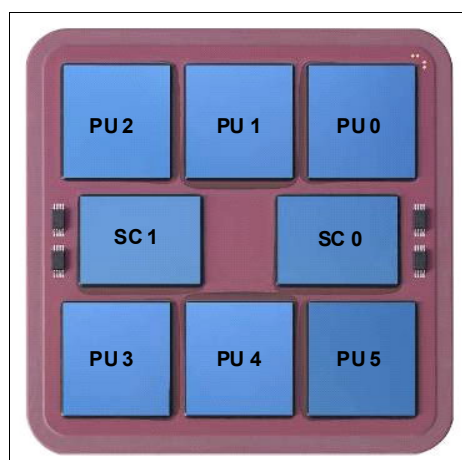


Figure 2-8 zEC12 multiple chip module

The zEC12 MCM has eight chip sites. All chip types on the MCM use Complementary Metal Oxide of Silicon (CMOS) 13S chip technology. CMOS 13S is a state-of-the-art microprocessor technology that is based on 15-layer copper interconnections and Silicon-On-Insulator (SOI) technologies. The chip lithography line width is 0.032 μm (32 nm). The processor unit chip contains close to 2,750,000,000 transistors in a 597.24 mm^2 die.

There is one MCM per book. The MCM contains all of the processor chips and the L1 - L4 caches of the book, as shown in Figure 2-9. The zEC12 has six PU chips for each MCM and each PU chip has up to six active PUs (cores). Two MCM options are offered: with 27 or 30 PUs. All the models employ an MCM size of 27 PUs except for the HA1 model, which has four books with 30 PU MCMs, for a total of 120 PUs.

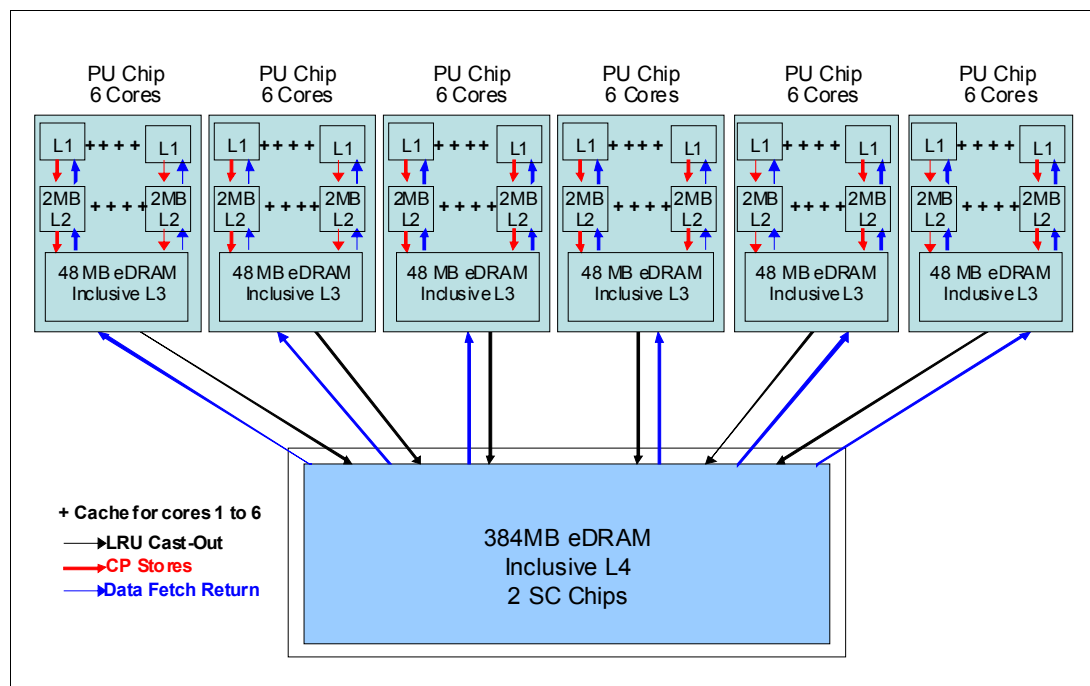


Figure 2-9 zEC12 MCM chips and cache structure

The MCM also has two storage control (SC) chips. Each SC chip packs 192 MB of eDRAM cache, interface logic for 30 cores, and SMP fabric logic, into 525.6 mm^2 . The two SC chips are configured to provide a single 384 MB L4 cache that is shared by all 27 or 30 cores on the module. This amount gives a total sum of 1536 MB if all four books are implemented, yielding outstanding SMP scalability on real-world workloads.

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

2.4 zBC12 processor drawer and single chip modules

The zBC12 central processor complex (CPC) uses a packaging concept for its processors based on processor drawers. A processor drawer contains single chip modules (SCM), memory, and connectors to I/O drawers, to PCIe I/O drawers, and other systems. The zBC12 H06 has one processor drawer installed and zBC12 H13 has two processor drawers installed. Figure 2-10 shows a processor drawer and its components.

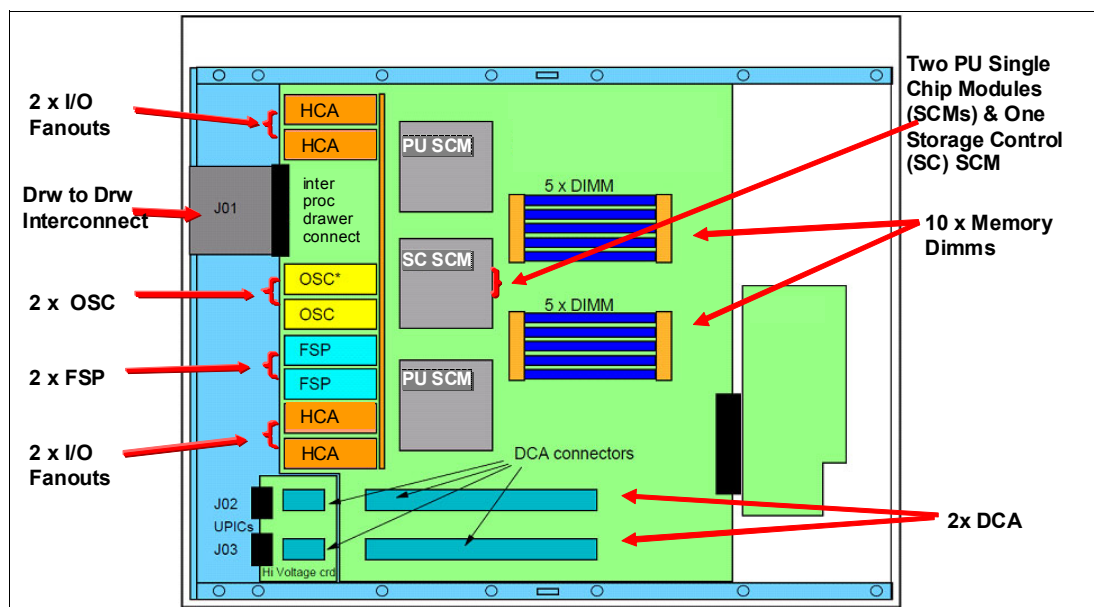


Figure 2-10 zBC12 processor drawer structure and components

Each processor drawer contains the following elements:

- ▶ One SC single chip module with 192 MB L4 cache.
- ▶ Two PU SCMs. One PU SCM has four active cores, while the other PU SCM has five active cores.
- ▶ Memory DIMMs plugged into 10 available slots, providing up to 320 GB of physical memory installed in a processor drawer.
- ▶ A combination of up to four fanouts. PCIe fanout connections are for links to the PCIe I/O drawers in the CPC, HCA2-Copper connections are for links to the I/O drawers in the CPC, and HCA2-Optical and HCA3-Optical connections are to external systems (coupling links).
- ▶ Two distributed converter assemblies (DCA) that provide power to the processor drawer. Loss of a DCA leaves enough power to satisfy the processor drawer's power requirements (N+1 redundancy). The DCAs can be concurrently maintained.
- ▶ Two flexible service processor (FSP) cards for system control.
- ▶ Two oscillator (OSC) cards.

Note: The OSC cards on the low processor drawer have BNC Connector for Pulse Per Second (PPS), while the Oscillator passthru cards on the high processor drawer (Model H13) do not have the PPS connector.

On the zBC12 H13, fabric book connectivity (FBC) provides the point-to-point connectivity between the two processor drawers of the CPC.

Both PU and storage control (SC) chips on the SCM use CMOS 13s chip technology. CMOS 13s is state-of-the-art microprocessor technology based on 15-layer copper interconnections and silicon-on insulator (SOI) technologies. The chip lithography line width is 0.032 μm (32 nm). On the SCM, two serial electrically erasable programmable ROM (SEEPROM) chips are rewriteable memory chips that hold data without power, use the same technology, and are used for retaining product data for the SCM, as well as relevant engineering information.

Each SCM is 50 x 50 mm in size. Figure 2-11 shows the cache structure on the zBC12.

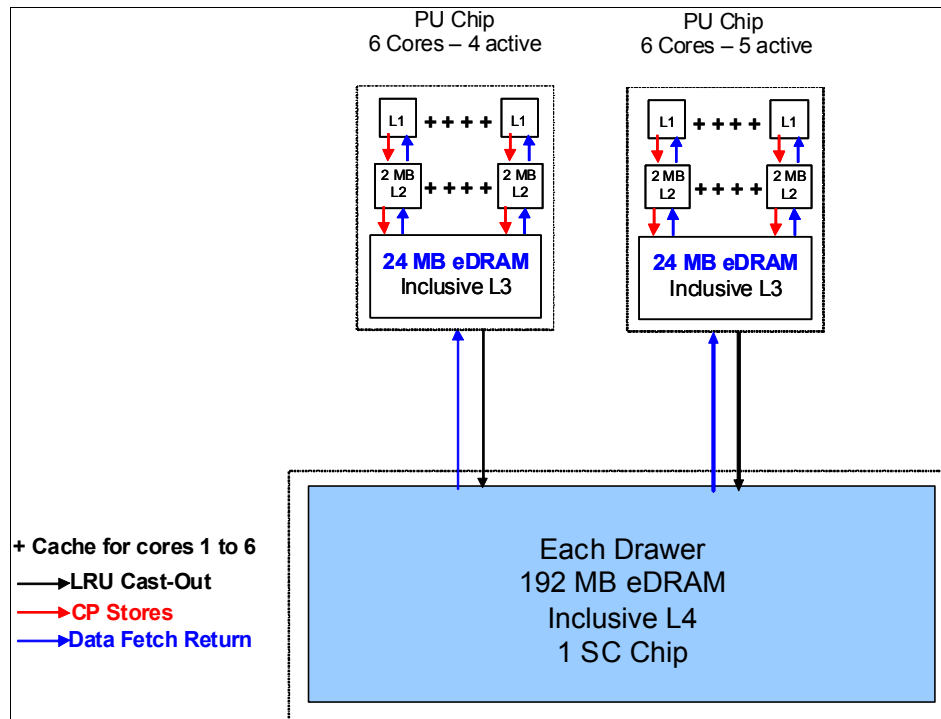


Figure 2-11 zBC12 processor drawer cache levels structure

2.5 Processor chip

The zEC12 and zBC12 feature a high-frequency hex-core processor chip, an advanced microprocessor design, a robust cache hierarchy, and an SMP design that is optimized for enterprise database and transaction processing workloads, and for workloads such as Java and Linux. It uses leading-edge technology and circuit design techniques while building on the rich heritage of mainframe system design, including industry-leading reliability, availability, and serviceability. Functions and features that are introduced with the zEC12 and zBC12 enable increased software efficiency and scalability, while maintaining full compatibility with existing software. Further details are given in 3.2.1, “Microprocessor” on page 76.

2.6 Processor unit

Processor unit (PU) is the generic term for a z/Architecture processor. A PU is embedded in a System z chip core. Each PU is a superscalar processor with the following attributes:

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

Each PU has an L1 cache that is divided into a 64 KB cache for instructions and a 96 KB cache for data. Each PU also has a private L2 cache, with 2 MB in size, split into 1 MB D-cache and 1 MB I-cache. In addition, each PU chip contains an L3 cache, which is shared by all six PUs on the chip. The shared L3 cache uses eDRAM. The zEC12 has a 48 MB L3 cache and zBC12 has a 24 MB L3 cache. The cache structures of zEC12 and zBC12 are shown, respectively, in Figure 2-9 on page 37 and Figure 2-11 on page 39. This implementation optimizes performance of the system for high-frequency, fast processors.

Each L1 cache has a *translation lookaside buffer (TLB)* of 512 entries that are 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 embedded throughout the design and combined with comprehensive instruction-level retry and dynamic CPU sparing. This provides the reliability and availability that is required for true mainframe quality.

Dedicated on-chip cryptographic hardware includes extended key and hash sizes for the Advanced Encryption Standard (AES), Secure Hash Algorithm (SHA) algorithms, and UTF8 to UTF16 conversion support.

The zEC12 and zBC12 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. Several hardware instructions that support more efficient code generation and execution are introduced. Examples are the Hardware Decimal Floating-Point, Transactional Execution Facility and the Runtime Instrumentation Facility, described in Chapter 3, “Key functions and capabilities of IBM zEnterprise EC12 and IBM zEnterprise BC12” on page 67

Processor unit characterization

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

At least one central processor (CP) must be purchased with, or before, a *System z Application Assist Processor (zAAP)* or *System z Integrated Information Processor (zIIP)* can be purchased. Clients can purchase up to two zAAPs and/or up to two zIIPs, for each purchased CP (assigned or unassigned) on the system. However, a logical partition definition can go behind the 1:2 ratio. For example, on a system with two CPs, a maximum of four zAAPs and four zIIPs can be installed. A logical partition definition for that system can contain up to two logical CPs, four logical zAAPs, and four logical zIIPs. Another possible configuration would be one logical CP, three logical zAAPs, and four logical zIIPs.

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

zAAP and zIIP processors: The addition of ICFs, IFLs, zAAPs, zIIPs, and system assist processors (SAP) to a CPC does not change the system capacity setting or its MSU rating (only CPs do). IBM does not impose any software charges on work that is dispatched on zAAP and zIIP processors.

Statement of Direction: The zEC12 and zBC12 are planned to be the last System z servers to offer support for zAAP specialty engine processors. IBM intends to continue support for running zAAP workloads on zIIP processors (“zAAP on zIIP”). This change is intended to help simplify capacity planning and performance management, while still supporting all the currently eligible workloads. IBM has provided a PTF for APAR OA38829 on z/OS V1R12 and V1R13. This PTF removes the restriction that prevents zAAP-eligible workloads from running on zIIP processors when a zAAP is installed on the server.

2.7 Memory

This section discusses memory for the zEC12 and zBC12.

2.7.1 zEC12 memory

Maximum physical memory sizes are directly related to the number of books in the system. Because part of the physically installed memory is used to implement the RAIM design, a zEC12 system has more memory installed than what was ordered. This configuration results in up to 768 GB of available memory per book and up to 3072 GB (3 TB) for each system. The HSA memory has a fixed amount of 32 GB and is managed separately from client memory. Therefore, theoretically, up to 736 GB on the one-book model and up to 3040 GB on the four-book models can be ordered. Because of some dependencies on the memory granularity, the maximum number of orderable memory can vary from the theoretical number.

Table 2-3 lists the maximum and minimum memory sizes for each zEC12 model.

Table 2-3 zEC12 models memory sizes

Model	Number of books	Client memory (GB)
H20	1	32 - 704
H43	2	32 - 1392
H66	3	32 - 2272
H89	4	32 - 3040
HA1	4	32 - 3040

On zEC12 systems, the granularity for memory orders varies from 32 GB up to 256 GB. Table 2-4 shows the memory granularity depending on installed client memory.

Table 2-4 zEC12 memory granularity

Granularity (GB)	Client memory (GB)
32	32 - 256
64	320 - 512
96	608 - 896
112	1008
128	1136 - 1520
256	1776 - 3056

Physically, memory is organized in the following ways:

- ▶ A book always contains a minimum of 64 GB of physically installed memory.
- ▶ A book can have more memory that is installed than enabled. The excess amount of memory can be enabled by a Licensed Internal Code load.
- ▶ Memory upgrades are satisfied from already-installed unused memory capacity until exhausted. When no more unused memory is available from the installed memory cards, either the cards must be upgraded to a higher capacity or the addition of a book with more memory is necessary.

When activated, a logical partition can use memory resources located in any book. No matter what book the memory is in, a logical partition has access to that memory if so allocated. Despite the book structure, the zEC12 is still a *symmetric multiprocessor (SMP)*.

A memory upgrade is concurrent when it requires no change of the physical memory cards. A memory card change is disruptive when no use is made of Enhanced Book Availability. See *IBM zEnterprise EC12 Technical Guide*, SG24-8049, 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 64 GB. During a model upgrade, the addition of a book is a concurrent operation. The addition of the physical memory that is in the added book is also concurrent.

2.7.2 zBC12 memory

A zBC12 CPC has more physical memory installed than what was ordered. Part of the physical installed memory is used to implement the redundant array of independent memory (RAIM) design, resulting in up to 256 GB of available memory for Model H06 and up to 512 GB for Model H13. As the HSA memory has a fixed amount of 16 GB and is managed separately from customer memory, up to 240 GB can be ordered for H06 and up to 496 GB for H13.

Table 2-5 shows the maximum and minimum memory sizes that customers can order for each zBC12 model, and the memory increments.

Table 2-5 zBC12 servers memory sizes

Model	Number of processor drawer	Increment (GB)	Customer memory (GB)
H06	1	8	8 - 112
H06	1	32	144 - 240
H13	2	8	16 - 112
H13	2	32	144 - 496

On zBC12 CPCs, the granularity of memory orders is 8 GB for customer memory sizes from 8 GB to 112 GB and 32 GB for systems with 144 GB to 496 GB of customer memory. Physically, memory is organized as follows:

- ▶ A processor drawer always contains a minimum of 40 GB of physically installed memory.
- ▶ A processor drawer can have more memory installed than enabled. The excess amount of memory can be enabled by a Licensed Internal Code load when required by the installation.

- Memory upgrades are satisfied from already-installed unused memory capacity until exhausted. When no more unused memory is available from the installed memory cards, either the cards must be upgraded to a higher capacity or the addition of a processor drawer with additional memory is necessary.

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

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

2.7.3 Concurrent memory upgrade

Memory can be upgraded concurrently by using *Licensed Internal Code Configuration Control (LICCC)* if physical memory is available, as described in the previous section. The *plan ahead memory function* that is available with the zEC12 and zBC12 is able to plan for nondisruptive memory upgrades by having in the system pre-plugged memory, which is based on a target configuration. Pre-plugged memory is enabled through an LICCC order that is placed by the client.

2.7.4 Redundant array of independent memory

The z196 introduced the *redundant array of independent memory (RAIM)* to System z, making the memory subsystem essentially a fully fault-tolerant N+1 design. RAIM design automatically detects and recovers from dynamic random access memory (DRAM), socket, memory channel, or dual inline memory module (DIMM) failures. The RAIM design is fully integrated in the zEC12 and zBC12 and requires the addition of one memory channel that is dedicated for reliability, availability, and serviceability (RAS).

2.7.5 Hardware system area

The *hardware system area (HSA)* is a fixed-size reserved area of memory, separate from the client-purchased memory. The HSA is used for several internal functions, but the bulk is used by channel subsystem functions. The HSA has grown with each successive mainframe generation. On older systems, model upgrades and also new logical partition definitions or changes required pre-planning and were sometimes disruptive. The fixed size 32 GB HSA of the zEC12 and 16 GB HSA of the zBC12 are large enough to accommodate any logical partition (LPAR) definitions or changes, thus eliminating those outage situations. In addition, planning needs are eliminated.

A fixed large HSA allows the Dynamic I/O capability to be enabled by default. It also enables the dynamic addition and removal, without planning, of the following features:

- New logical partition to new or existing channel subsystem (CSS)
- New CSS (up to four can be defined on zEC12, and up to two on zBC12)
- New subchannel set (up to three can be defined on zEC12, and up to two on zBC12)
- Devices, up to the maximum that is permitted, in each subchannel set
- Logical processors by type
- Cryptographic adapters

2.8 I/O system structure

The zEC12 and zBC12 support two internal I/O infrastructures:

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

The InfiniBand I/O infrastructure consists of the following features:

- ▶ InfiniBand fanouts in the CPC book (zEC12) or processor drawer (zBC12), which support the 6 GBps InfiniBand I/O interconnect
- ▶ InfiniBand I/O card domain multiplexers with redundant I/O interconnect in the following configurations:
 - The 14U, 28-slot, 7-domain I/O cage¹
 - The 5U, 8-slot, 2-domain I/O drawer
- ▶ I/O features

The PCIe I/O infrastructure consists of the following features:

- ▶ PCIe fanouts in the CPC book (zEC12) or processor drawer (zBC12), which support 8 GBps connectivity to the PCIe I/O drawer
- ▶ 7U PCIe I/O drawer with 32 slots (eight slots per I/O domain) for PCIe I/O features
- ▶ I/O features

Ordering of I/O features: I/O cages and I/O drawers, or PCIe I/O drawers, cannot be directly ordered. Ordering of I/O feature types determine the appropriate mix of I/O cages, I/O drawers, and PCIe I/O drawers.

Figure 2-12 shows a high-level view of the I/O system structure for the zEC12.

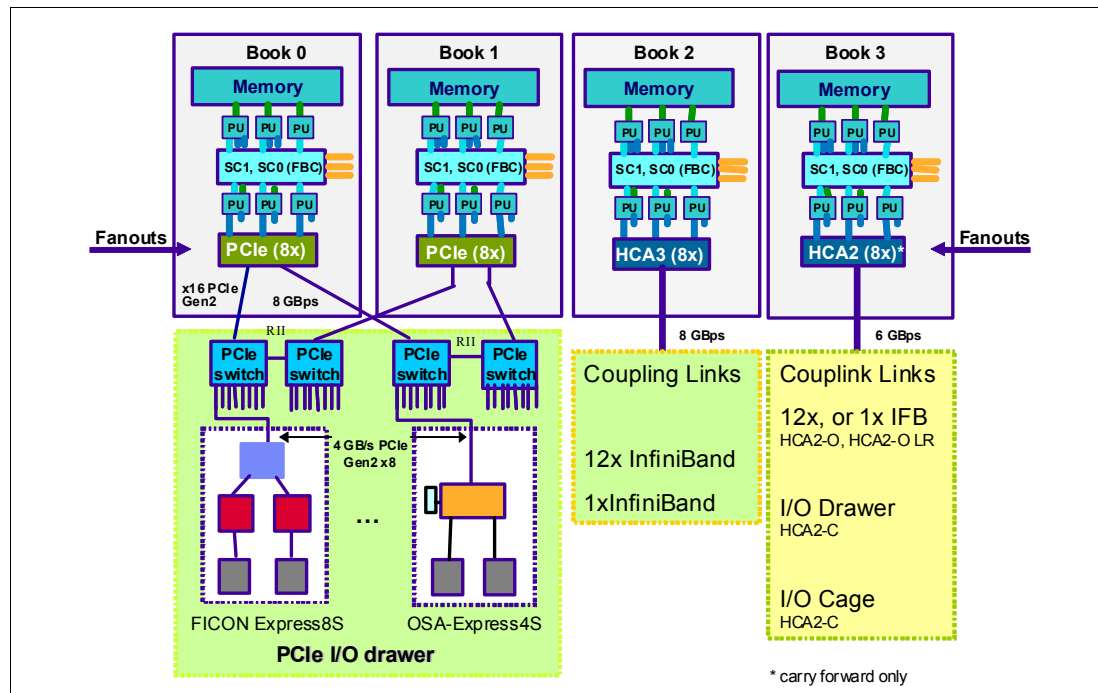


Figure 2-12 zEC12 I/O system structure

¹ Available on zEC12 only

The zEC12 and zBC12 have six fanout types, which are in the front of the book on zEC12 or processor drawer on zBC12:

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

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

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

- ▶ InfiniBand (IFB) connections to the I/O cages and I/O drawers from the Host Channel Adapter (HCA2-C) fanouts via copper cables. The two ports in the fanout are dedicated to connect to an InfiniBand multiplexer (IFB-MP) card in the I/O drawer, or in the I/O cage, which supports FICON Express8, FICON Express4, OSA-Express3, InterSystem Channel (ISC), and Crypto Express3 features.
- ▶ PCIe connections to the PCIe I/O drawers from the PCIe fanouts via copper cables. The two ports on the fanouts support FICON Express8S, OSA-Express5S, OSA-Express4S, Crypto Express4S, Flash Express, zEDC, and 10GbE RoCE Express features in the PCIe I/O drawers.

For coupling link connectivity in a Parallel Sysplex configuration, the zEC12 and zBC12 support the following fanouts:

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

The zEC12 can have up to eight fanouts (numbered D1, D2, and D5 - DA) for each book, which are used to connect to an I/O cage, I/O drawers, PCIe I/O drawers, or for Parallel Sysplex InfiniBand (PSIFB). The zBC12 can have up to 4 fanouts in each processor drawer. In a system that is configured for maximum availability, alternate paths maintain access to critical I/O devices, such as disks and networks.

Each I/O cage (zEC12 only) supports up to seven I/O domains and a total of 28 I/O feature slots. The IFB-MP card in the I/O cage supports 2 I/O domains, where each I/O domain supports four features (FICON Express8, FICON Express4, OSA-Express3, InterSystem Channel (ISC), and Crypto Express3). Up to four of the 32 slots in the I/O cage are occupied by the IFB-MB.

Each I/O drawer supports two I/O domains (A and B) for a total of eight I/O slots. Each I/O domain uses an IFB-MP card in the I/O drawer and a copper cable to connect to a Host Channel Adapter (HCA) fanout in the CPC book (zEC12) or processor drawer (zBC12). The 12x DDR² InfiniBand link between the HCA in the CPC and the IFB-MP in the I/O drawer supports a link rate of 6 GBps.

All features in the I/O drawer are installed horizontally. The two DCAs distribute power to the I/O drawer.

² DDR = Double Data Rate

The IFB-MP cards are installed at location 09 at the rear side of the I/O drawer. The I/O features are installed from the front and rear sides of the I/O drawer. Two I/O domains are supported. Each I/O domain has up to four features of any type (FICON Express8, FICON Express4, OSA-Express3, InterSystem Channel (ISC), and Crypto Express3). The I/O features are connected to the IFB-MP card through the backplane board.

The PCIe I/O drawer is a two-sided drawer (I/O features on both sides) that is 7U high (one half of the I/O cage). The drawer contains 32 slots, four switch cards (two in the front and two in the rear) to support four I/O domains, each of which contain eight features of any type (FICON Express8S, OSA-Express5S, OSA-Express4S, Crypto Express4S, Flash Express, zEDC Express, and 10GbE RoCE Express). To complete the parts of the drawer, there are two DCAs to provide the redundant power, and two air moving devices (AMD) for redundant cooling.

2.9 I/O features

The zEC12 and the zBC12 support two internal I/O infrastructures:

An InfiniBand-based³ infrastructure for I/O cages (zEC12 only) and I/O drawers, which supports the following I/O features:

- ▶ Crypto Express3
- ▶ Crypto Express3-1P(zBC12 only)
- ▶ FICON Express8
- ▶ FICON Express4
- ▶ ISC-3
- ▶ OSA-Express3

A PCIe-based infrastructure for PCIe I/O drawers, which supports the following I/O features:

- ▶ FICON Express8S
- ▶ OSA-Express5S
- ▶ OSA-Express4S
- ▶ 10GbE RoCE Express
- ▶ Crypto Express4S
- ▶ Flash Express
- ▶ zEDC Express

All new system builds, migration offerings, and exchange programs offer FICON Express8S, OSA-Express5S, OSA-Express4S (carry forward only), and Crypto Express4S features. The following features are no longer orderable for zEC12 and zBC12:

- ▶ ESCON
- ▶ FICON Express2
- ▶ FICON Express4
- ▶ FICON Express8
- ▶ OSA-Express2
- ▶ OSA-Express3
- ▶ ISC-3
- ▶ Crypto Express2

In addition, ICB-4 links are not supported. See Table B-1 on page 168 for more details.

³ Only available if carried forward on upgrades to zEC12 or zBC12

Features available when carried forward: FICON Express8, FICON Express4, OSA-Express4S, OSA-Express3, ISC-3, Crypto Express3, and Crypto Express3-1P (zBC12 only) features are available only when carried forward on upgrades.

2.9.1 FICON Express8S

Two types of *FICON Express8S* transceivers are supported on a new build zEC12 or zBC12, one long wavelength (LX) laser version, and one short wavelength (SX) laser version:

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

Each port supports attachment to the following elements:

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

FICON Express8S 10KM LX feature

The *FICON Express8S 10KM LX* feature occupies one I/O slot in the PCIe I/O drawer. It has two ports, each supporting an LC duplex connector and auto-negotiated link speeds of 2 Gbps, 4 Gbps, and 8 Gbps up to an unrepeated maximum distance of 10 km.

FICON Express8S SX feature

The *FICON Express8S SX* feature occupies one I/O slot in the PCIe I/O drawer. It has two ports, each supporting an LC duplex connector and auto-negotiated link speeds of 2 Gbps, 4 Gbps, and 8 Gbps up to an unrepeated maximum distance of up to 500 meters at 2 Gbps, 380 meters at 4 Gbps, or 150 meters at 8 Gbps.

2.9.2 FICON Express8

The *FICON Express8* features are available only when carried forward on upgrades. Two types of FICON Express8 transceivers are supported on zEC12 or zBC12:

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

FICON Express8 10KM LX feature

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

FICON Express8 SX feature

The *FICON Express8 SX* feature occupies one I/O slot in the I/O cage or I/O drawer. This feature has four ports, each supporting an LC duplex connector, and auto-negotiated link speeds of 2 Gbps, 4 Gbps, and 8 Gbps up to an unrepeated maximum distance⁴ of up to 500 meters at 2 Gbps, 380 meters at 4 Gbps, or 150 meters at 8 Gbps.

⁴ Distances valid for OM3 cabling

2.9.3 FICON Express4

The *FICON Express4* features are available only when carried forward on upgrades. Two types of FICON Express4 transceivers are supported on zEC12 or zBC12 systems:

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

Each port supports attachment to the following items:

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

Statement of Direction: The zEC12 and zBC12 are planned to be the last System z servers to offer support of the FICON Express4 features. Enterprises should continue migrating from the FICON Express4 features to the FICON Express8S features.

FICON Express4 10KM LX feature

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

FICON Express4 SX feature

The *FICON Express4 SX feature* occupies one I/O slot in the I/O cage or I/O drawer. This feature has four ports, each supporting an LC duplex connector. The feature also supports auto-negotiated link speeds of 1 Gbps, 2 Gbps, and 4 Gbps to an unrepeatable maximum distance⁴ of up to 860 meters operating at 1 Gbps, 500 meters operating at 2 Gbps, or 380 meters operating at 4 Gbps.

2.9.4 OSA-Express5S

This section describes the connectivity options that are offered by the *OSA-Express5S* features. The following OSA-Express5S features can be installed on zEC12 and zBC12:

- ▶ OSA-Express5S 10 Gigabit Ethernet (GbE) Long Reach (LR)
- ▶ OSA-Express5S 10 Gigabit Ethernet (GbE) Short Reach (SR)
- ▶ OSA-Express5S Gigabit Ethernet Long Wavelength (GbE LX)
- ▶ OSA-Express5S Gigabit Ethernet Short Wavelength (GbE SX)
- ▶ OSA-Express5S 1000BASE-T Ethernet

OSA-Express5S 10 GbE LR feature

The *OSA-Express5S 10 GbE LR* feature occupies one slot in a PCIe I/O drawer. It has one port that connects to a 10 Gbps Ethernet LAN through a 9 µm single mode fiber optic cable that is terminated with an LC Duplex connector. The feature supports an unrepeatable maximum distance of 10 km.

OSA-Express5S 10 GbE SR feature

The *OSA-Express5S 10 GbE SR* feature occupies one slot in the PCIe I/O drawer. This feature has one port that connects to a 10 Gbps Ethernet LAN through a 62.5 µm or 50 µm multimode fiber optic cable that is terminated with an LC Duplex connector. The maximum supported unrepeatable distance is 33 m on a 62.5 µm multimode fiber optic cable, and 300 m on a 50 µm multimode fiber optic cable.

OSA-Express5S GbE LX feature

The *OSA-Express5S GbE LX* occupies one slot in the PCIe I/O drawer. This feature has two ports, representing one *channel path identifier (CHPID)*, that connect to a 1 Gbps Ethernet LAN through a 9 µm single mode fiber optic cable. This cable is terminated with an LC Duplex connector, supporting an unrepeatable maximum distance of 5 km. A multimode (62.5 or 50 µm) fiber optic cable can be used with this feature. The use of these multimode cable types requires a *mode conditioning patch (MCP)* cable at each end of the fiber optic link. Use of the single mode to multimode MCP cables reduces the supported distance of the link to a maximum of 550 meters.

OSA-Express5S GbE SX feature

The *OSA-Express5S GbE SX* occupies one slot in the PCIe I/O drawer. This feature has two ports, representing one CHPID, that connect to a 1 Gbps Ethernet LAN through 50 or 62.5 µm multimode fiber optic cable. This cable is terminated with an LC Duplex connector over an unrepeatable distance of 550 meters (for 50 µm fiber) or 220 meters (for 62.5 µm fiber).

OSA-Express5S 1000BASE-T feature

The *OSA-Express5S 1000BASE-T* occupies one slot in the PCIe I/O drawer. It has two ports, representing one CHPID, that connect to a 1000 Mbps (1 Gbps), 100 Mbps, or 10 Mbps Ethernet LAN. Each port has an RJ-45 receptacle for UTP Cat5 or Cat6 cabling, which supports a maximum distance of 100 meters.

2.9.5 OSA-Express4S

This section describes the connectivity options that are offered by the *OSA-Express4S* features. The following OSA-Express4S features can be installed on zEC12 and zBC12:

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

The following OSA-Express4S feature can be installed on zEC12, and is not available on the zBC12:

- ▶ OSA-Express4S 1000BASE-T Ethernet

OSA-Express4S 10 GbE LR feature

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

OSA-Express4S 10 GbE SR feature

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

OSA-Express4S GbE LX feature

The *OSA-Express4S GbE LX* occupies one slot in the PCIe I/O drawer. This feature has two ports, representing one *channel path identifier (CHPID)*, that connect to a 1 Gbps Ethernet LAN through a 9 µm single mode fiber optic cable. This cable is terminated with an LC Duplex connector, supporting an unrepeatable maximum distance of 5 km. A multimode

(62.5 or 50 μ m) fiber optic cable can be used with this feature. The use of these multimode cable types requires a *mode conditioning patch (MCP)* cable at each end of the fiber optic link. Use of the single mode to multimode MCP cables reduces the supported distance of the link to a maximum of 550 meters.

OSA-Express4S GbE SX feature

The *OSA-Express4S GbE SX* occupies one slot in the PCIe I/O drawer. This feature has two ports, representing one CHPID, that connect to a 1 Gbps Ethernet LAN through 50 or 62.5 μ m multimode fiber optic cable. This cable is terminated with an LC Duplex connector over an unrepeat distance of 550 meters (for 50 μ m fiber) or 220 meters (for 62.5 μ m fiber).

OSA-Express4S 1000BASE-T feature

The *OSA-Express4S 1000BASE-T* occupies one slot in the PCIe I/O drawer. It has two ports, representing one CHPID, that connect to a 1000 Mbps (1 Gbps), 100 Mbps, or 10 Mbps Ethernet LAN. Each port has an RJ-45 receptacle for UTP Cat5 or Cat6 cabling, which supports a maximum distance of 100 meters.

Statement of Direction: The OSA-Express4S 1000BASE-T Ethernet feature is planned to be the last copper Ethernet feature to support half-duplex operation and a 10 Mbps link data rate. The IBM zEnterprise EC12 servers are planned to be the last IBM System z servers to support half-duplex operation and a 10 Mbps link data rate for copper Ethernet environments.

2.9.6 OSA-Express3

This section describes the connectivity options that are offered by the OSA-Express3 features. The OSA-Express3 features are available only if carried forward on an upgrade. The following OSA-Express3 features are supported on the zEC12 and zBC12:

- ▶ OSA-Express3 10 Gigabit Ethernet (GbE) Long Reach (LR)
- ▶ OSA-Express3 10 Gigabit Ethernet Short Reach (SR)
- ▶ OSA-Express3 GbE LX
- ▶ OSA-Express3 GbE SX
- ▶ OSA-Express3 1000BASE-T Ethernet
- ▶ OSA-Express3-2P 1000BASE-T (zBC12 only)
- ▶ OSA-Express3-2P GbE SX (zBC12 only)

OSA-Express features: Each OSA-Express3 feature installed in an I/O cage or I/O drawer reduces by two the number of allowed OSA-Express5S or OSA-Express4S features.

OSA-Express3 10 GbE LR feature

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

OSA-Express3 10 GbE SR feature

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

OSA-Express3 GbE LX feature

The *OSA-Express3 GbE LX* occupies one slot in the I/O cage or I/O drawer. This feature has four ports, representing two CHPIDs, that connect to a 1 Gbps Ethernet LAN through a 9 μ m single mode fiber optic cable. This cable is terminated with an LC Duplex connector, supporting an unrepeatable maximum distance of 5 km. A multimode (62.5 or 50 μ m) fiber optic cable can be used with this feature. The use of these multimode cable types requires an MCP cable at each end of the fiber optic link. Use of the single mode to multimode MCP cables reduces the supported distance of the link to a maximum of 550 meters.

OSA-Express3 GbE SX feature

The *OSA-Express3 GbE SX* occupies one slot in the PCIe I/O drawer. This feature has four ports that represent two CHPIDs that connect to a 1 Gbps Ethernet LAN through 50 or 62.5 μ m multimode fiber optic cable. This cable is terminated with an LC Duplex connector over an unrepeatable distance of 550 meters (for 50 μ m fiber) or 220 meters (for 62.5 μ m fiber).

OSA-Express3 1000BASE-T Ethernet feature

OSA-Express3 1000BASE-T occupies one slot in the I/O cage or I/O drawer. This feature has four ports, with two ports per CHPID, that connect to a 1000 Mbps (1 Gbps), 100 Mbps, or 10 Mbps Ethernet LAN. Each port has an RJ-45 receptacle for UTP Cat5 or Cat6 cabling, which supports a maximum distance of 100 meters.

OSA-Express3-2P 1000BASE-T

OSA-Express3-2P 1000BASE-T occupies one slot in the I/O drawer. It has two ports that belong to one CHPID and connect to a 1000 Mbps (1 Gbps), 100 Mbps, or 10 Mbps Ethernet LAN. Each port has an RJ-45 receptacle for UTP Cat5 or Cat6 cabling, which supports a maximum distance of 100 meters.

OSA-Express3-2P GbE SX

OSA-Express3-2P GbE SX occupies one slot in the I/O drawer. It has two ports that belong to one CHPID and connect to a 1 Gbps Ethernet LAN through 50 or 62.5 μ m multi-mode fiber optic cable terminated with an LC Duplex connector over an unrepeatable distance of 550 meters (for 50 μ m fiber) or 220 meters (for 62.5 μ m fiber).

Statement of Direction: The zEC12 and the zBC12 are planned to be the last IBM System z servers to offer support of the OSA-Express3 family of features. Enterprises should continue migrating from the OSA-Express3 features to the OSA-Express5S features.

2.9.7 Flash Express

Flash Express is an innovative optional feature introduced with the zEC12 and also available on the zBC12. It is intended to provide performance improvements and better availability for critical business workloads that cannot afford any hits to service levels. Flash Express is easy to configure, requires no special skills, and provides rapid time to value.

Flash Express implements storage-class memory (SCM) through internal NAND Flash solid-state drive (SSD), in a PCIe card form factor. The Flash Express feature is designed to allow each logical partition to be configured with its own SCM address space.

For availability, this feature is available in pairs of cards. Each feature offers a capacity of 1.4 TB of usable storage per pair of cards. A maximum of four pairs of cards can be installed on a zEC12 or zBC12, providing a maximum capacity of 5.6 TB of storage.

Flash Express is supported by z/OS V1R13 (or later). In 3.2.4, “Flash Express” on page 82, there are more details about the Flash Express feature and its exploitation.

2.9.8 zEDC Express

The *zEDC Express* is an optional feature, exclusive to the zEC12 and zBC12. It is designed to provide hardware-based acceleration for data compression and decompression for the enterprise, helping to improve cross platform data exchange, reduce CPU consumption, and save disk space.

A minimum of one feature can be ordered and a maximum of eight can be installed on the system, on the PCIe I/O drawer. Up to two zEDC Express features per domain can be installed. There is one PCIe adapter/compression coprocessor per feature which implements compression as defined by RFC1951 (DEFLATE)⁵. A zEDC Express feature can be shared by up to 15 LPARs.

The zEDC Express is a native PCI feature; the management functions are provided by Resource Groups (RG) running on the integrated firmware processor (IFP). See 3.2.3, “Native PCIe features and integrated firmware processor” on page 81 for more details about RGs and IFP.

For resilience, there are always two independent RGs on the system, sharing the IFP. Thus, it is recommended that a minimum of two zEDC Express features be installed, one per RG.

Consider also the total data throughput required and that, in the case of one feature becoming unavailable, the others should be able to absorb the load. Thus, for best data throughput and availability, it is suggested that at least two features per RG are installed.

2.9.9 10 Gigabit Ethernet RoCE Express

The 10 Gigabit Ethernet (10GbE) RoCE Express feature is designed to help reduce consumption of CPU resources for applications utilizing the TCP/IP stack and may also help to reduce network latency with memory-to-memory transfers utilizing Shared Memory Communications - Remote Direct Memory Access (SMC-R) in z/OS V2R1. It is transparent to applications and can be used for LPAR-to-LPAR communication on a single system or server-to-server communication in a multiple CPC environment.

This feature resides exclusively in the PCIe I/O drawer and is exclusive to the zEC12 and zBC12. The 10GbE RoCE Express feature has one PCIe adapter with two ports.

The 10GbE RoCE Express feature utilizes a short reach (SR) laser as the optical transceiver, and supports use of a multi-mode fiber optic cable terminated with an LC Duplex connector. Both point to point connection and switched connection with an enterprise-class 10 GbE switch are supported. Switch used by 10GbE RoCE Express feature must have Pause frame enabled as defined by the IEEE 802.3x standard.

The 10GbE RoCE Express feature does not use a CHPID. It is defined using the Input/Output Configuration Program (IOCP) **FUNCTION** statement or in the Hardware Configuration Definition (HCD).

A maximum of 16 features can be installed per system. Each feature must be dedicated to an LPAR. Only one of the two ports is supported by z/OS.

⁵ Refer to <http://www.ietf.org/rfc/rfc1951.txt>

2.9.10 Coupling links

Coupling links provide for communication in a Parallel Sysplex environment. They are further discussed under 2.10, “Coupling and clustering” on page 54.

2.9.11 Cryptographic features

Cryptographic coprocessor and accelerator functions can be provided by the PCIe cryptographic adapters. zEC12 and zBC12 support the Crypto Express4S feature and, in a carry forward only basis, also the Crypto Express3 features.

Statement of Direction: The zEC12 and zBC12 are planned to be the last IBM System z servers to support Crypto Express3 and Crypto Express3-1P features. Enterprises should begin migrating from the Crypto Express3 features to the Crypto Express4S feature.

Crypto Express4S

Crypto Express4S features provide the following capabilities:

- ▶ The Crypto Express4S feature occupies one I/O slot in a zEC12 or zBC12 PCIe I/O drawer.
- ▶ The Crypto Express4S feature has one PCIe adapter with one PCHID assigned to it according to its physical location in the PCIe I/O drawer.
- ▶ There is no need to define a CHPID for the Crypto Express4S feature in the HCD/IOCP. Care must be taken to avoid the use of Crypto Express4S associated PCHID by another device in the HCD/IOCP definition.
- ▶ On zEC12 and zBC12, each Crypto Express4S PCI Express adapter can be configured as one of the following environments:
 - Coprocessor with the following characteristic:
 - *Secure IBM Common Cryptographic Architecture (CCA) coprocessor* for Federal Information Processing Standard (FIPS) 140-2 Level 4 certification.
 - *Secure IBM Enterprise PKCS #11 (EP11) coprocessor* implements industry standardized set of services that adhere to the PKCS #11 specification v2.20. A Trusted Key Entry (TKE) Workstation with a smart card reader feature is required to support the administration of the Crypto Express4S, when configured as an Enterprise PKCS #11 coprocessor.
 - *Accelerator* for public key and private key cryptographic operations that are used with Secure Sockets Layer / Transport Layer Security (SSL/TLS) processing.

These modes can be configured by the Support Element, and the PCIe adapter must be configured offline to change the mode.

Crypto Express4S PCIe configuration: When the Crypto Express4S PCIe adapter is configured as a secure IBM CCA coprocessor, it still provides accelerator functions. However, up to three times better performance for those functions can be achieved if the Crypto Express4S PCIe adapter is configured as an accelerator.

- ▶ Up to 16 Crypto Express4S features are supported (16 PCI Express adapters per zEC12 or zBC12 system).

Crypto Express3

Crypto Express3 features are available only in a carry forward basis when upgrading from earlier generations to zEC12 or zBC12:

- ▶ The Crypto Express3 or the Crypto Express3-1P feature occupies one I/O slot in an I/O cage (zEC12 only) or in an I/O drawer.
- ▶ The Crypto Express3 feature has two PCI Express adapters with two PCHIDs assigned to it according to its physical location in the I/O cage or I/O drawer. The Crypto Express3-1P is available on zBC12 only in a carry forward basis. This feature has one PCI Express adapter with one PCHID assigned to it according to its physical location in the I/O cage or I/O drawer.
- ▶ There is no need to define a CHPID for the Crypto Express3 feature types in the HCD/IOCP. Care must be taken to avoid the use of Crypto Express3 features associated PCHIDs by another device in the HCD/IOCP definition.
- ▶ On zEC12 or zBC12, each Crypto Express3 or Crypto Express-1P PCI Express adapter can be configured as one of the following environments:
 - *Secure coprocessor* for Federal Information Processing Standard (FIPS) 140-2 Level 4 certification. This includes secure key functions and it is optionally programmable to deploy more functions and algorithms using User Defined Extension (UDX).
 - *Accelerator* for public key and private key cryptographic operations that are used with Secure Sockets Layer/Transport Layer Security (SSL/TLS) processing.

These modes can be configured by the Support Element, and the PCIe adapter must be configured offline to change the mode.

Crypto Express3S PCIe configuration: When the Crypto Express3 PCI Express adapter is configured as a secure coprocessor, it still provides accelerator functions. However, up to three times better performance for those functions can be achieved if the Crypto Express3 PCI Express adapter is configured as an accelerator.

- ▶ Up to eight Crypto Express3 features (16 PCI Express adapters) for each zEC12 or zBC12 system.

2.10 Coupling and clustering

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

Coupling connectivity for Parallel Sysplex on zEC12 and zBC12 uses the InfiniBand (IFB) technology. IFB links support longer distances between systems when compared with ICB connections which are not available on zEnterprise CPCs.

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

See the *Coupling Facility Configuration Options* white paper for a more specific explanation regarding the coupling links technologies. The white paper is available at the following web page:

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

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

2.10.1 InfiniBand coupling links

There are four types of *Host Channel Adapter (HCA)* fanouts that are used for IFB coupling links on the zEC12 and zBC12:

- ▶ HCA3-O fanout, which supports 12x InfiniBand (12x IFB)
- ▶ HCA3-O Long Reach (LR) fanout, which supports 1x InfiniBand (1x IFB)
- ▶ HCA2-O⁶ fanout, which supports 12x InfiniBand (12x IFB)
- ▶ HCA2-O⁶ Long Reach (LR) fanout, which supports 1x InfiniBand (1x IFB)

HCA3s are the latest generation of host channel adapters for coupling. The HCA3-O fanout for 12x InfiniBand (12x IFB) is designed for improved service times and is available for zEnterprise CPCs using the 12x InfiniBand3 (12x IFB3) protocol. The HCA3-O LR fanout for 1x InfiniBand (1x IFB) provides four ports of connectivity and optional additional subchannels for extended-distance solutions.

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

12x InfiniBand coupling links support double data rate (DDR) at 6 GBps for a zEC12 or zBC12 to zEnterprise CPCs, or a z10 connection:

- ▶ InfiniBand (HCA3-O) coupling links (12x IFB), used for z/OS-to-CF communication, CF-to-CF traffic, or STP messaging at distances up to 150 meters (492 feet) by using industry standard OM3 50 μ m fiber optic cables.

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

An HCA3-O to HCA2-O connection is supported, but the standard IFB protocol is used.

- ▶ InfiniBand (HCA2-O) coupling links (12x IFB), used for z/OS-to-CF communication, CF-to-CF traffic, or STP messaging at distances up to 150 meters (492 feet) by using industry standard OM3 50 μ m fiber optic cables.

1x InfiniBand coupling links (HCA3-O LR and HCA2-O LR) support up to 32 subchannels (devices) per CHPID, versus the current default value of seven devices per CIB type CHPID:

- ▶ InfiniBand (HCA3-O LR) coupling links (1x IFB) for z/OS-to-CF communication at unrepeated distances up to 10 km (6.2 miles) using 9 μ m single mode fiber optic cables and repeated distances up to 100 km (62 miles) using IBM System z qualified DWDM equipment. Connectivity to HCA2-O LR is supported.

The HCA3-O LR has four ports, as opposed to all other HCA-O fanouts, which have two ports. The number of supported CHPIDs remains at 16 for the fanout card.

- ▶ InfiniBand (HCA2-O LR) coupling links (1x IFB) for z/OS-to-CF communication at unrepeated distances up to 10 km (6.2 miles) using 9 μ m single mode fiber optic cables and repeated distances up to 100 km (62 miles) using System z qualified DWDM equipment. Connectivity to HCA3-O LR is supported.

⁶ Both HCA2-O and HCA2-O LR features are only available on zEC12 and zBC12 when carried forward during an upgrade.

HCA2-O and HCA2-O LR features: HCA2-O and HCA2-O LR features are only available on zEC12 or zBC12 when carried forward during an upgrade.

Time source for Server Time Protocol traffic

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

For details on all InfiniBand features, see *IBM System z Connectivity Handbook*, SG24-5444 or *Implementing and Managing InfiniBand Coupling Links on System z* SG24-7539.

2.10.2 Internal Coupling (IC)

Internal Coupling (IC) links are used for internal communication between LPARs on the same system running coupling facilities (CF) and z/OS images. The connection is emulated in Licensed Internal Code (LIC) and provides for very fast and secure memory-to-memory communications between LPARs within a single system. No physical cabling is required.

2.10.3 InterSystem Channel-3 (ISC-3)

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

The ISC-3 features supports a link data rate of 2 Gbps and are used for z/OS-to-CF communication at unrepeated distances up to 10 km. This rate is achieved by using 9 μ m single mode fiber optic cables and repeated distances up to 100 km by using System z qualified *dense wavelength division multiplexing (DWDM)* equipment. ISC-3s are supported exclusively in peer mode. STP message exchanges can flow over ISC-3.

The ISC-3 feature is only available on zEC12 and zBC12 systems when carried forward on an upgrade and supports a link data rate of 2 Gbps.

Statement of Direction: The zEC12 and zBC12 are planned to be the last IBM System z servers to support ISC-3 Links. Enterprises should continue migrating from ISC-3 features to InfiniBand Coupling Links.

2.10.4 Coupling Facility Control Code (CFCC)

This section discusses the various levels of Coupling Facility Control Code (CFCC).

CFCC Level 19

CFCC Level 19 is available on the zEC12 and zBC12 with driver level 15 and includes the following performance enhancements:

- ▶ Coupling Thin Interrupts:
 - Improves the performance in share CF engines environments.
 - Improves the response time of asynchronous CF requests.

Statement of Direction: CFCC Level 19 exploitation of Flash Express: IBM intends to provide exploitation of the Flash Express feature (#0402) on zEC12 and zBC12 servers with Coupling Facility Control Code (CFCC) Level 19 for certain Coupling Facility list structures in the first half of 2014.

CFCC Level 18

CFCC Level 18 is available on the zEC12 with driver level 12 and includes the following features:

- ▶ Performance improvements:
 - Dynamic alter for structure size or element/data ratio
 - DB2 GBP cache bypass
 - Cache structure management
- ▶ Coupling channel reporting improvement that enables the IBM RMF™ CF REPORT (Subchannel report):
 - Differentiate several IFB link types (z/OS did not have the possibility to discriminate among the underlying InfiniBand hardware, either PSIFB 1x or PSIFB 12x)
 - Detect whether coupling that uses InfiniBand (CIB) link is running “degraded”
- ▶ Improved shared CF Engine performance:
 - Efficient use of shared-processor CF images with good service times
 - Latency reduction for asynchronous CF operations and asynchronous CF notifications
- ▶ Serviceability enhancements:
 - Additional structure control info in CF dumps
 - Enhanced CFCC tracing support
 - Method to measure burst activity
 - Trigger non-disruptive dumping for other soft-failure cases beyond break-duplexing

CF structure size changes are expected to grow when going from CFCC Level 17 (or earlier) to CFCC Level 18 or CFCC Level 19 as well as from CFCC Level 18 to CFCC Level 19. We suggest reviewing the CF LPAR size by using the following tools:

- ▶ A web-based *CFSizer* tool, which is most useful when changing an existing workload or introducing a new one. CFSizer tool is available at the following web page:
<http://www.ibm.com/systems/support/z/cfsizer/>
- ▶ *Sizer Utility*, an authorized z/OS program download, most useful when upgrading a CF. The Sizer utility is available at the following web page:
<http://www.ibm.com/systems/support/z/cfsizer/altsize.html>

2.11 Time functions

Each server must have an accurate time source to maintain a time-of-day value.

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. These functions are critical for Parallel Sysplex operation.

IBM zEnterprise system support attachment to an External Time Source (ETS) for clock information, support the Server Time Protocol, and can participate in a coordinated timing network.

2.11.1 Server Time Protocol (STP)

Server Time Protocol (STP) is a system-wide facility that is implemented in the Licensed Internal Code. The STP presents a single view of time to PR/SM and provides the capability for multiple CPCs and CFs to maintain time synchronization with each other. The enablement for using this protocol with the zEC12 and zBC12 is ensured by activating the optional STP feature.

More detailed information about the implementation of STP on zEC12 and zBC12 can be found in 3.4.1, “Server Time Protocol (STP)” on page 102.

2.11.2 Network Time Protocol support

Network Time Protocol (NTP) support is available on IBM zEnterprise and z10 systems. This implementation answers the need for a single time source across the heterogeneous platforms in the enterprise. With this implementation, the STP is synchronized by the use of a NTP server as time source.

Pulse Per Second

The zEC12 and zBC12 provide a dual-path interface for *Pulse Per Second (PPS)* support. STP tracks the highly stable accurate PPS signal from the NTP server. PPS maintains accuracy of 10 μ s as measured at the PPS input of the zEC12 or zBC12 CPC. If STP uses an NTP server without PPS, a time accuracy of 100 ms to the ETS is maintained. A cable connection from the PPS port to the PPS output of an NTP server is required when the zEC12 or zBC12 is configured for using NTP with PPS as the external time source for time synchronization.

2.11.3 Time coordination for zBX components

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

The Battery Operated Clock (BOC) of the SE synchronizes to the Time-of-Day (TOD) clock of the system, every hour. This synchronization provides the capability for the components in the zBX to maintain an approximate time accuracy of 100 milliseconds to an NTP server if they synchronize to the NTP server of the SE at least once an hour.

2.12 Hardware Management Console and Support Element

The *Hardware Management Console (HMC)* and *Support Element (SE)* are appliances that together provide hardware platform management for IBM System z. Hardware platform management covers a complex set of configuration, operation, monitoring, service management tasks, and services that are essential to the use of the hardware platform product. The HMC is a stand-alone computer. See 3.5, “Hardware Management Console functions” on page 104 for more information about HMC capabilities.

The zEC12 and zBC12 is supplied with a pair of integrated notebooks (SEs). One, the primary SE, is always active while the other is an alternate. Power for the SEs is supplied by the CPCs power supply, and there are no additional power requirements. The SEs are connected to the bulk power hubs (BPH) for network connectivity with the CPC and the HMCs.

The SEs and HMCs are closed systems, and no other applications can be installed on them.

The HMCs are attached to a LAN, as are the SEs of the system. An HMC communicates with one or more System z systems and, through the SE, with the optional zBX controlled by the CPC, as shown in Figure 2-13. When tasks are performed on the HMC, the commands are sent to one or more SEs, which then issue commands to their CPCs and optional zBXs.

The HMC Remote Support Facility (RSF) provides communication with the IBM support network for hardware problem reporting and service.

Figure 2-13 shows an example of the HMC and SE connectivity.

RSF connection: RSF connection through a modem *is not supported* on the zEC12 and zBC12 HMC. An internet connection to IBM is needed to have hardware problem reporting and service.

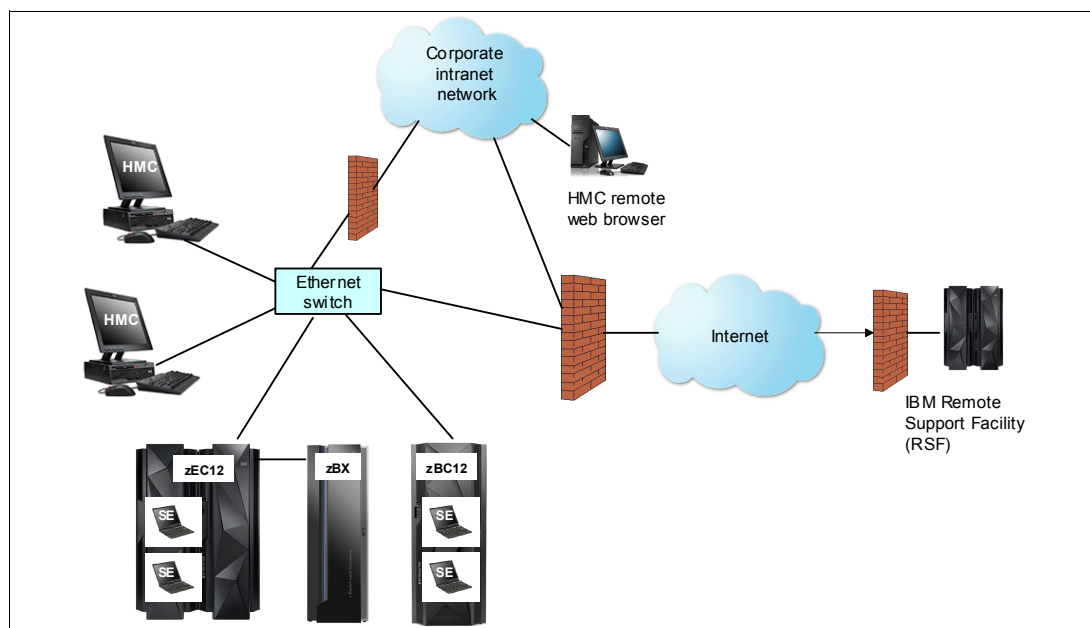


Figure 2-13 HMC and SE connectivity

2.13 Power and cooling

The power and cooling system of the zEC12 and zBC12 builds upon their predecessors (the z196 and z114), with the expansion of some significant newly developed technologies. The power service specifications of the zEC12 and zBC12 are almost same as those of their respective predecessors, and the total power consumption with the maximum configuration has increased by approximately 5% as compared to the previous generation.

In the zEC12, water cooling technology is used for MCM cooling. However, from a system level perspective, the zEC12 offers both air-cooled and water-cooled options. The zBC12 is always an air-cooled system.

zEC12 radiator cooling system

The cooling system in zEC12 is redesigned for better availability and lower cooling power consumption. Water cooling technology is now fully used in zEC12 MCMs.

The Modular Refrigeration Unit (MRU) technology for cooling the processor modules in previous System z servers is replaced with a “Radiator” design - a closed-loop water cooling pump system. The radiator has no connection to the chilled water required and the water is added to the closed-loop system during installation, with the Fill and Drain Tool.

The radiator cooling system is designed with two pumps and two blowers, but a single working pump and blower can handle the entire load. Replacement of a pump or blower is concurrent, without any performance impact.

In zEC12, the radiator is the primary cooling source of the MCM and is backed up by an air cooling system in the rare case of failure of the entire radiator. During the backup air cooling mode, hot air exits through the top of the system and the oscillator card is set to a slower cycle time. In this “cycle steering mode”, the system slows down 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, which are based on the temperature in the books.

zEC12 water cooling

The zEC12 continues to offer the possibility of using the building’s chilled water to cool the system, by employing the water cooling unit (WCU) technology. The MCM in the book is cooled by an internal, closed, water cooling loop. In the internal closed loop, water exchanges heat with building chilled water through a cold plate. The source of building chilled water is provided by the client.

In addition to the MCMs, the internal water loop also circulates through two heat exchangers that are in the path of the exhaust air in the rear of the frames. These heat exchangers remove approximately 60% - 65% of the residual heat from the I/O drawers, the air cooled logic in the books and the heat that is generated within the power enclosures. Almost two thirds of the total heat that is generated is removed from the room by the chilled water.

The zEC12 operates with two fully redundant water cooling units (WCU). One water cooling unit can support the entire load and the replacement of WCU is fully concurrent. If there is a total loss of building chilled water or if both water cooling units fail, the backup blowers are turned on to keep the system running, similarly to the radiator cooling system. At that time, cycle time degradation is required.

Unlike the z196, the zEC12 books are the same in both the air and water-cooled systems. However, conversion between air and water cooling is not available.

zBC12 cooling

The zBC12 is an air cooled system. Forced air flows from the front to the back and into the room. It requires chilled air, ideally coming from under the raised floor, to fulfill the air-cooling requirements. The chilled air is usually provided through perforated floor tiles. The amount of chilled air that is required for a variety of temperatures under the floor of the computer room is indicated in *zEnterprise BC12 Installation Manual - Physical Planning*, GC28-6923.

High Voltage Direct Current feature

With the High Voltage Direct Current (HV DC) power feature, the zEC12 and zBC12 can directly connect to DC power input and improve data center energy efficiency by removing the need for an additional DC to AC inversion step. In addition to the data center UPS and power distribution energy savings, a zEnterprise CPC running on HV DC power draws 1 - 3% less input power.

Power considerations

Power Sequence Controller: The zEC12 and zBC12 systems do not support the *Power Sequence Controller (PSC)* feature. PSC features cannot be ordered and cannot be carried forward on an upgrade to zEC12 or zBC12.

The zEC12 and zBC12 operate with two sets of redundant power supplies. Each set of the power supplies has its individual power cords or pair of power cords, depending on the number of Bulk Power Regulator (BPR) pairs installed. Power cords attach either 1-phase⁷, 3-phase, 50/60 Hz, 200 - 480 V AC power, or 380 - 520 V DC power. The total loss of one power supply has no effect on system operation.

There is a Balanced Power Plan Ahead feature available for future growth, also assuring adequate and balanced power for all possible configurations. With this feature, downtimes for upgrading a system are eliminated by including with the initial installation the maximum power requirements in terms of Bulk Power Regulators (BPR) and power cords. The Balance Power Plan Ahead feature is not available with DC and 1-phase line cords.

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

The power requirements depend on the cooling facility that is installed, number of books or drawers, and the number and kind of I/O units installed. Maximum power consumption tables for the various configurations and environments can be found in *Installation Manual for Physical Planning 2827*, GC28-6914 or *zEnterprise BC12 Installation Manual - Physical Planning*, GC28-6923. See also the power and weight estimation tool available in IBM Resource Link.

Top Exit Power

Both the zEC12 and the zBC12 support the *Top Exit Power* feature, which can be combined with the *Top Exit I/O Cabling* feature, providing more flexibility to planning the computer room cabling. The air-cooled zEC12 models and the zBC12 support installation on raised floor and non-raised floor environments. For water-cooled models, only the raised floor option is available.

Restriction: For both the air-cooled zEC12 as the zBC12, the installation on a non-raised floor is **only** supported in combination with the *Top Exit Power* feature and the *Top Exit I/O Cabling* feature.

⁷ Supported only by select number of zBC12 configurations

2.14 zEnterprise BladeCenter Extension

The zEnterprise BladeCenter Extension (zBX) Model 003 extends the System z qualities of service and management to integrate heterogeneous systems with high redundancy.

The zBX Model 003 (2458-003) connects to the zEC12 or zBC12 to become part of a node in an ensemble. That node, in turn, creates an integrated multi-platform system with advanced virtualization management (through the Unified Resource Manager) that supports diverse workloads.

The zBX is configured with the following key components:

- ▶ One to four standard 19-inch 42U IBM zEnterprise racks with required network and power infrastructure
- ▶ One to eight BladeCenter chassis with a combination of up to 112 blades⁸
- ▶ Redundant infrastructure for fault tolerance and higher availability
- ▶ Management support through the zEC12 or zBC12 HMC and SE

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

The zBX is managed through a private and physically isolated 1000BASE-T network (INMN), which interconnects all components in the zEC12 or zBC12 system (CPC and zBX). The OSA-Express for Unified Resource Manager (OSM) CHPID type supports the connectivity from the zEC12 or zBC12 to the Bulk Power Hubs (BPH). The BPHs are also connected to the INMN TOR switches in the zBX.

The IEDN provides private and secure 10 GbE high-speed data paths between all elements of an ensemble node through the IEDN TOR switches in the zBX. The IEDN connections use MAC addresses, not IP addresses (Layer 2 connection). The OSA-Express for zBX (OSX) CHPID type supports connectivity and access control from the zEC12 or zBC12 to the zBX.

⁸ Current maximum number of blades supported: 112 POWER7, 28 for DataPower XI50z, 56 for System x HX5.

Figure 2-14 shows the ensemble node connections through the OSA-Express5S 1000BASE-T, OSA-Express4S 1000BASE-T⁹, or OSA-Express3 1000BASE-T features (CHPID type OSM) and OSA-Express5S 10 GbE, OSA-Express4S 10 GbE, or OSA-Express3 10 GbE features (CHPID type OSX) in the zEC12.

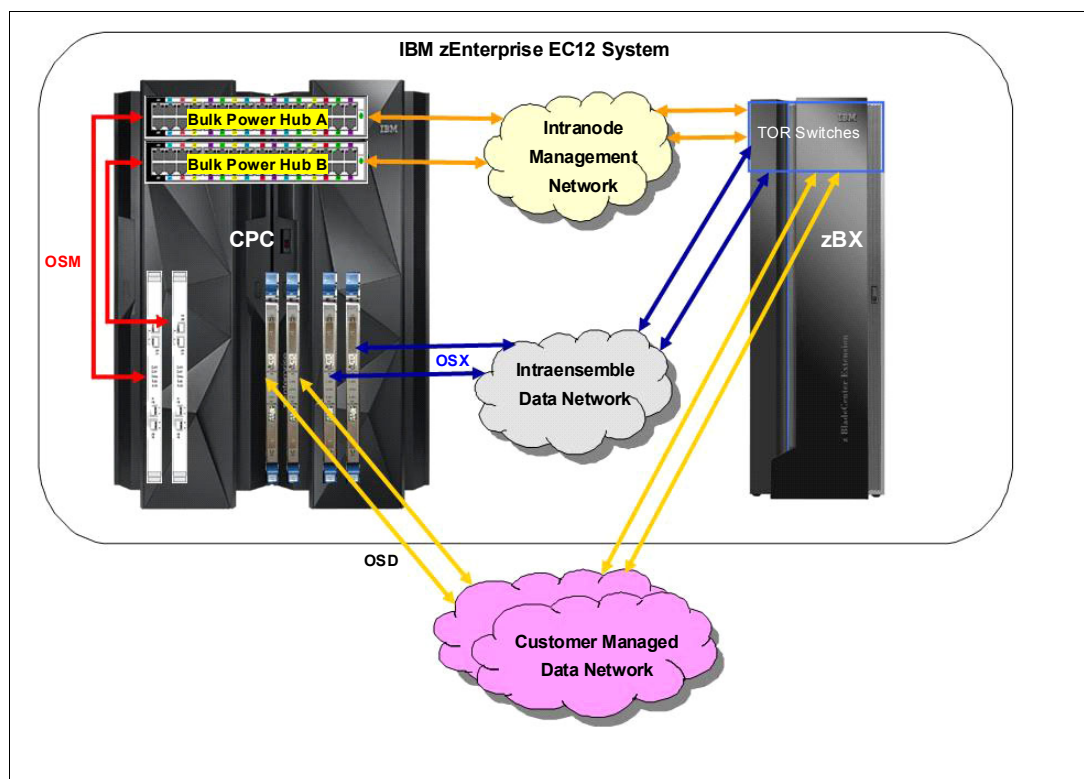


Figure 2-14 INMN, IEDN, and client-managed data networks in an ensemble

Optionally, as part of the ensemble, any OSA-Express5S, OSA-Express4S, or OSA-Express3 features (with CHPID type OSD) in the zEC12 or zBC12 can connect to the client-managed data network. The client-managed network can also be connected to the IEDN TOR switches in the zBX.

In addition, each BladeCenter chassis in the zBX has two Fibre Channel (FC) switch modules that connect to FC disk storage a SAN switch. Each FC switch supports up to six external FC links to connect to SAN switches.

⁹ Not available on zBC12

Figure 2-15 shows a rear view of a two-rack zBX configuration.

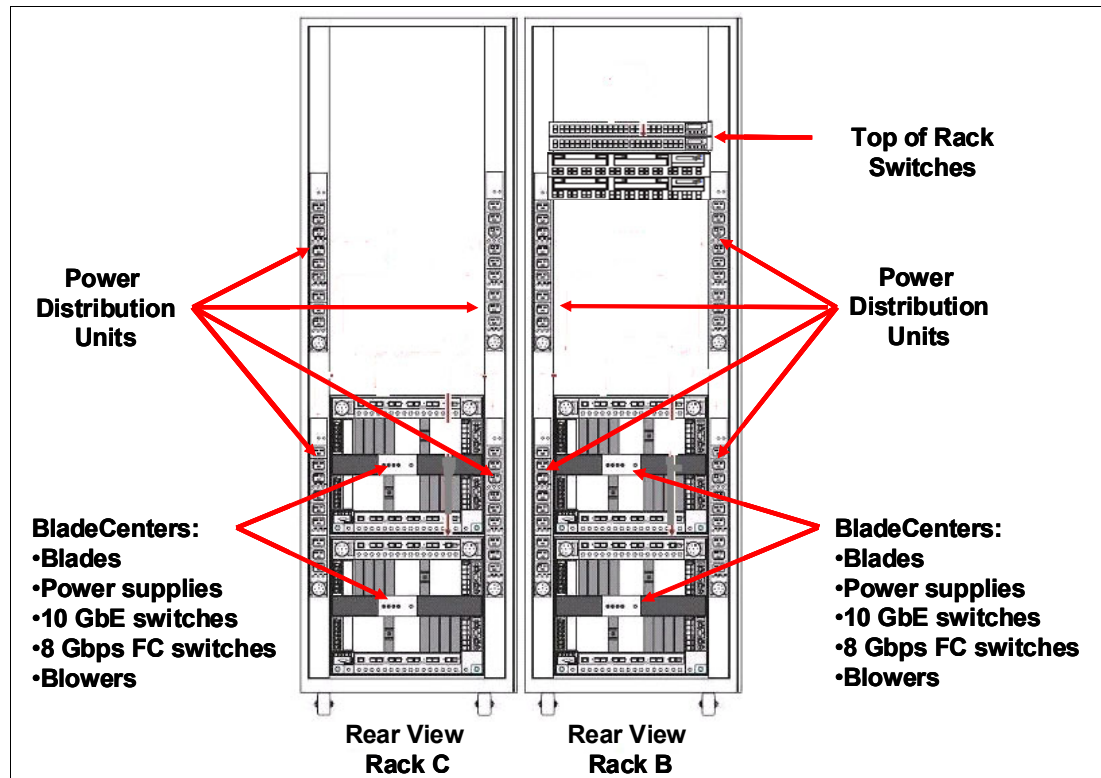


Figure 2-15 zBX rear view: two-rack configuration

The zBX racks include the following features:

- ▶ Two TOR 1000BASE-T switches (rack B only) for the INMN
- ▶ Two TOR 10 GbE switches (rack B only) for the IEDN
- ▶ Up to two BladeCenter chassis in each rack

Each BladeCenter consists of the following features:

- Up to 14 blade slots
- Two Advanced Management Modules (AMM)
- Two Ethernet Switch Modules (ESM)
- Two 10 GbE high speed switch (HSS) modules
- Two 8 Gbps Fibre Channel switch modules
- Two blower modules
- ▶ Power Distribution Units (PDU)

The following blade types are supported in zBX:

- ▶ IBM POWER7 PS701 Express blades
- ▶ IBM System x blades (HX5 7873 dual-socket 16-core)
- ▶ IBM WebSphere DataPower XI50 for zEnterprise blades (double-width)

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

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

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

See the *IBM zEnterprise EC12 Technical Guide*, SG24-8049 or the *IBM zEnterprise BC12 Technical Guide*, SG24-8138, for more information about the number of blades that are supported and configuration options.



Key functions and capabilities of IBM zEnterprise EC12 and IBM zEnterprise BC12

IBM zEnterprise EC12 is the follow-on to the IBM zEnterprise 196 and the flagship of the IBM Systems portfolio. Like its predecessor, the zEC12 offers five hardware models, but has a more powerful processor, more processor units, and new functions and features. The zBC12 is the follow-on to the IBM zEnterprise 114, offers two hardware models, and employs the same technologies as the zEC12.

The superscalar design allows the zEC12 to deliver a record level of capacity over the prior IBM System z® servers. It is powered by 120 of the world's most powerful microprocessors which run at 5.5 GHz. This extreme scalability provides up to 50% more total capacity than its predecessor, the z196. The zEC12 is the industry's premier enterprise infrastructure choice for large-scale consolidation, secure data serving, and transaction processing capabilities.

Likewise, the zBC12, which exploits the same microprocessor chip as the zEC12, running at 4.2 GHz, and which provides the same enhancements as the zEC12, is a leader in its class.

The zEC12 and zBC12 support heterogeneous platform requirements by introducing the *zEnterprise BladeCenter Extension (zBX) Model 003* and an updated Unified Resource Manager. This support allows extending the System z management strengths to other systems and workloads that run on AIX on POWER7, and Linux and Microsoft Windows on IBM System x servers. The *zBX Model 003* can house the IBM WebSphere DataPower Integration Appliance XI50 for zEnterprise (DataPower XI50z), select IBM BladeCenter PS701 Express blades, and IBM BladeCenter HX5 (7873) blades for increased flexibility in “fit for purpose” application deployment.

In this chapter, we highlight some of the zEC12 and zBC12 functions and capabilities and point out solution areas where they can be of special value:

- ▶ Virtualization
- ▶ zEC12 and zBC12 technology improvements
- ▶ Capacity and performance
- ▶ zEnterprise common time functions
- ▶ Hardware Management Console functions
- ▶ zEnterprise CPC power and cooling functions
- ▶ zEnterprise BladeCenter Extension Model 003
- ▶ Reliability, availability, and serviceability (RAS)
- ▶ High availability technology for zEnterprise

3.1 Virtualization

The zEC12 and zBC12 are fully virtualized, with the goal of maximizing the utilization of their resources, lowering the total amount of resources that are needed and their cost.

Virtualization is a key strength of the IBM System z[®] family. It is embedded in the architecture and built into the hardware, firmware, and operating systems.

Virtualization requires a *hypervisor*, which is the control code that manages multiple independent operating system images. Hypervisors can be implemented in software or hardware, and the zEC12 and zBC12 have both. The hardware hypervisor for the zEC12 and zBC12 is known as *IBM Processor Resource/Systems Manager™ (PR/SM)*. PR/SM is implemented in firmware as part of the base system, fully virtualizes the system resources, and does not require any additional software to run. The software hypervisor is implemented by the z/VM operating system. z/VM uses some PR/SM functions.

In the zBX, PowerVM Enterprise Edition is the hypervisor that offers a virtualization solution for any IBM Power Systems workload that runs on AIX. It allows use of the POWER7 processor-based PS blades and other physical resources, providing better scalability and reduction in resource costs. IBM System x blades have a Kernel Virtual Machine-based (KVM-based), integrated hypervisor with identical objectives.

Virtualization is key to the establishment of flexible infrastructures, with automated management and monitoring, such as those underpinning Cloud offerings, including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Application as a Service (AaaS). We will return to this subject after describing the hardware and software virtualization capabilities of the zEC12 and zBC12.

3.1.1 zEC12 and zBC12 hardware virtualization

Processor Resource/Systems Manager (PR/SM) was first implemented in the mainframe in the late 1980s. It allows defining and managing subsets of the server resources that are known as logical partitions (LPAR). PR/SM virtualizes processors, memory, and I/O features. Certain features are purely virtual implementations. For example, HiperSockets work like a LAN but do not use any I/O hardware.

PR/SM is always active on the system and has been enhanced to provide more performance and platform management benefits. PR/SM technology on zEC12 has received Common Criteria EAL5+ security certification. Each logical partition (LPAR) is as secure as a stand-alone system and can run any of the following supported operating systems:

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

The LPAR definition includes a number of logical PUs, memory, and I/O devices. The *z/Architecture* (inherent in the zEC12, zBC12 and its predecessors) is designed to meet those stringent requirements with low overhead and has achieved the highest security certification of the industry: Common Criteria EAL5+ with a Specific Target of Evaluation (Logical Partitions). This design has been proven in many client installations over several decades.

On zEC12 up to 60 LPARs can be defined, and hundreds or even thousands of virtual servers can be run under z/VM. zBC12 can have up to 30 LPARs and host many tens of servers under z/VM. Therefore, a high rate of context switching is to be expected, and accesses to the memory, caches, and virtual I/O devices must be kept isolated.

Logical processors

Logical processors are defined and managed by PR/SM and are perceived by the operating systems as real processors. These processors can be characterized as follows:

- ▶ Central processors (CP)
- ▶ System z Application Assist Processors (zAAP)
- ▶ System z Integrated Information Processors (zIIP)
- ▶ Integrated Facility for Linux (IFL)
- ▶ Internal Coupling Facility (ICF)

In addition, pre-characterized processors that are part of the system base configuration are always present: System Assist Processors (SAP) and Integrated Firmware Processors (IFP). They provide support for all LPARs but are never part of an LPAR configuration.

PR/SM is responsible for accepting requests for work on logical processors by dispatching logical processors on physical processors. Physical processors can be shared across LPARs, but can also be dedicated to an LPAR. However, an LPAR must have its logical processors either all shared or all dedicated.

The sum of logical processors (LPU) defined in all of the LPARs activated in a central processor complex (CPC) might be well over the number of physical processors (PPU). The maximum number of LPUs that can be defined in a single LPAR cannot exceed the total number PPUs that are available in the CPC. To achieve optimal ITR performance in sharing LPUs, it is suggested to keep the total number of online LPUs to a minimum. This action reduces both software and hardware overhead.

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

Logical processors can be dynamically added to and removed from LPARs. Operating system support is required to take advantage of this capability. Starting with z/OS V1R10, z/VM V5R4, and z/VSE V4R3 the ability to dynamically define and change the number and type of reserved PUs in an LPAR profile can be used for that purpose. No pre-planning is required. The new resources are immediately available to the operating systems and, in the case of z/VM, to its guests. Linux on System z provides the Standby CPU activation/deactivation function, which is implemented in SLES 11 and RHEL 6.

z/VM-mode partitions

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

- ▶ Operate Linux on System z on IFLs or CPs.
- ▶ Operate z/OS, z/VSE, and z/TPF on CPs.
- ▶ Operate z/OS while fully allowing *System z Application Assist Processor (zAAP)* and *System z Integrated Information Processor (zIIP)* usage by workloads (such as WebSphere and DB2) for an improved economics environment.

- Operate a complete Sysplex with ICF usage. This setup is especially valuable for testing and operations training; however, it is not recommended for production environments.

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

- CPs
- IFLs
- zIIPs
- zAAPs
- ICFs

If only Linux on System z is to be run under z/VM, then a z/VM-mode LPAR is not required, and we suggest that a Linux-only LPAR be used instead.

Memory

To ensure security and data integrity, memory cannot be concurrently shared by active LPARs. In fact, a strict isolation is maintained.

Using the plan-ahead capability, memory can be physically installed without being enabled. It can then be enabled when it is necessary. z/OS and z/VM support dynamically increasing the memory size of the LPAR.

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

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

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

The z/Architecture has a robust virtual storage architecture that allows, per LPAR, the definition of an unlimited number of address spaces and the simultaneous use by each program of up to 1023 of those address spaces. Each address space can be up to 16 EB (1 exabyte = 2^{60} bytes). Thus, the architecture has no real limits. Practical limits are determined by the available hardware resources, including disk storage for paging.

Isolation of the address spaces is strictly enforced by the Dynamic Address Translation hardware mechanism. The validation of a program’s right to read or write in each page frame is accomplished by comparing the page key with the key of the program that is requesting access. This mechanism has been in use since the System 370. Memory keys were part of, and used by, the original System 360 systems. Definition and management of the address spaces is under operating system control. Three addressing modes, 24-bit, 31-bit, and 64-bit, are simultaneously supported. This provides compatibility with earlier versions and investment protection.

zEC12 and zBC12 introduce 2 GB pages, in addition to the 4 KB and 1 MB pages, and an extension to the z/Architecture: the Enhanced Dynamic Address Translation-2 (EDAT-2).

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

I/O virtualization

The zEC12 supports four Logical Channel Subsystems (LCSS) each with 256 channels, for a total of 1024 channels. The zBC12 supports two LCSSs, for a total of 512 channels. In addition to the dedicated use of channels and I/O devices by an LPAR, I/O virtualization allows concurrent sharing of channels. This architecture also allows sharing the I/O devices that are accessed through these channels, by several active LPARs. This function is known as *Multiple Image Facility (MIF)*. The shared channels can belong to different channel subsystems, in which case they are known as *spanned channels*.

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

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

Many installations use the *parallel access volume (PAV)* function, which allows accessing a device by several addresses (normally one base address and an average of three aliases). This feature increases the throughput of the device by using more device addresses. *HyperPAV* takes the technology a step further by allowing the I/O Supervisor (IOS) in z/OS (and the equivalent function in the Control Program of z/VM) to create PAV structures dynamically. The structures are created depending on the current I/O demand in the system, lowering the need for manually tuning the system for PAV use.

In large installations, the total number of device addresses can be high. Thus, the concept of *channel sets* was introduced with the IBM System z9®. On the zEC12, up to three sets of approximately 64 K device addresses are available; on the zBC12 up to two sets are available. This availability allows the base addresses to be defined on set 0 (IBM reserves 256 subchannels on set 0) and the aliases on set 1 and set 2. In total, 196,350 subchannel addresses (130,815 on the zBC12) are available per channel subsystem. Channel sets are used by the *Metro Mirror* (also referred as synchronous Peer-to-Peer Remote Copy (PPRC)) function by the ability to have the Metro Mirror primary devices defined in channel set 0. Secondary devices can be defined in channel set 1 and 2, providing more connectivity through channel set 0.

To reduce the complexity of managing large I/O configurations further, starting with z/OS V1R10, System z introduced *extended address volumes (EAV)*. EAV is designed to provide large disk volumes by using virtualization technology. In addition to z/OS, both z/VM (starting with V5R4 with APARs) and Linux on System z support EAV.

By extending the disk volume size, potentially fewer volumes can be required to hold the same amount of data, making systems and data management less complex. EAV is supported by the IBM DS8000® series. Devices from other vendors should be checked for EAV compatibility.

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

The dynamic I/O configuration function is supported by z/OS and z/VM. It provides the capability of concurrently changing the currently active I/O configuration. Changes can be made to channel paths, control units, and devices. The existence of a fixed HSA area in the zEC12 and zBC12 greatly eases the planning requirements and enhances the flexibility and availability of these reconfigurations.

3.1.2 zEnterprise software virtualization

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

z/VM uses the resources of the LPAR in which it is running to create functional equivalents of real System z servers, which are known as *virtual machines (VM)* or *guests*. A z/VM virtual machine is the functional equivalent of a real server. In addition, z/VM can emulate I/O peripheral devices (for instance, printers) by using spooling and other techniques, and LAN switches and disks by using memory.

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

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

Although z/VM imposes no limits on the number of defined VMs, the number of active VMs is limited by the available resources. On a zBC12 system many tens of VMs can be supported. On a large server, such as the zEC12, thousands of VMs can be activated.

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

Virtualization provides hardware-enabled resource sharing, and can also be used for the following functions:

- ▶ Isolate production, test, training, and development environments.
- ▶ Support previous applications.
- ▶ Test new hardware configurations without actually buying the hardware.
- ▶ Enable parallel migration to new system or application levels, and provide easy back-out capabilities.

z/VM V6R2 introduced a new feature, *single system image (SSI)*. SSI enables improved availability, better management of planned outages, and capacity growth by creating clusters of z/VM systems with simplified management.

With SSI, it is possible to cluster up to four z/VM images in a single logical image. Highlights of SSI features include these:

- ▶ Live Guest Relocation (LGR) for Linux, the ability to move executing virtual servers without disruption from one z/VM system to another in the SSI.
- ▶ Management of resources with multi-system virtualization, to allow up to four z/VM instances to be clustered as a single system image.
- ▶ Horizontal scalability with up to four systems, even on mixed hardware generations.
- ▶ Availability, through non-disruptively moving work to available system resources and non-disruptively moving system resources to work.
- ▶ An SSI cluster can contain both 6.2 and 6.3 members, and a member can be upgraded from 6.2 to 6.3 using a new *upgrade in place* installation feature.

For more information about SSI, see *An introduction to z/VM Single System Image (SSI) and Live Guest Relocation (LGR)*, SG24-8006 and *Using z/VM v 6.2 Single System Image (SSI) and Live Guest Relocation (LGR)*, SG24-8039.

The Unified Resource Manager (zManager) uses the management *application programming interface (API)* of z/VM to provide a set of resource management functions for the z/VM V6R2 environment.

In light of the IBM cloud strategy and adoption of OpenStack, the management of z/VM environments in zManager is now stabilized and will not be further enhanced. Accordingly, zManager will not provide systems management support for z/VM V6R3 and later releases. However, zManager will continue to play a distinct and strategic role in the management of virtualized environments created by integrated firmware hypervisors (PR/SM, PowerVM, and System x hypervisor based on KVM) of zEnterprise.

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

3.1.3 zBX virtualized environments

On the zBX, the IBM POWER7 processor-based PS701 blades run PowerVM Enterprise Edition to create a virtualized environment that is similar to the one found in IBM Power Systems servers. The POWER7 processor-based LPARs run the AIX operating system.

PowerVM is EAL4+ certified and is isolated on the intranode management network, providing intrusion prevention, integrity, and secure virtual switches with integrated consolidation.

The IBM System x blades are also virtualized. The integrated System x hypervisor uses Kernel-based virtual machines. Support is provided for Linux and Microsoft Windows.

PowerVM, as well as the integrated hypervisor for the System x blades, is managed by the zEnterprise Unified Resource Manager, so it is shipped, deployed, monitored, and serviced at a single point.

Management of the zBX environment is done as a single logical virtualized environment by the Unified Resource Manager.

3.1.4 zEnterprise-based clouds

Cloud computing is a paradigm for providing IT services. It capitalizes on the ability to rapidly and securely deliver standardized offerings, while retaining the capacity for customizing the environment. Elasticity, allowing to accompany the ebbs and flows of demand, and exploiting just-in-time provisioning is another requirement. We make no distinction here between private and public clouds, as they are both well addressed by zEnterprise.

Virtualization is critical to the economic and financial viability of those offerings, because it allows to minimize over provisioning of resources. and reutilizing them at the end of the virtual server life-cycle.

Due to the extreme integration in the hardware, virtualization on zEC12 and zBC12 is highly efficient (the best in the industry) and encompasses computing as well as I/O resources, including the definition of *internal virtual networks with switches*. These are all characteristics of *Software Defined Environments*, and allow supporting on a single real server, dense sets of virtual servers and server networks, with up to 100% sustained resource utilization and the highest levels of isolation and security.

In other words, the cloud solution costs, whether hardware, software, or management, are minimized.

Cloud elasticity requirements are covered by the zEC12 and zBC12 granularity offerings, including capacity levels and Capacity on Demand. These and other technologic leadership characteristics that make the zEnterprise CPCs the server golden standard, are discussed in the remainder of this chapter.

z/VM V6R3 OpenStack support: If you want to get started with Cloud computing, the Extreme Cloud Administration Toolkit (xCAT), a scalable open source tool developed by IBM, can be used to provision, manage, and monitor physical and virtual machines. Because xCAT is integrated into z/VM V6R3, it no longer needs to be separately downloaded, installed, and configured. You can quickly deploy xCAT with a small amount of tailoring.

3.2 zEC12 and zBC12 technology improvements

zEC12 and zBC12 provide technology improvements in these areas:

- ▶ Microprocessor
- ▶ Memory
- ▶ Capacity and performance
- ▶ Flash Express feature
- ▶ 10GbE RoCE Express feature
- ▶ zEDC Express feature
- ▶ Cryptography
- ▶ I/O capabilities
- ▶ Cryptography

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

3.2.1 Microprocessor

The zEC12 and zBC12 have a newly developed microprocessor chip and a newly developed storage control chip. The chips use CMOS S13 (32nm) technology and represent a major step forward in technology use for the IBM System z® products, resulting in increased packaging density.

The microprocessor chip and the storage control chip for the zEC12 are packaged together on a multi-chip module (MCM). The MCM contains six microprocessor chips (each having four, five, or six active cores) and two storage control (SC) chips. The MCM is installed inside a book, and the zEC12 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 PCI Express (PCIe) I/O drawers, I/O drawers, and I/O cages through one or more cables.

On the zBC12, as with the z114, the microprocessor and storage control chip are packaged individually in single chip modules (SCM). The SCMs are mounted on a processor drawer and a zBC12 configuration can have one or two processor drawers. Each processor drawer also contains the memory arrays, I/O connectivity infrastructure, and various other mechanical and power controls.

I/O cables connect the processor drawer to PCIe I/O drawers and/or I/O drawers (I/O cages are not supported on the zBC12).

Standard PCIe and InfiniBand protocols are used for fast transfer of large volumes of data between the memory in the books or processor drawers and the I/O cards housed in the PCIe I/O drawers, I/O drawers, and I/O cages.

zEC12 and zBC12 processor chip

The zEC12 and zBC12 processor chips provide more functions per chip (six cores on a single chip) because of technology improvements that allow designing and manufacturing more transistors per unit of area. This configuration translates into using fewer chips to implement the needed functions, which helps enhance system availability.

The System z microprocessor development followed the same basic design set since the 9672-G4 (announced in 1997) until the z9. That basic design was stretched to its maximum, so a fundamental change was necessary. The z10 chip introduced a high-frequency design, which was improved with the z196 and z114 quad-core chip, and enhanced again with the hex-core zEC12 and zBC12 microprocessor chip.

Each core of the processor chip includes one co-processor for hardware acceleration of data compression and cryptography, I/O bus and memory controllers, and an interface to a separate storage controller/cache chip; see Figure 3-1.

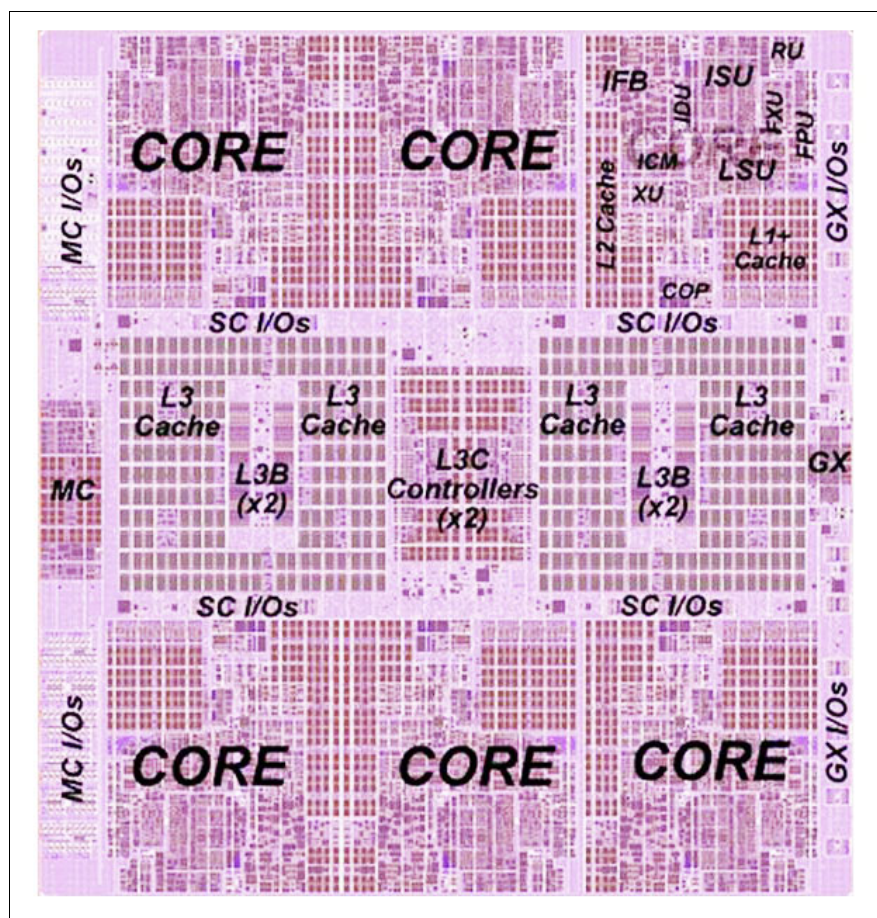


Figure 3-1 zEC12/zBC12 Hex-Core microprocessor chip

On-chip cryptographic hardware includes the full complement of the *Advanced Encryption Standard (AES)* algorithm, *Secure Hash Algorithm (SHA)*, and the *Data Encryption Standard (DES)* algorithm, as well as the protected key implementation.

zEC12 and zBC12 processor design highlights

The z/Architecture offers a rich *complex instruction set computer (CISC) Instruction Set Architecture (ISA)* supporting multiple arithmetic formats.

The z196 introduced 110 new instructions and offered a total of 984, out of which 762 were implemented entirely in hardware. The zEC12 and zBC12 also introduce new instructions, notably for the Transactional Execution and the EDAT-2 facilities.

Compared to z196 and z114, the zEC12 and zBC12 processor design improvements and architectural extensions include the following features:

- ▶ **Balanced Performance Growth**
 - 1.5 times more system capacity:
 - 50% more cores in a central processor chip.
 - Maximum number of cores increased from 96, on the z196, to 120 on the zEC12, and from 14, on the z114, to 18 on the zBC12.

- Third Generation High Frequency processor:
 - Frequency increased from 5.2 GHz, on the z196, to 5.5 GHz on zEC12, and from 3.8 GHz, on the z114, to 4.2 GHz on zBC12.
 - Up to 25% faster zEC12 uniprocessor performance as compared to z196.
 - Up to 36% faster zBC12 uniprocessor performance as compared to z114.
- ▶ Innovative Local Data-Cache design with larger caches but shorter latency:
 - Total L2 per core is 33% bigger.
 - Total on-chip shared L3 is 100% larger.
 - Unique private L2 cache (1 MB) design reduces L1 miss latency.
- ▶ Second Generation of Out of Order design (OOO) with increased resources and efficiency:
 - Numerous pipeline improvements that are based on z10 and z196 designs.
 - Number of instructions in flight is increased from five to seven.
- ▶ Improved Instruction Fetching Unit:
 - Improved capacity (24K branches) for large footprint workloads.
 - Improved sequential instruction stream delivery.
- ▶ Dedicated co-processor for each core with improved performance and more capability:
 - New hardware support for Unicode UTF8<>UTF16 bulk conversions (CU12/CU21).
 - Improved startup latency.
- ▶ Multiple innovative architectural extensions for software usage:
 - *Transactional Execution (TX)*, known in the academia as *Hardware Transactional Memory (HTM)*. This feature allows software-defined “lockless” sequences to be treated as an atomic “transaction” and improves efficiency on highly parallelized applications and multi-processor handling.
 - Runtime Instrumentation. Allows dynamic optimization of code generation as it is being executed.
 - Enhanced Dynamic Address Translation-2 (EDAT-2). Supports 2 GB page frames.
- ▶ Increased instruction issue, execution, and completion throughput:
 - Improved instruction dispatch and grouping efficiency.
 - Millicode handling.
 - Next Instruction Access Intent.
 - Load and Trap instructions.
 - Branch Prediction Preload.
 - Data prefetch.

Hardware decimal floating point function

Hardware decimal floating point (HDFP) support was introduced with the z9 EC and enhanced with a new decimal floating point accelerator feature in IBM zEnterprise 196. zEC12 and zBC12 facilitate better performance on traditional zoned-decimal operations with a broader usage of Decimal Floating Point facility by COBOL and PL/I programs.

This facility is designed to speed up such calculations and provide the necessary precision demanded mainly by the financial institutions sector. The decimal floating point hardware fully implements the new IEEE 754r standard.

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

Support and usage of HDFP varies with operating system and release. For a detailed description, see *IBM zEnterprise EC12 Technical Guide*, SG24-8049 or *IBM zEnterprise BC12 Technical Guide*, SG24-8138. See also “Decimal floating point (z/OS XL C/C++ considerations)” on page 143.

Transactional Execution (TX) facility

This capability, which is known in the industry as *hardware transactional memory*, allows issuing a group of instructions atomically, that is, either all the results of the instructions in the group are committed or none are, in true transactional way. The execution is optimistic: The instructions are issued, but previous state values are saved in a “transactional memory”. If the transaction succeeds, the saved values are discarded, otherwise they are used to restore the original values. Software can test the execution’s success and re-drive the code, if needed, using the same or a different path.

The TX facility provides several instructions, including declaring the beginning and end of a transaction, and to cancel the transaction. TX is expected to provide significant performance benefits and scalability to workloads by being able to avoid most of the locks. This ability is especially important for heavily threaded applications, such as Java.

Runtime Instrumentation Facility

The Runtime Instrumentation Facility is designed to provide managed run times and just-in-time compilers with enhanced feedback on application behavior, allowing dynamic optimization on code generation as it is being executed.

Statement of Direction: IBM plans for future maintenance roll-ups of 31-bit and 64-bit IBM SDK7 for z/OS Java to provide use of new IBM zEnterprise EC12 and BC12 features, including: Flash Express and pageable large pages, and Miscellaneous-Instruction-Extension Facility.

3.2.2 Memory

The zEC12 can have up to 3,040 GB of usable memory installed. The zBC12 can have up to 496 GB of usable memory, significantly more than its predecessor z114.

In addition, the zEC12 has doubled the size of the hardware system area (HSA) when compared with its predecessor, z196. The HSA is not included in the memory which the client orders, and has a fixed size of 32 GB. The zBC12 has a fixed size HSA of 16 GB.

z/Architecture addressing modes: The z/Architecture simultaneously supports 24-bit, 31-bit, and 64-bit addressing modes. These modes provide compatibility with earlier versions and investment protection.

Support of large memory varies with the operating system, as follows:

- ▶ z/OS V1R11 and later support up to 4 TB.
- ▶ z/VM V6R3 supports up to 1 TB.
- ▶ z/VM V5R4 and z/VM V6R2 support up to 256 GB.
- ▶ z/VSE V4R2 and later support up to 32 GB.
- ▶ z/TPF V1R1 supports up to 4 TB.
- ▶ SUSE SLES 11 supports 4 TB and Red Hat RHEL 6 supports 3 TB.

Regardless of an operating system’s ability to support more than 1 TB of memory, the maximum memory size per logical partition is restricted, on the zEC12, to 1 TB, and to 496 GB on the zBC12.

Plan-ahead memory

When a client can anticipate the requirements for future increases of the installed memory, the initial system order can contain both a starting and additional memory sizes. The additional memory is referred to as plan-ahead memory. A specific memory pricing model is available in support of this capability.

The starting memory size is activated at system installation time and the rest remains inactive. When more physical memory is required, it is fulfilled by activating the appropriate number of plan-ahead memory features. This activation is concurrent and can be nondisruptive to the applications depending on the operating system support. z/OS and z/VM support this function.

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

Flexible memory

Flexible memory was first introduced on the z9 EC as part of the design changes and offerings to support enhanced book availability. Flexible memory is used to temporarily replace the memory that becomes unavailable when performing maintenance on a book.

On zEC12, the additional resources that are required for the flexible memory configurations are provided through the purchase of planned memory features, along with the purchase of memory entitlement. Flexible memory configurations are available only on multi-book models H43, H66, H89, and HA1, and range from 32 GB to 2272 GB, depending on the model.

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

Tip: Flexible memory is not available on zBC12.

Large page support

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

Large pages enable 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. Users of large pages are better represented in the TLB and are expected to perform better.

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

The large page support function is not enabled without the required software support. Without the large page support, page frames are allocated at the current 4 KB size. At the time they were introduced, large pages were treated as fixed pages and were never paged out. With the availability of Flash Express, large pages might become pageable. Under z/OS, they are only available for 64-bit virtual private storage such as virtual memory located above 2 GB.

2 GB large page support

zEC12 and zBC12 introduce 2 GB page frames as an architectural extension. It is aimed at increasing efficiency for DB2 buffer pools, Java heap, and other large structures. Usage of 2 GB pages increases TLB coverage without proportionally enlarging the TLB size:

- ▶ A 2 GB memory page has the following characteristics:
 - It is 2048 times larger than a large page of 1 MB size.
 - It is 524,288 times larger than an ordinary base page with a size of 4 KB.
- ▶ A 2 GB page allows for a single TLB entry to fulfill many more address translations than either a large page or ordinary base page.
- ▶ A 2 GB page provides users with much better TLB coverage, and therefore provides better performance in the following ways:
 - By decreasing the number of TLB misses that an application incurs.
 - By spending less time on converting virtual addresses into physical addresses.
 - By using less real storage to maintain DAT structures.

2 GB Large Page exploitation: Exploitation of 2 GB pages is provided for the IBM 31-bit SDK for z/OS, Java™ Technology Edition, V7.0.0 (5655-W43) and SDK7 IBM 64-bit SDK for z/OS, Java Technology Edition, V7.0.0 (5655-W44).

3.2.3 Native PCIe features and integrated firmware processor

zEC12 and zBC12 introduced feature card types, known as native PCIe features, which require a different management design. The following native PCIe features are available:

- ▶ 10GbE RoCE Express
- ▶ zEDC Express

These features are plugged exclusively into a PCIe I/O drawer, where they co-exist with the other, non-native PCIe, I/O adapters and features, but they are managed in a different way from those other I/O adapters and features. The native PCIe feature cards have a PCHID assigned according to the physical location in the PCIe I/O drawer.

For non-native PCIe features, which are plugged into a PCIe I/O drawer, an I/O drawer, or an I/O cage, all adaptation layer functions have been integrated into the adapter hardware.

For the native PCIe features introduced by zEC12 and zBC12, there are drivers included in the operating system, and the adaptation layer is not needed. The adapter management functions (such as diagnostics and firmware updates) are provided by Resource Groups running on the integrated firmware processor (IFP).

The integrated firmware processor (IFP) is used to manage native PCIe adapters installed in a PCIe I/O drawer. The IFP is allocated from a pool of PUs available for the whole system. Because the IFP is exclusively used to manage native PCIe adapters, it is not taken from the pool of PUs that can be characterized for customer usage.

If a native PCIe feature is present in the system, the IFP is initialized and allocated during the system POR phase. Although the IFP is allocated to one of the physical PUs, it is not visible to the customer. In case of error or failover scenarios, the IFP will act like any other PU (that is, sparing is invoked).

3.2.4 Flash Express

zEC12 and zBC12 introduce the innovative *Flash Express* feature that helps to improve availability and performance to compete more effectively in today's service focused market. Flash Express capabilities enable the following features:

- ▶ Improved z/OS recovery and diagnostic times
- ▶ Handling of workload shifts and coping with dynamic environments more smoothly
- ▶ Use of pageable large pages yielding CPU performance benefits
- ▶ Offloading GBps of random I/O from the I/O Fabric
- ▶ Predictive paging

Flash Express is easy to configure, requires no special skills, and provides rapid time to value. This feature is designed to allow each logical partition to be configured with its own storage-class memory (SCM) address space and to be used for paging. One MB large pages can become pageable.

In Flash Express environment, the data privacy relies on a symmetric key that encrypts the data that temporarily is on the SSD. By using a smart card and an integrated smart card reader on the Support Element (SE), the encryption key is generated within the secure environment of the smart card. The key is tightly coupled to the SE serial number, which ensures that no other SE is able to share the key or the smart card that is associated with a specific SE. The generated key is replicated in a secure way to the alternate Support Element smart card. The key is transferred from the SE to the Flash Express adapter under the protection of a private and public key pair that is generated by the firmware that manages the Flash Express adapter.

Flash Express is an optional and priced feature. It is supported by z/OS V1R13 and above, with the z/OS V1R13 RSM Enablement Offering web deliverable installed. It is fully supported by z/OS V2R1.

Additional functions of Flash Express are expected to be supported later, including 2 GB large pages and dynamic reconfiguration for Flash Express.

Statement of Direction: CFCC Level 19 exploitation of Flash Express:

IBM intends to provide exploitation of the Flash Express feature on zEC12 and zBC12 servers with Coupling Facility Control Code (CFCC) Level 19 for certain Coupling Facility list structures in the first half of 2014. This new function is designed to allow list structure data to be migrated to Flash Express memory as needed when the consumers of data do not keep pace with its creators for some reason, and migrate it back to real memory to be processed.

3.2.5 zEDC Express

zEDC Express is an optional feature, exclusive to the zEC12 and zBC12, designed to help to improve cross platform data exchange, reduce CPU consumption, and save disk space by providing hardware-based acceleration for data compression and decompression for the enterprise. It provides data compression with lower CPU consumption than compression technology previously available on System z.

This capability is of special interest, for instance, to clients experiencing significant year-to-year growth in storage. Savings could be realized initially by making more efficient use of existing capacity, allowing more data to be kept active and online at lower cost, and longer term by elongating time frames for acquisitions for additional storage.

z/OS V2R1 zEnterprise Data Compression

Exploitation support of zEDC Express functionality is provided exclusively by z/OS V2R1 zEnterprise Data Compression for both data compression and decompression.

Statement of Direction: In a future z/VM deliverable, IBM plans to offer z/VM support for guest exploitation of the IBM zEnterprise Data Compression (zEDC) Express feature on the IBM zEnterprise EC12 and IBM zEnterprise BC12 systems.

Support for data recovery (decompression) in the case that zEDC Express is not installed, or installed but not available, on the system, is provided via software on z/OS V2R1, and on V1R13 and V1R12 with appropriate PTFs. Software decompression is slow and uses considerable processor resources, thus it is not recommended for production environments.

Initial support focuses on z/OS SMF log data where, using zEDC compression services to compress records before writing to log streams, higher write rates can be achieved. zEDC is also used to read back the records.

Additional support by z/OS DFSMS (BSAM/QSAM) is planned, providing a new type of policy-based compression support for non-VSAM extended format data sets. Benefits could result from improved disk utilization and economics of using flash for extended format BSAM/QSAM.

Statements of direction:

- ▶ IBM intends to provide exploitation of the zEDC Express feature for DFSMSdss and DFSMSHsm by the end of third quarter 2014.
- ▶ IBM plans for future updates of IBM 31-bit and 64-bit SDK7 for z/OS Java Technology Edition, Version 7 (5655-W43 and 5655-W44) (IBM SDK7 for z/OS Java) to provide exploitation of the following features: the zEDC Express feature and Shared Memory Communications - Remote Direct Memory Access (SMC-R), which is utilized by the 10GbE RoCE Express feature.

IBM System z Batch Network Analyzer

The IBM System z Batch Network Analyzer (zBNA) is a free, “as is” tool. It is available to clients, business partners, and IBM employees.

zBNA replaces the BWATOOL. It is Windows based, provides graphical and text reports, including Gantt charts, and support for alternate processors.

zBNA can be used to analyze customer provided SMF records, in order to identify jobs and data sets which are candidates for zEDC compression, across a specified time window, typically a batch window. zBNA is able to generate lists of data sets by job:

- ▶ Those which already do hardware compression and may be candidates for zEDC
- ▶ Those which may be zEDC candidates but are not in extended format

Thus, zBNA may help estimate utilization of zEDC features and help size number of features needed.

IBM Clients can obtain zBNA and other CPS tools via the Internet at this site:

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS5132>

IBM Business Partners can obtain zBNA and other CPS tools via the Internet at this site:

https://www.ibm.com/partnerworld/wps/servlet/mem/ContentHandler/tech_PRS5133

IBM Employees can obtain zBNA and other CPS tools via the IBM Intranet at this site:

<http://w3.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS5126>

3.2.6 10GbE RoCE Express

The 10 Gigabit Ethernet (10GbE) RoCE Express feature is designed to help reduce consumption of CPU resources for applications utilizing the TCP/IP stack (such as WebSphere accessing a DB2 database).

Use of the 10GbE RoCE Express feature may also help to reduce network latency with memory-to-memory transfers utilizing Shared Memory Communications - Remote Direct Memory Access (SMC-R) in z/OS V2R1. It is transparent to applications and can be used for LPAR-to-LPAR communication on a single z/OS system or server-to-server communication in a multiple CPC environment.

Statement of Direction: In a future z/VM deliverable, IBM plans to offer support for guest exploitation of the 10GbE RoCE Express feature on the IBM zEnterprise EC12 and IBM zEnterprise BC12 servers. This is designed to allow guests to use Shared Memory Communications - Remote Direct Memory Access (SMC-R) using Remote Direct Memory Access over Converged Ethernet (RoCE).

3.2.7 I/O capabilities

zEC12 and zBC12 have many I/O capabilities for supporting high-speed connectivity to resources inside and outside the system. The connectivity of the zEC12 and zBC12 is designed to maximize application performance and satisfy clustering, security, storage area network (SAN), and local area network (LAN) requirements.

Multiple subchannel sets

Multiple subchannel sets (MSS) are designed to provide greater I/O device configuration capabilities for large enterprises. Up to three subchannel sets for zEC12 and two for zBC12 can be defined to each channel subsystem (CSS). Up to four CSSs can be defined on the zEC12 and up to two on the zBC12.

For each additional subchannel set, the amount of addressable storage capacity is 64 K minus one subchannel, which enables a larger number of storage devices. This increase complements other functions (such as large or extended address volumes) and HyperPAV. This can also help facilitate consistent device address definitions, simplifying addressing schemes for congruous devices.

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

Initial program load from an alternate subchannel set

zEC12 and zBC12 support the *initial program load (IPL)* from subchannel set 1 (SS1) or subchannel set 2 (SS2) (zEC12 only). Devices that are used early during IPL processing can be accessed by using subchannel set 1 or subchannel set 2. This flexibility allows the users of Metro Mirror (PPRC) secondary devices that are defined using the same device number and a new device type in an alternate subchannel set to be used for IPL, input/output definition file (IODF), and stand-alone dump volumes, when needed.

Channel subsystem enhancement for I/O resilience

The zEC12 and zBC12 channel subsystem incorporates an improved load balancing algorithm that is designed to provide improved throughput and reduced I/O service times, even when abnormal conditions occur. For example, degraded throughput and response times can be caused by multi-system workload spikes. This reduction can also be caused by resource contention in storage area networks (SAN) or across control unit ports, SAN congestion, suboptimal SAN configurations, problems with initializing optics, dynamic fabric routing changes, and destination port congestion.

When such events occur, the channel subsystem is designed to dynamically select channels to optimize performance. The subsystem also minimizes imbalances in I/O performance characteristics (such as response time and throughput) across the set of channel paths to each control unit. This function is done by using the in-band I/O instrumentation and metrics of the System z FICON and zHPF protocols.

This channel subsystem enhancement is exclusive to zEC12 and zBC12 and is supported on all FICON channels when configured as CHPID type FC. In support of this new function, z/OS V1R12 and V1R13 with a *program temporary fix (PTF)* also provide an updated health check based on an I/O rate-based metric, rather than on initial control unit command response time.

This enhancement is transparent to operating systems. However, this feature requires an updated health check based on an I/O rate-based metric, rather than on initial control unit command response time, provided by z/OS V1R12 and V1R13 with a PTF and later.

FICON connectivity

The FICON (Fibre CONnection) features in the zEC12 and zBC12 can provide connectivity to servers, FC switches, and various devices (control units, disk, tape, printers) in a SAN environment. FICON improves upon the Fibre Channel Protocol (FCP) and continues to evolve, delivering improved throughput, reliability, availability, and serviceability.

High Performance FICON for System z

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

When used by the FICON channel, the z/OS operating system, and the control unit (appropriate levels of Licensed Internal Code are required), the FICON channel overhead can be reduced and performance can be improved. Additionally, the changes to the architecture provide end-to-end system enhancements to improve reliability, availability, and serviceability (RAS). The zHPF channel programs can be used, for instance, by z/OS OLTP I/O workloads, DB2, VSAM, PDSE, and zFS. zHPF requires matching support by the DS8000 series or similar devices from other vendors.

The zHPF is exclusive to zEnterprise System. The FICON Express8S, FICON Express8, and FICON Express4 (*channel path identifier (CHPID)* type FC) concurrently support both the existing FICON protocol and the zHPF protocol in the server Licensed Internal Code.

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

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

Modified Indirect Data Address Word (MIDAW) facility

The MIDAW facility is a system architecture and software usage that is designed to improve FICON performance. This facility was introduced with z9 servers and is used by the media manager in z/OS.

The MIDAW facility provides a more efficient structure for certain categories of data-chaining I/O operations resulting in improved FICON performance and I/O response times, in particular for extended format data-sets (DB2 is a major user). For more information about FICON, FICON channel performance, and MIDAW, see the I/O Connectivity website¹ and the IBM Redpaper™ publication, *How does the MIDAW Facility Improve the Performance of FICON Channels Using DB2 and other workloads?*². Also, see *DS8000 Performance Monitoring and Tuning*, SG24-7146.

Extended distance FICON

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

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

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

Extended Distance FICON is transparent to operating systems and applies to all the FICON Express8S, FICON Express8, and FICON Express4 features carrying basic FICON traffic (CHPID type FC). For usage, the control unit must support the new IU pacing protocol.

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

z/OS discovery and autoconfiguration

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

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

¹ <http://www.ibm.com/systems/z/connectivity/>

² <http://www.redbooks.ibm.com/abstracts/redp4201.html>

zDAC is designed to perform discovery for all systems in a sysplex that support the function. The proposed configuration incorporates the current contents of the I/O definition file (IODF) with additions for newly installed and changed control units and devices. zDAC is designed to simplify I/O configuration on zEC12 or zBC12 running z/OS and reduce complexity and setup time. zDAC applies to all FICON features supported on zEnterprise when configured as CHPID type FC.

FICON name server registration

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

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

This zEC12 and zBC12 function is transparent to operating systems and applicable to all FICON Express8S, FICON Express8, and FICON Express4 features (CHPID type FC). For more information, see *IBM System z Connectivity Handbook*, SG24-5444.

FCP connectivity

Fibre Channel Protocol is fully supported on the zEC12 and zBC12. It is commonly used with Linux on System z and supported by the z/VM, z/VSE and Linux on System z operating systems.

Fibre Channel Protocol enhancements for small block sizes

The *Fibre Channel Protocol (FCP)* Licensed Internal Code was modified to help provide increased I/O operations per second for small block sizes. This FCP performance improvement is transparent to operating systems and applies to all the FICON Express8S, FICON Express8, and FICON Express4 features, when configured as CHPID type FCP, communicating with SCSI devices.

For more information about FCP channel performance, see the performance technical papers on the System z I/O connectivity web page³.

FCP channels to support T10-DIF for enhanced reliability

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

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

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

³ http://www.ibm.com/systems/z/hardware/connectivity/fcp_performance.html

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. This is achieved by assigning a unique *worldwide port name (WWPN)* for each operating system that is 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 that are based on the assigned WWPN can be applied in the SAN environment. This function can be done by using standard mechanisms such as zoning in SAN switches and logical unit number (LUN) masking in the storage controllers.

WWPN tool

A part of the installation of your zEC12 or zBC12 server is the planning of the SAN environment. IBM has made a stand-alone tool available to assist with this planning before the installation. The tool, which is known as the *WWPN tool*, assigns WWPNs to each virtual Fibre Channel Protocol (FCP) channel/port. This function is done by using the same WWPN assignment algorithms that a system uses when assigning WWPNs for channels using NPIV. Thus, the SAN can be set up in advance, allowing operations to proceed much faster after the server is installed.

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

The WWPN tool is available for download at IBM Resource Link⁴ and is applicable to all FICON channels defined as CHPID type FCP (for communication with SCSI devices) on zEC12 or zBC12.

LAN connectivity

The zEC12 and zBC12 offer a wide range of functions that can help consolidate or simplify the LAN environment with the supported OSA-Express features, though also satisfying the demand for more throughput. Improved throughput (mixed inbound/outbound) is achieved by the data router function that was introduced in the OSA-Express3 and enhanced in OSA-Express5S, and OSA-Express4S features.

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

Queued direct I/O (QDIO) optimized latency mode (OLM)

QDIO OLM can help improve performance for applications that have a critical requirement to minimize response times for inbound and outbound data. OLM optimizes the interrupt processing as noted in the following configurations:

- ▶ For inbound processing, the TCP/IP stack looks more frequently for available data to process, ensuring that any new data is read from the OSA-Express3 or OSA-Express4S without requiring more program controlled interrupts (PCI).
- ▶ For outbound processing, the OSA-Express5S, OSA-Express4S, or OSA-Express3 looks more frequently for available data to process from the TCP/IP stack, thus not requiring a Signal Adapter (SIGA) instruction to determine whether more data is available.

⁴ <http://www.ibm.com/servers/resourceLink/>

Inbound workload queuing (IWQ)

IWQ is designed to help reduce overhead and latency for inbound z/OS network data traffic and implement an efficient way for initiating parallel processing. This is achieved by using an OSA-Express5S, OSA-Express4S, or OSA-Express3 feature in QDIO mode (CHPID types OSD and OSX) with multiple input queues and by processing network data traffic that is based on workload types. The data from a specific workload type is placed in one of four input queues (per device), and a process is created and scheduled to run on one of multiple processors, independent from the other three queues. This improves performance because IWQ can use the symmetric multiprocessor (SMP) architecture of the zEC12 and zBC12.

Virtual local area network (VLAN) support

VLAN is a function of OSA-Express features that takes advantage of the IEEE 802.1q 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, TCP/IP stacks can exist, potentially sharing OSA-Express features. VLAN provides a greater degree of isolation by allowing contact with a server from only the set of stations that comprise the VLAN.

Virtual MAC (VMAC) support

When sharing OSA port addresses across LPARs, VMAC support enables each operating system instance to have a unique virtual MAC (VMAC) address. All IP addresses associated with a TCP/IP stack are accessible by 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 used by z/OS and supported by z/VM for guest usage.

QDIO data connection isolation for the z/VM environment

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

The queued direct input/output (QDIO) data connection isolation enables the following elements:

- ▶ Adherence to security and HIPAA-security guidelines and regulations for network isolation between the instances that share physical network connectivity.
- ▶ Establishing security zone boundaries that are defined by the network administrators.
- ▶ A mechanism to isolate a QDIO data connection (on an OSA port) by forcing traffic to flow to the external network. This feature ensures that all communication flows only between an operating system and the external network.

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

QDIO data connection isolation (also known as VSWITCH port isolation) applies to the z/VM environment when using the Virtual Switch (VSWITCH) function and to all of the OSA-Express5S, OSA-Express4S and OSA-Express3 features (CHPID type OSD) on zEC12. z/OS supports a similar capability.

QDIO interface isolation for z/OS

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

QDIO interface isolation is supported by Communications Server for z/OS V1R11 and later and for all OSA-Express5S, OSA-Express4S, and OSA-Express3 features on zEC12 and zBC12.

Open Systems Adapter for NCP (OSN)

The OSN support is able to provide channel connectivity from System z Operating Systems to IBM Communication Controller for Linux on System z (CCL). This function is done by using the Open Systems Adapter for the Network Control Program (OSA for NCP) supporting the Channel Data Link Control (CDLC) protocol.

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

OSN: The OSN CHPID type is not supported on OSA-Express 5S GbE or OSA-Express4S GbE features.

Network management: Query and display OSA configuration

As more complex functions are added to OSA, the ability for the system administrator to display, monitor, and verify the specific current OSA configuration unique to each operating system is becoming more complex. OSA-Express5S, OSA-Express4S and OSA-Express3 have the capability for the operating system to query and display the current OSA configuration information (similar to OSA/SF) directly. z/OS uses this OSA capability by providing the TCP/IP operator command **Display OSAINFO**, which allows the operator to monitor and verify the current OSA configuration, helping to improve the overall management, serviceability, and usability of OSA-Express5S, OSA-Express4S, and OSA-Express3 features.

The **Display OSAINFO** command is exclusive to OSA-Express5S, OSA-Express4S, and OSA-Express3 (CHPID types OSD, OSM, and OSX), the z/OS operating system, and is supported on z/VM for guest usage.

zEnterprise ensemble connectivity

With the IBM zEnterprise System, two CHPID types were introduced to support the zEnterprise ensemble:

- ▶ OSA-Express for Unified Resource Manager (OSM) for the *intranode management network (INMN)*
- ▶ OSA-Express for zBX (OSX) for the *intraensemble data network (IEDN)*

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

The OSM connections are through the Bulk Power Hubs (BPH) in the zEnterprise CPC. The BPHs are also connected to the INMN TOR switches in the zBX. The INMN requires two OSA-Express5S 1000BASE-T, OSA-Express4S 1000BASE-T⁵, or OSA-Express3 1000BASE-T ports from separate features.

The IEDN is the ensemble's other private and secure internal network. IEDN is used for communications across the virtualized images (LPARs and virtual machines). The IEDN connections use MAC addresses, not IP addresses (Layer 2 connection).

The OSX connections are from the zEnterprise CPC to the IEDN TOR switches in zBX. The IEDN requires two OSA-Express5S 10 GbE, OSA-Express4S 10 GbE, or OSA Express3 10 GbE ports from separate features.

HiperSockets

HiperSockets have been called the "network in a box." HiperSockets simulates LANs entirely in the hardware. The data transfer is from LPAR memory to LPAR memory, mediated by microcode. The zEC12 and zBC12 support up to 32 HiperSockets. One HiperSockets network can be shared by up to 60 LPARs on a zEC12 and up to 30 LPARs on a zBC12. Up to 4096 communication paths support a total of 12,288 IP addresses across all 32 HiperSockets.

HiperSockets Layer 2 support

The HiperSockets internal networks can support two transport modes:

- ▶ Layer 2 (link layer)
- ▶ Layer 3 (network or IP layer)

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

Layer 2 support can help facilitate server consolidation. Complexity can be reduced, network configuration is simplified and intuitive, and LAN administrators can configure and maintain the mainframe environment the same way as they do for a non-mainframe environment. HiperSockets Layer 2 support is provided by Linux on System z, and by z/VM for guest usage.

HiperSockets Multiple Write Facility

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

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

zIIP-Assisted HiperSockets for large messages

In z/OS, HiperSockets are enhanced for zIIP usage. Specifically, the z/OS Communications Server allows the HiperSockets Multiple Write Facility processing for outbound large messages that originate from z/OS to be performed on a zIIP.

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

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

⁵ Not available on zBC12

zIIP Assisted HiperSockets for large messages is available on zEC12 and zBC12 with z/OS V1R10 (plus service) and later releases.

HiperSockets Network Traffic Analyzer (NTA)

HiperSockets NTA is a function available in the LIC of the zEC12 and zBC12. It can make problem isolation and resolution simpler by allowing Layer 2 and Layer 3 tracing of HiperSockets network traffic.

HiperSockets NTA allows Linux on System z to control tracing of the internal virtual LAN. It captures records into host memory and storage (file systems) that can be analyzed by system programmers and network administrators, using Linux on System z tools to format, edit, and process the trace records.

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

HiperSockets Completion Queue

The HiperSockets Completion Queue function allows both synchronous and asynchronous transfer of data between logical partitions. With the asynchronous support, during high volume situations, data can be temporarily held until the receiver has buffers available in its inbound queue. This provides end-to-end performance improvement for LPAR to LPAR communication and can be especially helpful in burst situations.

HiperSockets Completion Queue function is supported on the zEC12 and zBC12 running z/OS V1R13, z/VM V6R2 (with maintenance) and later, z/VSE V5R1 (with maintenance), Red Hat Enterprise Linux (RHEL) 6.2, or SUSE Linux Enterprise Server (SLES) 11 SP2.

HiperSockets integration with the intraensemble data network

The zEC12 and zBC12 servers provide the capability to integrate HiperSockets connectivity with the *intraensemble data network (IEDN)*. Thus the reach of the HiperSockets network is extended to outside the CPC to the entire ensemble, which is displayed as a single, Layer 2 network. Because HiperSockets and IEDN are both internal System z networks, the combination allows System z virtual servers to use an optimal path for communications.

The support of HiperSockets integration with the IEDN function is available starting with z/OS Communication Server V1R13 and with z/VM V6R2 with PTFs⁶.

HiperSockets Virtual Switch Bridge Support

The z/VM virtual switch is enhanced to transparently bridge a guest virtual machine network connection on a HiperSockets LAN segment. This bridge allows a single HiperSockets guest virtual machine network connection to also directly communicate with the following systems:

- ▶ Other guest virtual machines on the virtual switch
- ▶ External network hosts through the virtual switch OSA UPLINK port

z/VM V6R2 and later, TCP/IP, and Performance Toolkit APARs are required for this support.

A HiperSockets channel by itself is only capable of providing intra-CPC communications. The HiperSockets Bridge Port allows a virtual switch to connect z/VM guests by using real HiperSockets devices, the ability to communicate with hosts that reside externally to the CPC. The virtual switch HiperSockets Bridge Port eliminates the need to configure a separate next hop router on the HiperSockets channel to provide connectivity to destinations that are outside of a HiperSockets channel.

⁶ Not supported by z/VM V6R3.

z/VSE fast path to Linux support

Linux Fast Path (LFP) allows z/VSE TCP/IP applications to communicate with the TCP/IP stack on Linux without using a TCP/IP stack on z/VSE. LFP for use in a z/VM guest environment is supported on z/VSE V4R3 or higher. When LFP is used in an LPAR environment, it requires the HiperSockets Completion Queue function available on zEnterprise CPCs. LFP in an LPAR environment is supported on z/VSE V5R1.

3.2.8 Cryptography

zEC12 and zBC12 provide cryptographic functions that, from an application program perspective, can be grouped into the following functions:

- ▶ Synchronous cryptographic functions, provided by the CP Assist for Cryptographic Function (CPACF)
- ▶ Asynchronous cryptographic functions, provided by the Crypto Express features

CP Assist for Cryptographic Function (CPACF)

CPACF offers a set of symmetric cryptographic functions for high performance encryption and decryption with clear key operations for SSL/TLS, VPN, and data-storing applications that do not require FIPS⁷ 140-2 level 4 security. The CPACF is integrated with the compression unit in the coprocessor (CoP) in the System z microprocessor core.

The CPACF protected key is a function that is designed to facilitate the continued privacy of cryptographic key material while keeping the wanted high performance. CPACF ensures that key material is not visible to applications or operating systems during encryption operations. CPACF protected key provides substantial throughput improvements for large-volume data encryption and low latency for encryption of small blocks of data.

The cryptographic assist includes support for the following functions:

- ▶ Data Encryption Standard (DES) data encrypting and decrypting:
 - DES supports the following key types:
 - Single-length key DES
 - Double-length key DES
 - Triple-length key DES (T-DES)
- ▶ Advanced Encryption Standard (AES) for 128-bit, 192-bit, and 256-bit keys
- ▶ Pseudo random number generation (PRNG)
- ▶ Message Authentication Code (MAC)
- ▶ Hashing algorithms: SHA-1 and SHA-2 support for SHA-224, SHA-256, SHA-384, and SHA-512

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

Crypto Express4S

The *Crypto Express4S* represents the newest generation of the *Peripheral Component Interconnect Express (PCIe)* cryptographic coprocessors. It is an optional feature exclusive to the zEC12 or zBC12. This feature provides a secure programming and hardware environment wherein crypto processes are performed. Each cryptographic coprocessor includes a general-purpose processor, non-volatile storage, and specialized cryptographic electronics.

⁷ Federal Information Processing Standards (FIPS) 140-2 Security Requirements for Cryptographic Modules

The Crypto Express4S has one PCIe adapter per feature. For availability reasons, a minimum of 2 features is required. Up to 16 Crypto Express4S features are supported (16 PCI Express adapters per zEC12 or zBC12). The Crypto Express4S feature occupies one I/O slot in a zEC12 or zBC12 PCIe I/O drawer.

Each adapter can be configured as a Secure IBM CCA coprocessor, a Secure IBM Enterprise PKCS #11 (EP11) coprocessor or as an accelerator.

The accelerator function is designed for maximum-speed Secure Sockets Layer/Transport Layer Security (SSL/TLS) acceleration, rather than for specialized financial applications for secure, long-term storage of keys or secrets. The Crypto Express4S can also be configured as one of the following configurations:

- ▶ *Secure IBM CCA coprocessor* for Federal Information Processing Standard (FIPS) 140-2 Level 4 certification. This standard includes secure key functions and is optionally programmable to deploy more functions and algorithms using User Defined Extension (UDX).
- ▶ *Secure IBM Enterprise PKCS #11 (EP11) coprocessor*, implementing an industry standardized set of services that adheres to the PKCS #11 specification v2.20 and more recent amendments. It was designed for extended FIPS and Common Criteria evaluations to meet industry requirements.

This new cryptographic coprocessor mode introduced the PKCS #11 secure key function.

TKE feature: The Trusted Key Entry (TKE) Workstation feature is required for supporting the administration of the Crypto Express4S when configured as an Enterprise PKCS #11 coprocessor.

When the Crypto Express4S PCI Express adapter is configured as a secure IBM CCA coprocessor, it still provides accelerator functions. However, up to three times better performance for those functions can be achieved if the Crypto Express4S PCI Express adapter is configured as an accelerator.

Crypto Express3

The *Crypto Express3* features are available on zEC12 or zBC12 only when carried forward on an upgrade from earlier System z systems. The Crypto Express3 feature has two PCI Express adapters⁸ and each feature occupies one I/O slot in an I/O cage or in an I/O drawer.

Each adapter can be configured as a coprocessor or as an accelerator:

- ▶ *Secure coprocessor* for Federal Information Processing Standard (FIPS) 140-2 Level 4 certification. This standard includes secure key functions and is optionally programmable to deploy more functions and algorithms using User Defined Extension (UDX).
- ▶ *Accelerator* for public key and private key cryptographic operations that are used with Secure Sockets Layer/Transport Layer Security (SSL/TLS) processing.

When the Crypto Express3 PCI Express adapter is configured as a secure coprocessor, it still provides accelerator functions. But up to three times better performance for those functions can be achieved if the Crypto Express3 PCI Express adapter is configured as an accelerator.

Statement of Direction: The zEC12 and zBC12 are planned to be the last IBM System z servers to support Crypto Express3 feature. Enterprises should begin migrating from the Crypto Express3 feature to the Crypto Express4S feature.

⁸ The systems management Crypto Express3-1P feature has one PCIe adapter and just one PCHID associated to it.

Common Cryptographic Architecture (CCA) enhancements

The following functions were added to Crypto Express4S cryptographic feature and to the Crypto Express3 cryptographic feature when running on zEC12 or on zBC12, through the ICSF web deliverable FMID HCR77A0:

- ▶ Secure Cipher Text Translate
- ▶ Improved wrapping key strength for security and standards compliance
- ▶ DUKPT for derivation of Message Authentication Code (MAC) and encryption keys
- ▶ Compliance with new Random Number Generator standards
- ▶ EMV enhancements for applications which support American Express cards

Through the ICSF web deliverable FMID HCR77A1 the following enhancements were introduced:

- ▶ Diversified Key Generation TDES CBC support (EMV enhancement)
- ▶ DUKPT Initial PIN Encrypting Key (IPEK) Derivation
- ▶ Export/Import TDES key under AES transport key
- ▶ ATM Remote Key Export (RKX) key wrapping support
- ▶ Selected User Defined Extensions (UDX) functions integrated to the base CCA

Web deliverables

For z/OS downloads, see the z/OS website:

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

3.3 Capacity and performance

The zEC12 and the zBC12 offer significant increases in capacity and performance over their respective predecessors, the z196 and z114. Many factors contribute to this effect, including the larger number of processors, individual processor performance, memory caches, and machine instructions. Subcapacity settings continue to be offered.

3.3.1 Capacity settings

The zEC12 and the zBC12 expand the offer on subcapacity settings. Finer granularity in capacity levels allows the growth of installed capacity to more closely follow the enterprise growth, for a smoother, pay-as-you-go investment profile. There are many performance and monitoring tools that are available on System z environments that are coupled with the flexibility of the capacity on-demand options (see 3.7, “zEnterprise BladeCenter Extension Model 003” on page 108). These features help to manage growth by making capacity available when needed.

zEC12

Regardless of the installed model, the zEC12 offers four distinct capacity levels for the first 20 central processors (CP): full capacity and three subcapacities. These processors deliver the scalability and granularity to meet the needs of medium-sized enterprises, while also satisfying the requirements of large enterprises that have large-scale, mission-critical transaction and data-processing requirements.

A capacity level is a setting of each CP to a subcapacity of the full CP capacity. The clock frequency of those processors remains unchanged. The capacity adjustment is achieved through other means.

Full capacity CPs are identified as CP7. On the zEC12 server, 101 CPs can be configured as CP7. The three subcapacity levels are identified by CP6, CP5, and CP4, respectively, and are displayed in hardware descriptions as feature codes on the CPs.

If more than 20 CPs are configured to the system, then all must be full capacity because all CPs must be on the same capacity level. Granular capacity adds 60 subcapacity settings to the 101 capacity settings that are available with full capacity CPs (CP7). The 161 distinct capacity settings in the system, provide for a range of over 1:320 in processing power.

A processor that is characterized as anything other than a CP, such as a zAAP, a zIIP, an IFL, or an ICF, is always set at full capacity. There is, correspondingly, a separate pricing model for non-CP processors regarding purchase and maintenance prices, and various offerings for software licensing.

On zEC12, the CP subcapacity levels are a fraction of full capacity, as follows:

- ▶ Model 7xx = 100%
- ▶ Model 6xx = 63%
- ▶ Model 5xx = 42%
- ▶ Model 4xx = 16%

For administrative purposes, systems that have only ICF or IFL processors, are now given a capacity setting of 400. For either of these systems, it is possible to have up to 101 ICFs or IFLs, which always run at full capacity.

Figure 3-2 gives more details about zEC12 full capacity and subcapacity offerings.

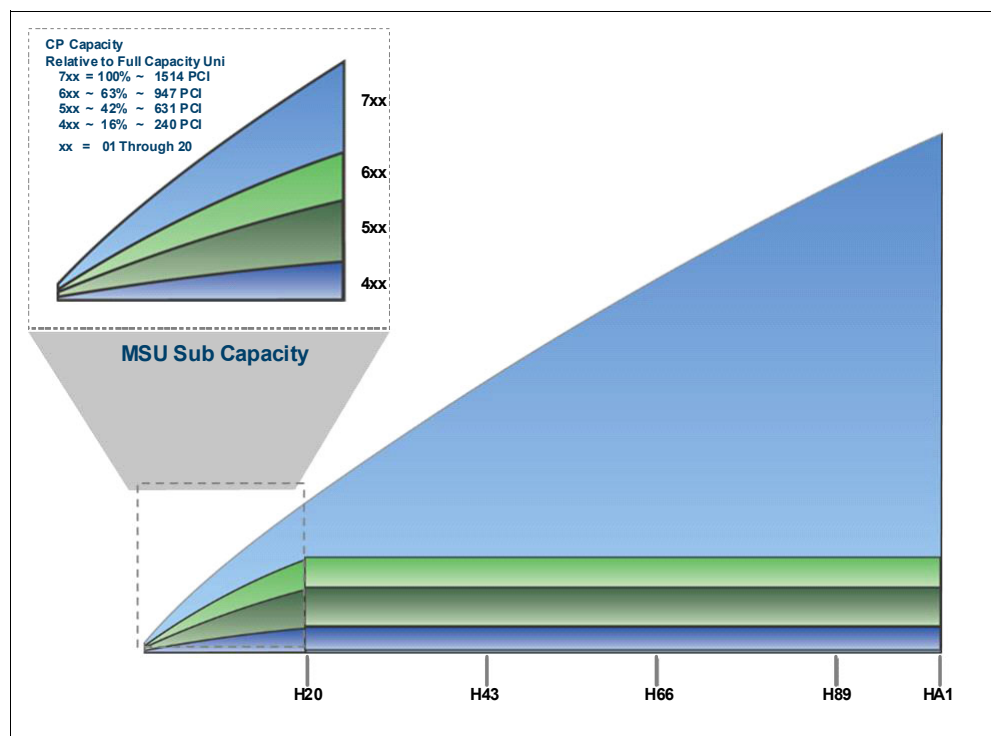


Figure 3-2 zEC12 full and subcapacity CP offerings

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

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

Also, see 3.3.7, “zEnterprise EC12 and BC12 performance” on page 100 for more information about LSPR measurements.

zBC12

The zBC12 offers twenty six (26) distinct capacity levels for each CP in the configuration. These processors deliver the scalability and granularity to meet the needs of small to medium-sized enterprises.

As in the zEC12, a zBC12 processor that is characterized as anything other than a CP, such as a zAAP, a zIIP, an IFL, or an ICF, is always set to full capacity.

Figure 3-3 gives more details about zBC12 capacities setting.

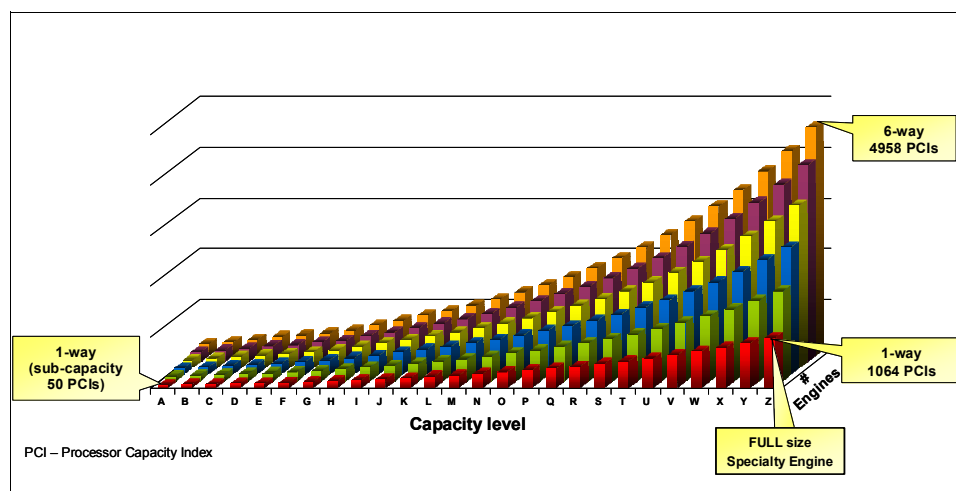


Figure 3-3 zBC12 capacity setting offerings

3.3.2 zEnterprise capacity on demand (CoD)

The zEC12 and zBC12 continue to offer on-demand offerings. The offerings provide flexibility and control to the client, ease the administrative burden in the handling of the offerings, and give the client finer control over resources that are needed to meet the resource requirements in various situations.

The zEC12 and zBC12 can perform concurrent upgrades, providing an increase of processor capacity with no server outage. In most cases, with operating system support, a concurrent upgrade can also be nondisruptive to the operating system. It is important to consider that these upgrades are based on the enablement of resources already physically present in the zEC12 or in the zBC12.

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

For more information about the CoD offerings, see the *IBM zEnterprise EC12 Technical Guide*, SG24-8049, or *IBM zEnterprise BC12 Technical Guide*, SG24-8138.

3.3.3 Permanent upgrades

Permanent upgrades of processors (CP, IFL, ICF, zAAP, zIIP, and SAP) and memory, or changes to a server's Model-Capacity Identifier, up to the limits of the installed processor capacity on an existing zEC12 or zBC12, can be performed by the client through the IBM Online Permanent Upgrade offering by using the CIU facility.

3.3.4 Temporary upgrades

Temporary upgrades of a zEC12 and zBC12 can be done by On/Off CoD, Capacity Backup (CBU), or Capacity for Planned Event (CPE) ordered from the CIU facility.

On/Off CoD function

On/Off CoD is a function that is available on the zEC12 and on the zBC12 that enables concurrent and temporary capacity growth of the CPC. On/Off CoD can be used for client peak workload requirements, for any length of time, has a daily hardware charge and can have an associated software charge. On/Off CoD offerings can be pre-paid or post-paid. Capacity tokens are available on zEC12 and on zBC12. Capacity tokens are always present in pre-paid offerings and can be present in post-paid if the client so desires. In both cases capacity tokens are being used to control the maximum resource and financial consumption.

When using the On/Off CoD function, the client can concurrently add processors (CP, IFL, ICF, zAAP, zIIP, and SAP), increase the CP capacity level, or both.

Capacity Backup (CBU) function

CBU allows the client to perform a concurrent and temporary activation of additional CP, ICF, IFL, zAAP, zIIP, and SAP, an increase of the CP capacity level, or both. This function can be used in the event of an unforeseen loss of System z capacity within the client's enterprise, or to perform a test of the client's disaster recovery procedures. The capacity of a CBU upgrade cannot be used for peak workload management.

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

Proper use of the CBU capability does not incur any additional software charges from IBM.

Capacity for Planned Event (CPE) function

CPE allows the client 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. This function can be used in the event of a planned outage of System z capacity within the client's enterprise (for example, data center changes or system maintenance). CPE cannot be used for peak workload management and can be active for a maximum of three days.

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

3.3.5 z/OS capacity provisioning

Capacity provisioning helps clients manage the CP, zAAP, and zIIP capacity of zEC12 or zBC12 that are running one or more instances of the z/OS operating system. Using z/OS Capacity Provisioning Manager (CPM) component, On/Off CoD temporary capacity can be activated and deactivated under control of a defined policy. Combined with functions in z/OS, the zEC12 or zBC12 provisioning capability gives the client a flexible, automated process to control the configuration and activation of On/Off CoD offerings.

3.3.6 Throughput optimization with zEC12 and zBC12

The z990 was the first server to use the concept of books. The memory and cache structure implementation in the books have been enhanced from the z990, through successive system generations to the zEC12 to provide sustained throughput and performance improvements. Despite the fact that the memory is distributed throughout the books and the books having individual levels of caches private to the cores and shared by the cores, all processors have access to the highest level of caches and all of the memory. Thus, the system is managed as a memory coherent symmetric multiprocessor (SMP).

The zBC12 uses the concept of processor drawers first introduced with the z10 BC. Despite the fact that the memory is distributed throughout the drawers and the drawers having individual levels of caches private to the cores and shared by the cores, all processors have access to the highest level of caches and all of the memory. Like the zEC12, the system is managed as a memory coherent symmetric multiprocessor (SMP).

Processors within the zEC12 book structure, or zBC12 drawer structure, have different distance-to-memory attributes. As described in 2.3, “zEC12 processor cage, books, and multiple chip modules” on page 34, books are connected in a star configuration to minimize the distance. Other non-negligible effects result from data latency when grouping and dispatching work on a set of available logical processors. To minimize latency, the system attempts to dispatch and later redispach work to a group of physical CPUs that share the same cache levels.

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

HiperDispatch

HiperDispatch, introduced with the z10 Enterprise Class, is enhanced in z114, z196, zBC12, and zEC12. It combines two functional enhancements, one in the z/OS dispatcher and one in PR/SM. This function is intended to improve efficiency both in the hardware and in z/OS. VM HiperDispatch is introduced by z/VM V6R3.

In general, the PR/SM dispatcher assigns work to the minimum number of logical processors that are needed for the priority (weight) of the LPAR. PR/SM attempts to group the logical processors into the same book on a zEC12 and, if possible, in the same chip. On the zBC12 PR/SM attempts to group the logical processors in the same drawer and, if possible, in the same chip. The result is to reduce the multi-processor effects, maximize use of shared cache, and lower the interference across multiple partitions.

The z/OS dispatcher is enhanced to operate with multiple dispatching queues, and tasks are distributed among these queues. Specific z/OS tasks can be dispatched to a small subset of logical processors. PR/SM ties these logical processors to the same physical processors, thus improving the hardware cache reuse and locality of reference characteristics, such as reducing the rate of cross communication.

To use the correct logical processors, the z/OS dispatcher obtains the necessary information from PR/SM through interfaces that are implemented on the zEC12 and on the zBC12. The entire zEC12 or zBC12 stack (hardware, firmware, and software) now tightly collaborates to obtain the full potential of the hardware. VM HiperDispatch provides support similar to the z/OS one.

The HiperDispatch function is enhanced on the zEC12 and zBC12 to use the new hex-core chip and improve computing efficiency. It is possible to dynamically switch on and off HiperDispatch without requiring an IPL.

3.3.7 zEnterprise EC12 and BC12 performance

The System z microprocessor chip of the zEC12 and zBC12 has a high-frequency design that uses IBM leadership technology and offers more cache per core than other chips. In addition, an enhanced instruction execution sequence delivers world-class per-thread performance. z/Architecture is enhanced which provides more instructions that are intended to deliver improved CPU-centric performance. For CPU-intensive workloads, more gains can be achieved by multiple compiler-level improvements. Improved performance of the zEC12 and zBC12 are a result of the enhancements that we described in Chapter 2, “Hardware overview” on page 25 and 3.2, “zEC12 and zBC12 technology improvements” on page 75.

The zEC12 Model HA1 is designed to offer approximately 50% more capacity than the largest z196 system. Uniprocessor performance has also increased significantly. A zEC12 Model 701 offers, based on an average workload, performance improvements of about 25% over the z196 Model 701.

The zBC12 Model H13 with 6 CPs characterized offer approximately 58% more capacity than the z114 Model M10 system with 5 CPs characterized. Uniprocessor performance has also increased significantly. A zBC12 Model Z01 offers, based on an average workload, performance improvements of about 36% over the z114 Model Z01.

However, variations on the observed performance increase are dependent upon the workload type.

To help in better understanding workload variations, IBM provides a no-cost tool that is called the *IBM Processor Capacity Reference for System z (zPCR)*⁹.

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

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

The MSU ratings are available at the following website:

<http://www-03.ibm.com/systems/z/resources/swprice/reference/exhibits/hardware.html>

⁹ <http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS1381>

LSPR workload suite: zEC12 and zBC12 changes

Historically, LSPR capacity tables, including pure workloads and mixes, have been identified with application names or a software characteristic. Examples are CICS, IMS, OLTP-T¹⁰, CB-L¹¹, LoIO-mix¹², and TI-mix¹³. However, capacity performance is more closely associated with how a workload uses and interacts with a particular processor *hardware design*.

Workload capacity performance is sensitive to three major factors:

- ▶ Instruction path length
- ▶ Instruction complexity
- ▶ Memory hierarchy

With the availability of the CPU measurement facility (MF) data, the ability to gain insight into the interaction of workload and *hardware design* in production workloads has arrived. CPU MF data helps LSPR to adjust workload capacity curves that are based on the underlying hardware sensitivities, in particular the processor access to caches and memory. This is known as *nest activity intensity*. With the IBM zEnterprise System, the LSPR introduced three new workload capacity categories that replace all prior primitives and mixes:

- ▶ LOW (relative nest intensity):

A workload category that represents light use of the memory hierarchy. This category is similar to past high scaling primitives.

- ▶ AVERAGE (relative nest intensity):

A workload category that represents average use of the memory hierarchy. This category is similar to the past LoIO-mix workload and is expected to represent most of the production workloads.

- ▶ HIGH (relative nest intensity):

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

These categories are based on the relative nest intensity, which is influenced by many variables such as application type, I/O rate, application mix, CPU usage, data reference patterns, LPAR configuration, and the software configuration that is running, among others. CPU MF data can be collected by z/OS System Measurement Facility on SMF 113 records.

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

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

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

¹⁰ Traditional online transaction processing workload (formerly known as IMS)

¹¹ Commercial batch with long-running jobs

¹² Low I/O Content Mix Workload

¹³ Transaction Intensive Mix Workload

The LSPR contains the internal throughput rate ratios (ITRR) for the zEC12 and the previous generations of processors that are based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user might experience varies depending on factors such as the amount of multiprogramming in the user's job stream, the I/O configuration, and the workload processed.

Experience demonstrates that System z servers can be run at up to 100% utilization levels, sustained, although most clients prefer to leave a bit of white space and run at 90% or slightly under. For any capacity comparison, using “one number” such as the MIPS or MSU metrics is not a valid method. That is why, while doing capacity planning, we suggest using zPCR and involving IBM technical support. For more information about zEC12 performance, see the *IBM zEnterprise EC12 Technical Guide*, SG24-8049; and for zBC12, see the *IBM zEnterprise BC12 Technical Guide*, SG24-8138.

3.4 zEnterprise common time functions

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

The zEC12 and the zBC12 support the Server Time Protocol and can participate in a coordinated timing network.

3.4.1 Server Time Protocol (STP)

Server Time Protocol (STP) is a message-based protocol in which timekeeping information is passed over data links between servers. The timekeeping information is transmitted over externally defined coupling links. The STP feature is the supported method for maintaining time synchronization between the zEC12, zBC12, and coupling facilities (CF) in Sysplex environments.

The STP design uses a concept called *Coordinated Timing Network (CTN)*. A CTN is a collection of CPCs that are time synchronized to a time value called *Coordinated Server Time (CST)*. Each CPC to be configured in a CTN must be STP-enabled. STP is intended for CPCs that are configured to participate in a Parallel Sysplex or CPCs that are not in a Parallel Sysplex, but must be time synchronized.

STP is implemented in LIC as a system-wide facility of zEC12, zBC12 and other System z servers. STP presents a single view of time to PR/SM and provides the capability for multiple CPCs to maintain time synchronization with each other. zEC12 and zBC12 servers are enabled for STP by installing the STP feature code. Additional configuration is required for a zEC12 or a zBC12 to become a member of a CTN.

Statement of Direction: IBM zEnterprise EC12 and zBC12 are the last System z servers to support connections to an STP Mixed CTN. This includes the IBM Sysplex Timer® (9037). After zEC12 and zBC12, servers that require time synchronization, such as to support a base or Parallel Sysplex, will require Server Time Protocol (STP), and all servers in that network must be configured in STP-only mode.

STP provides the following additional value over the former used-time synchronization method by a 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 protocol allows a Parallel Sysplex to span these distances and reduces the cross-site connectivity that is required for a multi-site Parallel Sysplex.
- ▶ The STP design allows more stringent synchronization between CPCs and CFs by using communication links that are already used for the sysplex connectivity. With the zEC12 and zBC12, STP supports coupling links over InfiniBand or ISC-3¹⁴ links.
- ▶ STP helps eliminate infrastructure requirements, such as power and space, needed to support the Sysplex Timers and helps eliminate maintenance costs that are associated with the Sysplex Timers.
- ▶ STP can reduce the fiber optic infrastructure requirements in a multi-site configuration since it can use the already in use coupling links.

Timing: Concurrent migration from an existing External Time Reference (ETR) network to a timing network using STP is supported only if a z10 EC or z10 BC is used for the Stratum 1 server. System z systems that precede the z10 cannot participate in the same CTN with zEC12 or zBC12.

Server Time Protocol (STP) recovery enhancement

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

The previously available STP recovery design is still available for the cases when a going away signal is not received or for other failures different from a system failure.

3.4.2 Network Time Protocol (NTP) client support

The use of Network Time Protocol (NTP) servers as an external time source (ETS) usually fulfills a requirement for a time source or common time reference across heterogeneous platforms and for providing a higher time accuracy.

NTP client support is available in the Support Element (SE) code of the zEC12 and zBC12. The code interfaces with the NTP servers. This interaction allows an NTP server to become the single time source for zEC12, zBC12 and for other servers that have NTP clients. NTP can be used only for an STP-only CTN environment.

ETS access: ETS access through *Modem is not supported* on the zEC12 or zBC12 HMC.

Pulse per second (PPS) support

Two External Clock Facility (ECF) cards are shipped as a standard feature of the zEC12 and the zBC12 CPCs and provide a dual-path interface for the PPS signal. The redundant design allows continuous operation, in case of failure of one card, and concurrent card maintenance.

¹⁴ ISC-3 features are only available on zEC12 and zBC12 when carried forward during an upgrade.

The zEC12 and zBC12 provide a dual-path interface for Pulse Per Second (PPS) support. STP tracks the highly stable accurate PPS signal from the NTP server. PPS maintains accuracy of 10 μ s as measured at the PPS input of the zEC12 or zBC12 CPC. If STP uses an NTP server without PPS, a time accuracy of 100 ms to the ETS is maintained. A cable connection from the PPS port to the PPS output of an NTP server is required when the zEC12 or zBC12 is configured for using NTP with PPS as the external time source for time synchronization.

NTP server on HMC with security enhancements

The NTP server capability on the HMC addresses the potential security concerns that users can have for attaching NTP servers directly to the HMC/SE LAN. When using the HMC as the NTP server, the pulse per second capability is not available.

HMC NTP broadband authentication support for zEC12 and zBC12

The HMC NTP authentication capability is provided by the HMC Level 2.12.0. SE NTP support stays unchanged. To use this option for STP, configure the HMC as the NTP server for the SE.

The authentication support of the HMC NTP server can be set up in either of two ways:

- ▶ NTP requests are UDP socket packets and cannot pass through the proxy. If a proxy to access outside corporate data center is used, then this proxy must be configured as an NTP server to get to target servers on the web. Authentication can be set up on the client's proxy to communicate to the target time sources.
- ▶ If a firewall is used, HMC NTP requests must pass through the firewall. Clients in this configuration should use the HMC authentication to ensure untampered time stamps.

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

3.5 Hardware Management Console functions

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

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

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

Order these Unified Resource Manager features to equip an HMC to manage an ensemble:

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

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

HMC/SE Version 2.12.1 is the current version available for the zEC12 and zBC12. See *Building an Ensemble Using IBM zEnterprise Unified Resource Manager*, SG24-7921, for more information about these HMC functions and capabilities.

3.5.1 Hardware Management Console key enhancements for zEC12 and zBC12

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

- ▶ Tasks and panels are updated to support configuring and managing the zEC12 and zBC12 introduced Flash Express, IBM zAware, zECD Express and 10GbE RoCE Express features.
- ▶ OSA/SF is now available on the HMC for specific OSA-Express features.
- ▶ For STP NTP broadband security, authentication is added to the HMC's NTP communication with NTP time servers.
- ▶ Modem support is removed from HMC. The Remote Support Facility (RSF) for IBM support, service, and configuration update is only possible through an Ethernet broadband connection.
- ▶ The Monitors Dashboard on the HMC and SE is enhanced with an adapter table for zEC12 and zBC12. The Crypto Utilization percentage is displayed on the Monitors Dashboard according to the PCHID number. The associated Crypto number (Adjunct Processor Number) for this PCHID is also shown in the table. It provides information about utilization rate on a system-wide basis. The adapter table also displays Flash Express.
- ▶ The Environmental Efficiency Statistic Task provides historical power consumption and thermal information for zEC12 and zBC12 on the HMC. This task provides similar data along with a historical summary of processor and channel use. The initial chart display shows the 24 hours that precede the current time so that a full 24 hours of recent data is displayed. The data is presented in table form, graphical (histogram) form, and it can also be exported to a .csv formatted file so that it can be imported into a spreadsheet.
- ▶ The microcode update to a specific bundle is now possible.

For more information about the key capabilities and enhancements of the HMC, refer to *IBM zEnterprise EC12 Technical Guide*, SG24-8049 and *IBM zEnterprise BC12 Technical Guide*, SG24-8138.

3.5.2 Considerations for multiple Hardware Management Consoles

Often, multiple HMC instances are deployed to manage an overlapping collection of systems. Before the announcement of zEnterprise System, all HMCs were peer consoles to the managed systems. Furthermore, all management actions are possible to any of the CPCs that are reachable while logged in to a session on any of the HMCs (subject to access control).

With the definition of an ensemble, this resource management paradigm changes. Management actions that target a node of an ensemble can be done only from the primary HMC for that ensemble.

3.6 zEnterprise CPC power and cooling functions

As environmental concerns raise the focus on energy consumption, zEnterprise CPCs offer a holistic focus on the environment. New efficiencies and functions¹⁵, such as an improved integrated cooling system, static power save mode, and cycle steering (as radiator and water-cooled systems are being backed-up), enable a dramatic reduction of energy usage and floor space when consolidating workloads from distributed servers.

3.6.1 High voltage DC power

In today's data centers, many businesses are paying increasing electric bills and are also running out of power. The High Voltage Direct Current (HV DC) power feature adds nominal 380 - 520 Volt DC input power capability to the existing 1 phase¹⁶, 3 phase, 50/60 hertz, totally redundant power capability (nominal 200–240VAC or 380–415VAC or 480VAC).

This feature allows CPCs to directly use the high voltage DC distribution in new, green data centers. A direct HV DC data center power design can improve data center energy efficiency by removing the need for a DC-to-AC inversion step. The zEnterprise CPCs bulk power supplies have been modified to support HV DC, so the only difference in shipped HW to implement the option is the DC power cords.

Because HV DC is a new technology, there are multiple proposed standards. The zEnterprise supports both ground-referenced and dual-polarity HV DC supplies, such as +/-190V or +/-260V, or +380V. Beyond the data center uninterruptible power supply (UPS) and power distribution energy savings, a zEnterprise CPC running on HV DC power draws 1 - 3% less input power. HV DC does not change the number of power cords that a system requires.

3.6.2 Integrated battery feature (IBF)

IBF is an optional feature on the zEnterprise CPCs. See Figure 2-3 on page 32 or Figure 2-5 on page 34 for a pictorial view of the location of IBF. 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. The feature provides battery power to preserve processor data if there is a total loss of power from the utility company. The IBF can hold power briefly during a brownout, or for orderly shutdown in a longer outage.

3.6.3 Power capping and power saving

zEnterprise CPCs support power capping, which gives the ability to control the maximum power consumption and reduce cooling requirements (especially with zBX). To use power capping, the Automate Firmware Suite feature must be ordered. This feature is used to enable the Automate suite of functionality that is associated with the Unified Resource Manager. The Automate suite includes representation of resources in a workload context, goal-oriented monitoring and management of resources, and energy management. A static power-saving mode is also available for the zEC12 when the Automate Firmware Suite feature is installed. It uses frequency and voltage reduction to reduce energy consumption and can be set up ad hoc or as a scheduled operation. It means, for example, in periods of low utilization or on CBU systems, that clients can set the system in a static power-saving mode. Power Saving functions are also provided for the blades in the zBX.

¹⁵ Some new efficiencies and functions are only available on certain CPCs, where applicable this will be described.

¹⁶ Supported only by select configurations of the zBC12.

3.6.4 Power estimation tool

The power estimation tool for zEnterprise CPCs is a web based tool available to registered users of IBM Resource Link. The tool allows entering the exact server configuration to produce an estimate of power consumption.

Log in to IBM Resource Link and go to **Planning → Tools → Power Estimation Tools**. Specify the quantity for the features that are installed in the machine. The tool estimates the power consumption for the specified configuration. The tool does not verify whether the specified configuration can be physically built.

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

3.6.5 IBM Systems Director Active Energy Manager

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

AEM runs on Windows, Linux on System x, AIX, Linux on IBM System p®, and Linux on System z. For more information, see its documentation¹⁷.

How AEM works

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

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

¹⁷ <http://www.ibm.com/systems/software/director/resources.html>

3.6.6 Top Exit Power

The zEC12 and zBC12 support the optional *Top Exit Power* feature. This feature enables installing a radiator (air) cooled zEC12 or a zBC12 on a non-raised floor, when the optional top exit I/O cabling feature is also installed. Water-cooled zEC12 models cannot be installed on a non-raised floor as top exit support for water cooling systems is not available. On a raised floor, either radiator or water cooling is supported.

3.7 zEnterprise BladeCenter Extension Model 003

zEC12 introduces the zEnterprise BladeCenter Extension (zBX) Model 003, also available with the zBC12.

The zEnterprise BladeCenter Extension (zBX) Model 003 continues to support workload optimization and integration. As an optional feature that is attached to the zEC12 or a zBC12 by a secure high-performance private network, the zBX can house multiple environments that include AIX, Linux on System x, and Windows, supporting a “fit for purpose” application deployment.

The zBX is tested and packaged together at the IBM manufacturing site and shipped as one unit, relieving complex configuration and set up requirements. With a focus on availability, the zBX has hardware redundancy that is built in at various levels: the power infrastructure, rack-mounted network switches, power and switch units in the BladeCenter chassis, and redundant cabling for support and data connections. The zEnterprise BladeCenter Extension (zBX) Model 003 components are configured, managed, and serviced the same way as the other components of a System z server.

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

GDPS/PPRC and GDPC/GM support zBX hardware components, providing workload failover for automated multi-site recovery. These capabilities can help facilitate the management of planned and unplanned outages across IBM zEnterprise EC12 or zEnterprise BC12.

Statement of Direction: IBM intends to deliver new functionality with IBM Systems Director offerings to support the IBM zBX. Such planned new capabilities will be designed to provide virtual image management and enhanced energy management functions for IBM Power Systems™ and System x blades.

3.7.1 IBM blades

IBM offers select IBM BladeCenter PS701 Express blades that can be installed and operated on the zBX Model 003. These blades are virtualized by PowerVM Enterprise Edition. The virtual servers in PowerVM run the AIX operating system.

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

A select set of IBM BladeCenter HX5 (7873) blades is available for the zBX. These blades have an integrated hypervisor, and their virtual machines run Linux on System x and Microsoft Windows.

Also available on the zBX is the IBM WebSphere DataPower XI50 for zEnterprise appliance. The DataPower XI50z is a multifunctional appliance that can help provide multiple levels of XML optimization, streamline and secure valuable service-oriented architecture (SOA) applications, and provide drop-in integration for heterogeneous environments by enabling core *enterprise service bus (ESB)* functionality. These functions include: routing, bridging, transformation, and event handling. It can help to simplify, govern, and enhance the network security for XML and web services.

Software that is supported on blades is described in more detail in 4.3, “Software support for zBX Model 003” on page 139.

3.8 Reliability, availability, and serviceability (RAS)

The IBM zEnterprise System family presents numerous enhancements in the *RAS* areas. In the availability area, focus was given to reduce the planning requirements, while continuing to improve the elimination of planned, scheduled, and unscheduled outages.

Enhanced driver maintenance (EDM) helps reduce the necessity and the eventual duration of a scheduled outage. One of the contributors to scheduled outages is LIC Driver updates that are performed in support of new features and functions. When properly configured, the zEC12 or the zBC12 can concurrently activate a new LIC Driver level. Concurrent activation of the select new LIC Driver level is supported at specifically released synchronization points. However, there are certain LIC updates where a concurrent update or upgrade is not possible.

On the zEC12 only, with enhanced book availability, the affect of book replacement is minimized. In a multiple book system, a single book can be concurrently removed and reinstalled for an upgrade or repair. To ensure that the zEC12 configuration supports removal of a book with minimal affect to the workload, consider the flexible memory option.

The zEC12 and zBC12 provide a way to increase memory availability, called Redundant Array of Independent Memory (RAIM), where a fully redundant memory system can identify and correct memory errors without stopping. The implementation is similar to the RAID concept used in storage systems for a number of years. See *IBM zEnterprise EC12 Technical Guide*, SG24-8049, or *IBM zEnterprise BC12 Technical Guide*, SG24-8138 for a detailed description of the RAS features.

To prevent both scheduled and unscheduled outages, there are several availability improvements in different components of zEC12 and zBC12. These enhancements include error detection and recovery improvements in both caches and memory, IBM zAware, Flash Express, Fibre Channel Protocol support for T10-DIF, a fixed HSA with its size doubled to 32 GB on the zEC12 and 16 GB on the zBC12, OSA firmware changes to increase the capability of concurrent maintenance change level (MCL) updates, radiator cooling system with N+1 redundancy, corrosion sensors, new CFCC level, RMF reporting, zBX connectivity.

zEC12 and zBC12 continue to support concurrent addition of resources, such as processors or I/O cards, to an LPAR to achieve better serviceability. If an additional system assist processor (SAP) is required on a zEC12 or zBC12 (for example, as a result of a disaster recovery situation), the SAPs can be concurrently added to the CPC configuration.

It is possible to concurrently add CP, zAAP, zIIP, IFL, and ICF processors to an LPAR. This function is supported by z/VM V5R4 and later, and also (with appropriate PTFs) by z/OS and z/VSE V4R3 and later. Previously, proper planning was required to add CP, zAAP, and zIIP to a z/OS LPAR concurrently. It is possible to concurrently add memory to an LPAR. This ability is supported by z/OS and z/VM.

zEC12 and zBC12 support dynamically adding Crypto Express features to an LPAR by being able to change the cryptographic information in the image profiles without outage to the LPAR. Users can also dynamically delete or move Crypto Express features. This enhancement is supported by z/OS, z/VM, and Linux on System z.

3.8.1 IBM System z Advanced Workload Analysis Reporter (IBM zAware)

Introduced with the zEC12 and also available with the zBC12, the *IBM zAware* feature is an integrated expert solution that uses sophisticated analytics to help clients identify potential problems and improve overall service levels.

IBM zAware runs analytics in firmware and intelligently examines z/OS message logs for potential deviations, or inconsistencies, or variations from the norm, providing out-of-band monitoring and machine learning of operating system health.

IBM zAware can accurately identify system anomalies in minutes. This feature analyzes massive amounts of processor data to identify problematic messages and provides information that can feed other processes or tools. The IBM zAware virtual appliance monitors the z/OS operations log (OPERLOG), which contains all messages that are written to the z/OS console, including application-generated messages. IBM zAware provides a graphical user interface (GUI) for easy drill-down into message anomalies, which can lead to faster problem resolution.

Statement of Direction fulfillment: You can get more from the zAware feature by integrating with Tivoli Service Management. Tivoli utilizes the zAware API to integrate log analysis with existing service management capabilities:

- ▶ Provide visibility into IBM zAware anomalies via Event Management.
- ▶ Improve mean time to repair (MTTR) through integration with existing problem determination and performance monitoring tools.

Identify system errors and eliminate subsequent occurrences through automation and more sophisticated analysis.

More detail about IBM zAware can be found in *IBM zEnterprise EC12 Technical Guide*, SG24-8049, *IBM zEnterprise BC12 Technical Guide*, SG24-8138, *IBM zAware Concept Guide*, SG24-8070, and *Advanced Workload Analysis Reporter (IBM zAware)*, SC27-2623.

3.8.2 RAS capability for the HMC

In an ensemble environment, the Unified Resource Manager routines are run in the HMC.

The Unified Resource Manager is an active part of the ensemble infrastructure. Thus, the HMC has a stateful environment that needs high-availability features to ensure survival of the system in case of an HMC failure.

Each ensemble requires two HMC workstations: a primary and a backup, called alternate. The contents and activities of the primary are kept synchronously updated on the alternate HMC so that the alternate can automatically take over the activities of the primary should the primary fail. Although the primary HMC can perform the classic HMC activities in addition to the Unified Resource Manager activities, the alternate HMC can only be a backup. No additional tasks or activities can be performed at the backup HMC.

3.8.3 RAS capability for zBX

The zBX was built following the traditional System z hardware QoS to include RAS capabilities. The zBX offering provides extended service capability through the hardware management structure. The HMC/SE functions of the zEC12 and the zBC12 provide management and control functions for the zBX solution.

Independently on the number of racks installed, the zBX is configured to provide N+1 components. The zBX components are replaced concurrently. In addition, zBX configuration upgrades can be performed concurrently.

Installed only on zBX's first rack are four Top of Rack (TOR) switches, two per each network (INMN and IEDN). These switches provide N + 1 connectivity for the private networks between the zEC12 or zBC12 and the zBX for monitoring, controlling, and managing the zBX components.

zBX firmware

The testing, delivery, installation, and management of the zBX firmware is handled the same way as for the zEC12 and zBC12. The same zEC12 and zBC12 processes and controls are used. Any fixes to the zBX machine are downloaded to the controlling zEC12's and zBC12's SEs and are applied to the zBX.

The MCLs for the zBX are concurrent and their status can be viewed at the zEC12's and zBC12's HMCs.

These and additional features are further described in *IBM zEnterprise EC12 Technical Guide*, SG24-8049 and *IBM zEnterprise BC12 Technical Guide*, SG24-8138.

3.9 High availability technology for zEnterprise

System z is renowned for its reliability, availability, and serviceability capabilities, of which Parallel Sysplex is an exponent. Extended availability technology with IBM PowerHA® for Power is available for blades in the zBX. First, we describe the System z Parallel Sysplex technology and then the PowerHA technology.

3.9.1 High availability for zEnterprise with Parallel Sysplex

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

A Parallel Sysplex has the following minimum components:

- Coupling Facility (CF):

This is the cluster center. It can be implemented either as an LPAR of a stand-alone System z system or as an additional LPAR of a System z system where other loads are running. Processor units that are characterized as either CPs or ICFs can be configured to this LPAR. ICFs are often used because they do not incur any software license charges. Two CFs are recommended for availability.

- Coupling Facility Control Code (CFCC):

This IBM Licensed Internal Code is both the operating system and the application that runs in the CF. No other code runs in the CF. The code is used to create and maintain the structures which are exploited under z/OS by software components such as z/OS itself, DB2 for z/OS, WebSphere MQ, amongst others.

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

- Coupling links:

These are high-speed links that connect the several system images (each running in its own logical partition) that participate in the Parallel Sysplex. At least two connections between each physical server and the CF must exist. When all of the system images belong to the same physical server, internal coupling links are used.

On the software side, the z/OS operating system uses the hardware components to create a Parallel Sysplex. One example of z/OS and CF collaboration is the System-managed CF structure duplexing, which provides a general-purpose, hardware-assisted, easy-to-exploit mechanism for duplexing structure data held in CFs. This function provides a robust recovery mechanism for failures (such as loss of a single structure on CF or loss of connectivity to a single CF). The recovery is done through rapid failover to the other structure instance of the duplex pair.

Clients that are interested in deploying system-managed CF structure duplexing can read the technical paper *System-Managed CF Structure Duplexing*, ZSW01975USEN, which can be accessed by selecting **Learn More** on the Parallel Sysplex website:

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

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

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

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

In short, a Parallel Sysplex comprises one or more z/OS operating system images that are coupled through one or more coupling facilities. A properly configured Parallel Sysplex cluster is designed to maximize availability *at the application level*. Rather than a quick recovery of a failure, the Parallel Sysplex design objective is *zero failure*.

The major characteristics of a Parallel Sysplex include the following features:

- ▶ 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 users of the data sharing capability are DB2, WebSphere MQ, WebSphere ESB, IMS, and CICS. Because these are major infrastructure components, applications that use them inherently benefit from sysplex characteristics. For instance, many large SAP implementations have the database component on DB2 for z/OS, in a Parallel Sysplex.

- ▶ Continuous (application) availability:

Changes, such as software upgrades and patches, can be introduced one image at a time, while the remaining images continue to process work. For more details, see *Parallel Sysplex Application Considerations*, SG24-6523.

- ▶ High capacity:

Parallel Sysplex scales from two to 32 images. Remember that each image can have from one to 100 processor units. CF scalability is near-linear. This structure contrasts with other forms of clustering that employ n-to-n messaging, which leads to rapidly degrading performance with a growing 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:

This architecture provides the infrastructure to satisfy a client 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 use CF shared storage. This usage enables the 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 is displayed as a single entity to the operator, user, database administrator, and so on. A single system image ensures reduced complexity from both operational and definition perspectives.

- ▶ N-2 support:

Multiple hardware generations (normally three, the current and the two previous ones) are supported in the same Parallel Sysplex. This configuration provides for a gradual evolution of the systems in the Sysplex, without forcing changing all simultaneously. Similarly, software support for multiple releases or versions is supported.

Figure 3-4 illustrates the components of a Parallel Sysplex as implemented within the System z architecture. The diagram shows one of many possible Parallel Sysplex configurations.

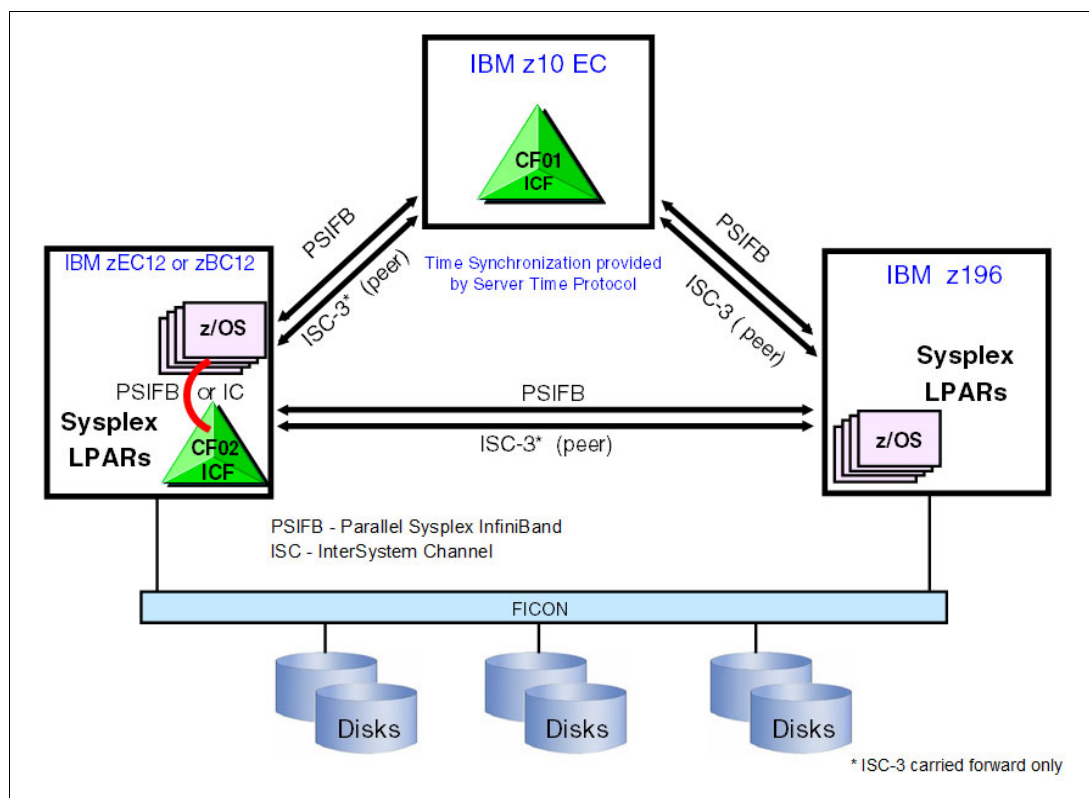


Figure 3-4 Sysplex hardware overview

Figure 3-4 shows a zEC12 or a zBC12 system that contains multiple z/OS sysplex partitions and an internal coupling facility (CF02), a z10 EC server containing a stand-alone CF (CF01), and a z196 containing multiple z/OS sysplex partitions. STP over coupling links provides time synchronization to all servers. Appropriate CF link technology (1x IFB or 12x IFB) selection depends on server configuration and how distant they are physically located. ISC-3 links can be carried forward to zEC12 or to zBC12 only when upgrading from either z196 or z10 EC or from a z114 or a z10 BC, respectively.

3.9.2 PowerHA in zBX environment

High availability for applications running on AIX is provided by the IBM PowerHA IBM SystemMirror® for AIX (formerly known as IBM HACMP™¹⁸). PowerHA is easy to configure (menu driven) and helps define and manage resources (required by applications) running on AIX by providing service/application continuity through platform resources and application monitoring, and automated actions (start/manage/monitor/restart/move/stop).

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

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

A PowerHA configuration or *cluster* consists of two or more servers¹⁹ (up to 32) that have their resources managed by PowerHA cluster services to provide automated service recovery for the applications managed. Servers can have physical or virtual I/O resources, or a combination of both.

PowerHA performs the following functions at the cluster level:

- ▶ Manage and monitor operating systems and hardware resources.
- ▶ Manage and monitor application processes.
- ▶ Manage and monitor network resources.
- ▶ Automate applications (start, stop, restart, move).

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

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

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

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

¹⁸ High Availability Cluster Multi-Processing

¹⁹ Servers can be also virtual servers. One server equals one instance of the AIX Operating System.

Figure 3-5 shows a typical PowerHA cluster.

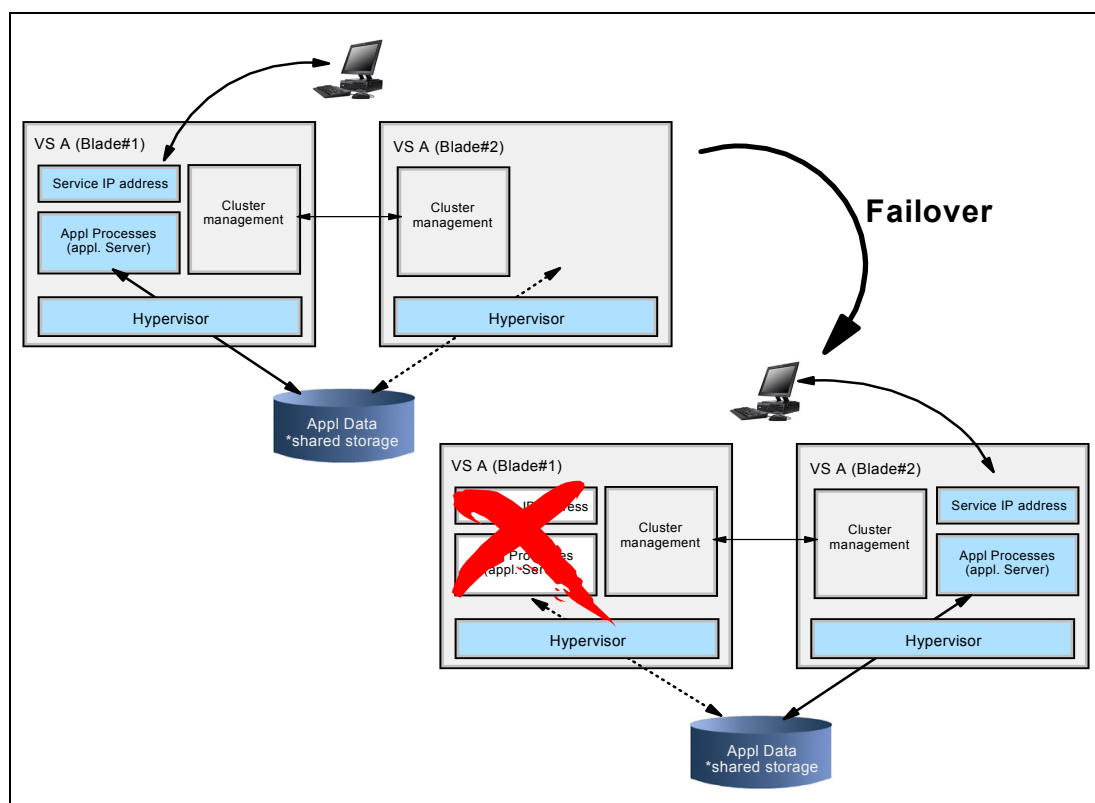


Figure 3-5 Typical PowerHA cluster diagram

For more information about IBM PowerHA SystemMirror for AIX, see the following web page:
http://pic.dhe.ibm.com/infocenter/aix/v6r1/index.jsp?topic=%2Fcom.ibm.aix.powerha.navigation%2Fpowerha_main.htm

3.10 zEnterprise and emerging paradigms

Having reviewed the most recent and important characteristics of zEnterprise System, we would like to conclude this discussion with some observations on the role zEC12 and zBC12 can play in today's leading IT initiatives.

We are witnessing a transformation of the interaction between users and systems, increasingly based on mobile devices, and instrumented devices ("the Internet of things"). This front office transformation requires highly responsive and dynamic transaction systems, and demands very high security. As discussed in "zEnterprise-based clouds" on page 75, and evidenced by the descriptions of zEC12 and zBC12, these systems can answer the infrastructure hardware requirements, whether I/O bandwidth, computing dynamic scalability, or security.

In addition, software requirements are covered as well. We note that several software licensing offerings are available on System z to cater for different environments and workloads. In particular, Linux on System z closely follows the distributed paradigm; see Appendix A, "Software licensing" on page 159.

Several transactional servers are available on the Linux on z and z/OS environments, which can be exploited by mobile applications such as those developed with the state-of-the-art IBM Worklight software. Those applications can benefit from the unmatched availability afforded by a Parallel Sysplex setup.

Data sharing in a Parallel Sysplex can also be exploited from Linux on System z. This includes specialized solutions such as the IBM DB2 Analytics Accelerator, which can transparently, that is without application modification, benefit from the radical acceleration of very complex queries such as those used by Business Intelligence and Data Analytics, enabling their insertion into on-line applications.

Finally, but no less important, the need for a coherent security landscape across the enterprise is increasingly being recognized. The zEC12 and zBC12 are, of themselves very secure and specialized offerings such as the Enterprise Key Management Foundation enable their security features, such as the Crypto Express and secure key management, to be exploited by the larger enterprise.

For further discussion on how to benefit from IBM zEnterprise System, see Chapter 5, “A smarter infrastructure” on page 145.



Operating system support and considerations

This chapter contains operating system requirements and support considerations for the IBM zEnterprise EC12, IBM zEnterprise BC12 and its features.

This chapter describes the following topics:

- ▶ Software support summary
- ▶ Support by operating system
- ▶ References
- ▶ z/OS considerations
- ▶ z/OS considerations
- ▶ Coupling Facility and CFCC considerations
- ▶ Input/output configuration program (IOCP) considerations
- ▶ IBM Device Support Facilities (ICKDSF) considerations

Support of the IBM zEnterprise EC12 and IBM zEnterprise BC12 functions depends on the operating system version and release. This information is subject to change. Therefore, for the most current information, see the *Preventive Service Planning (PSP)* bucket for 2827DEVICE and 2828DEVICE.

4.1 Software support summary

The software portfolio for the IBM zEnterprise EC12 (zEC12) and IBM zEnterprise BC12 (zBC12) includes a large variety of operating systems and middleware that support the most recent and significant technologies. Continuing the mainframe-rich tradition, five major operating systems are supported:

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

For software supported on the zBX Model 003, see 4.2.6, “References” on page 138.

Operating systems summary

Table 4-1 summarizes the current and minimum operating system levels that are required to support the zEC12 and zBC12. Operating system levels that are no longer in service are not covered in this publication. These older levels can provide support for certain features.

Table 4-1 zEC12 and zBC12 operating system requirements

Operating system	ESA/390 (31-bit mode)	z/Architecture (64-bit mode)	End of service	Notes
z/OS V2R1	No	Yes	September 2018 ^a	See the z/OS, z/VM, z/VSE, and z/TPF subsets of the 2827DEVICE and 2828DEVICE Preventive Service Planning (PSP) buckets before installing zEC12 and zBC12.
z/OS V1R13	No	Yes	September 2016 ^a	
z/OS V1R12	No	Yes	September 2014 ^a	
z/OS V1R11	No	Yes	September 2012 ^b	
z/VM V6R3 ^c	No ^d	Yes	September 2016 ^a	
z/VM V6R2 ^c	No ^d	Yes	April 2015 ^a	
z/VM V5R4	No ^d	Yes	December 2014 ^a	
z/VSE V5R1 ^e	No	Yes ^f	Not announced	
z/VSE V4R3	No ^g	Yes ^h	May 2014 ^a	
z/TPF V1R1	Yes	Yes	Not announced	
Linux on System z	No ⁱ	See Table 4-6 on page 135.	See footnote ^j	

a. Planned date. All statements regarding IBM plans, directions, and intent are subject to change or withdrawal without notice. Any reliance on these Statements of General Direction is at the relying party's sole risk and will not create liability or obligation for IBM.

b. With the announcement of IBM Lifecycle Extension for z/OS V1.11, fee-based corrective service can be ordered for up to two years after the withdrawal of service for z/OS V1R11.

c. z/VM V6R2 and V6R3 require an architectural level set exclusive to z10 and successors.

d. z/VM supports both ESA/390 mode and z/Architecture mode virtual machines.

e. z/VSE V5R1 requires an architectural level set exclusive to z9 and successors.

f. z/VSE V5R1 supports 64-bit virtual addressing both for applications and ISV products.

g. ESA/390 is not supported. However, 31-bit mode is supported.

h. z/VSE V4R3 support 64-bit real addressing only. It does not support 64-bit virtual addressing for user, system, or vendor applications.

i. 64-bit distributions included a 31-bit emulation layer to run 31-bit software products.

j. For information about support-availability of Linux on System z distributions, see:
SUSE:
<http://www.suse.com/support/>
Red Hat:
<http://www.redhat.com/security/updates/errata/>

PTFs and PSP buckets: The use of several features depends on a particular operating system. In all cases, program temporary fixes (PTF) might be necessary with the operating system level indicated.

Preventive Service Planning (PSP) buckets are continuously updated and are reviewed regularly when planning for installation of a new system. They contain the latest information about installation, hardware and software service levels, service recommendations, and cross-product dependencies.

For Linux on System z distributions, consult the distributor's support information.

Middleware

Middleware offerings for the IBM zEnterprise EC12 and IBM zEnterprise BC12 environments include the following environments:

- ▶ 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
 - IBM DB2 for z/OS
 - IBM DB2 for Linux
 - IBM DB2 Connect™

Operations

The IBM Tivoli® brand has a large product set that includes the following offerings:

- ▶ IBM Tivoli Service Management Center
- ▶ IBM Tivoli Information Management for z/OS
- ▶ IBM Tivoli Workload Scheduler
- ▶ IBM Tivoli OMEGAMON® XE
- ▶ IBM Tivoli System Automation

Security

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

- ▶ IBM Security zSecure suite

- ▶ The Security Server component of z/OS and z/VM; includes Resource Access Control Facility (IBM RACF®)
- ▶ IBM Tivoli Directory Server for z/OS
- ▶ z/OS Communications Server and Policy Agent; for policy-based network security
- ▶ The z/OS Cryptographic Services component of z/OS; includes the Integrated Cryptographic Service Facility (ICSF)

Application development and languages

Many languages are available for the IBM zEnterprise EC12 and IBM zEnterprise BC12 environment. Because the Linux environment is similar to Linux on other servers, we focus on the z/OS environment.

In addition to the traditional COBOL, PL/I, Fortran, and Assembler languages, others are available such as C, C++, and Java (including Java Platform, Enterprise Edition and batch environments).

Development can be conducted by using the latest software engineering technologies and advanced *integrated development environments (IDE)*. The extensive tool set uses a workstation environment for development and testing, with final testing and deployment performed on z/OS. Application development tools, many of which have components that are based on the Eclipse platform, are provided through the following offerings:

- ▶ IBM Rational® Application Developer for WebSphere
- ▶ IBM Rational Developer for System z
- ▶ IBM WebSphere Developer for System z
- ▶ IBM Rational Rose® product line
- ▶ IBM Rational Software Architect and Software Modeler

For more information about software for System z, see this website:

<http://www.ibm.com/software/products/us/en/category/sysz?pgel=1nav>

We cannot overemphasize the importance of using the most recent versions of the compilers. The compilers enable the use of the latest technologies that are implemented on the system and take advantage of the performance benefits that are introduced. Examples of benefits include new cache structures, new machine instructions, and instruction execution enhancements:

- ▶ *Runtime Instrumentation Facility*, which is designed to provide managed run times and just-in-time compilers with enhanced feedback on application behavior, allowing dynamic optimization on code generation as it is being executed. Java is planned to use a significant set of the new instructions available on zEC12 and zBC12.
- ▶ *Transactional Execution Facility*, which is designed to help eliminate software locking overhead that can affect performance, offering increased scalability and parallelism to drive higher transaction throughput. Transactional memory provides atomic processing for multiple storage areas, which can reduce serialization overhead for users.

Java exploitation of Transactional Execution: Java exploitation of Transactional Execution has the following minimum requirements:

- ▶ IBM 31-bit and 64-bit SDK7 for z/OS Java Technology Edition, Version 7 SR3 (5655-W43 and 5655-W44).
- ▶ IBM Java 7 SR3 with Linux on System z distributions:
 - SLES 11 SP3.
 - RHEL 6.4.

IBM compilers

Each new version of IBM z/OS compilers (Enterprise COBOL, Enterprise PL/I, XL C/C++) underscores the continuing IBM commitment to the COBOL, PL/I, and C/C++ programming languages on the z/OS platform.

The latest version of Enterprise COBOL delivers enhanced XML parsing support, facilitates compiler message severity customization, uses system-determined block size for QSAM files, supports the underscore (_) character in COBOL user-defined words, provides compiler listings that display CICS options in effect, and supports Java 5 and Java 6 SDKs for Java interoperability.

The latest version of Enterprise PL/I uses the latest z/Architecture for application performance improvements, improves SQL preprocessing, and uses user-requested productivity enhancements with new options, built-in functions, and statements.

The latest version of z/OS XL C/C++ delivers application performance improvements by using the latest advancements in optimization and hardware technology. This feature provides easier access to debug information for debug files, enabling easier porting of C/C++ applications to z/OS, and reduces application development effort with new programming features.

IBM Enterprise COBOL and Enterprise PL/I support are strategic components (separately orderable products) for IBM Rational Developer for IBM System z software. These features provide a robust, integrated development environment (IDE) for COBOL and PL/I and connecting web services, Java Platform, Enterprise Edition (Java EE) applications, and traditional business processes.

z/OS XL C/C++ programmers can also tap into Rational Developer for System z to boost their productivity by easily editing, compiling, and debugging z/OS XL C and XL C++ applications right from their workstation.

4.2 Support by operating system

In this section, we list the support by in-service operating systems of selected functions of the zEC12 and zBC12. For a detailed description of zEC12 and its features, see the *IBM zEnterprise EC12 Technical Guide*, SG24-8049. For a detailed description of zBC12 and its features, see the *IBM zEnterprise BC12 Technical Guide*, SG24-8138. For an in-depth description of all I/O features, see the *IBM System z Connectivity Handbook*, SG24-5444.

4.2.1 z/OS

z/OS Version 1 Release 12 is the earliest in-service release that supports the zEC12 and the zBC12. The IBM Lifecycle Extension program provides a fee-based extension for defect support for up to two years after end of service. This program is available for V1R11 and extends the support to September 2014.

Table 4-2 summarizes the support requirements of selected zEC12 and zBC12 functions for the currently supported z/OS releases. It uses the following conventions:

- Y** The function is supported.
- N** The function is not supported.

Table 4-2 z/OS support summary

Function	V2R1	V1R13	V1R12	V1R11 ^a
zEC12 and zBC12	Y	Y ^b	Y ^b	Y ^b
Support of Unified Resource Manager	Y	Y	Y ^f	Y ^f
Support of 100 PUs by a single system image	Y	Y	Y	Y
Dynamic add of logical central processors (CP)	Y	Y	Y	Y
Support of IBM zAware	Y	Y ^f	N	N
zAAP on zIIP	Y	Y	Y	Y
Large memory > 128 GB	Y	Y	Y	Y
Large page support	Y	Y ^{cd}	Y	Y
Hardware decimal floating point ^e	Y	Y	Y	Y
Out-of-order execution	Y	Y	Y	Y
Java Exploitation of Transactional Execution	Y	Y ^f	N	N
Capacity Provisioning Manager	Y	Y	Y	Y
LPAR group capacity limit	Y	Y	Y	Y
LPAR physical capacity limit	Y ^f	Y ^f	Y ^f	N
Enhanced flexibility for CoD	Y	Y	Y ^m	Y ^m
HiperDispatch	Y	Y	Y	Y
CPACF	Y	Y	Y	Y
CPACF AES-128, AES-192, and AES-256	Y	Y	Y	Y
CPACF SHA-1, SHA-224, SHA-256, SHA-384, SHA-512	Y	Y	Y	Y
CPACF protected key ^f	Y	Y	Y	Y ^g
Crypto Express4S	Y	Y ^{hi}	Y ^{hi}	Y ^h
Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode	Y	Y ⁱ	Y ⁱ	N
Crypto Express3 ^f	Y	Y	Y	Y ^g
Elliptic Curve Cryptography (ECC)	Y	Y	Y ^g	Y ^g
HiperSockets Completion Queue	Y	Y ^f	N	N
HiperSockets integration with IEDN	Y	Y ^f	N	N
HiperSockets multiple write facility	Y	Y	Y	Y
Flash Express	Y ^f	Y ^{fgj}	N	N
zEDC Express	Y	Y ^k	Y ^k	N
10GbE RoCE Express	Y	Y ^l	Y ^l	Y ^l
High Performance FICON (zHPF)	Y	Y	Y	Y
FICON Express8S ^m	Y	Y	Y ^f	Y ^f

Function	V2R1	V1R13	V1R12	V1R11 ^a
FICON Express8 ^m	Y	Y	Y	Y ^f
FICON Express4	Y	Y	Y	Y
FICON Express4-2C ⁿ	Y	Y	Y	Y
OSA-Express5S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y	Y	Y ^f
OSA-Express5S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y	Y	Y	Y ^f
OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	Y	Y	Y	Y ^f
OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	Y	Y	Y	Y ^f
OSA-Express5S 1000BASE-T CHPID type OSC (using one or two ports per CHPID)	Y	Y	Y	Y ^f
OSA-Express5S 1000BASE-T CHPID type OSD (using two ports per CHPID)	Y	Y	Y	Y ^f
OSA-Express5S 1000BASE-T CHPID type OSD (using one port per CHPID)	Y	Y	Y	Y ^f
OSA-Express5S 1000BASE-T CHPID type OSE ^o (using one or two ports per CHPID)	Y	Y	Y	Y ^f
OSA-Express5S 1000BASE-T CHPID type OSM ^p	Y	Y	Y	Y ^f
OSA-Express5S 1000BASE-T CHPID type OSN ^q (using one or two ports per CHPID)	Y	Y	Y	Y ^f
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y	Y	Y ^f
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y	Y	Y ^f	Y ^f
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	Y	Y	Y	Y ^f
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	Y	Y	Y	Y ^f
OSA-Express4S 1000BASE-T ^r CHPID type OSC (using one or two ports per CHPID)	Y	Y	Y	Y ^f
OSA-Express4S 1000BASE-T ^r CHPID type OSD (using two ports per CHPID)	Y	Y	Y	Y ^f
OSA-Express4S 1000BASE-T ^r CHPID type OSD (using one port per CHPID)	Y	Y	Y	Y ^f
OSA-Express4S 1000BASE-T ^r CHPID type OSE ^o (using one or two ports per CHPID)	Y	Y	Y	Y ^f
OSA-Express4S 1000BASE-T ^r CHPID type OSM ^p	Y	Y	Y ^f	Y ^f

Function	V2R1	V1R13	V1R12	V1R11 ^a
OSA-Express4S 1000BASE-T ^f CHPID type OSN (using one or two ports per CHPID)	Y	Y	Y	Y ^f
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y	Y	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSX	Y	Y	Y ^f	Y ^f
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN ^q (using four ports)	Y	Y	Y	Y
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN ^q (using two ports)	Y	Y	Y	Y
OSA-Express3-2P Gigabit Ethernet SX ⁿ CHPID types OSD and OSN ^q	Y	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID types OSC and OSD (using four ports)	Y	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID types OSE ^o and OSN ^q (using four ports)	Y	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet using two ports CHPID types OSC, OSD, OSE ^o , and OSN ^q	Y	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID type OSM ^p (using two ports)	Y	Y	Y ^f	Y ^f
OSA-Express3-2P 1000BASE-T Ethernet ⁿ CHPID types OSC, OSD, OSE ^o , and OSN ^q	Y	Y	Y	Y
OSA-Express3-2P 1000BASE-T Ethernet ⁿ CHPID type OSM ^p	Y	Y	Y	Y
Coupling using InfiniBand CHPID type CIB	Y	Y	Y	Y
InfiniBand coupling links (12x IFB-SDR or 12x IFB-DDR) at a distance of 150 m	Y	Y	Y	Y ^f
InfiniBand coupling links (1x IFB-SDR or 1x IFB-DDR) at an unrepeat distance of 10 km	Y	Y	Y	Y ^f
Server Time Protocol	Y	Y	Y	Y
CFCC Level 19	Y ^f	Y ^f	Y ^f	N
CFCC Level 18	Y	Y ^f	Y ^f	N
Assembler instruction mnemonics	Y	Y	Y	Y
C/C++ exploitation of hardware instructions	Y	Y	Y	Y
Layer 3 VMAC	Y	Y	Y	Y
Large dumps	Y	Y	Y	Y
CPU measurement facility	Y	Y ^f	Y ^f	Y ^f

- a. With the announcement of IBM Lifecycle Extension for z/OS V1.11, fee-based corrective service can be ordered for up to two years after the withdrawal of service for z/OS V1R11
- b. Service is required for support of zEC12 or zBC12.

- c. A web deliverable is required for Pageable 1M Large Page Support.
- d. A web deliverable is required for 2G Large Page Support.
- e. The level of decimal floating-point exploitation varies with z/OS release and PTF level.
- f. Service is required.
- g. FMIDs are shipped in a web deliverable.
- h. Crypto Express4S Toleration requires a web deliverable and PTFs.
- i. Crypto Express4S Exploitation requires a web deliverable.
- j. A web deliverable is required for Dynamic Reconfiguration Support for Flash Express.
- k. Software decompression only.
- l. Toleration support only.
- m. Support varies with operating system and level.
- n. Not available on zEC12.
- o. A CHPID Type OSE supports both SNA (LLC2) and IP connectivity over Ethernet (802.3 or DIX V2).
- p. One port is configured for OSM. The second port of the pair is unavailable.
- q. CHPID type OSN does not use ports. LPAR-to-LPAR communication is used.
- r. Not available on zBC12.

4.2.2 z/VM

At general availability, z/VM V6R3, z/VM V6R2, and z/VM V5R4 provide compatibility support with use of some new zEC12 and zBC12 functions.

Statement of Direction: The IBM zEnterprise EC12 and IBM zEnterprise BC12 are planned to be the last System z servers supported by z/VM V5R4 and the last System z servers that will support z/VM V5R4 running as a guest (second level). z/VM V5R4 will continue to be supported until December 31, 2014, or until the IBM System z9 EC and IBM System z9 BC are withdrawn from support, whichever is later. Refer to Withdrawal Announcement 912-144, (RFA56762) dated August 7, 2012.

Table 4-3 summarizes the support requirements of selected functions for the currently supported z/VM releases. It uses the following conventions:

- Y** The function is supported.
- N** The function is not supported.

Table 4-3 z/VM support summary

Function	V6R3	V6R2	V5R4
zEC12 ^a and zBC12 ^a	Y	Y	Y
Support of Unified Resource Manager	N	Y	N
Support of up to 32 PUs for single system image ^b	Y	Y	Y
Dynamic add of logical CPs	Y	Y	Y
zAAP on zIIP ^{c,g}	Y	Y	Y
Large memory > 128 GB ^d	Y ^e	Y	Y
Large page support ^f	N	N	N
Hardware decimal floating point ^g	Y	Y	Y
Out-of-order execution	Y	Y	Y
Capacity Provisioning Manager ^f	N	N	N
CPU measurement facility counter	Y	Y	Y ^h

Function	V6R3	V6R2	V5R4
HiperDispatch	Y ^f	N	N
CPACF ^g	Y	Y	Y
CPACF AES-128, AES-192, and AES-256 ^g	Y	Y	Y
CPACF SHA-1, SHA-224, SHA-256, SHA-384, SHA-512 ^g	Y	Y	Y
CPACF protected key ^g	Y	Y	Y ^h
Crypto Express4S ^g	Y	Y ^h	Y ^h
Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode ^g	Y	Y ^h	Y ^h
Crypto Express3 ^g	Y	Y	Y ^h
Elliptic Curve Cryptography (ECC) ^g	Y	Y ^h	Y ^h
Execute relative guest exploitation ^g	Y	Y	Y
Restore subchannel facility	Y	Y	Y
HiperSockets integration with IEDN	N	Y ^h	N
HiperSockets Virtual Switch Bridge	Y	Y ^h	N
HiperSockets multiple write facility	N ^f	N ^f	N ^f
Flash Express	N	N	N
zEDC Express ⁱ	N	N	N
10GbE RoCE Express ⁱ	N	N	N
High Performance FICON (zHPF)	Y ^g	Y ^{gh}	N ^f
FICON Express8S	Y	Y	Y ^h
FICON Express8 ^j	Y	Y	Y ^h
FICON Express4	Y	Y	Y
FICON Express4-2C ^k	Y	Y	Y
OSA-Express QDIO data connection isolation for z/VM environments	Y	Y	Y ^h
OSA-Express5S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y	Y
OSA-Express5S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y ^p	Y	Y ^p
OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	Y	Y	Y
OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	Y	Y	Y
OSA-Express5S 1000BASE-T CHPID type OSC (using one or two ports per CHPID)	Y	Y	Y
OSA-Express5S 1000BASE-T CHPID type OSD (using two ports per CHPID)	Y	Y	Y
OSA-Express5S 1000BASE-T CHPID type OSD (using one port per CHPID)	Y	Y	Y

Function	V6R3	V6R2	V5R4
OSA-Express5S 1000BASE-T CHPID type OSE ^l (using one or two ports per CHPID)	Y	Y	Y
OSA-Express5S 1000BASE-T CHPID type OSM ^m	Y ^p	Y	Y ^p
OSA-Express5S 1000BASE-T CHPID type OSN ⁿ (using one or two ports per CHPID)	Y	Y	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y ^p	Y	N ^p
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	Y	Y	Y ^h
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	Y	Y	Y
OSA-Express4S 1000BASE-T ^o CHPID type OSC (using one or two ports per CHPID)	Y	Y	Y
OSA-Express4S 1000BASE-T ^o CHPID type OSD (using two ports per CHPID)	Y	Y	Y ^h
OSA-Express4S 1000BASE-T ^o CHPID type OSD (using one port per CHPID)	Y	Y	Y
OSA-Express4S 1000BASE-T ^o CHPID type OSE ^l (using one or two ports per CHPID)	Y	Y	Y
OSA-Express4S 1000BASE-T ^o CHPID type OSM ^m	Y ^p	Y	Y ^p
OSA-Express4S 1000BASE-T ^o CHPID type OSN ⁿ (using one or two ports per CHPID)	Y	Y	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSX	Y ^p	Y	Y ^p
OSA-Express3 Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	Y	Y	Y ^h
OSA-Express3 Gigabit Ethernet LX and SX CHPID type OSN ⁿ (using four ports)	Y	Y	Y ^h
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN ⁿ (using two ports)	Y	Y	Y
OSA-Express3-2P Gigabit Ethernet SX ^k CHPID types OSD and OSN ⁿ			
OSA-Express3 1000BASE-T Ethernet using four ports CHPID type OSD (using four ports)	Y	Y	Y ^h
OSA-Express3 1000BASE-T Ethernet using four ports CHPID types OSC, OSE ^l , and OSN ⁿ (using four ports)	Y	Y	Y ^h

Function	V6R3	V6R2	V5R4
OSA-Express3 1000BASE-T Ethernet CHPID types OSC, OSD, OSE ^l , and OSN ⁿ (using two ports)	Y	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID type OSM ^{h,m} (using two ports)	Y ^p	Y	Y ^p
OSA-Express3-2P 1000BASE-T Ethernet ^k CHPID types OSC, OSD, OSE ⁿ , and OSN ⁿ	Y	Y	Y
OSA-Express3-2P 1000BASE-T Ethernet ^k CHPID type OSM ^m	Y ^p	Y	Y ^p
Dynamic I/O support for InfiniBand CHPIDs	Y ^p	Y ^p	Y ^p
InfiniBand coupling links (12x IFB-SDR or 12x IFB-DDR) at a distance of 150 m	Y ^p	Y ^p	Y ^p
InfiniBand coupling links (1x IFB-SDR or 1x IFB-DDR) at an unrepeatd distance of 10 km	Y ^p	Y ^p	Y ^p
CFCC Level 19	Y ^g	Y ^{gh}	Y ^{gh}
CFCC Level 18	Y ^g	Y ^{gh}	Y ^{gh}

- a. Service is required for support of zEC12 or zBC12.
- b. z/VM supports up to 32 real PUs and up to 64 logical PUs per guest.
- c. Available for z/OS on virtual machines without virtual zAAPs defined when the z/VM LPAR does not have zAAPs defined to it.
- d. 256 GB of central memory are supported by z/VM V5R4 and later. z/VM V5R4 and later support more than 1 TB of virtual memory in use for guests.
- e. z/VM V6R3 supports 1TB of real memory and 1TB of virtual memory per guest.
- f. Not available to guests.
- g. Supported for guest use only.
- h. Service is required.
- i. In a future z/VM deliverable IBM plans to offer z/VM support for guest exploitation of this feature. All statements regarding IBM plans, directions, and intent are subject to change or withdrawal without notice.
- j. Support varies with operating system and level.
- k. Not available on zEC12.
- l. A CHPID Type OSE supports both SNA (LLC2) and IP connectivity over Ethernet (802.3 or DIX V2).
- m. One port is configured for OSM. The second port of the pair is unavailable.
- n. CHPID type OSN does not use ports, it uses LPAR-to-LPAR communication.
- o. Not available on zBC12.
- p. Support is for dynamic I/O configuration only.

z/VM logical partitions: zEC12 and zBC12 CPs and IFLs have increased capacity over that of their predecessors. Therefore, we suggest that the capacity of z/VM logical partitions and of any guests, in terms of the *number* of IFLs and CPs (real or virtual), be reviewed and adjusted to achieve the required capacity. Virtual machine shares might also need to be adjusted.

4.2.3 z/VSE

Table 4-4 summarizes the support requirements of selected zEC12 and zBC12 functions for the currently supported z/VSE releases. It uses the following conventions:

- Y** The function is supported.
- N** The function is not supported.

Table 4-4 z/VSE support summary

Function	V5R1 ^a	V4R3 ^b
zEC12 ^c and zBC12 ^c	Y	Y
Support for up to four CPs ^d	Y	Y
Dynamic add of logical CPs	Y	Y
Large page support for data spaces	Y	Y
Out-of-order execution	Y	Y
CPACF	Y	Y
CPACF AES-128, AES-192, and AES-256	Y	Y
CPACF SHA-1, SHA-224, SHA-256, SHA-384, SHA-512	Y	Y
CPACF protected key	N	N
Crypto Express4S Toleration ^f	Y ^e	N
Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode	N	N
Crypto Express3 ^f	Y	Y
Elliptic Curve Cryptography (ECC)	N	N
FICON Express8S	Y	Y
FICON Express8	Y	Y
FICON Express4	Y	Y
FICON Express4-2C ^g	Y	Y
OSA-Express5S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y
OSA-Express5S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y	N
OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	Y	Y
OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	Y	Y
OSA-Express5S 1000BASE-T CHPID type OSC (using one or two ports per CHPID)	Y	Y
OSA-Express5S 1000BASE-T CHPID type OSD (using two ports per CHPID)	Y	Y
OSA-Express5S 1000BASE-T CHPID type OSD (using one port per CHPID)	Y	Y
OSA-Express5S 1000BASE-T CHPID type OSE ^h (using one or two ports per CHPID)	Y	Y
OSA-Express5S 1000BASE-T CHPID type OSM ⁱ	N	N
OSA-Express5S 1000BASE-T CHPID type OSN (using one or two ports per CHPID)	Y	Y

Function	V5R1 ^a	V4R3 ^b
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y	N
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	Y	Y
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using one port)	Y	Y
OSA-Express4S 1000BASE-T ^j CHPID type OSC (using one or two ports per CHPID)	Y	Y
OSA-Express4S 1000BASE-T ^j CHPID type OSD (using two ports per CHPID)	Y	Y
OSA-Express4S 1000BASE-T ^j CHPID type OSD (using one port per CHPID)	Y	Y
OSA-Express4S 1000BASE-T ^j CHPID type OSE ^h (using one or two ports per CHPID)	Y	Y
OSA-Express4S 1000BASE-T ^j CHPID type OSM ⁱ	N	N
OSA-Express4S 1000BASE-T ^j CHPID type OSN ^k (using one or two ports per CHPID)	Y	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSD	Y	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSX	Y	N
OSA-Express3 Gigabit Ethernet LX and SX CHPID type OSD and OSN ^k (using four ports)	Y	Y
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN ^k (using two ports)	Y	Y
OSA-Express3-2P Gigabit Ethernet SX ^g CHPID types OSD and OSN ^k	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID types OSC, OSD, OSE ^h , and OSN ^k (using four ports)	Y	Y
OSA-Express3 1000BASE-T Ethernet CHPID types OSC, OSD, OSE ^h , and OSN ^k (using two ports)	Y	Y
OSA-Express3-2P 1000BASE-T Ethernet ^g CHPID types OSC, OSD, OSE ^h , and OSN ^k	Y	Y
OSA-Express3-2P 1000BASE-T Ethernet ^g CHPID type OSM ⁱ	N	N
HiperSockets Completion Queue for Linux Fast Path function in LPAR	Y	N

a. z/VSE V5 executes in z/Architecture mode and supports 64-bit real and 64-bit virtual memory addressing.

b. z/VSE V4 executes in z/Architecture mode and supports 64-bit real memory addressing. It does not support 64-bit virtual memory addressing.

- c. PTFs are required for zEC12 and zBC12 support.
- d. z/VSE Turbo Dispatcher supports up to four CPs and tolerates up to 10-way LPARs
- e. Crypto Express4S Toleration requires PTFs.
- f. z/VSE supports clear key RSA operations only.
- g. Not available on zEC12.
- h. A CHPID Type OSE supports both SNA (LLC2) and IP connectivity over Ethernet (802.3 or DIX V2).
- i. One port is configured for OSM. The other port is unavailable.
- j. Not available on zBC12.
- k. CHPID type OSN does not use ports. All communication is LPAR to LPAR.

4.2.4 z/TPF

Table 4-5 summarizes the support requirements of selected zEC12 and zBC12 functions for the currently supported z/TPF release. It uses the following conventions:

- Y** The function is supported.
- N** The function is not supported.

Table 4-5 TPF and z/TPF support summary

Function	z/TPF V1R1
zEC12 and zBC12	Y
Support for up to 86 PUs for single system image	Y
Large memory > 128 GB (4 TB)	Y
CPACF	Y
CPACF AES-128, AES-192, and AES-256	Y ^a
CPACF SHA-1, SHA-224, SHA-256, SHA-384, SHA-512	Y ^b
CPACF protected key	N
Crypto Express4S ^c	Y ^d
Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode	N
Crypto Express3	Y
Elliptic Curve Cryptography (ECC)	N
FICON Express8S	Y
FICON Express8	Y
FICON Express4	Y
FICON Express4-2C ^e	Y
OSA-Express5S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y
OSA-Express5S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y
OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	Y
OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	Y

Function	z/TPF V1R1
OSA-Express5S 1000BASE-T CHPID type OSC (using one or two ports per CHPID)	Y
OSA-Express5S 1000BASE-T CHPID type OSD (using two ports per CHPID)	Y
OSA-Express5S 1000BASE-T CHPID type OSD (using one port per CHPID)	Y
OSA-Express5S 1000BASE-T CHPID type OSE ^f (using one or two ports per CHPID)	Y
OSA-Express5S 1000BASE-T CHPID type OSM ^g	Y
OSA-Express5S 1000BASE-T CHPID type OSN ^h (using one or two ports per CHPID)	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSD	Y
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSX	Y ⁱ
OSA-Express4S Gigabit Ethernet LX and SX CHPID types OSD and OSN ^h (using four ports)	Y ⁱ
OSA-Express4S Gigabit Ethernet LX and SX CHPID types OSD and OSN ^h (using two ports)	Y
OSA-Express4S 1000BASE-T ^j CHPID type OSC (using one or two ports per CHPID)	N
OSA-Express4S 1000BASE-T ^j CHPID type OSD (using two ports per CHPID)	Y ⁱ
OSA-Express4S 1000BASE-T ^j CHPID type OSD (using one port per CHPID)	Y
OSA-Express4S 1000BASE-T ^j CHPID type OSE ^f (using one or two ports per CHPID)	N
OSA-Express4S 1000BASE-T ^j CHPID type OSM ^g	N
OSA-Express4S 1000BASE-T ^j CHPID type OSN ^h (using one or two ports per CHPID)	Y ⁱ
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSD	Y
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSX	N
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN ^h (using four ports)	Y ⁱ
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN ^h (using two ports)	Y
OSA-Express3-2P Gigabit Ethernet SX ^e CHPID types OSD and OSN ^h	Y

Function	z/TPF V1R1
OSA-Express3 1000BASE-T Ethernet using four ports CHPID types OSC, OSD, and OSN ^h	Y
OSA-Express3 1000BASE-T Ethernet using two ports CHPID types OSC, OSD, and OSN ^h	Y
OSA-Express3-2P 1000BASE-T Ethernet ^e CHPID types OSC, OSD, OSE ^f , and OSN ^h	Y
OSA-Express3-2P 1000BASE-T Ethernet ^e CHPID type OSM ^g	N
Coupling over InfiniBand CHPID type CIB	Y ^k
CFCC Level 19	N
CFCC Level 18	Y

- a. Supports only AES-128 and AES-256.
- b. Supports only SHA-1 and SHA-256.
- c. Service is required.
- d. Accelerator mode only.
- e. Not available on zEC12.
- f. A CHPID Type OSE supports both SNA (LLC2) and IP connectivity over Ethernet (802.3 or DIX V2).
- g. One port is configured for OSM. The other port is unavailable.
- h. CHPID type OSN does not use ports. All communication is LPAR to LPAR.
- i. Requires PUT 4 with PTFS.
- j. Not available on zBC12.
- k. Compatibility is supported.

4.2.5 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 (SUSE SLES 10 and SLES 11; Red Hat RHEL 5 and RHEL 6) require toleration support; therefore, any release of these distributions can run on the zEC12 and zBC12.

Table 4-6 lists the most recent service levels of the current SUSE and Red Hat releases at the time of writing.

Table 4-6 Current Linux on System z distributions, by z/Architecture mode

Linux distribution	z/Architecture (64-bit mode)
SUSE SLES 11 SP2	Yes
SUSE SLES 10 SP4	Yes
Red Hat RHEL 5.8	Yes
Red Hat RHEL 6.2	Yes

Table 4-7 lists selected zEC12 and zBC12 features, showing the minimum level of SUSE and Red Hat distributions that support each feature.

Table 4-7 Linux on System z support summary

Function	SUSE	Red Hat
zEC12 and zBC12	SLES 10	RHEL 5
Maximum number of CPs or IFLs	64	80
Large page support	SLES 10 SP2	RHEL 5.3
Hardware decimal floating point	SLES 11	N
CPACF	SLES 9	RHEL 4.4
CPACF AES-128, AES-192, and AES-256	SLES 10 SP2	RHEL 5.2
CPACF SHA-1, SHA-224, SHA-256, SHA-384, SHA-512	SLES 10 SP2	RHEL 5.2
CPACF protected key	N	N
Crypto Express4S ^a	N ^a	N ^a
Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode	N	N
Crypto Express3	SLES 10 SP3 ^b SLES 11 SP1 ^c	RHEL 5.4 ^b RHEL 6.1 ^c
Elliptic Curve Cryptography (ECC)	N	N
HiperSockets Completion Queue	SLES 11 SP2	RHEL 6.2
HiperSockets Virtual Switch Bridge	SLES 10 SP4	RHEL 5.8
HiperSockets Layer 2 support	SLES 10 SP2	RHEL 5.3
10GbE RoCE Express ^a	N	N
High Performance FICON (zHPF)	Note ^d	Note ^d
FICON Express8S CHPID type FC	SLES 11 ^d SLES 10	RHEL 6 ^d RHEL 5
FICON Express8, FICON Express4 ^e CHPID types FC and FCP	SLES 10	RHEL 5
FICON Express4-2C ^f CHPID type FC	SLES 10	RHEL 5
OSA-Express5S 10 Gigabit Ethernet LR and SR CHPID type OSD	SLES 10	RHEL 5
OSA-Express5S 10 Gigabit Ethernet LR and SR CHPID type OSX	SLES 10	RHEL 5
OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	SLES 10	RHEL 5
OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	SLES 10	RHEL 5
OSA-Express5S 1000BASE-T CHPID type OSC (using one or two ports per CHPID)	SLES 10	RHEL 5

Function	SUSE	Red Hat
OSA-Express5S 1000BASE-T CHPID type OSD (using two ports per CHPID)	SLES 10	RHEL 5
OSA-Express5S 1000BASE-T CHPID type OSD (using one port per CHPID)	SLES 10	RHEL 5
OSA-Express5S 1000BASE-T CHPID type OSE ^g (using one or two ports per CHPID)	SLES 10	RHEL 5
OSA-Express5S 1000BASE-T CHPID type OSM ^h	SLES 10	RHEL 5
OSA-Express5S 1000BASE-T CHPID type OSN ⁱ (using one or two ports per CHPID)	SLES 10	RHEL 5
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSD	SLES 10	RHEL 5
OSA-Express4S 10 Gigabit Ethernet LR and SR CHPID type OSX	SLES 10 SP4	RHEL 5.6
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using four ports)	SLES 10 SP2	RHEL 5.2
OSA-Express4S Gigabit Ethernet LX and SX CHPID type OSD (using two ports)	SLES 10	RHEL 5
OSA-Express4S 1000BASE-T ^j CHPID type OSD (using two ports per CHPID)	SLES 10 SP2	RHEL 5.2
OSA-Express4S 1000BASE-T ^j CHPID type OSD (using one port per CHPID)	SLES 10	RHEL 5
OSA-Express4S 1000BASE-T ^j CHPID type OSM ^h	SLES 10 SP4	RHEL 5.6
OSA-Express4S 1000BASE-T ^j CHPID type OSN ⁱ (using one or two ports per CHPID)	SLES 10	RHEL 5
OSA-Express3 10 Gigabit Ethernet LR and SR CHPID type OSD	SLES 10	RHEL 5
OSA-Express3 Gigabit Ethernet LX and SX CHPID type OSD and OSN ⁱ (using four ports)	SLES 10 SP2	RHEL 5.2
OSA-Express3 Gigabit Ethernet LX and SX CHPID types OSD and OSN ⁱ (using two ports)	SLES 10	RHEL 5
OSA-Express3-2P Gigabit Ethernet SX ^j CHPID types OSD and OSN ⁱ	SLES 10	RHEL 5
OSA-Express3 1000BASE-T Ethernet using four ports CHPID types OSD and OSN ⁱ	SLES 10 SP2	RHEL 5.2
OSA-Express3 1000BASE-T Ethernet using two ports CHPID types OSD, and OSN ⁱ	SLES 10	RHEL 5
OSA-Express3 1000BASE-T Ethernet CHPID type OSM ^h (using two ports)	SLES 10 SP4	RHEL 5.6

Function	SUSE	Red Hat
OSA-Express3-2P 1000BASE-T Ethernet ^j CHPID types OSC, OSD, OSE ^g , and OSN ⁱ	SLES 10	RHEL 5
OSA-Express3-2P 1000BASE-T Ethernet ^j CHPID type OSM ^h	N	N

- a. IBM is working with its Linux distribution partners to include support in future Linux on System z distribution releases.
- b. Crypto Express3 toleration mode.
- c. Crypto Express3 exploitation.
- d. For support of zHPF single track operations.
- e. All FICON Express4, FICON Express2 and FICON features are withdrawn from marketing.
- f. Not available on zEC12.
- g. A CHPID Type OSE supports both SNA (LLC2) and IP connectivity over Ethernet (802.3 or DIX V2).
- h. One port is configured for OSM. The second port of the pair is unavailable.
- i. CHPID type OSN does not use ports. All communication is LPAR-to-LPAR.
- j. Not available on zBC12.

IBM is working with its Linux distribution partners so that use of further IBM zEnterprise EC12 and IBM zEnterprise BC12 functions are provided in future Linux on System z distribution releases. We suggest the following updates:

- SUSE SLES 11 or Red Hat RHEL 6 should be used in any new projects for the zEC12 and zBC12.
- Any Linux distributions should be updated to their latest service level before migration to zEC12 or zBC12.
- Adjust the capacity of any z/VM or Linux logical partitions, and of any z/VM guests, in terms of the *number* of IFLs and CPs, real or virtual, in face of the increased processor unit (PU) capacity of the zEC12 and zBC12.

4.2.6 References

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

- z/OS:
<http://www.ibm.com/systems/z/os/zos>
- z/VM:
<http://www.ibm.com/systems/z/os/zvm>
- z/TPF:
<http://www.ibm.com/systems/z/os/tpf>
- z/VSE:
<http://www.ibm.com/systems/z/os/zvse>
- Linux on System z:
<http://www.ibm.com/systems/z/os/linux>

4.3 Software support for zBX Model 003

The *zEnterprise BladeCenter Extension (zBX) Model 003* offers the following types of application environments:

- ▶ Special purpose, dedicated environments such as the *IBM WebSphere DataPower XI50 for zEnterprise*. In this case, support is dictated by the solution.
- ▶ POWER7 blades. The blades are virtualized by PowerVM Enterprise Edition, and their LPARs run either AIX Version 5 Release 3 TL12 (IBM POWER6® mode), AIX Version 6 Release 1 TL5 (POWER7 mode), and AIX Version 7 Release 1, and subsequent releases. Applications that are supported on AIX can be deployed to blades.
- ▶ Also offered are selected IBM System x HX5 blades. Virtualization is provided by an integrated hypervisor, using Kernel-based virtual machines, and supporting the following operating systems:
 - Linux on System x (64-bit only):
 - Red Hat: RHEL 5.5 and up, and RHEL 6.0 and up
 - SUSE: SLES 10 (SP4) and up, and SLES 11 SP1 and up
 - Microsoft Windows (64-bit only, Datacenter Edition is recommended):
 - Microsoft Windows Server 2012
 - Microsoft Windows Server 2008 R2
 - Microsoft Windows Server 2008 (SP2)

Applications that are supported on Linux on System x and on Windows can be deployed to blades.

4.4 z/OS considerations

zEC12 or zBC12 base processor support is required in z/OS. With that exception, software changes do not require the new zEC12 or zBC12 functions and, equally, the new functions do not require functional software. The approach is to (where applicable) automatically decide to *enable* or *disable* a function that is based on the presence (enable) or absence (disable) of the required hardware and software.

General recommendations

The zEC12 and zBC12 introduce the latest System z technology, notable cases are Flash Express, zEDC Express and 10GbE RoCE Express support. Although support for zEC12 and zBC12 is provided by z/OS starting with z/OS V1R11, the use of zEC12's and zBC12's functions is dependent on the z/OS release.

In general, we suggest the following approaches:

- ▶ Do not migrate software releases and hardware at the same time.
- ▶ Keep members of the sysplex at the same software level other than during brief migration periods.
- ▶ Migrate to a Server Time Protocol (STP) or Mixed-CTN network before introducing a zEC12 or zBC12 into a Sysplex.

Statement of Direction: IBM zEnterprise EC12 and IBM zEnterprise BC12 will be the last System z servers to support connections to an STP-Mixed CTN. This includes the Sysplex Timer (9037). After zEC12 and zBC12, servers that require time synchronization, such as to support a base or Parallel Sysplex, will require STP. Additionally, all servers in that network must be configured in the STP-only mode.

- Review zEC12 or zBC12 restrictions and considerations before creating an upgrade plan.

Flash Express

IBM zEnterprise EC12 introduces the Flash Express feature, also available on the zBC12, which can help improve resilience and performance of the z/OS system. Flash Express is designed to assist with the handling of workload spikes or increased workload demand that might occur at the opening of the business day, or in the event of a workload shift from one system to another.

z/OS is the first exploiter to use Flash Express storage as Storage Class Memory (SCM) for paging store and SVC dump. SVC dump data capture time is expected to be substantially reduced. As a paging store, Flash Express storage is suitable for workloads that can tolerate paging and does not benefit workloads that cannot afford to page. The z/OS design for Flash Express storage does not completely remove the virtual storage constraints that are created by a paging spike in the system.

Flash Express storage is allocated to logical partition similarly to main memory. The initial and maximum amount of Flash Express Storage that is available to a particular logical partition is specified at the SE or HMC via a new flash memory allocation panel. The amount of Flash Express storage in the partition can be changed dynamically, between the initial and the maximum amount at the SE or HMC. For z/OS, this can also be done by an operator command. Each partition's Flash Express storage is isolated similarly to main storage and sees only its own space in the flash memory space.

zEDC Express

zEDC Express, an optional feature exclusive to zEC12 and zBC12, addresses data growth requirements by providing hardware-based acceleration for data compression and decompression. zEDC provides data compression with lower CPU consumption than previously existing compression technology on System z.

The z/OS V2R1 zEnterprise Data Compression capability exploits zEDC Express and has the following minimum requirements:

- z/OS V2R1 with PTFs and the zEDC for z/OS feature
- z/OS V1R13 with PTFs (software decompression support only, no compression)
- z/OS V1R12 with PTFs (software decompression support only, no compression)

Initial support focuses on SMF log data where, using zEDC compression services to compress records before writing to log streams, higher write rates can be achieved. zEDC is also used when reading back the records.

Additional support by DFSMS (BSAM/QSAM) is planned, providing a new type of policy-based compression support for non-VSAM extended format data sets.

Statement of direction: IBM plans for future updates of IBM 31-bit and 64-bit SDK7 for z/OS Java Technology Edition, Version 7 (5655-W43 and 5655-W44) (IBM SDK7 for z/OS Java) to provide exploitation of the following: the zEDC Express feature and Shared Memory Communications - Remote Direct Memory Access (SMC-R), which is utilized by the 10GbE RoCE Express feature.

IBM System z Batch Network Analyzer

The IBM System z Batch Network Analyzer (zBNA) is a free, “as is” tool. It is available to clients, business partners and IBM employees.

zBNA replaces the BWATOOL. It is Windows based, provides graphical and text reports, including Gantt charts, and support for alternate processors.

zBNA can be used to analyze customer provided SMF records, in order to identify jobs and data sets which are candidates for zEDC compression, across a specified time window, typically a batch window. “IBM System z Batch Network Analyzer” on page 83 provides information about how to get the zBNA tool.

zAAP on zIIP capability

This capability enables, under defined circumstances, workloads that are eligible to run on *IBM System z Application Assist Processor (zAAP)* to run on *IBM System z Integrated Information Processor (zIIP)*. A maximum of two zAAPs and two zIIPs per one CP installed now applies. It is valid for zEC12 and zBC12 only, making migrate from zAAPs to zIIPs easier.

Statement of Direction: IBM zEnterprise EC12 and IBM zEnterprise BC12 are planned to be the last System z servers to offer support for zAAP specialty engine processors. IBM intends to continue support for running zAAP workloads on zIIP processors (“zAAP on zIIP”). This configuration is intended to help simplify capacity planning and performance management, while still supporting all the currently eligible workloads.

IBM has released PTF for APAR OA38829 on z/OS V1R12 and V1R13. This PTF allows zAAP-eligible workloads to be dispatched on zIIPs even when there are active zAAPs installed. This PTF is intended to help facilitate migration and testing of zAAP workloads on zIIP processors.

Because z/VM can dispatch both virtual zAAPs and virtual zIIPs on real CPs¹, the z/VM partition does not require any real zIIPs defined to it. However, in general, real zIIPs should be used due to software licensing reasons.

Large page support

On z/OS, memory that is reserved for large page support needs to be defined in the IEASYSxx member of SYS1.PARMLIB. The definition cannot be dynamically changed.

HiperDispatch

On z/OS, a parameter of the IEAOPTxx member of SYS1.PARMLIB controls whether *HiperDispatch* is enabled or disabled for the z/OS image. It can be dynamically changed (without an initial program load (IPL) or any outage).

¹ The z/VM system administrator can use the SET CPUAFFINITY command to influence the dispatching of virtual specialty engines on CPs or real specialty engines.

To use HiperDispatch effectively, adjustment of defined *Workload Manager (WLM)* goals and policies might be required. We suggest that WLM policies and goals are reviewed and updated as necessary. Installations might want to operate with the new policies and HiperDispatch active for a period, turn it off, and use the older WLM policies while analyzing the results of using HiperDispatch. Readjust the new policies, and repeat the cycle as needed. To change WLM policies, turning HiperDispatch off, then on, is not necessary.

A health check is provided to verify whether HiperDispatch is enabled on zEC12 and on zBC12 systems.

Capacity provisioning

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

1. Setting up and customizing z/OS RMF, including the Distributed Data Server (DDS)
2. Setting up the z/OS CIM Server (a z/OS base element with z/OS V1R9 and later)
3. Performing capacity provisioning customization as described in the *z/OS MVS Capacity Provisioning User's Guide*, SC33-8299

Use of the capacity provisioning function requires the following elements:

- ▶ TCP/IP connectivity to observed systems
- ▶ TCP/IP connectivity from the observing system to the HMC of observed systems
- ▶ IBM Resource Measurement Facility™ (RMF) Distributed Data Server must be active
- ▶ Common Information Model (CIM) Server must be active
- ▶ Security and CIM customization
- ▶ Capacity Provisioning Manager customization

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

Customization of the capacity provisioning function is required on the operating system that observes other z/OS systems in one or multiple sysplexes. For a description of the capacity provisioning domain, see the *z/OS MVS Capacity Provisioning User's Guide*, SC33-8299. Also, see *IBM System z10 Enterprise Class Capacity on Demand*, SG24-7504, which describes capacity provisioning in more detail.

Integrated Cryptographic Service Facility (ICSF)

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

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

See the *IBM zEnterprise EC12 Technical Guide*, SG24-8049 or *IBM zEnterprise BC12 Technical Guide*, SG24-8138 for a table listing the ICSF FMIDs and web-deliverable codes for z/OS V1R10 through V2R1. Later FMIDs include the functions of previous ones.

The Cryptographic Support for z/OS V1R12-V1R13 web deliverable (ICSF FMID HCR77A0) is required to use the new functions that are available with the zEC12 and zBC12. Crypto Express4S Toleration is available through maintenance at ICSF FMID HCR7770 or higher.

InfiniBand coupling links

Each system can use, or not use, *InfiniBand coupling links* independently of what other systems are doing, and do so with other link types. zEC12 and zBC12 do not support participating in a Parallel Sysplex with System z9 and earlier systems.

InfiniBand coupling connectivity is only available when other systems also support InfiniBand coupling. We suggest that, when planning to use the InfiniBand coupling technology, the Coupling Facility Configuration Options white paper be consulted. The white paper is available at the following website:

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

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

z/OS V1R13 with program temporary fixes (PTFs) or newer is required to use the latest level (10) of the following two C/C++ compiler options:

- ▶ *ARCHITECTURE*. This option selects the minimum level of machine architecture on which the program runs. Certain features that are provided by the compiler require a minimum architecture level. ARCH(10) uses instructions that are available on the zEC12 and zBC12.
- ▶ *TUNE*. This option allows optimization of the application for a specific machine architecture, within the constraints that are imposed by the ARCHITECTURE option. The TUNE level must not be lower than the setting in the ARCHITECTURE option.

For more information about the ARCHITECTURE and TUNE compiler options, see the *z/OS V1R13.0 XL C/C++ User's Guide*, SC09-4767.

C/C++: A C/C++ program that is compiled with the ARCH(10) or TUNE(10) options runs only on zEC12 and zBC12. Otherwise, an operation exception results. This is a consideration for programs that might need to run on different-level servers during development, test, production, and fallback or disaster recovery.

IBM System z Advanced Workload Analysis Reporter (IBM zAware)

IBM zAware is designed to offer a real-time, continuous learning, diagnostics and monitoring capability that is intended to help pinpoint and resolve potential problems quickly enough to minimize their effects to the businesses.

IBM zAware runs analytics in firmware and intelligently examines the message logs for potential deviations, inconsistencies, or variations from the norm. Many z/OS environments have a large volume of OPERLOG messages, which makes it difficult for operations personnel to easily use and analyze them. IBM zAware provides a simple graphical user interface (GUI) for easy drill-down and identification of message anomalies, which can facilitate faster problem resolution.

IBM zAware is ordered through specific features of zEC12 and zBC12 and requires z/OS V1R13 with IBM zAware exploitation support or newer to collect specific log stream data. It also requires a properly configured LPAR.

See 3.8.1, “IBM System z Advanced Workload Analysis Reporter (IBM zAware)” on page 110 to learn more about zAware and Tivoli Service Management integration.

4.5 Coupling Facility and CFCC considerations

Coupling Facility connectivity to a zEC12 or zBC12 is supported on the z196, z114, System z10, or another zEC12 or zBC12. The logical partition running the Coupling Facility Control Code (CFCC) can be located on any of these supported systems.

Because coupling link connectivity to System z9 and previous systems is not supported, this might affect the introduction of zEC12 or zBC12 into existing installations and require more planning. For more information, see the companion book *IBM zEnterprise EC12 Technical Guide*, SG24-8049 and *IBM zEnterprise BC12 Technical Guide*, SG24-8138.

zEC12 supports CFCC Level 18 and CFCC Level 19. zBC12 supports CFCC Level 19. To support migration from one CFCC level to the next, different levels of CFCC can be run concurrently if the Coupling Facility logical partitions are running on different central processor complexes (CPCs). CF logical partitions running on the same CPC share the CFCC level.

For more information about CFCC code levels, see the Parallel Sysplex website:

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

4.6 Input/output configuration program (IOCP) considerations

All System z servers require a description of their I/O configuration. This description is stored in *Input/Output Configuration Data Set (IOCDS)* files. The *Input/Output Configuration Program (IOCP)* allows creation of the IOCDS file from a source file that is 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 zEC12 is *V3 R3 L0* (IOCP 3.3.0) and for zBC12 is *V4 R1 L0* (IOCP 4.1.0), or later with PTF. See the *Input/Output Configuration Program User's Guide*, SB10-7037, for details.

4.7 IBM Device Support Facilities (ICKDSF) considerations

The *ICKDSF* Release 17 device support facility program is required on all systems that share disk subsystems with a zEC12 and zBC12 processor.

ICKDSF supports a modified format of the CPU information field, which contains a 2-digit logical partition identifier. ICKDSF uses the CPU information field instead of CCW reserve/release for concurrent media maintenance. It prevents multiple systems from running ICKDSF on the same volume, and at the same time allows user applications to run while ICKDSF is processing. To prevent any possible data corruption, ICKDSF must determine all sharing systems that can potentially run ICKDSF. Therefore, this support is required for the zEC12 and zBC12.

ICKDSF Release 17: 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.



A smarter infrastructure

We started this book by reviewing current IT infrastructures and the most pressing problems that they present. Also noted is the need to better align IT with business. In this chapter, we further explore how the zEC12 and zBC12 can help establish a smarter infrastructure, better able to rapidly respond to business needs: flexible, resilient, and secure.

This chapter describes the following topics:

- ▶ Integrated hybrid infrastructures
- ▶ Benefiting from a smart infrastructure
- ▶ Cloud computing

5.1 Integrated hybrid infrastructures

IBM System z has long been an integrated diverse platform, with specialized hardware, and dedicated computing capabilities. Recall, for instance, in the mid-1980s, the IBM 3090, and its vector facility (occupying a separate frame). Or note that the cryptographic processors and all the I/O cards, which are specialized dedicated hardware, run non System z code on non System z processors. This design allows System z processor units (PU) to concentrate on application's computational tasks, while specialized processors take care of the other tasks, thus providing more value through lower cost of computing.

All of these specialized hardware components have been seamlessly integrated within the mainframe for over a decade. In creating a hybrid infrastructure, IBM has extended, wherever possible, the same integration and simplification philosophy to other servers outside of the mainframe itself, creating a logical environment of shared resources.

It would seem that increased flexibility inevitably leads to increased complexity. However, it does not need to be that way. IT operational simplification greatly benefits from the zEC12 and zBC12 intrinsic autonomic characteristics and the ability to consolidate and reduce the number of system images. There are also benefits from management best practices and products that were developed and are available for the mainframe, in particular for the z/OS environment.

5.1.1 A cornerstone of a smart IT infrastructure

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

- ▶ Access management, to authenticate and authorize who can access specific business services and associated IT resources.
- ▶ Use management, to drive maximum utilization of the system. Unlike other classes of servers, System z systems are designed to run at 100% utilization all the time, which is based on the varied demands of its users. Clustering can be used to achieve 99,999% availability *at the application level*.
- ▶ Just-in-time capacity, to deliver more processing power and capacity when needed.
- ▶ Virtualization security, to enable clients to allocate resources on demand without fear of security risks.
- ▶ Enterprise-wide operational management and automation, leading to a more autonomic environment.

zEC12 and zBC12 are the result of the IBM sustained and continuous investment and development policies. Commitment to IBM systems design means that zEC12 and zBC12 bring all this innovation while helping clients to optimize their current investment in the mainframe and improve the economics of IT.

The zEC12 and zBC12 continue the evolution of the mainframe, building upon the z/Architecture definitions. IBM mainframes traditionally provide an advanced combination of reliability, availability, security, scalability, and virtualization. The zEC12 and zBC12 are designed to extend these capabilities into heterogeneous resources and are optimized for today's business needs.

The zEC12 and zBC12 are emerging as a choice platform for mobile computing, which can improve the integration of people, processes, and technology to help run the business more cost effectively while also supporting business growth and innovation.

A number of enterprises are reaching the limits of available physical space and electrical power at their data centers. The extreme virtualization capabilities of the zEC12 and zBC12 enable the creation of dense and simplified infrastructures that are highly secure and can lower operational costs. They are, therefore, the most powerful tools available to reduce cost, energy, and complexity in enterprise data centers.

Further simplification is possible by using the HiperSockets¹ and z/VM virtual switch functions. These functions can be used, at no additional cost, to replace physical routers, switches, and their cables, while eliminating security exposures and simplifying configuration and administration tasks. In actual simplification cases, measured at customer sites, cables have been reduced by 97%.

Further ahead, we discuss how these capabilities play a key role in cloud computing. Summing up the following characteristics leads us to an interesting result:

Capacity range and flexibility offer these benefits:

- ▶ A processor equally able to handle compute-intensive and I/O-intensive workloads
- ▶ Specialty engines for improved price/performance
- ▶ Extreme virtualization
- ▶ Secure access to data (and the network)
- ▶ Additional platforms and the Unified Resource Manager

All of this equals a flexible infrastructure that is based on an integrated heterogeneous environment, on which a wide range of workloads can be seamlessly deployed and managed.

The zEC12 and zBC12 are the platforms of choice for the integration of the new generations of applications with existing applications and data. They truly are cornerstones of a smart IT infrastructure.

5.2 Benefiting from a smart infrastructure

When distilled to a single central theme, embracing a heterogeneous infrastructure and creating an ensemble has the single objective of providing an optimized infrastructure. In this smart infrastructure, multiple workloads can be deployed across heterogeneous environments and managed under a common umbrella. The computing resources of different hardware platforms are managed as a single system with these characteristics

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

Adopting this smarter infrastructure and extracting benefit from it can be done gradually, and does not require disruptive moves such as those popularized under the names *lift and shift* and *rip and replace*. In the IBM solution, protection of the client's investment in all IT aspects is maximized.

¹ For a description of HiperSockets, see "HiperSockets" on page 91. The z/VM virtual switch is a z/VM system function that uses memory to emulate network switching hardware.

The clients' IT infrastructures are all unique and different in areas such as virtualization, monitoring, and automation. A single approach does not fit all, and there are multiple points of entry to the adoption of an integrated, heterogeneous, virtualized infrastructure.

Figure 5-1 shows a simplified view of a progressive adoption process, where building blocks are added to achieve an enterprise class infrastructure for heterogeneous workload deployment. Again, one can notice a resemblance with cloud computing.

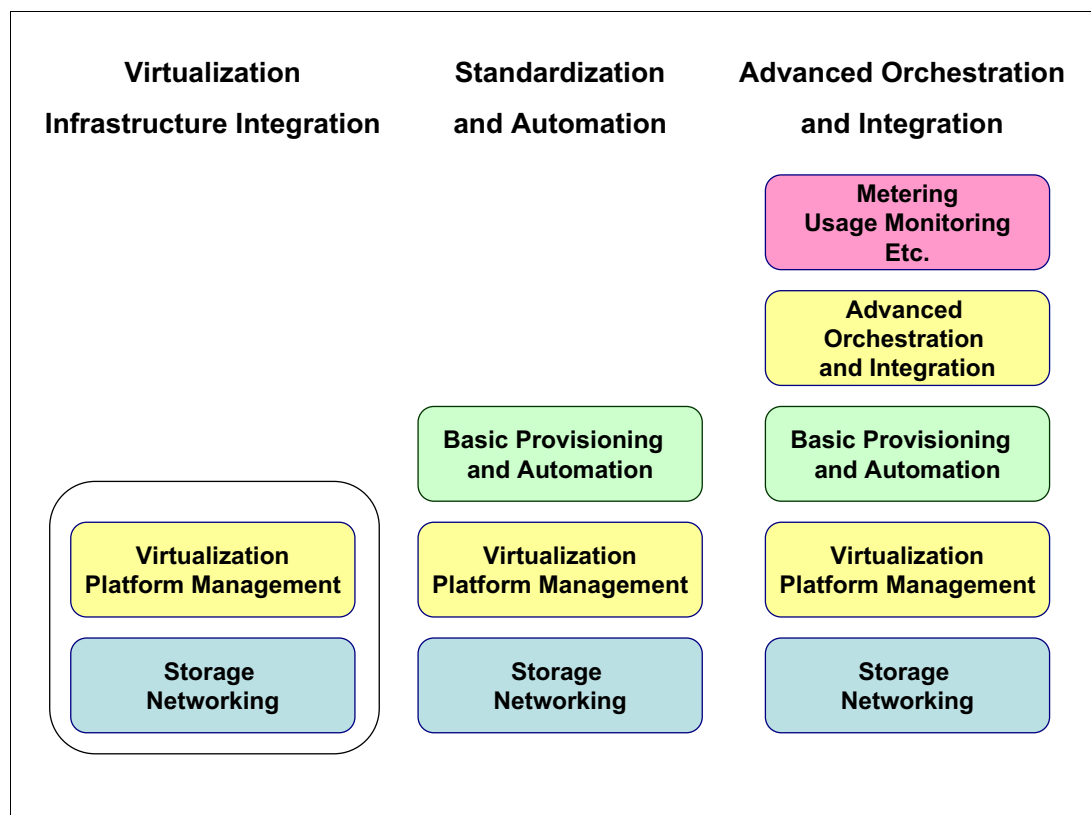


Figure 5-1 Enterprise class heterogeneous platform deployment building blocks

Workloads

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

We now define the terms *application* and *workload*:

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

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

Workloads require computing resources, which can be classified as the following functions:

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

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

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

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

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

- ▶ Response time: The ability to return results in a specified time.
- ▶ Throughput: The amount of data that can be processed in a specified time interval

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

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

- ▶ I/O co-processors, which are designed to offload the main processors from the I/O-related tasks
- ▶ Specialized cryptographic co-processors, which are designed to offload the computational intensive mathematical functions required to encrypt and decrypt data
- ▶ Specialized co-processors for accelerating compressing and decompressing of data
- ▶ Specialized units inside the processor core, such as the BCD Arithmetic Logic Unit or vector units
- ▶ Special purpose units, such as the IBM DB2 Analytics Accelerator, for accelerating complex queries such as the ones found in Data Warehousing workloads

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

- ▶ Custom code
- ▶ Commercial off the shelf (COTS)

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

- ▶ Platform availability (development costs always must consider this aspect)
- ▶ Platform fit-for-the-purpose

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

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

- ▶ In *banking*, there are components across retail and wholesale banking that employ several architectures to run, but the core of most banking applications relies on System z and z/OS.
- ▶ *Insurance* typically maintains claims processing on System z but reaches out to the internet for interaction with consumers, using Linux, UNIX, Power, and x86. Fraud detection can benefit from exploiting, the vast amounts of data managed by System z, using System z based computing.
- ▶ The *public sector* is relying more on the web-based capability to reach out to citizens and improve the rate of return for taxes, accurate payment of social benefits, election process, and even census-based reporting. Here, too, fraud detection and tax evasion can benefit from System z capabilities.
- ▶ The *retail industry* can effectively use a heterogeneous infrastructure. For instance, by benefiting from System z large I/O capacity to implement large databases, using Business Intelligence to characterize their clients.
- ▶ *Airline reservation systems* are one example of extreme online transaction processing. zEC12 provides the z/TPF operating system and application environment specifically for this situation. Another example of z/TPF utilization is in banking with credit card and ATM processing.

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

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

Attributes of workload components

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

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

- ▶ Transaction processing and database (OLTP):
 - High transaction rates
 - High quality of service
 - Peak workloads
 - Resiliency and security
- ▶ Analytics and high performance:
 - Compute or I/O intensive
 - High memory bandwidth
 - Floating point and vector processing (SIMD²)
 - Scale-out capable (horizontal scalability)

² Single Instruction Multiple Data (SIMD): same instruction that is run on a vector of data

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

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

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

In addition, today's applications can rarely be classified as only one of the aforementioned types. See Figure 5-2. In most cases, an application consists of components with varying requirements and workload profiles. Thus, proper platform choice is key in obtaining the wanted results (SLA).

For example, a batch job requires fast movement of data through the processor, and does not, of itself, use any multi-threading capabilities that a platform might have. Thus, the higher the CPU speed, the faster the job is processed. Conversely, a highly threaded application (web services, for example) performs better on a multi-threaded capable platform.

With zEC12 and zBC12, and in addition to the advanced workload management capabilities of z/OS, the zEnterprise Unified Resource Manager also offers workload management capabilities of zBX environments:

- ▶ For virtual servers on POWER7 blades, through the POWER/VM hypervisor.
- ▶ For System x in the zBX, by using a function in the KVM hypervisor known as *cgroups*. This allows monitoring the availability of workload resources to satisfy a defined workload service level policy. The Unified Resource Manager will exploit cgroups by assigning cgroup to virtual servers and dynamically managing the CPU share of a virtual server based on policy goals.

Figure 5-2 shows an image of workload characterization.

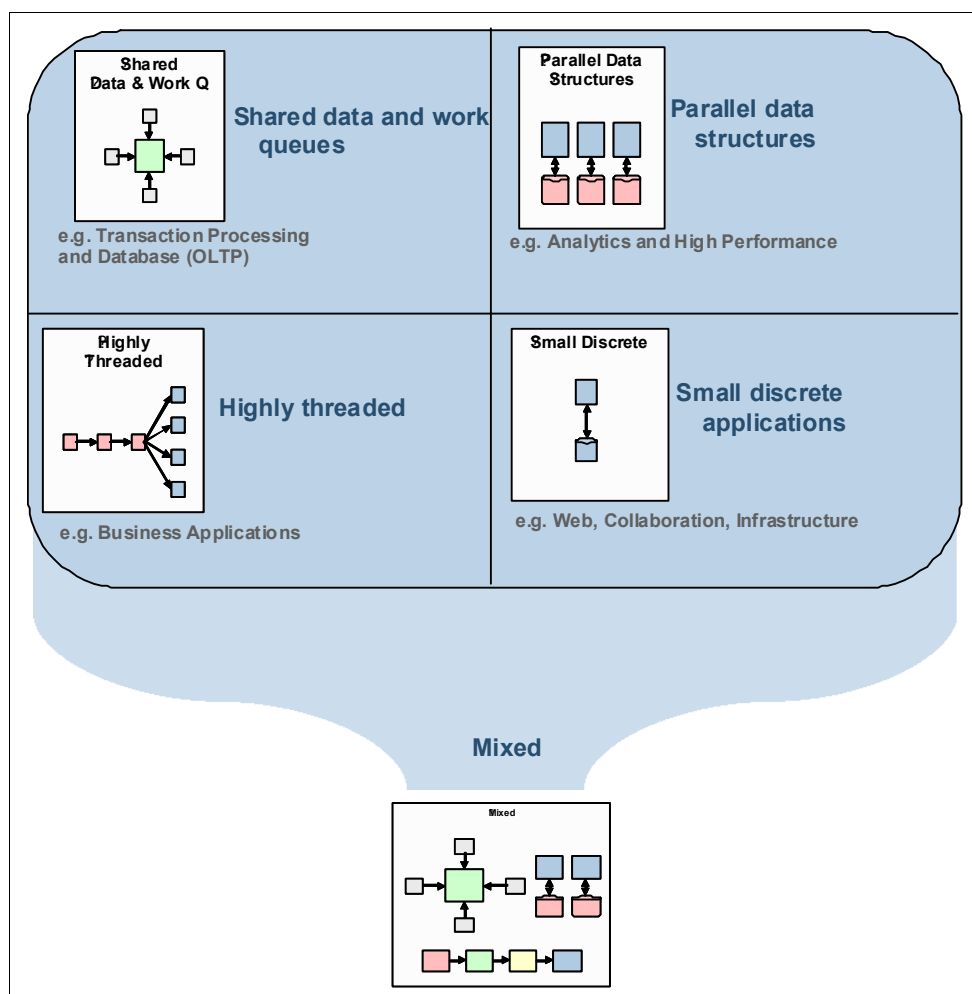


Figure 5-2 Workload characterization

5.3 Cloud computing

“Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.”³ A cloud can be public, private, or a hybrid of both. With cloud computing, the application can be running on a server anywhere in the world. That flexibility is why it is changing the way companies provide services to their clients and suppliers.

³ US National Institute of Standards and Technology. Read more at this website:
<http://csrc.nist.gov/groups/SNS/cloud-computing/index.html>

Figure 5-3 shows the IBM Cloud Computing Reference Architecture (CC RA⁴), which defines the fundamental architectural elements which constitute a cloud computing environment. It is required that all of these infrastructure components be managed from a single, central Common Cloud Management Platform with the ability to place instances of each cloud service on the corresponding infrastructure. This requirement perfectly fits the zEC12-based or zBC12-based heterogeneous infrastructure, with its end-to-end management capabilities for flexible delivery of high-value services.

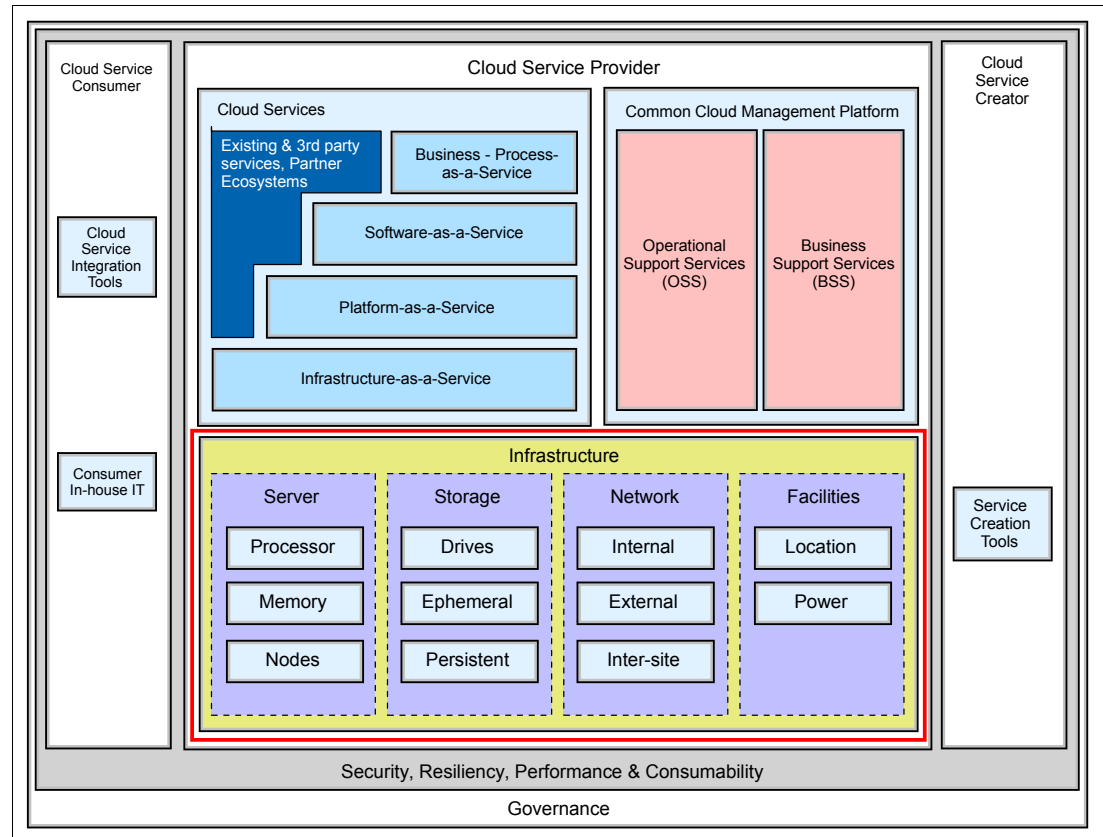


Figure 5-3 The IBM Cloud Computing Reference Architecture infrastructure details

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

⁴ <https://www.opengroup.org/cloudcomputing/uploads/40/23840/CCRA.IBMSubmission.02282011.doc>

Deploying a cloud infrastructure is not a simple process, but there is a defined path that can be followed. Figure 5-4 depicts the path from Standard Managed Services to Cloud.

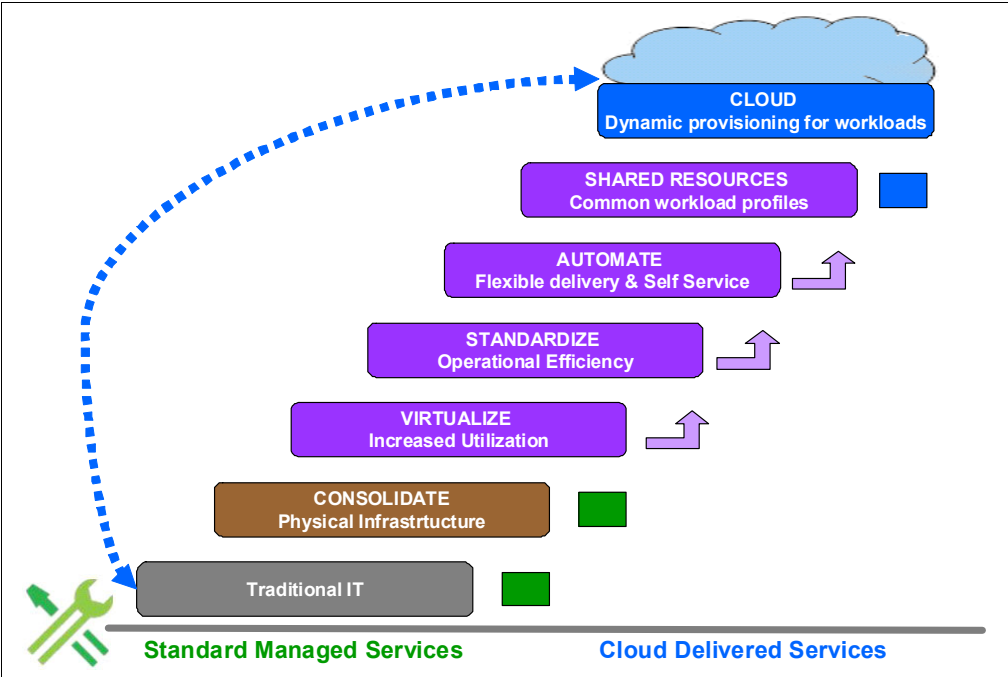


Figure 5-4 Cloud delivered services

Because zEC12 and zBC12, with their “shared everything infrastructure,” offer a fully virtualized system, it becomes easier to integrate a cloud computing deployment as part of the existing IT optimization strategy and roadmap. Table 5-1 summarizes potential benefits that are provided through cloud computing.

Table 5-1 Benefits with cloud computing

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

Cloud computing on zEC12 and zBC12 builds on the industry's leading virtualization technology that uses virtualization, standardization, and automation to free operational budget for new investments. This technology also allows you to optimize new investments for direct business benefits. zEC12 and zBC12 provide the following features:

- ▶ A highly scalable heterogeneous pool of virtualized resources that are managed in a single system
- ▶ On demand activation, allocation, prioritization, and retiring of resources, and automation of service delivery
- ▶ Maximizing utilization of resources for improved ROI and lower cost of service delivery
- ▶ Increased levels of security, resiliency, and manageability to create a cloud environment that is enterprise ready

Building upon all the previous resource management capabilities and core functionality, application-supporting middleware, such as transactional servers, are adding function to enable existing applications to become cloud accessible, without the need to rewrite and replace them. True to IBM commitment, and similarly to web enablement, applications can become naturally immersed in the cloud, while preserving the client's investment.

CICS Transaction Server support for Cloud computing: CICS Transaction Server for z/OS V5.1 provides new application, platform, and policy capabilities that can help clients build private clouds from new and existing CICS applications. This capability is intended to assist CICS clients to deploy new and updated CICS applications faster, more easily, and with greater levels of confidence.

As depicted in Figure 5-5, IBM zEnterprise enables the following attributes, being the IT industry's first multi-architecture cloud solution:

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

Figure 5-5 shows a zEC12 for IT optimization, consolidation, and cloud computing.

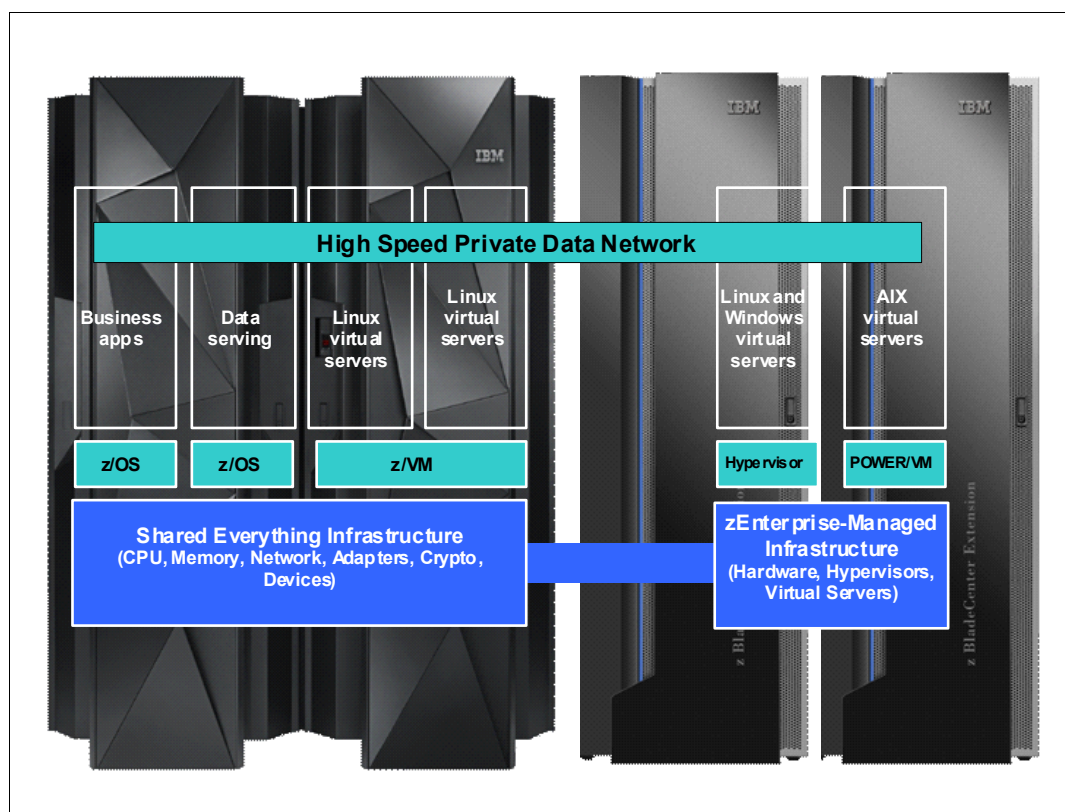


Figure 5-5 zEC12 for IT optimization, consolidation, and cloud computing

Quality of service improvements

In an ensemble, the qualities for which System z is renowned are extended to the non-System z components of the ensemble, providing support for mission-critical workloads running on the ensemble's heterogeneous infrastructure. Compared to other heterogeneous infrastructures, the ensemble provides the following benefits:

- ▶ Potentially lower the cost of enterprise computing by implementing a single management and policy framework across web serving, transaction, database, and servers
- ▶ Simplified operations through integration of multiplatform management capabilities through extended functionality in the well-known mainframe HMC
- ▶ Improved infrastructure reliability by extending the mainframe systems management and service to the zBX environment
- ▶ Improved service through dynamic resource management of the mainframe to all devices within a multitier architecture
- ▶ Simplified and improved infrastructure management through monitoring and management of a heterogeneous solution as a single, logical virtualized solution
- ▶ Better alignment of IT with business objectives by managing the platform's resources in accordance with specified business-service-level objectives
- ▶ Improved infrastructure manageability through management of virtual servers as part of the overall deployed business workload
- ▶ Dramatically simplified infrastructure, improved application performance, and simplified management by using IEDN, the secure and managed Layer 2 network which connects the zBX blades with the CPC

Cloud computing is one of the key ways to address the challenges of today and build an IBM Smarter Planet®. As more companies embrace cloud computing, zEC12 becomes more obviously the perfect platform for delivering large-scale *software as a service* (SaaS) application software services.⁵

⁵ More information about cloud on an IBM System z can be found in *Deploying a Cloud on IBM System z*, REDP-4711. Available at: <http://www.redbooks.ibm.com/abstracts/redp4711.html>



A

Software licensing

This appendix briefly describes the software licensing options that are available for the zEC12 and zBC12. Basic information about software licensing for the zBX blade environments is also covered.

Software licensing considerations

The zEC12's and zBC12's IBM software portfolio includes operating system software¹ (that is, z/OS, z/VM, z/VSE, and z/TPF) and middleware that runs on these operating systems. The portfolio also includes middleware for Linux on System z environments.

zBX software products are covered by the International Program License Agreement (IPLA) and additional agreements, such as the IBM International Passport Advantage® Agreement, similarly to other AIX, Linux on System x, and Windows environments. POWER/VM Enterprise Edition licenses must be ordered for IBM POWER7 blades.

For the zEC12 and zBC12, two metric groups for software licensing are available from IBM, depending on the software product:

- ▶ Monthly license charge (MLC)
- ▶ International Program License Agreement (IPLA)

MLC pricing metrics have a recurring charge that applies each month. In addition to the right to use the product, the charge includes access to IBM product support during the support period. MLC metrics, in turn, include various offerings.

IPLA metrics have a single, up-front charge for an entitlement to use the product. An optional and separate annual charge, called *subscription and support*, entitles clients to access IBM product support during the support period. With this option, you can also receive future releases and versions at no additional charge.

For more details about software licensing, consult the following websites and web pages:

- ▶ Learn about Software licensing:
http://www-01.ibm.com/software/lotus/passportadvantage/about_software_licensing.html
- ▶ Base license agreements:
<http://www-03.ibm.com/software/sla/sladb.nsf/viewbla>
- ▶ IBM System z Software Pricing Reference Guide:
<http://www.ibm.com/systems/z/resources/swprice/reference/index.html>
- ▶ IBM System z Software Pricing:
<http://www.ibm.com/systems/z/resources/swprice/index.html>
- ▶ The IBM International Passport Advantage Agreement can be downloaded from the “Learn about Software licensing” web page:
ftp://ftp.software.ibm.com/software/passportadvantage/PA_Agreements/PA_Agreement_International_English.pdf

The remainder of this appendix describes the software licensing options that are available for the zEC12 and zBC12.

Monthly License Charge pricing metrics

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

¹ Linux on System z distributions are not IBM products.

Charge models

There are various *Workload License Charges (WLC)* pricing structures that support two charge models:

- ▶ Variable charges (several pricing metrics):

Variable charges apply to products such as z/OS, z/VSE, z/TPF, DB2, IMS, CICS, and WebSphere MQ. There are several pricing metrics that employ the following charge types:

- Full-capacity license charges:

The total number of MSUs of the central processor complex (CPC) is used for charging. Full-capacity licensing is applicable when the CPC of the client is not eligible for subcapacity.

- Subcapacity license charges:

Software charges that are based on the utilization of the logical partitions where the product is running.

- ▶ Flat charges:

Software products that are licensed under flat charges are not eligible for subcapacity pricing. There is a single charge for each CPC on the zEC12 or zBC12.

Subcapacity license charges

For eligible programs, subcapacity licensing allows software charges that are based on the measured utilization by logical partitions instead of the total number of MSUs of the CPC. Subcapacity licensing removes the dependency between the software charges and CPC (hardware) installed capacity.

The subcapacity licensed products are charged monthly based on the highest observed 4-hour rolling average utilization of the logical partitions in which the product runs (except for products that are licensed by using the select application license charge (SALC) pricing metric). This type of charge requires measuring the utilization and reporting it to IBM.

The 4-hour rolling average utilization of the logical partition can be limited by a defined capacity value on the image profile of the partition. This value activates the soft capping function of the Processor Resource/Systems Manager (PR/SM), limiting the 4-hour rolling average partition utilization to the defined capacity value. Soft capping controls the maximum 4-hour rolling average usage (the last 4-hour average value at every 5-minute interval), but does not control the maximum instantaneous partition use.

Also available is a logical partition (LPAR) group capacity limit, which sets soft capping by PR/SM for a group of logical partitions running z/OS.

Even by using the soft capping option, the use of the partition can reach up to its maximum share based on the number of logical processors and weights in the image profile. Only the 4-hour rolling average utilization is tracked, allowing utilization peaks above the defined capacity value.

Some pricing metrics apply to stand-alone System z servers. Others apply to the aggregation of multiple System z server workloads within the same Parallel Sysplex.

For more information about WLC and details about how to combine logical partition utilization, see the publication *z/OS Planning for Workload License Charges*, SA22-7506, available from the following website:

http://www.ibm.com/systems/z/os/zos/bkserv/find_books.html

IBM zEnterprise EC12

Metrics that are applicable to a stand-alone zEC12 include the following charges:

- ▶ Advanced Workload License Charges (AWLC)
- ▶ System z New Application License Charges (zNALC)
- ▶ Parallel Sysplex License Charges (PSLC)

Metrics that are applicable to a zEC12 in an actively coupled Parallel Sysplex include the following charges:

- ▶ *Advanced Workload License Charges (AWLC)*, when all nodes are zEnterprise CPCs (zEC12, z196, or z114)
- ▶ *Variable Workload License Charges (VWLC)*, allowed only under the AWLC Transition Charges for Sysplexes when not all of the nodes are zEnterprise CPCs
- ▶ *System z New Application License Charges (zNALC)*
- ▶ *Parallel Sysplex License Charges (PSLC)*

IBM zEnterprise BC12

Metrics that are applicable to a stand-alone zBC12 include the following charges:

- ▶ Advanced Entry Workload License Charges (AEWLC)
- ▶ System z New Application License Charges (zNALC)

Metrics that are applicable to a zBC12 in an actively coupled Parallel Sysplex include the following charges:

- ▶ *Advanced Workload License Charges (AWLC)*, when all nodes are zEnterprise CPCs (zEC12, zBC12, z196, or z114)
- ▶ *Variable Workload License Charges (VWLC)*, allowed only under the AWLC Transition Charges for Sysplexes when not all of the nodes are zEnterprise CPCs
- ▶ *System z New Application License Charges (zNALC)*
- ▶ *Parallel Sysplex License Charges (PSLC)*

Advanced Workload License Charges

Advanced Workload License Charges (AWLC) were introduced with the IBM zEnterprise 196. They use the measuring and reporting mechanisms, as well as the existing MSU tiers, from VWLCs, although the prices for each tier were lowered.

Similar to Workload License Charges, AWLCs can be implemented in full-capacity or subcapacity mode. The AWLC applies to z/OS and z/TPF and their associated middleware products such as DB2, IMS, CICS, and IBM WebSphere MQ, and IBM Lotus® IBM Domino®.

With zEC12, there are three new Technology Transition Offerings available which extend the software price and performance of the AWLC pricing metric:

- ▶ Technology Update Pricing for AWLC applicable for clients which run on a stand-alone zEC12 server or in an aggregated Parallel Sysplex consisting exclusively of zEC12 servers.
- ▶ AWLC Transition Charges for Sysplexes (TC2) when two or more servers exist in an actively coupled Parallel Sysplex consisting of one or more zEC12 servers with one or more z196 or z114 servers.

- ▶ AWLC Transition Charges for Sysplexes (TC1) for Parallel Sysplexes consisting of one or more zEnterprise CPCs with one or more z10 EC or z10 BC servers. In this case, all servers remain on VWLC pricing, but receive a reduction to VWLC pricing across the aggregated Parallel Sysplex.

For more information, see the AWLC website:

<http://www.ibm.com/systems/z/resources/swprice/mlc/awlc.html>

Advanced Entry Workload License Charges

Advanced Entry Workload License Charges (AEWLC) were introduced with the IBM zEnterprise 114. They use the measuring and reporting mechanisms, as well as the existing MSU tiers, from Entry Workload Licence Charges (EWLC) pricing metric and Midrange Workload Licence Charges (MWLC) pricing metric.

AEWLC also offers improved price performance, as compared with EWLC and MWLC, for most clients.

Similar to Workload License Charges, AEWLCs can be implemented in full-capacity or subcapacity mode. The AEWLC applies to z/OS, z/TPF and z/VSE in z/Architecture (64-bit) mode and their associated middleware products such as DB2, IMS, CICS, and IBM WebSphere MQ, and Lotus® Domino®, when running on a zBC12.

For more information, see the AEWLC website:

<http://www.ibm.com/systems/z/resources/swprice/mlc/aewlc.html>

System z New Application License Charges

System z New Application License Charges (zNALC) offer a reduced price for the z/OS operating system on logical partitions which run a qualified *new workload* application. An example includes Java language business applications that run under the WebSphere Application Server for z/OS, Domino, SAP, PeopleSoft, and Siebel.

z/OS with zNALCs provides a strategic pricing model that is 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 more information, see the zNALC website:

<http://www.ibm.com/systems/z/resources/swprice/mlc/znalc.html>

Select Application License Charges

Select Application License Charges (SALC) apply only to IBM WebSphere MQ for System z. This type of charge allows a WLC client to license WebSphere MQ under the product use rather than the subcapacity pricing that is provided under WLC.

WebSphere MQ is typically a low-usage product that runs pervasively throughout the environment. Clients who run WebSphere MQ at a low usage can benefit from SALC. Alternatively, you can still choose to license WebSphere MQ under the same metric as the z/OS software stack.

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

1. Determine the highest daily usage of a program family, which is the highest of 24 hourly measurements recorded each day. *Program* refers to all active versions of WebSphere MQ.
2. Determine the monthly usage of a program family, which is the fourth highest daily measurement that is recorded for a month.
3. Use the highest monthly usage that is determined for the next billing period.

For more information about SALC, see the Other Monthly License Charge Metrics website:

<http://www.ibm.com/systems/z/resources/swprice/mlc/other.html>

Midrange Workload License Charges

Midrange Workload License Charges (MWLC) apply to z/VSE V4 and above when running on zEC12, z196, z114, System z10, and z9 servers. The exceptions are: The z10 BC and z9 BC servers at the capacity setting A01, to which zELC applies; and zBC12 where MWLC is *not* available.

Similar to workload license charges, MWLCs can be implemented in full-capacity or subcapacity mode. An MWLC applies to z/VSE V4 and above, and several IBM middleware products for z/VSE. All other z/VSE programs continue to be priced as before.

The z/VSE pricing metric is independent of the pricing metric for other systems (for instance, z/OS) that might be running on the same server. When z/VSE is running as a guest of z/VM, z/VM V5R4 or later is required.

To report usage, the subcapacity report tool is used. One subcapacity reporting tool (SCRT) report per server is required.

For more information, see the MWLC website:

<http://www.ibm.com/systems/z/resources/swprice/mlc/mwlc.html>

Parallel Sysplex License Charges

Parallel Sysplex License Charges (PSLC) apply to a large range of mainframe servers. The list can be obtained from this website:

<http://www.ibm.com/systems/z/resources/swprice/reference/exhibits/hardware.html>

Although it can be applied to stand-alone CPCs, the metric provides aggregation benefits only when applied to a group of CPCs in an actively coupled Parallel Sysplex cluster according to IBM terms and conditions.

Aggregation allows charging a product that is based on the total MSU value of the systems where the product runs (as opposed to all the systems in the cluster). In an uncoupled environment, software charges are based on the MSU capacity of the machine.

For more information, see the PSLC web page:

<http://www.ibm.com/systems/z/resources/swprice/mlc/pslc.html>

System z International Program License Agreement

For System z systems, the following types of products are generally in the International Program License Agreement (IPLA) category:

- ▶ Data management tools
- ▶ DB2 for z/OS VUE
- ▶ CICS TS VUE V5 and CICS tools
- ▶ IMS DB VUE V12 and IMS tools
- ▶ Application development tools
- ▶ Certain WebSphere for z/OS products
- ▶ Linux middleware products
- ▶ z/VM V5 and V6

Generally, three pricing metrics apply to IPLA products for zEC12 and System z:

- ▶ Value Unit (VU):

Value Unit pricing, which applies to the IPLA products that run on z/OS. Value Unit pricing is typically based on the number of MSUs and allows for lower cost of incremental growth. Examples of eligible products are IMS Tools, CICS Tools, DB2 Tools, application development tools, and WebSphere products for z/OS.

- ▶ Engine-Based Value Unit (EBVU):

Engine-Based Value Unit pricing enables a lower cost of incremental growth with more engine-based licenses purchased. Examples of eligible products include z/VM V5 and V6, and certain z/VM middleware, which are priced based on the number of engines.

- ▶ Processor Value Unit (PVU)²:

Processor Value Units are determined from the number of engines, under the Passport Advantage terms and conditions. Most Linux middleware is also priced based on the number of engines.

For more information, see the System z IPLA website:

<http://www.ibm.com/systems/z/resources/swprice/zipla/index.html>

zBX licensed software

The software licensing for the zBX select System x and POWER7 blades and DataPower XI50z follows the same rules as licensing for blades installed outside of zBX.

Note that POWER/VM Enterprise Edition *must be* licensed for POWER7 blades at the time of ordering the blades.

The hypervisor for the select System x blades for zBX is provided as part of the zEnterprise Unified Resource Manager.

zEnterprise Unified Resource Manager

The zEnterprise Unified Resource Manager is available through zEC12 and zBC12 *hardware* features, either ordered with the system or later. No separate software licensing is required.

² See http://www.ibm.com/software/lotus/passportadvantage/pvu_licensing_for_customers.html for PVUs per core ratings.



Channel options

Table B-1 lists the attributes of the channel options that are supported on zBC12 and zEC12 systems, the required cable type, the maximum unrepeated distance, and the bit rate.

For all optical links, the connector type is LC Duplex, except the 12xIFB connection is established with an MPO connector. The electrical Ethernet cable for the Open Systems Adapter (OSA) connectivity is connected through an RJ45 jack.

Statement of direction:

- ▶ The zBC12 and the zEC12 are planned to be the last IBM System z[®] server to support ISC-3 Links. Enterprises should continue migrating from ISC-3 features to InfiniBand Coupling Links.
- ▶ The zBC12 and the zEC12 are planned to be the last IBM System z[®] server to support Ethernet half-duplex operation and 10-Mbps link data rate on 1000BASE-T Ethernet features. Any future 1000BASE-T Ethernet feature will support full-duplex operation and auto-negotiation to 100 or 1000 Mbps exclusively.
- ▶ The zBC12 and zEC12 are planned to be the last IBM System z[®] server to support OSA-Express3 family of features. Enterprises should continue migrating from the OSA-Express3 features to the OSA-Express5S features.
- ▶ The zBC12 and zEC12 are planned to be the last IBM System z[®] server to support FICON Express4 features. Enterprises should continue migrating from the FICON Express4 features to the FICON Express8S features.

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

Table B-1 zBC12 and zEC12 feature support

I/O feature	Feature codes	Bit rate in Gbps (or stated)	Cable type	Maximum unrepeat distance ^a	Ordering information	Remark
Fibre Connection (FICON)						
FICON Express8S 10KM LX	0409	2, 4, or 8	SM 9 μm	10 km	New build	
FICON Express8 10KM LX	3325				Carry forward	
FICON Express4 10KM LX	3321	1, 2, or 4			Carry forward	
FICON Express8S SX	0410	2, 4, or 8	OM1,OM2, OM3	See Table B-2 on page 169.	New build	
FICON Express8 SX	3326				Carry forward	
FICON Express4 SX	3322	1, 2, or 4			Carry forward	
FICON Express4-2C SX	3318				Carry forward	zBC12 only
Open Systems Adapter (OSA)						
OSA-Express5S 10 GbE LR	0415	10	SM 9 μm	10 km	New build	
OSA-Express4S 10 GbE LR	0406				Carry forward	
OSA-Express3 10 GbE LR	3370				Carry forward	
OSA-Express5S 10 GbE SR	0416	10	MM 62.5 μm	33 m (200)	New build	
OSA-Express4S 10 GbE SR	0407		MM 50 μm	300 m (2000) 82 m (500)	Carry forward	
OSA-Express3 10 GbE SR	3371				Carry forward	
OSA-Express5S GbE LX	0413	1	SM 9 μm	5 km	New build	
OSA-Express4S GbE LX	0404		MCP 50 μm	550 m (500)	Carry forward	
OSA-Express3 GbE LX	3362				Carry forward	
OSA-Express5S GbE SX	0414	1	MM 62.5 μm	220 m (166) 275 m (200)	New build	
OSA-Express4S GbE SX	0405		MM 50 μm	550 m (500)	Carry forward	
OSA-Express3 GbE SX	3363				Carry forward	
OSA-Express3-2P GbE SX	3373				Carry forward	zBC12 only
OSA-Express5S 1000BASE-T	0417	10, 100, or 1000 Mbps	Cat 5, Cat 6 copper		New build	
OSA-Express4S 1000BASE-T	0408				Carry forward	not on zBC12
OSA-Express3 1000BASE-T	3367				Carry forward	
OSA-Express3-2P 1000BASE-T	3369				Carry forward	zBC12 only
10GbE RoCE Express	0411	10	OM3 50 μm	300 m	New build	
Coupling links						
HCA3-O (12x IFB)	0171	6 GBps	OM3	150 m	New build	
HCA3-O LR (1x IFB)	0170	2.5 or 5 Gbps	SM 9 μm	10 km	New build	
HCA2-O (12x IFB)	0163	6 GBps	OM3	150 m	Carry forward	

I/O feature	Feature codes	Bit rate in Gbps (or stated)	Cable type	Maximum unrepeated distance ^a	Ordering information	Remark
HCA2-O LR (1x IFB)	0168	2.5 or 5 Gbps	SM 9 μm	10 km	Carry forward	
IC	N/A		N/A	N/A	N/A	
ISC-3 (peer mode)	0217	2	SM 9 μm	10 km	Carry forward	
ISC-3 (RPQ 8P2197 Peer mode at 1 Gbps) ^b	0218 0219	1	SM 9 μm	20 km	Carry forward	
Special purpose features						
Crypto Express4S	0865	N/A	N/A	N/A	New build	
Crypto Express3	0864	N/A	N/A	N/A	Carry forward	
Flash Express	0402	N/A	N/A	N/A	New build	
zEDC Express	0420	N/A	N/A	N/A	New build	

a. Minimum fiber bandwidth distance product in MHz·km for multimode fiber optic links are included in parentheses, where applicable.

b. RPQ 8P2197 enables the ordering of a daughter card supporting 20 km unrepeated distance for 1 Gbps peer mode. Request for price quotation (RPQ) 8P2262 is a requirement for that option, and other than the normal mode, the channel increment is two (that is, both ports (FC 0219) at the card must be activated).

Table B-2 shows the maximum unrepeated distance for the FICON SX features.

Table B-2 Maximum unrepeated distance for FICON SX features

Cable type and bit rate	1 Gbps	2 Gbps	4 Gbps	8 Gbps
OM1 (62.5 μm at 200 MHz·km)	300 meters	150 meters	70 meters	21 meters
	984 feet	492 feet	230 feet	69 feet
OM2 (50 μm at 500 MHz·km)	500 meters	300 meters	150 meters	50 meters
	1640 feet	984 feet	492 feet	164 feet
OM3 (50 μm at 2000 MHz·km)	860 meters	500 meters	380 meters	150 meters
	2822 feet	1640 feet	1247 feet	492 feet

Related publications

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

IBM Redbooks publications

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only:

- ▶ *IBM zEnterprise EC12 Technical Guide*, SG24-8049
- ▶ *IBM zEnterprise BC12 Technical Guide*, SG24-8138
- ▶ *IBM System z Connectivity Handbook*, SG24-5444
- ▶ *Building an Ensemble Using IBM zEnterprise Unified Resource Manager*, SG24-7921
- ▶ *IBM zEnterprise EC12 Configuration Setup*, SG24-8034
- ▶ *IBM zAware Concept Guide*, SG24-8070
- ▶ *IBM System z10 Enterprise Class Capacity On Demand*, SG24-7504

You can search for, view, download, or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Other publications

These publications are also relevant as further information sources:

- ▶ *z/Architecture Principles of Operation*, SA22-7832
- ▶ *System z Functional Matrix*, ZSW0-1335
- ▶ *Planning for Fiber Optic Links*, GA23-1406
- ▶ *Ensemble Planning and Configuring Guide*, GC27-2608
- ▶ *Installation Manual for Physical Planning 2827*, GC28-6914
- ▶ *Advanced Workload Analysis Reporter (IBM zAware)*, SC27-2623

Online resources

These websites are also relevant as further information sources:

- ▶ IBM Resource Link:
<http://www.ibm.com/servers/resourceLink>
- ▶ Large Systems Performance Reference (LSPR) for IBM System z:
<https://www-304.ibm.com/servers/resourceLink/lib03060.nsf/pages/lsprindex?>
- ▶ IBM System z Software Contracts: MSU Ratings:
<http://www-03.ibm.com/systems/z/resources/swprice/reference/exhibits/hardware.html>
- ▶ System z HMC and SE (Version 2.12.1) Information Center:
<http://pic.dhe.ibm.com/infocenter/hwmca/v2r12m1/index.jsp>

Help from IBM

IBM Support and downloads:

ibm.com/support

IBM Global Services:

ibm.com/services

IBM zEnterprise System Technical Introduction

(0.2"spine)
0.17"<->0.473"
90<->249 pages



IBM zEnterprise System Technical Introduction



**Explains why a
smarter
infrastructure is
needed**

**Describes the
zEnterprise System
and related features
and functions**

**Discusses hardware
and software
capabilities**

In a smarter planet, information-centric processes are exploding in growth. The mainframe has always been the IT industry's leading platform for transaction processing, consolidated and secure data serving, and support for available enterprise-wide applications. IBM has extended the mainframe platform to help large enterprises reshape their client experiences through information-centric computing and to deliver on key business initiatives.

IBM zEnterprise is recognized as the most reliable and trusted system, and the most secure environment for core business operations. The new zEnterprise System consists of the IBM zEnterprise EC12 (zEC12) or IBM zEnterprise BC12 (zBC12), the IBM zEnterprise Unified Resource Manager, and the IBM zEnterprise IBM BladeCenter Extension (zBX) Model 003.

This IBM Redbooks publication describes the zEC12 and zBC12, with their improved scalability, performance, security, resiliency, availability, and virtualization. The zEnterprise System has no peer as a trusted platform that also provides the most efficient transaction processing and database management. With efficiency at scale delivering significant cost savings on core processes, resources can be freed up to focus on developing new services to drive growth.

This book provides a technical overview of the zEC12, zBC12, zBX Model 003, and Unified Resource Manager. This publication is intended for IT managers, architects, consultants, and anyone else who wants to understand the elements of the zEnterprise System. For this introduction to the zEnterprise System, readers are not expected to be familiar with current IBM System z technology and terminology.

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

SG24-8050-01

ISBN 0738439134