

IBM z16 A02 and IBM z16 AGZ Technical Guide

Ewerson Palacio

Octavian Lascu

Andre Spahni

John Troy

Martijn Raave

Martin Sollig



IBM zSystems



IBM Redbooks

IBM z16 A02 and IBM z16 AGZ Technical Guide

April 2023

Note: Before using this information and the product it supports, read the information in “Notices” on page 1.

First Edition (April 2023)

This edition applies to IBM z16 A02 and IBM z16 AGZ Machine Type 3932.

This document was created or updated on April 17, 2023.

Contents

Notices	1
Trademarks	2
Preface	1
Authors	2
Now you can become a published author, too!	3
Comments welcome	3
Stay connected to IBM Redbooks	3
Chapter 1. Introduction and overview.....	1
1.1 Design considerations	3
1.1.1 Predicting and automating with accelerated AI	4
1.1.2 A cyber-resilient system	5
1.1.3 IBM Modernize for hybrid cloud	5
1.1.4 Platform Sustainability	6
1.2 IBM z16 A02 and IBM z16 AGZ highlights	6
1.2.1 Supported upgrade paths	8
1.2.2 Capacity and performance	9
1.2.3 Supported operating systems	9
1.2.4 Supported IBM compilers	10
1.3 IBM z16 A02 and IBM z16 AGZ technical overview	10
1.3.1 IBM z16 A02 - IBM factory frame	11
1.3.2 IBM z16 AGZ	12
1.3.3 CPC drawers	12
1.3.4 I/O subsystem and I/O drawers	13
1.3.5 Storage connectivity	14
1.3.6 Network connectivity	15
1.3.7 Clustering connectivity	16
1.3.8 Cryptography	16
1.3.9 Supported connectivity and crypto features	17
1.3.10 Special-purpose features and functions	18
1.4 Hardware management	18
1.5 Reliability, availability, and serviceability	19
Chapter 2. Central processor complex hardware components	21
2.1 System overview: frames and drawers	22
2.1.1 IBM z16 A02	22
2.1.2 Top and bottom exit I/O and cabling	23
2.1.3 IBM z16 A02 system features	25
2.1.4 IBM z16 AGZ system features	26
2.1.5 System configurations for IBM z16 A02	27
2.1.6 System configurations for IBM z16 AGZ	28
2.1.7 PCIe I/O Drawers	28
2.2 CPC drawer	29
2.2.1 CPC drawer interconnect topology	32
2.2.2 Oscillator	33
2.2.3 System control	35
2.2.4 CPC drawer power	36
2.3 Dual chip modules	37

2.3.1 Processor unit (core)	39
2.3.2 PU characterization	40
2.3.3 Cache level structure	41
2.4 PCIe+ I/O drawer	41
2.5 Memory	43
2.5.1 Memory subsystem topology	44
2.5.2 Redundant array of independent memory	45
2.5.3 Memory Encryption	45
2.5.4 Memory configurations	45
2.5.5 Memory Offerings (initial order)	48
2.5.6 Drawer replacement and memory	52
2.5.7 Virtual Flash Memory	52
2.6 Reliability, availability, and serviceability	52
2.7 Connectivity	53
2.7.1 Redundant I/O interconnect	54
2.7.2 Enhanced drawer availability (EDA)	56
2.7.3 CPC drawer upgrade	56
2.8 Processor configurations	57
2.8.1 Upgrades	58
2.8.2 Model capacity identifier	60
2.8.3 Capacity Backup Upgrade	61
2.8.4 On/Off Capacity on Demand and CPs	64
2.9 Power and cooling	65
2.9.1 Power considerations- IBM z16 A02 (factory frame)	65
2.9.2 Power considerations - IBM z16 AGZ	66
2.9.3 Power estimation tool	67
2.9.4 Cooling	67
2.10 Summary	67
Chapter 3. Central processor complex design	69
3.1 Overview	70
3.2 Design highlights	71
3.3 CPC drawer design	73
3.3.1 Cache levels and memory structure	75
3.3.2 CPC drawer interconnect topology	77
3.4 Processor unit design	78
3.4.1 Simultaneous multithreading	80
3.4.2 Single-instruction multi-data	81
3.4.3 Out-of-Order execution	82
3.4.4 Superscalar processor	85
3.4.5 On-chip coprocessors and accelerators	86
3.4.6 IBM Integrated Accelerator for Artificial Intelligence (on-chip)	90
3.4.7 Decimal floating point accelerator	93
3.4.8 IEEE floating point	95
3.4.9 Processor error detection and recovery	95
3.4.10 Branch prediction	95
3.4.11 Wild branch	96
3.4.12 Translation lookaside buffer	96
3.4.13 Instruction fetching, decoding, and grouping	97
3.4.14 Extended Translation Facility	97
3.4.15 Instruction set extensions	97
3.4.16 Transactional Execution	98
3.4.17 Runtime Instrumentation	98

3.5 Processor unit functions	98
3.5.1 Overview	98
3.5.2 Central processors	100
3.5.3 Integrated Facility for Linux (FC 1959)	101
3.5.4 Internal Coupling Facility (FC 1960)	102
3.5.5 IBM Z Integrated Information Processor (FC 1961)	105
3.5.6 System assist processors	110
3.5.7 Reserved processors	110
3.5.8 Integrated Firmware Processors	110
3.5.9 Processor unit assignment	111
3.5.10 Sparing rules	112
3.5.11 CPC drawer numbering	112
3.6 Memory design	113
3.6.1 Overview	113
3.6.2 Main storage	116
3.6.3 Hardware system area	116
3.6.4 Virtual Flash Memory (FC 0644)	117
3.7 Logical partitioning	117
3.7.1 Overview	117
3.7.2 Storage operations	125
3.7.3 Reserved storage	126
3.7.4 Logical partition storage granularity	127
3.7.5 LPAR dynamic storage reconfiguration	127
3.8 Intelligent Resource Director	128
3.9 Clustering technology	129
3.9.1 CF Control Code	131
3.9.2 Coupling Thin Interrupts	131
3.9.3 Dynamic CF dispatching	131
3.10 Virtual Flash Memory	132
3.10.1 Overview	132
3.10.2 VFM feature	133
3.10.3 VFM administration	133
3.11 Secure Service Container	133
Chapter 4. I/O structure	137
4.1 Introduction to I/O infrastructure	138
4.1.1 I/O infrastructure	138
4.1.2 PCIe Generation 3	139
4.2 I/O system overview	140
4.2.1 Characteristics	140
4.2.2 Supported I/O features	141
4.3 PCIe+ I/O drawer	142
4.3.1 PCIe+ I/O drawer offering	144
4.4 CPC drawer fanouts	145
4.4.1 PCIe+ Generation 3 fanout (FC 0175)	146
4.4.2 Integrated Coupling Adapter (FC 0172 and FC 0176)	146
4.4.3 Fanout considerations	147
4.5 I/O features	148
4.5.1 I/O feature card ordering information	149
4.5.2 Physical channel ID report	150
4.6 Connectivity	152
4.6.1 I/O feature support and configuration rules	152
4.6.2 Storage connectivity	155

4.6.3 Network connectivity	162
4.6.4 Parallel Sysplex connectivity	175
4.7 Cryptographic functions	180
4.7.1 CPACF functions (FC 3863)	180
4.7.2 Crypto Express8S feature (FC 0908 and FC 0909)	180
4.7.3 Crypto Express7S feature (FC 0898 and FC 0899) as carry forward only	181
4.7.4 Crypto Express6S feature (FC 0893) as carry forward only	183
4.8 Integrated Firmware Processor	183
Chapter 5. Logical I/O - Channel Subsystem	185
5.1 Channel subsystem	186
5.1.1 Multiple logical channel subsystems	187
5.1.2 Multiple subchannel sets	188
5.1.3 Channel path spanning	191
5.2 I/O configuration management	194
5.3 Channel subsystem summary	195
Chapter 6. Cryptographic features	197
6.1 Cryptography enhancements on IBM z16 A02 and IBM z16 AGZ	199
6.2 Cryptography overview	200
6.2.1 Modern cryptography	200
6.2.2 Kerckhoffs' principle	201
6.2.3 Keys	201
6.2.4 Algorithms	203
6.3 Cryptography on IBM z16 A02 and IBM z16 AGZ	204
6.4 CP Assist for cryptographic functions	209
6.4.1 Cryptographic synchronous functions	210
6.4.2 CPACF protected key	211
6.5 Crypto Express8S	213
6.5.1 Cryptographic asynchronous functions	215
6.5.2 Crypto Express8S as a CCA coprocessor	217
6.5.3 Crypto Express8S as an EP11 coprocessor	223
6.5.4 Crypto Express8S as an accelerator	224
6.5.5 Managing Crypto Express8S	224
6.6 Trusted Key Entry workstation	228
6.6.1 Logical partition, TKE host, and TKE target	229
6.6.2 Optional smart card reader	229
6.6.3 TKE hardware support and migration information	230
6.7 Cryptographic functions comparison	231
6.8 Cryptographic operating system support for IBM z16 A02 and IBM z16 AGZ	233
6.8.1 Crypto Express8S Exploitation	233
6.8.2 Crypto Express8S support of VFPE	234
6.8.3 Crypto Express8S support of greater than 16 domains	234
6.9 Further use of cryptography on IBM z16 A02 and IBM z16 AGZ	235
6.9.1 Validated boot	235
6.9.2 Secure Boot for ECKD devices	238
6.9.3 z/VM 7.3 Guest Secure-IPL	239
Chapter 7. Operating system support	241
7.1 Operating systems summary	242
7.2 Support by operating system	242
7.2.1 z/OS	243
7.2.2 z/VM	243
7.2.3 z/VSE	244

7.2.4 21 st Century Software z/VSE ⁿ V6.3	245
7.2.5 z/TPF	245
7.2.6 Linux on IBM Z	245
7.2.7 KVM hypervisor.....	246
7.3 IBM z16 A02 and IBM z16 AGZ features and functions support overview	246
7.3.1 Supported CPC functions	247
7.3.2 Coupling and clustering	250
7.3.3 Storage connectivity	250
7.3.4 Network connectivity.....	253
7.3.5 Cryptographic functions	257
7.4 Support by features and functions	258
7.4.1 LPAR Configuration and Management	259
7.4.2 Base CPC features and functions.....	262
7.4.3 Coupling and clustering features and functions	274
7.4.4 Storage connectivity-related features and functions	279
7.4.5 Networking features and functions	290
7.4.6 Cryptography Features and Functions Support	301
7.5 z/OS migration considerations	307
7.5.1 General guidelines	308
7.5.2 Hardware Fix Categories (FIXCATs)	309
7.5.3 z/OS V2.R5	310
7.5.4 z/OS V2.R4	310
7.5.5 z/OS V2.R3	311
7.5.6 Coupling links	311
7.5.7 z/OS XL C/C++ considerations	312
7.6 IBM z/VM migration considerations.....	313
7.6.1 IBM z/VM 7.3	313
7.6.2 IBM z/VM 7.2	313
7.6.3 Capacity	313
7.7 z/VSE migration considerations	313
7.8 Software licensing	314
7.9 References	316
Chapter 8. System upgrades	317
8.1 Permanent and Temporary Upgrades.....	318
8.1.1 Overview	318
8.1.2 CoD for IBM z16 A02 and IBM z16 AGZ systems-related terminology	319
8.1.3 Concurrent and nondisruptive upgrades	321
8.1.4 Permanent upgrades	321
8.1.5 Temporary upgrades	322
8.2 PU upgrades	323
8.2.1 CPC drawer feature and PU capacity upgrades	324
8.2.2 Customer Initiated Upgrade facility	326
8.2.3 Concurrent upgrade functions summary	330
8.3 Miscellaneous equipment specification (MES) upgrades	330
8.3.1 MES upgrade for processors	331
8.3.2 MES upgrades for memory	332
8.3.3 MES upgrades for I/O and CPC drawers	333
8.4 Permanent upgrade by using the CIU facility	334
8.4.1 Ordering	336
8.4.2 Retrieval and activation	337
8.5 On/Off Capacity on Demand	338
8.5.1 Overview	338

8.5.2 Capacity Provisioning Manager	339
8.5.3 Ordering	340
8.5.4 On/Off CoD testing	342
8.5.5 Activation and deactivation	343
8.5.6 Termination	343
8.6 z/OS Capacity Provisioning	344
8.7 Capacity for Planned Event	349
8.8 Capacity Backup	349
8.8.1 Ordering	349
8.8.2 CBU activation and deactivation	351
8.8.3 Automatic CBU enablement for GDPS	353
8.9 Flexible Capacity Cyber Resiliency	353
8.10 Planning for nondisruptive upgrades	355
8.10.1 Components	356
8.10.2 Concurrent upgrade considerations	357
8.11 Summary of Capacity on-Demand offerings	361
Chapter 9. Reliability, availability, and serviceability	363
9.1 RAS strategy	364
9.2 Technology	364
9.2.1 Processor Unit chip	364
9.2.2 Main memory	368
9.2.3 I/O and service	368
9.3 Structure	369
9.4 Reducing complexity	370
9.5 Reducing touches	370
9.6 IBM z16 A02 and IBM z16 AGZ availability characteristics	370
9.7 IBM z16 A02 and IBM z16 AGZ RAS functions	374
9.7.1 Scheduled outages	375
9.7.2 Unscheduled outages	376
9.8 Enhanced drawer availability	378
9.8.1 EDA planning considerations	378
9.8.2 Enhanced drawer availability processing	380
9.9 Concurrent Driver Maintenance	385
9.9.1 Resource Group and native PCIe features MCLs	386
9.10 RAS capability for the HMA and SE	387
9.11 IBM z16 AGZ specifics	388
Chapter 10. Hardware Management Console and Support Element	391
10.1 Introduction and overview	392
10.2 HMC and SE new features and changes	393
10.2.1 Driver 51/Version 2.16.0 HMC and SE new features	393
10.2.2 Hardware Management Appliance (HMA)	404
10.2.3 HMC and SE servers	406
10.2.4 USB support for HMC and SE	408
10.2.5 SE Driver/Version support with the HMC Driver 51/Version 2.16.0	408
10.3 HMC and SE connectivity	409
10.3.1 Hardware Management Appliance (HMA) HMC/SE connectivity	410
10.3.2 Legacy standalone HMC connectivity	410
10.3.3 Network planning for the HMC, SE, and ETS	411
10.3.4 Hardware considerations	413
10.3.5 TCP/IP Version 6 on the HMC and SE	413
10.3.6 Assigning TCP/IP addresses to the HMC, SE, and ETS	414

10.3.7 HMC Multi-factor authentication	415
10.4 Remote Support Facility	416
10.4.1 Security characteristics	416
10.4.2 RSF connections to IBM and Enhanced IBM Service Support System	416
10.5 HMC and SE capabilities	417
10.5.1 Central processor complex management	417
10.5.2 LPAR management	418
10.5.3 HMC and SE remote operations	418
10.5.4 Operating system communication	420
10.5.5 HMC and SE Microcode	421
10.5.6 Monitoring	424
10.5.7 Capacity on-demand support	427
10.5.8 Server Time Protocol (STP) support	427
10.5.9 NTP client and server on the HMC	429
10.5.10 Security and user ID management	430
10.5.11 System Input/Output Configuration Analyzer on the SE and HMC	432
10.5.12 Automated operations	432
10.5.13 Cryptographic support	433
10.5.14 Installation support for z/VM that uses the HMC	434
10.5.15 Dynamic Partition Manager (DPM)	434
Chapter 11. Environmentals	437
11.1 Introduction	438
11.2 IBM z16 A02 environmental considerations	438
11.2.1 Power infrastructure	438
11.2.2 Cooling requirements	439
11.2.3 Physical specifications	440
11.2.4 Physical planning	440
11.3 IBM z16 AGZ environmental considerations	442
11.3.1 Power infrastructure	443
11.3.2 Cooling requirements	444
11.3.3 Physical specifications	445
11.3.4 Physical planning	446
11.4 Energy Management	447
11.4.1 Environmental monitoring	448
Chapter 12. Performance	453
12.1 IBM z16 A02 and IBM z16 AGZ performance overview	454
12.1.1 IBM z16 A02 and IBM z16 AGZ single-thread capacity	454
12.1.2 IBM z16 A02 and IBM z16 AGZ SMT capacity	454
12.1.3 IBM Integrated Accelerator for zEnterprise Data Compression	455
12.1.4 Primary performance improvement drivers with z16	455
12.2 IBM z16 Large System Performance Reference ratio	455
12.2.1 LSPR workload suite	456
12.3 Fundamental components of workload performance	457
12.3.1 Instruction path length	457
12.3.2 Instruction complexity	457
12.3.3 Memory hierarchy and memory nest	458
12.4 Relative Nest Intensity	459
12.5 LSPR workload categories based on RNI	460
12.6 Relating production workloads to LSPR workloads	461
12.7 CPU MF counter data and LSPR workload type	461
12.8 Workload performance variation	462

12.9 Capacity planning for z16 A02 and AGZ.....	463
12.9.1 Collect CPU MF counter data.....	463
12.9.2 Creating EDF file with CP3KEXTR.....	463
12.9.3 Loading EDF file to the capacity planning tool	463
12.9.4 Tips to maximize IBM z16 A02 and IBM z16 AGZ capacity	464
Appendix A. Channel options	467
Appendix B. IBM Z Integrated Accelerator for AI	471
Overview	472
NNPA and IBM z16 A02 and IBM z16 AGZ Hardware	473
How to leverage IBM Z Integrated AI Accelerator in your enterprise	473
References	474
Appendix C. IBM Integrated Accelerator for zEnterprise Data Compression	475
Client value of IBM zSystems compression	476
IBM z16 A02 and IBM z16 AGZ IBM Integrated Accelerator for zEDC	476
Eliminating adapter sharing by using Nest Compression Accelerator	477
Compression modes	477
IBM z16 A02 and IBM z16 AGZ migration considerations	478
All z/OS configuration stay the same	478
Consider fail-over and DR sizing.....	478
Performance metrics	478
zEDC to IBM z16 A02 and IBM z16 AGZ zlib Program Flow for z/OS	478
Software support	478
C.0.1 IBM Z Batch Network Analyzer.....	479
Compression acceleration and Linux on IBM Z	479
Appendix D. Rack configurations	481
IBM z16 A02 (Factory Frame) configurations	482
IBM z16 AGZ configurations	484
Related publications	487
IBM Redbooks	487
Other publications	487
Online resources	488
Help from IBM	488
Additional material	489
Locating the web material	489
Using the web material	489
System requirements for downloading the web material	489
Downloading and extracting the web material	489

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

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, MD-NC119, Armonk, NY 10504-1785, US

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 jurisdictions 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 provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

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 actual people or business enterprises 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. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at <http://www.ibm.com/legal/copytrade.shtml>

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

Redbooks (logo) ®

Trademarking:

Refresh the IBM trademark list.

Run the **Toolkit** → **RXFM** → **Maintenance** → **Refresh-Toolkit-Rxfm-Rex-Scripts.rex** tool.

Trademark search and mark first use of a trademark:

Open the book file.

Run the **Toolkit** → **RXFM** → **Editor_tools** → **Trademark-Search.rex** tool.

The following terms are trademarks of other companies:

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



Preface

This IBM® Redbooks® publication describes the features and functions of the latest IBM zSystems® platform members, built with the IBM Telum processor, the IBM z16 A02 and IBM z16 AGZ™. It includes information about the IBM z16 A02 and IBM z16 AGZ processors design, I/O innovations, security features, and supported operating systems.

The IBM zSystems platform is recognized for its security, resiliency, performance, and scale, and it is relied on for mission-critical workloads and as an essential element of hybrid cloud infrastructures. The new members of the IBM z16 A02 and IBM z16 AGZ generation adds more capabilities and value with innovative technologies that are needed to accelerate the digital transformation journey.

The IBM z16 A02 and IBM z16 AGZ are a state-of-the-art data and transaction systems that deliver advanced capabilities, which are vital to any digital transformation. The IBM z16 A02 and IBM z16 AGZ are designed for enhanced modularity in an industry standard footprint, offering clients the choice of factory frame (IBM z16 A02) or rack mount configuration (IBM z16 AGZ) to install in their own datacenter rack infrastructure. These systems excel at the following tasks:

- ▶ AI inference with Integrated Accelerator for Artificial Intelligence
- ▶ Making use of multicloud integration services
- ▶ Securing data with pervasive encryption
- ▶ Accelerating digital transformation with agile service delivery
- ▶ Transforming a transactional platform into a data powerhouse
- ▶ Getting more out of the platform with IT Operational Analytics
- ▶ Accelerating digital transformation with agile service delivery
- ▶ Revolutionizing business processes
- ▶ Blending open source and IBM zSystems technologies

This book explains how these systems use new innovations and traditional IBM zSystems strengths to satisfy growing demand for cloud, analytics, and open source technologies. With the IBM z16 A02 and IBM z16 AGZ as the base, applications can run in a trusted, reliable, and secure environment that improves operations and lessens business risk.

Authors

This book was produced by a team of specialists from around the world working at IBM Redbooks, Poughkeepsie Center.

Ewerson Palacio is an IBM Redbooks Project Leader. He holds Bachelor's degree in Math and Computer Science. Ewerson worked for IBM Brazil for over 40 years and retired in 2017 as an IBM Distinguished Engineer. Ewerson co-authored many IBM zSystems Redbooks, and created and presented ITSO seminars around the globe.

Octavian Lascu is an IBM Redbooks Project Leader and a Senior IT Consultant with over 30 years of experience. He specializes in designing, implementing, and supporting complex IT infrastructure environments (systems, storage, and networking), including high availability and disaster recovery (HADR) solutions and high-performance computing deployments. He has developed materials for and taught over 50 workshops for technical audiences around the world. He is the author of several IBM publications.

Martijn Raave is an IBM zSystems and LinuxONE Client Architect and Hardware Technical Specialist for IBM Northern Europe. Over a period of over 20 years, his professional career has revolved around the mainframe platform, supporting several large Dutch customers in their technical and strategic journey on IBM zSystems. His focus areas are hardware, resiliency, availability, architecture, and any other topics about IBM zSystems.

Martin Sollig is a Consultant IT Specialist in Germany. He has 32 years of experience working in the IBM zSystems field. He holds a degree in mathematics from the University of Hamburg. His areas of expertise include IBM z/OS and IBM zSystems hardware, specifically in Parallel Sysplex and GDPS environments, and also in cryptography on IBM zSystems.

Andre Spahni is an IBM zSystems Client Technical Specialist based in Zurich, Switzerland. He has over 20 years of experience working with and supporting IBM zSystems clients. André has worked for EMEA 2nd level supporter and national Top Gun. His areas of expertise include IBM zSystems hardware, HMC/SE, and connectivity.

John Troy is an IBM zSystems and storage hardware National Top Gun in the northeast area of the United States. He has over 40 years of experience in the service field. His areas of expertise include IBM zSystems servers and high-end storage systems technical and customer support and services. John has also been an IBM zSystems hardware technical support course designer, developer, and instructor for the last eight generations of IBM high-end servers.

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

Rhonda Sundlof
IBM Poughkeepsie-US

Ronald Geiger, Frank J. Miele, Gary Sullivan, Chris Rayns
IBM Poughkeepsie-US

David Evans
IBM Fort Mitchell-US

Volker Urban
IBM Boeblingen-DE

Bob Haimowitz, Bill White
IBM Redbooks, Poughkeepsie Center

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 IBM Redbooks 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, IBM Redbooks
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- ▶ Find 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>



Introduction and overview

The IBM zSystems platform is recognized for its long-standing commitment to delivering best-of-breed security, resiliency, performance, and scalability. IBM zSystems platform is relied on for mission-critical workloads and as an essential element of hybrid cloud infrastructures.

This IBM® Redbooks® publication introduces the latest members of the IBM zSystems family, available in two new configuration options:

- ▶ IBM z16 A02
- ▶ IBM z16 AGZ

The IBM z16 A02 and IBM z16 AGZ align with the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) Class A3 data center guidelines and is available in two different configuration options:

- ▶ **The IBM z16 A02** is built with an IBM 19-inch format single frame. There are four orderable features: Max5, Max16, Max32, and Max68.
- ▶ **The IBM z16 AGZ** is a rack mount configuration that allows the core compute, I/O and networking features to be installed into and powered by a client-designated rack with power distribution units (PDUs), respectively. The rack mount options are under a combined AGZ warranty umbrella and orderable as: Max5, Max16, Max32, and Max68.

The IBM z16 A02 and IBM z16 AGZ ensure continuity and upgradeability from the IBM z15 T02 and IBM z14 ZR1.

The IBM z16 families (IBM z16 A01, IBM z16 A02 and IBM z16 AGZ) are the first IBM zSystems built with the *IBM Telum processor*¹. It is designed to help businesses:

- ▶ Create value in every interaction and to optimize decision making, with the on-chip Artificial Intelligence (AI) accelerator. The Accelerator for AI can deliver the speed and scale required to infuse AI inferencing into workloads with no impact on service delivery.
- ▶ Act now to protect today's data against current and future threats with quantum-safe protection out of the box through quantum-safe cryptography APIs and crypto discovery tools.

¹ IBM Telum Processor: the next-gen microprocessor for IBM zSystems and IBM LinuxONE

- ▶ Enhance resiliency with flexible capacity to dynamically shift system resources across locations to proactively avoid disruptions.
- ▶ Modernize and integrate applications and data in a hybrid cloud environment with consistent and flexible deployment options to innovate with speed and agility.
- ▶ Reduce cost and keep up with changing regulations through a solution that helps simplify and streamline compliance tasks.

Businesses worldwide in every industry are investing in digital transformation, with the rate and pace increasing over the past several years. The IBM z16 A02 and IBM z16 AGZ are built for hybrid cloud. It can help expedite your transformation with new on-chip AI acceleration to enable decision velocity, quantum-safe technologies designed to help protect your business now and into the future, a flexible infrastructure to meet the resiliency and compliance demands of a constantly changing environment, and capabilities to accelerate modernization and delivery of additional services.

The pandemic accelerated the rate and pace of the digital transformation journey, which caused most companies to find novel ways to sustain operations while unlocking new opportunities and increasing innovation.

To move the business forward while continuing to maintain the necessary levels of resiliency, compliance, and sustainability, a secure infrastructure with more flexibility and agility is needed. The latest members of the IBM zSystems family, the IBM z16 A02 and IBM z16 AGZ, can help with these new demands through *accelerated AI*, *cyber resiliency*, a *modernized hybrid cloud*, and *sustainability*.

This chapter describes the basic concepts and design considerations around IBM z16 A02 and z16 AGZ:

- ▶ 1.1, “Design considerations” on page 3
- ▶ 1.2, “IBM z16 A02 and IBM z16 AGZ highlights” on page 6
- ▶ 1.3, “IBM z16 A02 and IBM z16 AGZ technical overview” on page 10
- ▶ 1.4, “Hardware management” on page 18
- ▶ 1.5, “Reliability, availability, and serviceability” on page 19

1.1 Design considerations

More than any other platform, the IBM z16 A02 and IBM z16 AGZ offer a high-value architecture that can satisfy the growing demands that are driven by the pace of digital transformation, such as:

- ▶ **IBM Secure Execution for Linux®:** A second-generation hardware-based security technology designed to provide scalable isolation for individual workloads to help protect them from not only external attacks, but also insider threats. It can help protect and isolate workloads on-premises, or on IBM zSystems hybrid cloud environments.
- ▶ **New processor Chip design:** The IBM z16 A02 and z16 AGZ, built with the IBM Telum (TM) processor, provide dedicated on-chip accelerators to enable real-time AI inferencing imbedded at scale in transactional workloads.
- ▶ **IBM Single Frame or Customer designated rack:** The new IBM z16 is available as a 19" IBM single frame with IBM PDUs (IBM z16 A02), or as a rack mount for client-defined data center racks for new location spaces (IBM z16 AGZ). The power option for the rack mount configuration are client-provided PDUs / power management with DCIM sustainability tool integrations that meets systems specifications.
- ▶ **Quantum Safe:** Protect data with IBM z16 A02 and IBM z16 AGZ, the industry's first quantum-safe system. The Crypto Express8S card offers quantum-safe APIs that provide access to quantum-safe algorithms. Quantum-safe cryptography refers to efforts to identify algorithms that are resistant to attacks by both classical and quantum computers.
- ▶ **Flexible Capacity:** Proactively avoid disruptions and manage capacity and workloads with enhanced Flexible Capacity for Cyber Resiliency, enabling the shift of production workloads between participating IBM z16 A02 and IBM z16 AGZ systems at different sites that remain operational at the target site for up to one year.
- ▶ **Hybrid Cloud platform:** Accelerate application modernization and IT automation with IBM zSystems and Cloud Modernization Stack to connect, provision, and manage z/OS systems across the hybrid cloud with Red Hat OpenShift.
- ▶ **System Recovery Boost:** The enhancements can provide boosted processor capacity and parallelism for specific events. Client-selected middleware starts and restarts may be boosted to expedite recovery for middleware regions and restore steady-state operations as soon as possible. SVC dump processing and HyperSwap configuration load and reload may be boosted to minimize the impact to running workloads.
- ▶ **z/OS Validated Boot:** With IBM z16 A02 and IBM z16 AGZ and accompanying z/OS V2.5 operating system support, IBM is providing optional basic support for performing a Validated Boot (IPL) of z/OS systems, using IPL volumes defined and built on ECKD DASD devices.
- ▶ **Linux® Secure Boot:** IBM z16 A02 and IBM z16 AGZ are designed to extend Linux secure boot capabilities to Linux IPL Volumes built on ECKD devices, in addition to existing support for secure boot from SCSI/FBA devices and allows client-provided validation certificates provided through the SE/HMC to be used for validation purposes during Linux secure boot. z/VM 7.3 has been enhanced to add support to securely boot a Linux guest.
- ▶ **IBM Z Security and Compliance Center:** Simplify compliance and help reduce risk with IBM Z Security and Compliance Center, a centralized, interactive dashboard with out-of-the-box profiles specifically built for regulatory requirements that you can use or customize to accommodate or establish your regulatory framework.
- ▶ **Sustainability:** The IBM z16 A02 and IBM z16 AGZ are key contributors to a sustainable data center with transparency through PAIA product carbon footprint reports²,

partition-level power monitoring through the new HMC Environmental Dashboard and enhanced Web Service API, and power reporting through HMC integration with Instana.

- ▶ **Parallel Sysplex and Coupling:** Parallel Sysplex enhancements include improved Integrated Coupling Adapter Short Reach (ICA SR) performance and Coupling Facility (CF) image scalability, technology and protocol upgrades for coupling links, simplified Dynamic CF Dispatching (DYNDISP) support, and resiliency enhancements for CF cache and lock structures. IBM z16 A02 and IBM z16 AGZ coupling hardware and firmware, including Coupling Facility Control Code (CFCC) CFLEVEL25, provide several coupling facility (CF) and coupling link enhancements.
- ▶ **Crypto Express8S:** IBM z16 A02 and IBM z16 AGZ with the Crypto Express8S card offers quantum-safe APIs that provide access to quantum-safe algorithms, which have been selected as finalists during the PQC standardization process conducted by [NIST Information Technology Laboratory](#). Quantum-safe cryptography refers to efforts to identify algorithms that are resistant to attacks by both classical and quantum computers, to keep information assets secure even after a large-scale quantum computer has been built. Source: [ETSI Quantum-Safe Cryptography \(QSC\)](#). These algorithms are also used to help ensure the integrity of several of the firmware and boot processes. IBM z16 A02 and IBM z16 AGZ are the first industry-system protected by quantum-safe technology across multiple layers of firmware.

1.1.1 Predicting and automating with accelerated AI

An approach where data gravity and transaction gravity intersect, that co-collocates data, transactional systems, and AI inferencing, can deliver insights at speed and scale to enable decision velocity. Decision velocity means delivering insights faster to make decisions to help identify new business opportunities improve customer experience, and reduce operational risk.

Consider the following points:

- ▶ The on-chip Integrated Accelerator for AI is designed for high-speed, real-time inferencing at scale. It is designed to add up to 5.8 TFLOPS of processing power shared by all cores on the chip. This centralized AI design is intended to provide extremely high performance and consistent low-latency inferencing for processing a mix of transactional and AI workloads at speed and scale.

Now, complex neural network inferencing that uses real-time data can be run and delivers insights within high throughput enterprise workloads in real-time while still meeting stringent SLAs.

- ▶ A robust ecosystem of frameworks and open source tools, combined with the IBM Deep Learning Compiler that generates inferencing programs that are highly optimized for the IBM zSystems architecture and the Integrated Accelerator for AI, help enable rapid development and deployment of deep learning and machine learning models on IBM zSystems to accelerate time to market.
- ▶ The IBM z16 A02 and IBM z16 AGZ support 16 TB per system (with up to 8TB per CPC drawer). IBM z16 A02 and IBM z16 AGZ memory is designed with a new memory buffer chip that provides up to DDR4-3200 memory speed, depending on memory size, delivering 50% more memory bandwidth per drawer than IBM z15 T02. This design improves overall workload performance, particularly for data-intensive analytics and AI applications. The new memory interface uses transparent memory encryption technology to protect all data leaving the processor chips before it gets stored in the memory DIMMs.

² See <https://www.ibm.com/downloads/cas/75EX0KDJ>

1.1.2 A cyber-resilient system

A cyber-resilient system can help protect against risks, vulnerabilities, attacks, and failures that might happen while digitally transforming your business.

With the opportunity that is created by quantum computing comes the threat to today's public key cryptography. Businesses must start now to prepare for the time when a quantum computer can break today's cryptography. In fact, today's data is at risk for future exposure through "harvest now, decrypt later" attacks.

IBM z16 A01, IBM z16 A02 and IBM z16 AGZ are the industry-first quantum-safe system, which is protected by quantum-safe technology across multiple layers of firmware.

Quantum-safe secure boot technology helps protect IBM z16 A02 and IBM z16 AGZ firmware from quantum attacks through a built-in dual signature scheme with no configuration changes that are required for enablement.

With the new Crypto Express8S, IBM z16 A02 and IBM z16 AGZ help deliver quantum-safe APIs that position businesses to begin the use of quantum-safe cryptography along with classical cryptography as they begin modernizing applications and building new applications.

Discovering where and what kind of cryptography is being used is a key first step along the journey to quantum-safety. IBM z16 A02 and IBM z16 AGZ provide instrumentation that can be used to track cryptographic instruction execution in the CP Assist for Cryptographic Functions (CPACF).

Additionally, IBM Application Discovery and Delivery Intelligence (ADDI) was enhanced with new crypto discovery capabilities.

1.1.3 IBM Modernize for hybrid cloud

IBM z16 A02 and IBM z16 AGZ deliver technology innovation in AI, security, and resiliency on a flexible infrastructure that is designed for mission-critical workloads in a hybrid cloud environment. IBM z16 A02 and IBM z16 AGZ continue to deliver new and improved cloud capabilities on the platform.

IBM z16 A02 and IBM z16 AGZ provide the foundation for application modernization and hybrid cloud velocity by delivering leading hybrid cloud infrastructure to support the optimization of mission-critical applications and data.

IBM z16 A02 and IBM z16 AGZ and the accompanying IBM zSystems and cloud software, which is developed to support a cloud-native experience, delivers a broad set of open and industry-standard tools, including an agile DevOps methodology to accelerate modernization. These capabilities deliver speed to market and agility for development and operational teams as IBM z16 A02 and IBM z16 AGZ integrate as a critical component of hybrid cloud.

Businesses can accelerate modernization and delivery of new services by using the following key software offerings along with IBM z16 A02 and IBM z16 AGZ:

- ▶ [IBM zSystems and Cloud Modernization Stack](#) to help empower developers to modernize and integrate z/OS applications with services across the hybrid cloud. This solution provides a flexible and integrated platform to support z/OS-based cloud-native development, application, and data modernization and infrastructure automation.
- ▶ [Red Hat OpenShift](#) and [IBM Cloud Paks](#) running on IBM z16 A02 and IBM z16 AGZ infrastructure provide the combination of infrastructure, hybrid cloud container platform, and middleware to modernize applications and develop cloud-native applications that

integrate, extend, and supply data and workloads from IBM z16 A02 and IBM z16 AGZ across the hybrid cloud with Red Hat OpenShift.

1.1.4 Platform Sustainability

The IBM z16 A02 and IBM z16 AGZ mark a distinct sustainability focus across product lifecycle, from the improved energy efficiency, enhancement of manufacturing and material sourcing, to the improved packaging strategies for shipment, to material recycling at product end-of-life.

IBM has a long-standing commitment to building a more sustainable, equitable future. In 1971, IBM formalized its environmental programs and commitment to leadership with the issuance of its Corporate Policy on IBM's Environmental Responsibilities. This was a quarter century before the first International Organization for Standardization (ISO) 14001 environmental management systems standard was published. IBM's activities between then and 2021, when IBM committed to reaching net zero greenhouse gas emissions by 2030 in all 175 countries in which it operates and beyond, make it an ideal partner for the increasing number of businesses that consider sustainability a strategic direction. For more information, see the [IBM Commits To Net Zero Greenhouse Gas Emissions By 2030](#) web page.

The IBM z16 A02 and IBM z16 AGZ are the latest in a long line of machines that are designed for system and data center energy efficiency with differentiated architectural advantages, including on-chip compression, and encryption designed to sustain 90% utilization along with new embedded on-chip AI acceleration to seamlessly integrate real-time AI insights into business-critical transactions.

IBM z16 A02 and IBM z16 AGZ build on our more than 24-year history of improving the system performance per watt – a key metric for improving the data center carbon footprint. Beginning with the first CMOS mainframe processor and continuing through the IBM z16 A02 and IBM z16 AGZ, IBM zSystems has a 27-year history of improved mainframe system capacity per watt.

The biggest opportunity for energy savings with the IBM z16 A02 and IBM z16 AGZ come through workload modernization and consolidation for distributed x86 systems. Many clients can't easily grow their data center size and that inhibits dealing with workload surges inherent in planned or reactive digital transformation on a distributed platform. The vertical scalability of the IBM zSystems architecture can address this while improving your carbon footprint dramatically.

The IBM z16 A02 and IBM z16 AGZ continue to prioritize how we contribute to the circular economy. New for IBM z16 A02 and IBM z16 AGZ is the publication [Product carbon footprint](#) reports that shows what attributes of the product lifecycle have the biggest impact on the carbon footprint. The IBM z16 A02 and IBM z16 AGZ continue the platform's long history of focusing on the product lifecycle, from the improved energy

1.2 IBM z16 A02 and IBM z16 AGZ highlights

Each new IBM zSystems platform continues to deliver innovative technologies. The IBM z16 A02 and IBM z16 AGZ are no exception. IBM z16 A02 and IBM z16 AGZ implement a new processor chip design with each processor unit (PU) running at 4.6 GHz.

The new processor chip design has a new cache hierarchy, on-chip AI accelerator shared by the PU cores, transparent memory encryption, and increased uniprocessor capacity (single thread and SMT similar).

The on-chip AI scoring logic provides sub-microsecond AI inferencing for deep learning and complex neural network models.

The redesigned cache structure has the following cache sizes:

- ▶ 256 KB L1 per PU core
- ▶ 32 MB semi-private L2 per PU core
- ▶ 256 MB (logical) shared victim virtual L3 per chip
- ▶ 2 GB³ (logical) shared victim virtual L4 per CPC drawer

The result is improved system performance and scalability with 1.5x more cache capacity per core over the IBM z15 T02 and reduced average access latency through a flatter topology.

The IBM z16 A02 is delivered in a factory frame (19-inch standard) while the IBM z16 AGZ is delivered as a bundle installed on site in client supplied standard 19-inch rack. Both share the same feature names (maximum number of characterizable processor units (PUs)): Max5, Max16, Max32, and Max68.

The number of characterizable PUs, spare PUs, and System Assist Processors (SAPs) are included with each feature of IBM z16 A02 and IBM z16 AGZ (see Table 1-1).

Table 1-1 IBM z16 A02 and IBM z16 AGZ processor unit (PU) configurations

Feature name ^a	Number of CPC drawers	Feature code	Characterizable PUs	SAPs	Spare PUs
Max5	1	0672	1 - 5	2	2
Max16	1	0673	1 - 16	2	2
Max32	1	0674	1 - 32	4	2
Max68	2	0675	1 - 68	8	2

a. IBM z16 AGZ supports the same features as IBM z16 A02.

The IBM z16 A02 and IBM z16 AGZ memory subsystem uses proven redundant array of independent memory (RAIM) technology to ensure high availability. Up to 16TB (8 TB per CPC drawer) of addressable memory per system can be ordered.

The IBM z16 A02 and IBM z16 AGZ also have unprecedented capacity to meet consolidation needs with innovative I/O features for transactional and hybrid cloud environments.

The IBM z16 A02 and IBM z16 AGZ (maximum configuration) can support up to 3 PCIe+ I/O drawers. Each I/O drawer can support up to 16 I/O or special purpose features for storage, network, and clustering connectivity, as well as cryptography. The following features were introduced with the IBM z16 A02 and IBM z16 AGZ:

- ▶ FICON Express32S
- ▶ OSA-Express7S 1.2
- ▶ RoCE Express3 (Long Reach and Short Reach)
- ▶ Coupling Express2 Long Reach
- ▶ Crypto Express8S

The IBM z16 A02 and IBM z16 AGZ are more flexible and have simplified on-demand capacity to satisfy peak processing demands and quicker recovery times with built-in resiliency capabilities. The Capacity on Demand (CoD) function can dynamically change available system capacity. This function can help respond to new business requirements with flexibility and precise granularity. The IBM Tailored Fit Pricing for IBM Z options are designed

³ The size of virtual L4 for the Max5 and Max16 is 1024MB.

to deliver unmatched simplicity and predictability of hardware capacity and software pricing, even in the constantly evolving era of hybrid cloud.

- ▶ The IBM z16 A02 and IBM z16 AGZ enhancements in resiliency include a capability called IBM Z Flexible Capacity for Cyber Resiliency. With Flexible Capacity for Cyber Resiliency, you can remotely shift capacity and production workloads between IBM z16 A02 and IBM z16 AGZ systems at different sites on demand with no on-site personnel or IBM intervention. This capability is designed to help you proactively avoid disruptions from unplanned events, as well from planned scenarios such as site facility maintenance.
- ▶ IBM z16 A02 and IBM z16 AGZ System Recovery Boost enhancements provide boosted processor capacity and parallelism for specific events. Client-selected middleware starts and restarts to expedite recovery for middleware regions and restore steady-state operations as soon as possible. SVC dump processing and HyperSwap configuration load and reload are boosted to minimize the impact to running workloads.
- ▶ On IBM z16 A02 and IBM z16 AGZ, the enhanced ICA-SR coupling link protocol provides up to 10% improvement for read requests and lock requests, and up to 25% for write requests and duplexed write requests, compared to CF service times on IBM z15 T02 systems. The improved CF service times for CF requests can translate into better Parallel Sysplex coupling efficiency and therefore, may reduce software costs for the attached z/OS images in the Parallel Sysplex.
- ▶ IBM z16 A02 and IBM z16 AGZ provide improved CF processor scalability for CF images. Compared to IBM z15 T02, the relative scaling of a CF image beyond a 9-way is significantly improved, meaning that the effective capacity of IBM z16 A02 and IBM z16 AGZ CF images continue to increase all the way up to the maximum of 16 processors in a CF image.

The IBM z16 A02 and IBM z16 AGZ also added functions to protect today's data now, and from future cyber attacks that can be initiated by quantum computers. The IBM z16 A02 and IBM z16 AGZ generation provides the following quantum-safe capabilities:

- ▶ Key generation
- ▶ Encryption
- ▶ Key encapsulation mechanisms
- ▶ Hybrid key exchange schemes
- ▶ Dual digital signature schemes

In addition to the quantum-safe cryptographic capabilities, tools such as IBM Application Discovery and Delivery Intelligence (ADDI), Integrated Cryptographic Service Facility (ICSF), and IBM Crypto Analytics Monitor (CAT) can help you discover where and what cryptography is used in applications. This can aid in developing a cryptographic inventory for migration and modernization planning.

1.2.1 Supported upgrade paths

Upgrades from previous server generations are available as "Frame Roll MES". Frame roll MES upgrades are disruptive. See Table 1-2.

Table 1-2 Supported Frame Roll MES upgrades

From/To System	IBM z16 A02 and IBM z16 AGZ
IBM z14 ZR1	Y
IBM z15 T02	Y

Concurrent upgrades are available for CPs, IFLs, ICFs, and Z Integrated Information Processors (zIIPs). However, concurrent processor unit upgrades require that more processor units are physically installed, but not activated previously. Upgrades from previous IBM zSystems generations are disruptive via frame roll MES.

Feature upgrades within IBM z16 A02 and IBM z16 AGZ are concurrent from Max5 to Max16 and from Max32 to Max68, while upgrades from Max5 or Max16 to Max32 are disruptive (require system downtime). IBM z16 AGZ upgrade from Max32 to Max68 are concurrent only when plan-ahead feature is ordered.

For more information see also 2.8.1, “Upgrades” on page 58.

1.2.2 Capacity and performance

The IBM z16 A02 and IBM z16 AGZ offer 156 capacity levels for z/OS. In all, there are 26 subcapacity levels for up to 6 CPs ($26 \times 6 = 156$ subcapacity levels) for z/OS workloads.

The IBM z16 A02 and IBM z16 AGZ provide increased processing and enhanced I/O capabilities over the predecessor, IBM z15 T02. This capacity is achieved by increasing the number of PUs per system, redesigning the system cache and introducing new I/O technologies.

IBM z16 A02 and IBM z16 AGZ provide 14% more z/OS processor capacity (6 CPs) and up to 25% full system capacity (Max68) compared to the IBM z15 T02 Max65. Uniprocessor performance has also increased 13% for full speed processor over its predecessor, IBM z15 T02 (IBM z16 A02 and IBM z16 AGZ capacity model Z01 over z15 T02 capacity model Z01⁴).

The Integrated Facility for Linux (IFL) and IBM Z Integrated Information Processor (zIIP) processor units on the IBM z16 A02 and IBM z16 AGZ can be configured to run two simultaneous threads in a single processor (SMT). SMT increases the capacity of these processors with 25% in average over the same processors running single thread. SMT is also enabled by default on System Assist Processors (SAPs).

This comparison is based on the Large System Performance Reference (LSPR) mixed workload analysis. The range of performance ratings across the individual LSPR workloads is likely to have a large spread. More performance variation of individual logical partitions (LPARs) is available when an increased number of partitions and more PUs are available. For more information, see Chapter 12, “Performance” on page 453.

For more information about performance, [see the LSPR website](#).

For more information about millions of service units (MSUs) ratings, see the [IBM zSystems Software Contracts](#) website.

1.2.3 Supported operating systems

The IBM z16 A02 and IBM z16 AGZ are supported by a large set of software products and programs, including independent software vendor (ISV) applications. Use of various features might require the latest releases.

The following operating systems are supported on the IBM z16 A02 and IBM z16 AGZ:

- ▶ z/OS Version 2 Release 5 with PTFs
- ▶ z/OS Version 2 Release 4 with PTFs

⁴ Observed performance increases vary depending on the workload types.

- ▶ z/OS Version 2 Release 3 with PTFs
- ▶ z/OS Version 2 Release 2 with PTFs (toleration support only)
- ▶ z/VM Version 7 Release 3
- ▶ z/VM Version 7 Release 2 with PTFs
- ▶ z/VM Version 7 Release 1 with PTFs
- ▶ z/VSE Version 6 Release 2 with PTFs
- ▶ z/TPF Version 1 Release 1 (compatibility support)

IBM plans to support 21st Century Software VSEⁿ V6.3 on IBM z16 A02 and IBM z16 AGZ, see 7.2.4, “21st Century Software z/VSEⁿ V6.3” on page 245.

IBM plans to support the following Linux on IBM Z distributions on IBM z16 A02 and IBM z16 AGZ:

- ▶ SUSE SLES 15 SP4 and SUSE SLES 12 SP5
- ▶ Red Hat RHEL 8.7 and Red Hat RHEL 9.1
- ▶ Ubuntu 22.04 LTS, and Ubuntu 20.04.1 LTS

The support statements for the IBM z16 A02 and IBM z16 AGZ also cover the KVM hypervisor on distribution levels that have KVM support.

For more information about the features and functions that are supported on IBM z16 A02 and IBM z16 AGZ by operating system, see Chapter 7, “Operating system support” on page 241.

1.2.4 Supported IBM compilers

The following IBM compilers for IBM zSystems can be used with the IBM z16 A02 and IBM z16 AGZ:

- ▶ Enterprise COBOL for z/OS
- ▶ Enterprise PL/I for z/OS
- ▶ Automatic Binary Optimizer
- ▶ z/OS XL C/C++
- ▶ XL C/C++ for Linux on IBM Z

The compilers increase the return on your investment in IBM zSystems hardware by maximizing application performance by using the compilers’ advanced optimization technology for z/Architecture. Through their support of web services, XML, and Java, they allow for the modernization of assets in web-based applications. They also support the latest IBM middleware products (CICS, Db2, and IMS), which allows applications to use their latest capabilities.

To fully use the capabilities of the IBM z16 A02 and IBM z16 AGZ, you must compile it by using the minimum level of each compiler. To obtain the best performance, you must specify an architecture level of 14 by using the **ARCH(14)** option.

For more information, see 7.5.7, “z/OS XL C/C++ considerations” on page 312.

1.3 IBM z16 A02 and IBM z16 AGZ technical overview

This section briefly reviews the following main elements of the IBM z16 A02 and IBM z16 AGZ (M/T 3932):

- ▶ Packaging: IBM z16 A02 and IBM z16 AGZ;
- ▶ CPC drawers

- ▶ I/O subsystem and I/O drawers
- ▶ Storage connectivity
- ▶ Network connectivity
- ▶ Clustering connectivity
- ▶ Cryptography
- ▶ Supported connectivity and crypto features
- ▶ Special-purpose features and functions

1.3.1 IBM z16 A02 - IBM factory frame

The IBM z16 A02 configuration is delivered in an industry standard 19-inch frame, installed and cabled at the IBM factory. The z16 A02 is a single frame, air-cooled system.

The frame forms the IBM z16 A02 CPC and contains one or two CPC drawers and supports up to three PCIe+ I/O drawers. PCIe+ I/O drawers can be added concurrently⁵.

In addition, the IBM z16 A02 (new builds and MES orders) offers top-exit options for the fiber optic and copper cables (used for I/O and power). These options (*Top Exit Power* and *Top Exit I/O Cabling*) give you more flexibility in planning where the system is installed. This flexibility potentially frees you from running cables under a raised floor, which increases air flow over the system.

The IBM z16 A02 supports installation on raised floor and non-raised floor environments.

The internal, front, and rear views of the IBM z16 A02 with two CPC drawers and the maximum of three PCIe+ I/O drawers are shown in Figure 1-1.

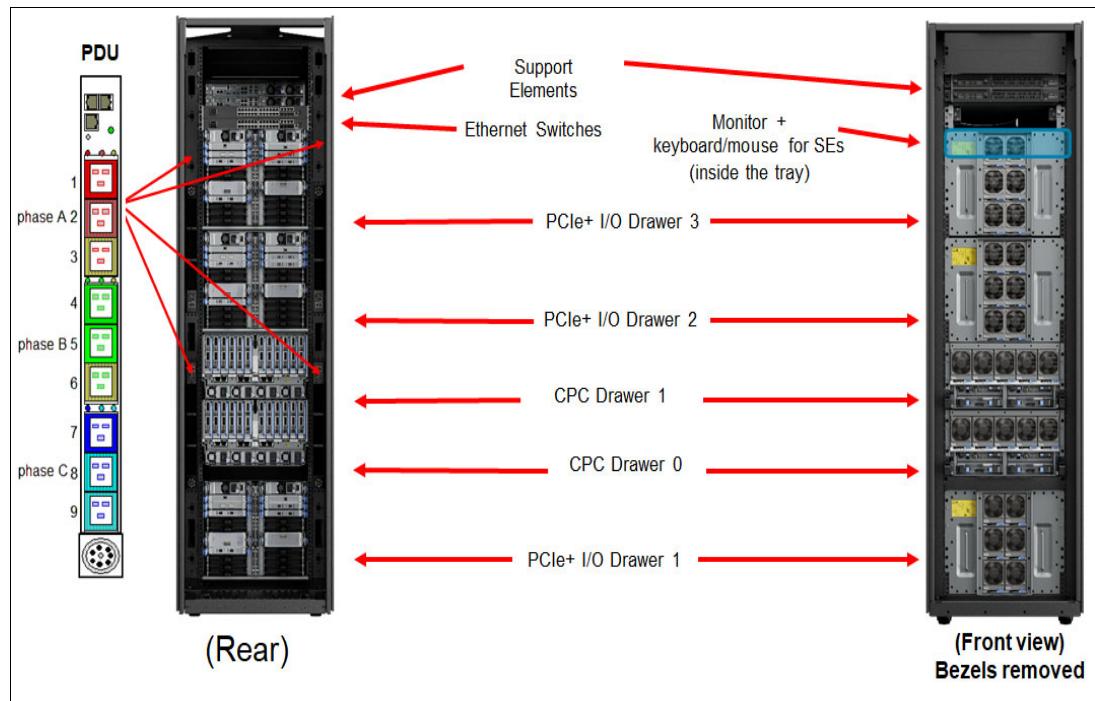


Figure 1-1 z16 A02 (full configuration) front and rear views with four PCIe+ I/O drawers

⁵ The number of available PCIe fanout slots depends on the CPC drawer feature (Max5, Max16, Max32, and Max68) and the number of short reach (SR) coupling features installed in the system (each SR feature occupies one fanout slot).

1.3.2 IBM z16 AGZ

The IBM z16 AGZ is delivered as a bundle (CPC drawer(s), PCIe+ I/O drawer(s), Support Elements, internal network switches, and associated cables) which is installed in a client supplied 19-inch rack. Power to the system is supplied using client supplied PDUs.

The IBM z16 AGZ is installed by the IBM System Service Representative (SSR) in the client supplied 19-inch" rack. Additional planning considerations are needed for IBM z16 AGZ (components' location, cabling and power feeds).

Note: It is clients' responsibility to provide rack and power that meet the z16 specifications for size, service clearance, and power needs per the planning tool instructions in the *IBM z16 and LinuxONE Rockhopper 4 Rack Mount Bundle Installation Manual for Physical Planning (IMPP)*, GC28-7035. required to install an IBM z16 AGZ rack mount system. Also, Refer to Chapter 11 for power and environmental requirements.

A sample configuration for an IBM z16 AGZ featuring one CPC drawer and one PCIe+ I/O drawer is shown in Figure 1-2.

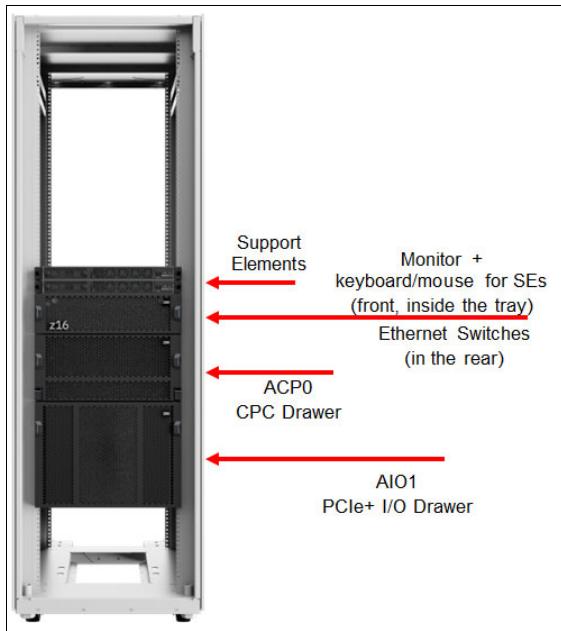


Figure 1-2 IBM z16 AGZ installed in standard 19" rack

1.3.3 CPC drawers

The IBM z16 AGZ can be configured with one or two CPC drawers that contain the following elements:

- ▶ Max5 and Max16 (single CPC drawer):
 - Two Dual Chip Modules (DCMs) containing 22 active processor unit cores (air-cooled). The processor unit cores run at 4.6 GHz each
- ▶ Max32 and Max68 (single- and dual- CPC drawer):
 - Four Dual Chip Modules (DCMs) with 40 active cores per CPC drawer, air cooled
- ▶ Memory:

- A minimum of 64 GB and a maximum of 8 TB per CPC drawer for a total of 16 TB per system (excluding 160 GB HSA) is available for client use.
 - Eight Channel RAIM (eight equal size DIMMs per channel, DIMM sizes include RAIM overhead)
 - Eight DIMMs of equal size in each memory feature
 - Two, four, five, or six features (up to 48 DIMMs) are plugged in each drawer
 - Features of different DIMMs sizes can be mixed in the same drawer
- Fanouts:
- Each CPC drawer provides up to 12 PCIe Gen4 fanout slots to accommodate PCIe+ Gen3 I/O fanouts (to connect to the PCIe+ I/O drawers) and Integrated Coupling Adapter Short Reach (ICA SR) coupling links. The number of fanouts that can be installed depends on the system feature.
- Each fanout feature includes the following configurations:
 - Two-port PCIe+ Gen3 (16 GBps each port) I/O fanout, each port supporting one domain in a 16-slot PCIe+ I/O drawer.
 - Two-port ICA SR or ICA SR1.1 PCIe fanout for coupling links (two links, 8 GBps each).
- Four Power Supply Units (PSUs) that provide power to the CPC drawer, hot swappable, accessible from the rear.

The loss of one PSU leaves enough power to satisfy the power requirements of the entire drawer. The PSUs can be concurrently maintained.

- Two combined Baseboard Management Cards (BMCs) and Oscillator cards that provide redundant interfaces to the internal management network and clock synchronization for the CPC. The combined BMC/OSC are located in the front of the CPC drawer.

With the IBM z16 A02 and IBM z16 AGZ, up to four Virtual Flash Memory (VFM) features are offered from the main memory capacity in 0.5 TB units to increase granularity for the feature. VFM can provide much simpler management and better performance by reducing the I/O to the adapters in the PCIe+ I/O drawers.

1.3.4 I/O subsystem and I/O drawers

The IBM z16 A02 and IBM z16 AGZ implement PCIe Generation 4 switch cards, which are used to connect the dual port fanout features in the CPC drawers to the I/O features in the I/O drawers. The I/O infrastructure is designed to reduce processor usage and I/O latency, and provide increased throughput and availability.

For a two CPC drawer system, up to 24 PCIe+ fanout slots can be populated with fanout cards for data communications between the CPC drawers and the I/O infrastructure, and for coupling. The multiple channel subsystem (CSS) architecture allows up to three CSSs, each with 256 channels.

PCIe+ I/O drawer

The *PCIe+ I/O drawer*, together with the PCIe features, offers finer granularity and capacity over previous I/O infrastructures. It can be concurrently added and removed in the field⁶, which eases planning. Only PCIe cards (features) are supported, in any combination.

⁶ Plan ahead is required to concurrently add a PCIe+ I/O drawer to an IBM z16 AGZ.

The IBM z16 A02 and IBM z16 AGZ support Generation 3 PCIe-based infrastructure by using PCIe+ I/O drawers (PCIe Gen3) for PCIe features (adapters). The number of supported PCIe+ I/O drawers is listed in Table 1-3.

Table 1-3 IBM z16 A02 and IBM z16 AGZ fanouts per feature

Feature name ^a	PU DCMs	Max. PCIe fanouts	Max. PCIe+ I/O drawers
Max5 (FC 0672)	2	6	3
Max16 (FC 0673)	2	6	3
Max32 (FC 0674)	4	12	3
Max68 (FC 0675)	8	24	3

a. IBM z16 AGZ supports the same features as IBM z16 A02.

1.3.5 Storage connectivity

Storage connectivity is provided by FICON Express and the IBM zHyperLink Express features.

FICON Express

FICON Express features follow the established Fibre Channel (FC) standards to support data storage and access requirements, along with the latest FC technology in storage and access devices. FICON Express features support the following protocols:

- ▶ FICON

This enhanced protocol (as compared to FC) provides for communication across channels, channel-to-channel (CTC) connectivity, and with FICON devices, such as disks, tapes, and printers. It is used in z/OS, z/VM®, z/VSE® (Virtual Storage Extended), z/TPF (Transaction Processing Facility), and Linux on IBM Z environments.
- ▶ Fibre Channel Protocol (FCP)

This standard protocol is used for communicating with disk and tape devices through FC switches and directors. The FCP channel can connect to FCP SAN fabrics and access FCP/SCSI devices. FCP is used by z/VM, KVM, z/VSE, and Linux on IBM Z environments.

FICON Express32S features are implemented by using PCIe cards, and offer better port granularity and improved capabilities over the previous FICON Express features. FICON Express32S features support a link data rate of 32 gigabits per second (Gbps) (8, 16, or 32 Gbps auto-negotiate), and it is the preferred technology for new systems.

zHyperLink Express

zHyperLink was created to provide fast access to data by way of low-latency connections between the IBM zSystems platform and storage.

The zHyperLink Express1.1 feature allows you to make synchronous requests for data that is in the storage cache of the IBM DS8900F. This process is done by directly connecting the zHyperLink Express1.1 port in the IBM z16 A02 or IBM z16 AGZ to an I/O Bay port of the IBM DS8000®. This short distance (up to 150 m [492 feet]), direct connection is designed for low-latency reads and writes, such as with IBM DB2® for z/OS synchronous I/O reads and log writes.

Working with the FICON SAN Infrastructure, zHyperLink can improve application response time, which cuts I/O-sensitive workload response time in half without requiring application changes.⁷

Note: The zHyperLink channels complement FICON channels, but they do *not* replace FICON channels. FICON remains the main data driver and is mandatory for zHyperLink usage.

1.3.6 Network connectivity

The IBM z16 A02 and IBM z16 AGZ are a fully virtualized platform that can support many system images at once. Therefore, network connectivity covers not only the connections between the platform and external networks with Open Systems Adapter-Express (OSA-Express) and RoCE Express features, it supports specialized internal connections for intra-system communication through IBM HiperSockets and Internal Shared Memory (ISM).

OSA-Express

The OSA-Express features provide local area network (LAN) connectivity and comply with IEEE standards. In addition, OSA-Express features assume several functions of the TCP/IP stack that normally are performed by the PU, which allows significant performance benefits by offloading processing from the operating system.

OSA-Express7S 1.2 features continue to support copper and fiber optic (single-mode and multi-mode) environments.

HiperSockets

IBM HiperSockets is an integrated function of the IBM zSystems platforms that supplies attachments to up to 32 high-speed virtual LANs, with minimal system and network overhead.

HiperSockets is a function of the Licensed Internal Code (LIC). It provides LAN connectivity across multiple system images on the same IBM zSystems platform by performing memory-to-memory data transfers in a secure way. The HiperSockets function eliminates the use of I/O subsystem operations. It also eliminates having to traverse an external network connection to communicate between LPARs in the same IBM zSystems platform. In this way, HiperSockets can help with server consolidation by connecting virtual servers and simplifying the enterprise network.

RoCE Express

The 25 GbE and 10 GbE RoCE Express3 features⁸ use Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE) to provide fast memory-to-memory communications between two IBM zSystems platforms.

These features are designed to help reduce consumption of CPU resources for applications that use the TCP/IP stack (such as IBM WebSphere® that accesses an IBM Db2® database). They can also help reduce network latency with memory-to-memory transfers by using Shared Memory Communications over RDMA (SMC-R).

With SMC-R, you can transfer huge amounts of data quickly and at low latency. SMC-R is transparent to the application and requires no code changes, which enables rapid time to value.

⁷ The performance results can vary depending on the workload. Use zBNA tool for the zHyperLink planning.

⁸ RoCE Express features can also be used as general-purpose Network Interface Cards (NICs) with Linux on IBM Z.

The RoCE Express3 features can also provide local area network (LAN) connectivity for Linux on IBM Z, and comply with IEEE standards. In addition, RoCE Express features assume several functions of the TCP/IP stack that normally are performed by the PU, which allows significant performance benefits by offloading processing from the operating system.

Internal Shared Memory

ISM is a virtual Peripheral Component Express (PCI) network adapter that enables direct access to shared virtual memory, providing a highly optimized network interconnect for IBM zSystems platform intra-communications. Shared Memory Communications-Direct Memory Access (SMC-D) uses ISM. SMC-D optimizes operating systems communications in a way that is transparent to socket applications. It also reduces the CPU cost of TCP/IP processing in the data path, which enables highly efficient and application-transparent communications.

SMC-D requires no extra physical resources (such as RoCE Express features, PCIe bandwidth, ports, I/O slots, network resources, or Ethernet switches). Instead, SMC-D uses LPAR-to-LPAR communication through HiperSockets or an OSA-Express feature for establishing the initial connection.

z/OS and Linux on IBM Z support SMC-R and SMC-D. Now, data can be shared by way of memory-to-memory transfer between z/OS and Linux on IBM Z.

1.3.7 Clustering connectivity

A Parallel Sysplex is an IBM zSystems clustering technology that is used to make applications that are running on logical and physical IBM zSystems platforms highly reliable and available. The IBM zSystems platforms in a Parallel Sysplex are interconnected by way of coupling links.

Coupling connectivity on the IBM z16 A02 and IBM z16 AGZ uses Coupling Express2 Long Reach (CE2 LR) and Integrated Coupling Adapter Short Reach (ICA SR and ICA SR1.1) features. The ICA SR feature supports distances up to 150 meters (492 feet), while the CE2 LR feature supports unrepeatable distances of up to 10 Kim (6.21 miles) between IBM zSystems platforms. ICA SR features provide sysplex/timing connectivity direct to the CPC drawer, while Coupling Express2 LR features connect into the PCIe+ I/O Drawer.

Coupling links can also carry timing information (Server Time Protocol - STP) for synchronizing time across multiple IBM zSystems CPCs in a Coordinated Time Network (CTN).

For more information about coupling and clustering features, see 4.5, “I/O features” on page 148.

1.3.8 Cryptography

IBM z16 A02 and IBM z16 AGZ provide two main cryptographic functions: CP Assist for Cryptographic Functions (CPACF) and Crypto-Express8S.

CPACF

CPACF is a high performance, low-latency coprocessor that performs symmetric key encryption operations and calculates message digests (hashes) in hardware. The following algorithms are supported:

- ▶ AES
- ▶ Data Encryption Standard (DES) and Triple Data Encryption Standard (TDES)

- ▶ Secure Hash Algorithm (SHA)-1
- ▶ SHA-2
- ▶ SHA-3

CPACF supports Elliptic Curve Cryptography (ECC) clear key, improving the performance of Elliptic Curve algorithms. The following algorithms are supported:

- ▶ EdDSA (Ed448 and Ed25519)
- ▶ ECDSA (P-256, P-384, and P-521)
- ▶ ECDH (P-256, P-384, P521, X25519, and X448)
- ▶ Support for protected key signature creation

Crypto-Express8S

The tamper-sensing and tamper-responding Crypto-Express8S features provide acceleration for high-performance cryptographic operations and support up to 40 domains with IBM z16 A02 and IBM z16 AGZ. This specialized hardware performs AES, DES and TDES, RSA, Elliptic Curve (ECC), SHA-1, and SHA-2, and other cryptographic operations.

It supports specialized high-level cryptographic APIs and functions, including those functions that are required with quantum-safe cryptography and in the banking industry.

Crypto-Express8S features are designed to meet the Federal Information Processing Standards (FIPS) 140-2 Level 4 and PCI HSM security requirements for hardware security modules.

IBM z16 A02 and IBM z16 AGZ are the industry's first quantum-safe systems⁹.

- ▶ IBM z16 A02 and IBM z16 AGZ quantum-safe secure boot technology helps to protect IBM zSystems firmware from quantum attacks through a build-in dual signature scheme with no changes required.
- ▶ IBM z16 A02 and IBM z16 AGZ quantum-safe technology and key management services, were developed to help you protect data and keys against a potential future quantum attack like harvest now, decrypt later.
- ▶ IBM z16 A02 and IBM z16 AGZ will position clients to use quantum-safe cryptography along with classical cryptography as they begin modernizing existing applications and building new applications.

For more information about cryptographic features and functions, see Chapter 6, “Cryptographic features” on page 197.

1.3.9 Supported connectivity and crypto features

The IBM z16 A02 and IBM z16 AGZ provide a PCIe-based infrastructure for the PCIe+ I/O drawers to support the following features:

- ▶ Storage connectivity:
 - zHyperLink Express1.1 (new build and carry forward)
 - zHyperLink Express (carry forward only)
 - FICON Express32S (new build only)

⁹ DISCLAIMER: IBM z16 A02 and IBM z16 AGZ with the Crypto Express 8S card provide quantum-safe APIs providing access to quantum-safe algorithms which have been selected as finalists during the PQC standardization process conducted by NIST.

<https://https://csrc.nist.gov/Projects/post-quantum-cryptography/post-quantum-cryptography-standardization/round-3-submissions>. Quantum-safe cryptography refers to efforts to identify algorithms that are resistant to attacks by both classical and quantum computers, to keep information assets secure even after a large-scale quantum computer has been built. Source: <https://www.etsi.org/technologies/quantum-safe-cryptography>. These algorithms are used to help ensure the integrity of a number of the firmware and boot processes. IBM z16 A02 and IBM z16 AGZ, are the Industry-first system protected by quantum-safe technology across multiple layers of firmware.

- FICON Express16S+ (carry forward only)
- ▶ Network connectivity:
 - OSA-Express7S 1.2 (new build only)
 - OSA-Express6S (carry forward only)
 - RoCE Express3 (new build only)
 - RoCE Express2.1 (carry forward only)
 - RoCE Express2 (carry forward only)
- ▶ Cryptographic features:
 - Crypto Express8S, one or two HSMs¹⁰ (new build only)
 - Crypto Express7S, 1-port or 2-port¹¹ (carry forward only)
 - Crypto Express6S (carry forward only)
- ▶ Clustering connectivity¹²:
 - ICA SR1.1 (new build or carry forward)
 - ICA SR (carry forward only)
 - Coupling Express2 Long Reach (new build only)

1.3.10 Special-purpose features and functions

When it comes to designing and developing the IBM zSystems platform, IBM takes a *total systems* view. The IBM zSystems stack is built around digital services, agile application development, connectivity, and system management. This design approach creates an integrated, diverse platform with specialized hardware and dedicated computing capabilities.

The IBM z16 A02 and IBM z16 AGZ deliver a range of features and functions, allowing PUs to concentrate on computational tasks, while distinct, specialized features take care of the rest. For more information about these features and other IBM z16 A02 and IBM z16 AGZ features, see 3.5, “Processor unit functions” on page 93.

1.4 Hardware management

The Hardware Management Consoles (HMCs) and Support Elements (SEs) are appliances that together provide platform management for IBM zSystems.

The HMC is an appliance that is designed to provide a single point of control for managing local or remote hardware elements.

For IBM z16 A02 and IBM z16 AGZ new builds, IBM zSystems Hardware Management Appliance (FC 0129) is the only Hardware Management Console available. Both Hardware Management Console Appliance and Support Element Appliance run virtualized on the Support Element hardware.

Older standalone HMCs (rack mounted or tower) can be carried forward during an MES upgrade to IBM z16 A02 or IBM z16 AGZ.

For more information, see Chapter 10, “Hardware Management Console and Support Element” on page 391.

¹⁰ The Crypto Express8S is available with either one or two hardware security modules (HSM). The HSM is the IBM 4770 PCIe Cryptographic Coprocessor (PCIeCC).

¹¹ The Crypto Express7S comes with either one (1-port) or two (2-port) hardware security modules (HSM). The HSM is the IBM 4769 PCIe Cryptographic Coprocessor (PCIeCC).

¹² ICA SR and ICA SR1.1 features connect directly into the CPC (processor) Drawer, while Coupling Express2 Long Reach connects into the PCIe+ I/O Drawer.

1.5 Reliability, availability, and serviceability

This section provides an overview of the IBM z16 A02 and IBM z16 AGZ RAS characteristics. For detailed RAS information, see Chapter 9, “Reliability, availability, and serviceability” on page 363.

System reliability, availability, and serviceability (RAS) is an area of continuous IBM focus and a defining IBM zSystems platform characteristic. The RAS objective is to reduce, or eliminate if possible, all sources of planned and unplanned outages while providing adequate service information if an issue occurs. Adequate service information is required to determine the cause of an issue without the need to reproduce the context of an event.

IBM zSystems platforms are designed to enable highest availability and lowest downtime. These facts are recognized by various IT analysts, such as ITIC¹³ and IDC¹⁴.

Comprehensive, multi-layered strategy includes the following features:

- ▶ Error Prevention
- ▶ Error Detection and Correction
- ▶ Error Recovery
- ▶ System Recovery Boost

With a properly configured IBM z16 A02 and IBM z16 AGZ, further reduction of outages can be attained through First Failure Data Capture (FFDC), which is designed to reduce service times and avoid subsequent errors. It also improves nondisruptive replace, repair, and upgrade functions for memory, drawers, and I/O adapters. IBM z16 A02 and IBM z16 AGZ support the nondisruptive download and installation of LIC updates.

IBM z16 A02 and IBM z16 AGZ RAS features provide unique high-availability and nondisruptive operational capabilities that differentiate IBM zSystems in the marketplace. IBM z16 A02 and IBM z16 AGZ RAS enhancements are made on many components of the CPC (processor chip, memory subsystem, I/O, and service) in areas, such as error checking, error protection, failure handling, error checking, faster repair capabilities, sparing, and cooling.

The ability to cluster multiple systems in a Parallel Sysplex takes the commercial strengths of the z/OS platform to higher levels of system management, scalable growth, and continuous availability.

The IBM z16 A02 and IBM z16 AGZ builds on the RAS of the IBM z15 family with the following RAS improvements:

- ▶ System Recovery Boost
 - System Recovery Boost was introduced with IBM z15. It offers customers more Central Processor (CP) capacity during system recovery operations to accelerate the startup (IPL), shutdown or stand-alone dump operations (at image level - LPAR¹⁵). System Recovery Boost requires operating system support. No other IBM software charges are made during the boost period.

System Recovery Boost can be used during LPAR shutdown or start to make the running operating system and services available in a shorter period.

The System Recovery Boost provides the following options for the capacity increase:

- Subcapacity CP speed boost: During the boost period, subcapacity engines allocated to the boosted LPAR are transparently activated at their full capacity (CP engines).

¹³ For more information, see [ITIC Global Server Hardware, Server OS Reliability Report](#).

¹⁴ For more information, see [Quantifying the Business Value of IBM zSystems](#).

¹⁵ Logical partition (LPAR) running an Operating System image.

- zIIP Capacity Boost: During the boost period, all active zIIPs that are assigned to an LPAR are used to extend the CP capacity (CP workload is dispatched to zIIP processors during the boost period).
- System Recovery Boost enhancements delivered with the IBM z16 A02 and IBM z16 AGZ maximize service availability by using tailored short-duration boosts to mitigate the impact of these recovery processes:
 - SVC Dump boost will boost the system on which SVC dump is taken, to reduce system impact and expedite diagnostic capture. It is possible to enable/disable/set thresholds for this option.
 - Middleware restart/recycle boost will boost the system on which a middleware instance is being restarted, to expedite resource recovery processing, release retained locks, and so on. It is applicable to planned restarts, or restarts after failure, automated, or ARM-driven restarts. System Recovery Boost will not boost any system address spaces by default and must be explicitly configured by the WLM policy specification.
 - HyperSwap configuration load boost will boost the system in which the HyperSwap configuration and policy information are being loaded/re-loaded. This applies to both Copy Services Manager (CSM) and GDPS. HyperSwap Configuration Load boost is enabled by default. There are no thresholds or criteria applied to the boost request based on the size or number of devices present in the HyperSwap configuration.

Through System Recovery Boost, the IBM z16 A02 and IBM z16 AGZ continue to offer more CP capacity during particular system recovery operations to accelerate system (operating system and services) start when the system is being started or shutdown. System Recovery Boost is operating system-dependent. No other hardware, software, or maintenance charges are required during the boost period for the base functions of System Recovery Boost.

- At the time of this writing, the main System Recovery Boost users are z/OS (running in an LPAR), z/VM, z/VSE, and z/TPF, as well as standalone dump (SADMP).

z/VM uses the System Recovery Boost if it runs on subcapacity CP processors only (IFLs are always at their full clock speed). Second-level z/VM guest operating systems¹⁶ can inherit the boost if they are running on CPs.

For more information about RAS and System Recovery Boost, see *Introducing IBM Z System Recovery Boost*, REDP-5563.

- ▶ Level 2 (physical), Level 3 and Level 4 (virtual) cache enhancements include the use symbol ECC to extend the reach of prior IBM zSystems generations cache and memory improvements for augmented availability. The L2, L3, and L4 cache powerful symbol ECC is designed to make it resistant to more failure mechanisms. Preemptive DRAM marking is added to the main memory to isolate and recover failures more quickly.

¹⁶ z/OS configured as a guest system under z/VM does not use the boost.



Central processor complex hardware components

This chapter provides information about the new IBM z16 A02 and IBM z16 AGZ, their hardware building blocks, and how these components physically interconnect. This information is useful for planning purposes and can help in defining configurations that meet your requirements.

This chapter includes the following topics:

- ▶ 2.1, “System overview: frames and drawers” on page 22
- ▶ 2.2, “CPC drawer” on page 29
- ▶ 2.3, “Dual chip modules” on page 37
- ▶ 2.4, “PCIe+ I/O drawer” on page 41
- ▶ 2.5, “Memory” on page 43
- ▶ 2.7, “Connectivity” on page 53
- ▶ 2.8, “Processor configurations” on page 57
- ▶ 2.9, “Power and cooling” on page 65
- ▶ 2.10, “Summary” on page 67

2.1 System overview: frames and drawers

The IBM z16 A02 is delivered factory installed in an IBM 19-inch frame - or as IBM z16 AGZ or z16 rack mount, that can be easily installed in a *client supplied* 19-inch (industry standard) form factor frame in any data center.

The IBM z16 A02 and IBM z16 AGZ can be configured with one or two central processor complex (CPC) drawers, and up to three Peripheral Component Interconnect Express+ (PCIe+) I/O drawers. Figure 2-1 shows the differences between the available IBM z16 A02 and IBM z16 AGZ air cooled systems.



Figure 2-1 IBM z16 A02 and IBM z16 AGZ

Similar to the predecessor z15 T02, the redesigned CPC drawer and I/O infrastructure also lowers power consumption, reduces the footprint, and allows installation in virtually any data center. The IBM z16 A02 and IBM z16 AGZ are rated for ASHRAE class A3¹ data center operating environment.

2.1.1 IBM z16 A02

The IBM z16 A02 continues previous IBM zSystems generations improvements through the following significant characteristics of the modular hardware:

- ▶ All external cabling (power, I/O, and management) is performed at the rear of the system.
- ▶ Flexible configurations: Feature based CPC sizing to fit capacity requirements.
- ▶ Power Distribution Unit (PDU) with single-phase or three-phase utility AC power (configuration dependent).
- ▶ PCIe+ Gen3 I/O drawers (19-inch format) supporting 16 PCIe adapters each.

The 8U and 16U Reserved space features discontinued: previously offered on z14 ZR1 and z15 T02 systems are not offered on IBM z16 A02 and cannot be carried forward. All components for the z16 A02 have fixed location in the IBM factory frame. No other hardware can be installed in the IBM z16 A02.

¹ For more information, see Chapter 2, Environmental specifications in the *3932 Single Frame Installation Manual for Physical Planning*, (Models A02/LA2) GC28-7040.

The air cooled IBM z16 A02 includes the following basic hardware building blocks:

- ▶ 19-inch 42u frame (factory installed)
- ▶ CPC (Processor) drawers (one or two)
- ▶ PCIe+ Gen3 I/O drawers (up to three)
- ▶ Air cooled system
- ▶ Power:
 - Single- or three-phase Power Distribution Units (PDU) pairs (two or four PDUs), for systems with a single CPC drawer.
 - Three-phase Power Distribution Units (PDU) pairs (two or four PDUs) for systems with two CPC drawers.
- ▶ Redundant (two) Support Elements:
 - Single KMM² device (USB-C connection)
 - Optional feature code for IBM Hardware Management Appliance (FC0129)
- ▶ Two 24-port 1GbE Switches
- ▶ Hardware for cable management at the rear of the system.

An example of a dual-CPC drawer system (Max68 w/ three PCIe+ I/O drawers) is shown in Figure 2-2

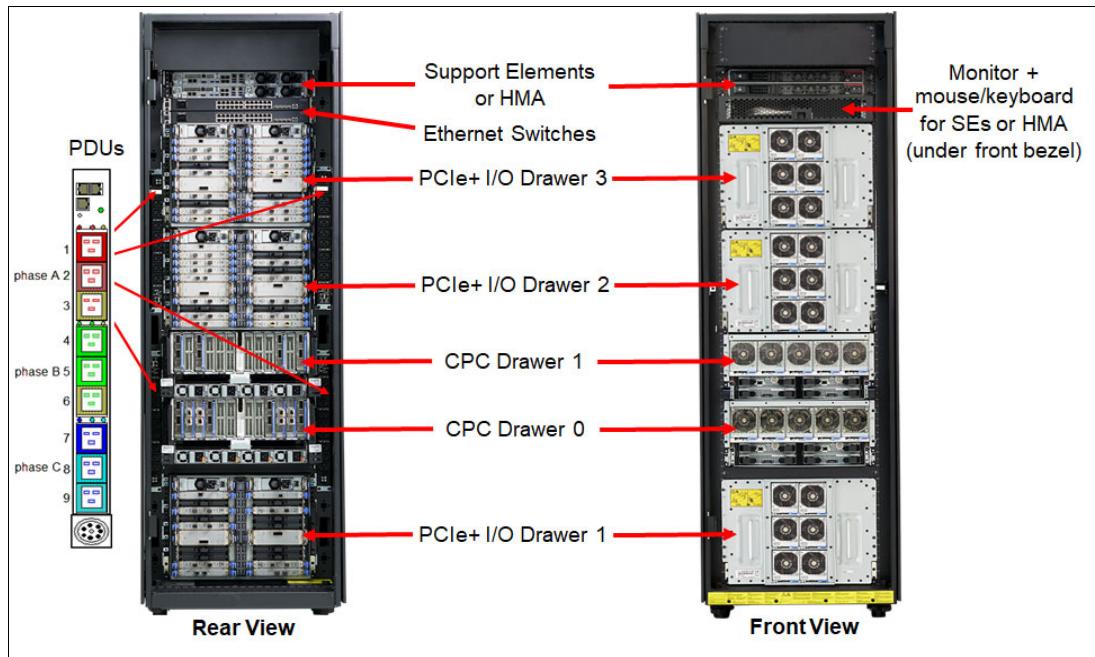


Figure 2-2 IBM z16 A02 - Maximum configuration, dual-CPC drawer system

2.1.2 Top and bottom exit I/O and cabling

For the z16 A02, the top exit of all cables for I/O or power is always an option with no feature codes required. Adjustable cover plates are available for the openings at the top rear of each frame.

All external cabling enters the system at the rear of the frames for all I/O adapters, management LAN, and power connections.

The Top Exit feature code (FC 7802) provides an optional Top Exit *cover enclosure*. The optional Top Exit cover enclosure provides a fiber cable organizer to optimize the cables

² KMM - Keyboard, Mouse, Monitor

storage and strain relief. It also provides mounting locations to secure Fiber Quick Connector (FQC) MPO³ brackets (FC 5827) on the top of the frames.

- FC 7804 provides bottom exit cabling/power support and when ordered with FC 5827 will provide harness brackets for the bottom tailgate and fiber bundle organizer hardware
- FC 7803 is for top exit cabling without the Top Exit *cover enclosure*.
- FC 5827 will provide the MPO mounting brackets that cable harnesses connect to in the Top and/or Bottom Exit tailgates

Overhead I/O cabling is contained within the frames. Extension “chimneys” that were featured with legacy systems are no longer used.

A view of the top rear of the frame and the openings for top exit cables and power is shown in Figure 2-3. When FC 7802 is installed, the plated adjustable shields are removed and the top exit enclosure is installed. Also shown is the top exit encloser installed with the power cables exiting out the top, and the MPO brackets and cabling. If desired, the top exit encloser can be installed with the cable exit facing the front of the system.

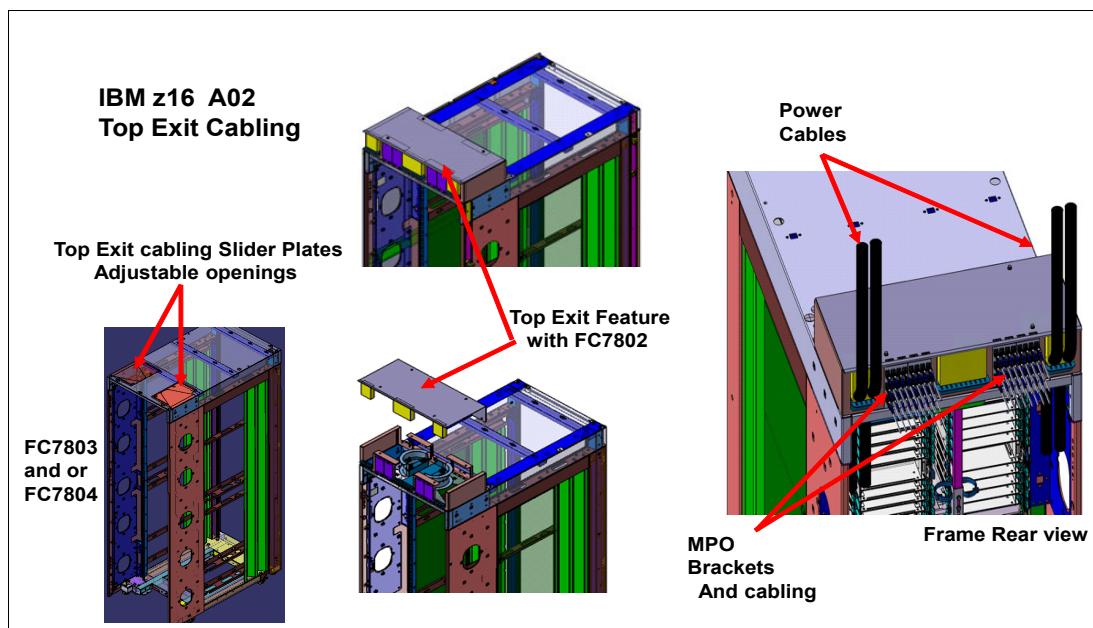


Figure 2-3 IBM z16 A02 Top Exit with and without Feature Codes

Care should be taken when ordering the correct feature codes when cables enter the frames from above the floor, below the floor or both. Also, if the Top Exit feature is desired, which will add the Top Hat (“top exit enclosure”).

Table 2-1 IBM z16 A02 Feature Code Combinations

Environment	Bottom Exit	Top Exit	Features to order
Raised Floor	Yes	No	7804 only
Raised Floor	Yes	Yes but no Top Hat H/W	7804 and 7803
Raised Floor	Yes	Yes but with Top Hat H/W	7804 and 7802
Raised Floor	No	Yes but no Top Hat H/W	7803

³ MPO - Multi-fiber Push On connector

Environment	Bottom Exit	Top Exit	Features to order
Raised Floor	No	Yes but with Top Hat H/W	7802
Non-Raised Floor	No (not supported)	Yes but no Top Hat H/W	7998 and 7803
Non-Raised Floor	No (not supported)	Yes but with Top Hat H/W	7998 and 7802

A vertical cable management guide (“spine”) can assist with proper cable management for fiber, copper, and coupling cables. The full length cable management spine is installed with the presence of the 2nd CPC drawer, 3rd IO drawer or the 1st 8U Reserve feature.

The cable retention clips can be relocated for best usage. All external cabling to the system (from top or bottom) can use the spines to minimize interference with the PDUs that are mounted on the sides of the rack.

The rack with the spine mounted and new optional fiber cable organizer hoops are shown in Figure 2-4. If necessary, the spine, and organizer hoops can be easily relocated for service procedures.

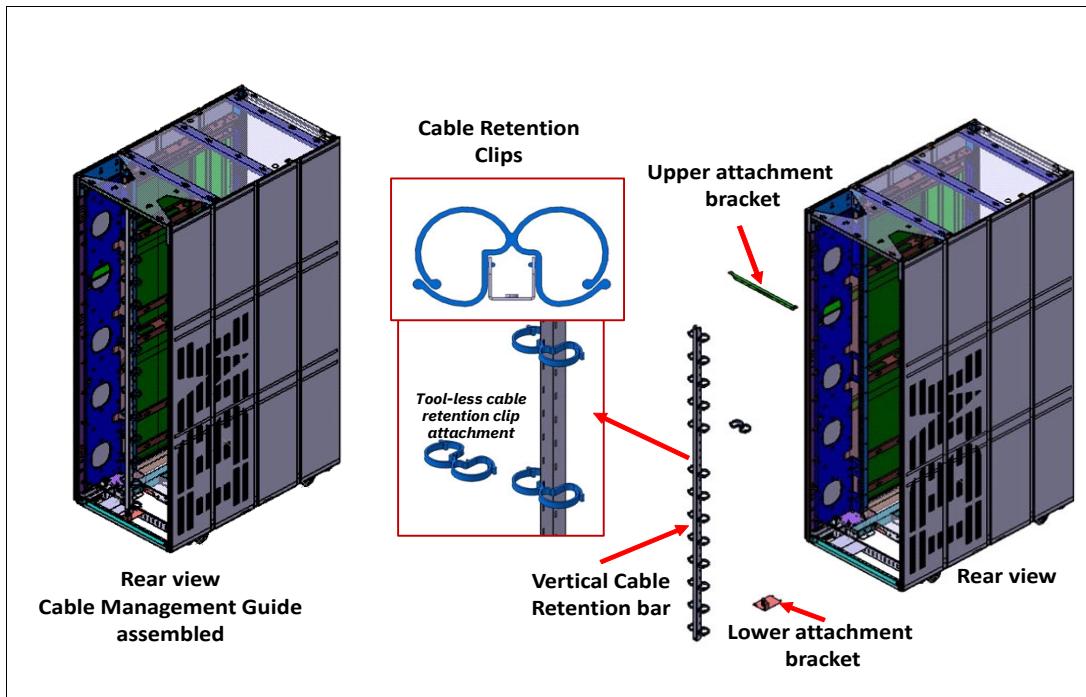


Figure 2-4 IBM z16 A02 Vertical Cable Retention Bar and Cable Retention Clips

2.1.3 IBM z16 A02 system features

The key features that are used to build the IBM z16 A02 configuration are listed in Table 2-2.

Table 2-2 Key features that influence the system configurations

Feature Code	Description	Comments
0513	3932-A02	Supports CPs and specialty engines
0672	One CPC Drawer, one PU DCM	Feature Max5
0673	One CPC Drawer, two PU DCMs	Feature Max16

Feature Code	Description	Comments
0674	One CPC Drawer, four PU DCMs	Feature Max32
0675	Two CPC Drawers, eight PU DCMs	Feature Max68
4010	CPC Drawer	Maximum two (feature dependant)
4014	A Frame	Air cooled
Power		
0510	200-208V 60/30A 3 Phase (Delta - "Δ")	North America and Japan
0511	200-240V 32A single- or 380-415V 3 Phase (Wye - "Y")	Single Phase Worldwide (3 Phase except North America and Japan)
I/O		
4023	PCIe+ I/O drawer	Maximum three (feature dependant)
5827	FQC Bracket & Mounting Hdw	Maximum eight Harness plates for trunking installed in top hat or bottom tailgate
7802	Top Exit Cabling with Tophat	Includes cable management top hat
7803	Top Exit Cabling without Tophat	Uses rear slide plates at top of frame
7804	Bottom exit cabling	Tailgate hardware for cables routing

2.1.4 IBM z16 AGZ system features

The IBM z16 AGZ configurations comes with 1-2 CPC drawers and 0-3 I/O drawers and is sold and delivered without and IBM frame or PDUs. The key features that are used to build the IBM z16 AGZ configuration are listed in Table 2-3.

Table 2-3 Key feature codes that influence the system configurations

Feature Code	Description	Comments
0515	g identifier	Supports CPs and specialty engines
0672	One CPC Drawer, two PU DCMs	Feature Max5
0673	One CPC Drawer, two PU DCMs	Feature Max16
0674	One CPC Drawer, four PU DCMs	Feature Max32
0675	Two CPC Drawers, eight PU DCMs	Feature Max68
2332	CPC1 reserve	Reserve 5u immediately above CPC0 location for future add CPC1
2333	AIO1 reserve	Reserve 8u immediately below CPC location for future add
2334	AIO2 reserve	Reserve 8u immediately above CPC location for future add
2335	AIO3 reserve	Reserve 8u immediately above AIO2 location for future add
4010	CPC Drawer	Maximum two (feature dependant)

Feature Code	Description	Comments
Power		
0534	PWR Jumper 1m w/C13	
0535	PWR Jumper 1m w/C19	
0536	PWR Jumper 2m w/C13	
0537	PWR Jumper 2m w/C19	
0538	PWR Jumper 3m w/C13	
0539	PWR Jumper 3m w/C19	
I/O		
4023	PCIe+ I/O drawer	Maximum three (feature dependant)

2.1.5 System configurations for IBM z16 A02

All system components are designed and integrated in an IBM 19-inch frame for the IBM z16 A02. Sample configuration of the system (rear view) is shown in Figure 2-2 on page 23 (systems with two CPC drawers).

The IBM z16 A02 is built in a 19-inch form factor, 42 EIA units frame (A-frame). The base frame is 40 EIA units high with a 2U removable top. FC 9975 is available if a height reduction is necessary.

Single CPC drawer system

PCIe+ I/O drawers are provided as required by the number of I/O adapters ordered. For systems with a single CPC drawer, the PCIe+ I/O drawers (1, 2, and 3) are installed in order in the following EIA locations: A01B, A20B, and A28B.

The possible configurations for an IBM z16 A02 (viewed from the rear of the system) are shown in “IBM z16 A02 (Factory Frame) configurations” on page 482.

The number of PDUs depends on the system configuration. A system with a single CPC drawer can be powered either from single-phase or three-phase utility power. For more information see Chapter 11, “Environmentals” on page 433 and 3932 *Single Frame Installation Manual for Physical Planning (Models A02/LA2)*, GC28-7040.

Important: IBM z16 A02 *does not* support the installation any other equipment in the rack (as previous systems allowed - i.e. IBM z14 ZR1 and IBM z15 T02).

Dual CPC drawer system or system with FC 0675

For systems with two CPC drawers, the PCIe+ I/O drawers (1, 2, and 3) are installed in order in the following EIA locations: A01B, A20B, and A28B.

There is no need to define Reserved space feature codes for future CPC or I/O drawer empty locations, as all placements in the rack are defined and will be filled according to the features ordered.

The possible configurations for a dual CPC drawer system with the view from the rear of the system are shown in “IBM z16 A02 (Factory Frame) configurations” on page 482.

The number of PDUs depends on the system configuration. A system with a dual CPC drawer can only be powered from three-phase utility power. For more information, see Chapter 11, “Environmentals” on page 437 and *3932 Single Frame Installation Manual for Physical Planning (Models A02/LA2)*, GC28-7040.

2.1.6 System configurations for IBM z16 AGZ

All system components are designed and integrated in a client provided 19-inch rack for the IBM z16 AGZ configurations.

- The client is expected to provide rack level PDU Power.
- IBM will provide drawer requirements for power input.
- IBM firmware will have no control of customer PDUs.
- All components must be installed in the same rack together with the associated PDUs
- PDUs must have enough C19 connectors to support the CPC drawers.
- All other components can be supported using C19 or C13 connectors. The length of these connectors (1m, 2m, or 3m) is ordered by feature code.
- PDUs must be within the same rack as the components but their locations within the racks are not defined by IBM.
- Client provided PDU/power management with DCIM sustainability tool integrations that meets systems specifications.
- The feature codes ordered will dictate the specific order the components are mounted in the rack, starting from the bottom up, in specified EIU locations.

For more information see *IBM z16 and LinuxONE Rockhopper 4 Rack Mount Bundle Installation Manual for Physical Planning (IMPP)*, GC28-7035.

The IBM z16 AGZ allows CPC0 to be placed anywhere within the client rack as indicated by location 'U' (per client planning⁴). When CPC1 is installed, it must be placed in the rack at the location which is 5U higher than CPC0's location. The components are installed in a specific order and must be installed in the same rack, together with the power distribution units. Note that for an IBM z16 AGZ to have an I/O drawer, it requires that CPC0 be placed no lower than location A09 (9U).

The system components are installed from the bottom and stacking vertically upwards.

The possible configurations for IBM z16 AGZ (rear view) are shown in “IBM z16 AGZ configurations” on page 484

2.1.7 PCIe I/O Drawers

I/O features are installed in PCIe+ I/O drawers. Both IBM z16 A02 and IBM z16 AGZ share the same set of I/O features and PCIe+ I/O drawers:

- ▶ Both IBM z16 A02 and IBM z16 AGZ configurations support up to three PCIe+ I/O drawers.
- ▶ I/O PCHID numbering starts with 0100 and goes up, depending on the number of features that are ordered, starting from the bottom I/O drawer, and working upwards.

Table 2-4 lists the common plugging limits between the IBM z16 A02 and IBM z16 AGZ.

⁴ 'U' is the 19-inch rack EIA unit number designated in the planning.

Table 2-4 Components for IBM z16 A02 and IBM z16 AGZ features

Feature	CPC Drawers	DCMs	Max Fanouts	Max I/O Drawers
Max 5	1	2	6	3
Max16	1	2	6	3
Max32	1	4	12	3
Max68	2	8	24	3

2.2 CPC drawer

The IBM z16 A02 and IBM z16 AGZ, have some changes to the design of IBM z16 A01 primarily with the processor modules and memory packaging in the drawers. Their CPC drawer includes the following features:

- ▶ 2 or 4 Dual chip modules (DCMs)
- ▶ Up to 48 Memory DIMMs (16, 32, 40, and 48 DIMMs configurations)
- ▶ Max 12 PCIe fanout cards to support PCIe+ Gen3 fanout cards for PCIe+ I/O drawers or coupling fanouts for coupling links to other CPCs
- ▶ Symmetric multiprocessor (SMP-9) connectivity (cables) when the 2nd CPC drawer exists.

The IBM z16 A02 and IBM z16 AGZ include one or two CPC drawers. A CPC drawer and its components are shown in Figure 2-5.

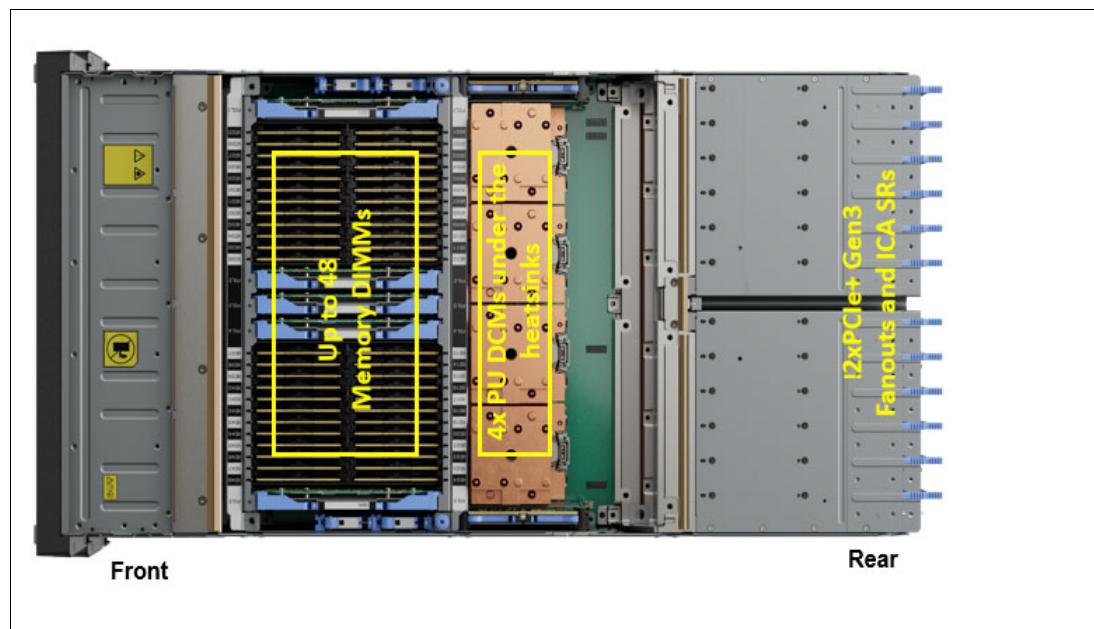


Figure 2-5 CPC drawer components (top view)

The first 5U CPC drawer contains two or four Processor Unit (PU) DCMs, and up to 48 DIMMs available in capacities of 32GB, 64GB, 128GB, or 256GB per DIMM. The second CPC drawer always contains four DCMs.

Depending on the feature, IBM z16 A02 or IBM z16 AGZ have the following CPC components:

- ▶ The number of CPC drawers installed is driven by the following feature codes:
 - FC 0672: One CPC drawer, Max5, two PU DCMs up to 5 characterizable PUs
 - FC 0673: One CPC drawer, Max16, two PU DCMs up to 16 characterizable PUs
 - FC 0674: One CPC drawer, Max32, four PU DCMs up to 32 characterizable PUs
 - FC 0675: Two CPC drawers, Max68, eight PU DCMs up to 68 characterizable PUs
- ▶ The following DCMs are used:
 - PU DCM contains two PU chips (Telum) using 7nm FinFET silicon wafer technology, 22.5 Billion transistors, core running at 4.6 GHz: (with 8 cores per chip / 16 cores per PU DCM design)
- ▶ Memory plugging:
 - Up to six memory controllers per drawer (two on DCM2/DCM1) (one on DCM3/DCM0)
 - Each memory controller supports 8 DIMM slots
 - Two, four or six memory controllers per drawer are populated (up to 48 DIMMs)
 - Different memory controllers can have different size DIMMs
- ▶ Up to 12 PCIe+ Gen3 fanout slots that can host:
 - 2-Port PCIe+ Gen3 I/O fanout for PCIe+ I/O drawers (ordered and used in pairs for availability)
 - ICA SR and ICA SR1.1 PCIe fanout for coupling (two ports per feature)
- ▶ Management components: Two dual function Baseboard Management Controllers (BMC) and Oscillator Cards (OSC) for system control and to provide system clock (N+1 redundancy)
- ▶ CPC drawer power infrastructure consists of the following components:
 - Four 2kW Power Supply Units (PSUs) that provide power to the CPC drawer. The loss of one power supply leaves enough power to satisfy the drawer's power requirements (N+1 redundancy). The power supplies can be concurrently removed and replaced (one at a time).
 - 5x 12v distribution point-of-load (POL) that plug in slots that divide the memory banks, N+2
 - 3x Voltage Regulator Modules that plug outside of the memory DIMMs, N+2
 - Two Processor Power Control cards (PPC) that contain the ambient temperature sensor and PU temperature change monitoring.
- ▶ One SMP9 connector (dual CPC drawer) that provides for CPC drawer to CPC drawer communication.

The front view of the CPC drawer, which includes the 80mm CPC cooling fans, BMC/OSC and Processor Power Control Cards (PPC), is shown in Figure 2-6 on page 31.

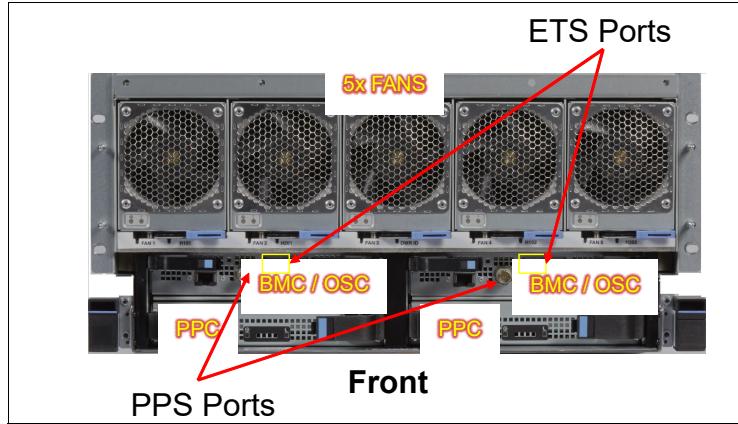


Figure 2-6 Front view of the CPC drawer

The rear view of a fully populated CPC Drawer is shown in Figure 2-7. Dual port I/O fanouts and ICA SR adapters are plugged in specific slots for best performance and availability. Redundant power supplies and the SMP9 ports also are shown (only the center two SMP9 ports are used for the IBM z16 A02 and IBM z16 AGZ for dual CPC configurations). The pair of SMP9 ports are redundant. In the event of a single cable failure, the repair can be performed concurrently.

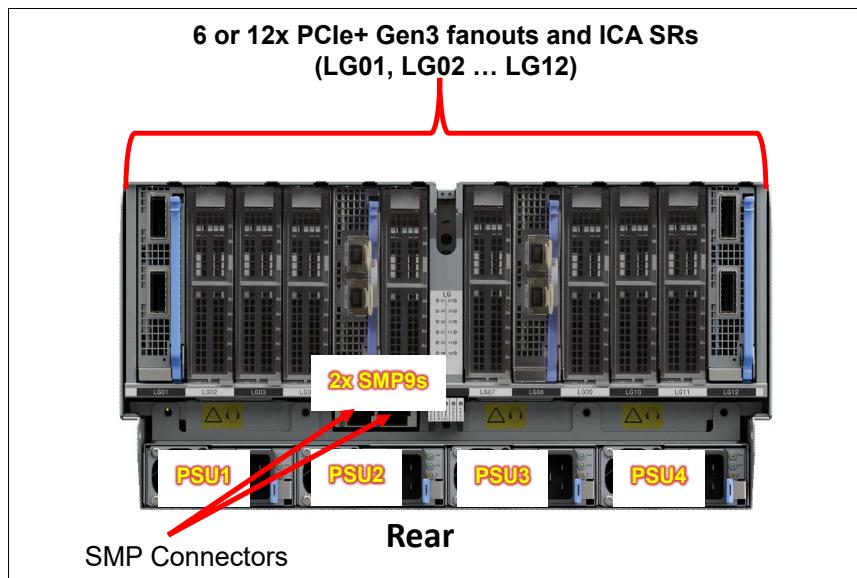


Figure 2-7 Rear view of the CPC drawer

The CPC drawer logical structure, component connections are shown in Figure 2-8 on page 32.

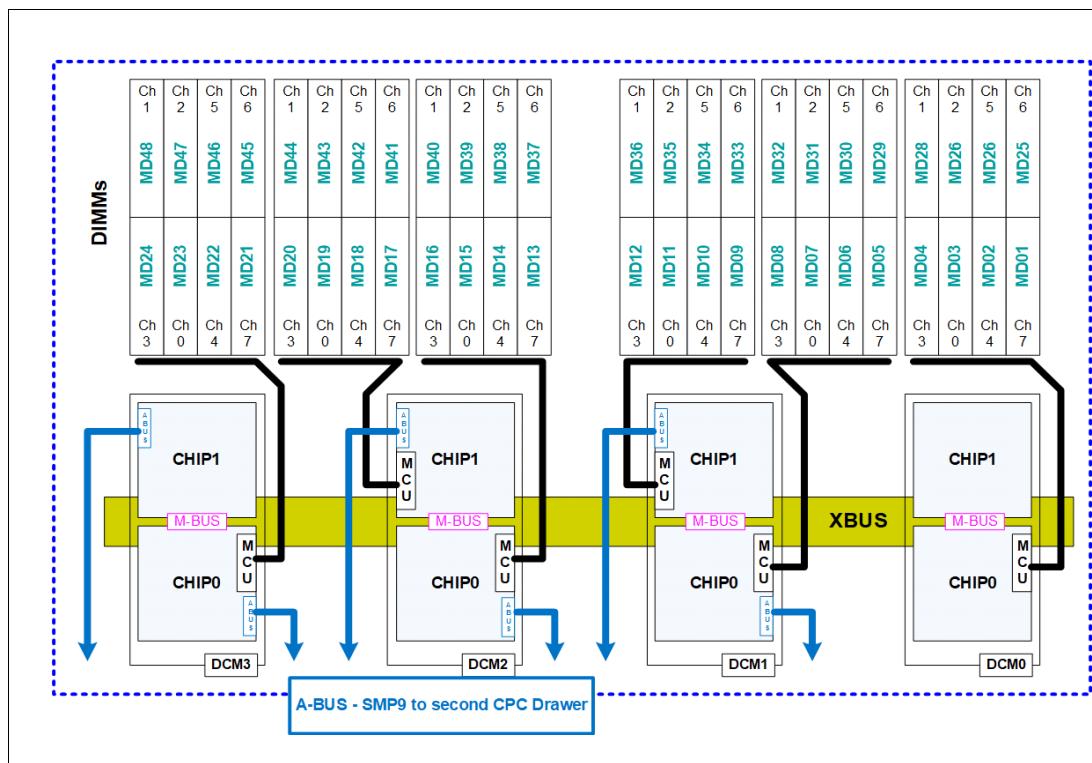


Figure 2-8 CPC drawer logical structure

Memory is connected to the DCMs through memory control units (MCUs). Up to six MCUs are available in a CPC drawer (one or two per DCM) and provide the interface to the DIMM controller. A memory controller uses eight DIMM slots.

The buses are organized in the following configurations:

- ▶ The M-bus provides interconnects between PUs chips in the same DCM
- ▶ The X-bus provides interconnects between PUs chips to each other, in the same drawer

The A-bus provides interconnects between CPC drawers using SMP9 cables.

2.2.1 CPC drawer interconnect topology

The point-to-point SMP9 connection topology for CPC drawers is shown in Figure 2-8. Each CPC drawer communicates directly to the other CPC drawers' DCMs by using point-to-point links. Drawer-to-drawer SMP9 connections are in pairs where both are active. A failure on one cable of the pair can be repaired concurrently.

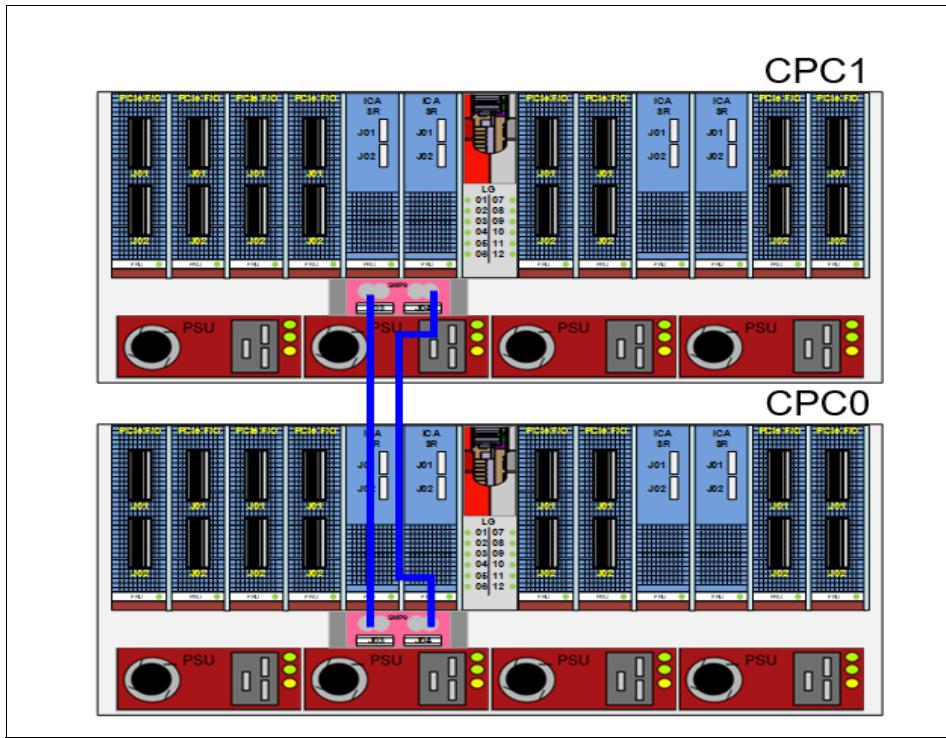


Figure 2-9 CPC drawer logical structure

The CPC drawers that are installed from bottom to top in the 19-inch rack. The order of CPC drawer installation for IBM z16 A02 is listed in Table 2-5.

Table 2-5 CPC drawer installation order and position for IBM z16 A02

CPC drawer	CPC0	CPC1
Installation order	First	Second
Position in Frame A	A10B	A15B

For the IBM z16 A02, a second CPC drawer can be installed concurrently when not present at initial installation provided that the system is powered by three-phase PDUs. For the IBM z16 AGZ, a non-disruptive addition of a CPC drawer is possible in the field (MES upgrade) if the CPC1 reserve feature (FC 2332) is included with the initial system order. Concurrent drawer repair requires a minimum of two CPC drawers.

2.2.2 Oscillator

With IBM z16 A02 and IBM z16 AGZ the oscillator card⁵, design and signal distribution scheme are new; however, the RAS strategy for the redundant clock signal and the dynamic switchover is unchanged. One primary OSC card and one backup are used. If the primary OSC card fails, the backup detects the failure, takes over transparently, and continues providing the CPC with the clock signals.

The OSC card also provides the infrastructure for CPC time synchronization to an external time source (ETS). External time synchronization is provided via the IBM zSystems Server Time Protocol (STP), orderable feature FC1021, free of charge. Time synchronization

⁵ Oscillator card (OSC) is combined (single FRU P/N) with the Baseboard Management Controllers (BMC); installed in pairs for each CPC Drawer

functionality is configured and managed through the Hardware Management Console. The HMC provides the **Manage System Time** task. For managing time on IBM z16 A02 and IBM z16 AGZ, the HMC must be at level 2.16.0 (Driver 51).

The IBM z16 A02 and IBM z16 AGZ supports the following timing information synchronization protocols:

- ▶ Network Time Protocol (NTP)
- ▶ Network Time protocol w/ Pulse Per Second (PPS)
- ▶ Precision Time Protocol (PTP)
- ▶ Precision Time Protocol w/ Pulse Per Second

With IBM z16 A02 and IBM z16 AGZ, the STP External Time Source connects directly to the CPC (dedicated LAN ports) through the OSC card. This is different from the previous IBM zSystem where the Support Element was connected to the External Time Source

For additional information about Server Time Protocol, see the Redbook *IBM zSystems Server Time Protocol Guide*, SG24-8480.

Consider the following points:

- ▶ A new card that combines the BMC and OSC is implemented with IBM z16 A02 and IBM z16 AGZ. Internally, the physical cards (BMC and OSC) are separated but incorporated as a single FRU because of a packaging design.
- ▶ Two local redundant oscillator cards are available per CPC drawer, each with one PPS port and one ETS port (RJ45 Ethernet, for both PTP and NTP).
- ▶ The current design requires Pulse Per Second to provide maximum time accuracy for NTP and PTP.
- ▶ An enhanced precision oscillator (20 PPM⁶ versus 50 PPM on previous systems) is used.
- ▶ The following PPS plugging rules apply (see Figure 2-10 on page 35):
 - Single CPC drawer plug left and right OSC PPS coaxial connectors.
 - Multi-drawer plug CPC0 left OSC PPS and CPC1 left OSC PPS coaxial connectors.
 - Cables are routed from rear to front using a pass-through hole in the frame and under the CPC bezel using a right-angle Bayonet Neill-Concelman (BNC) connector that provides the pulse per second (PPS) input for synchronization to an external time source with PPS output.
- ▶ Cables are supplied by the customer.
- Connected PPS ports must be assigned in the Manage System Time menus on the HMC.

⁶ PPM - Parts Per Million

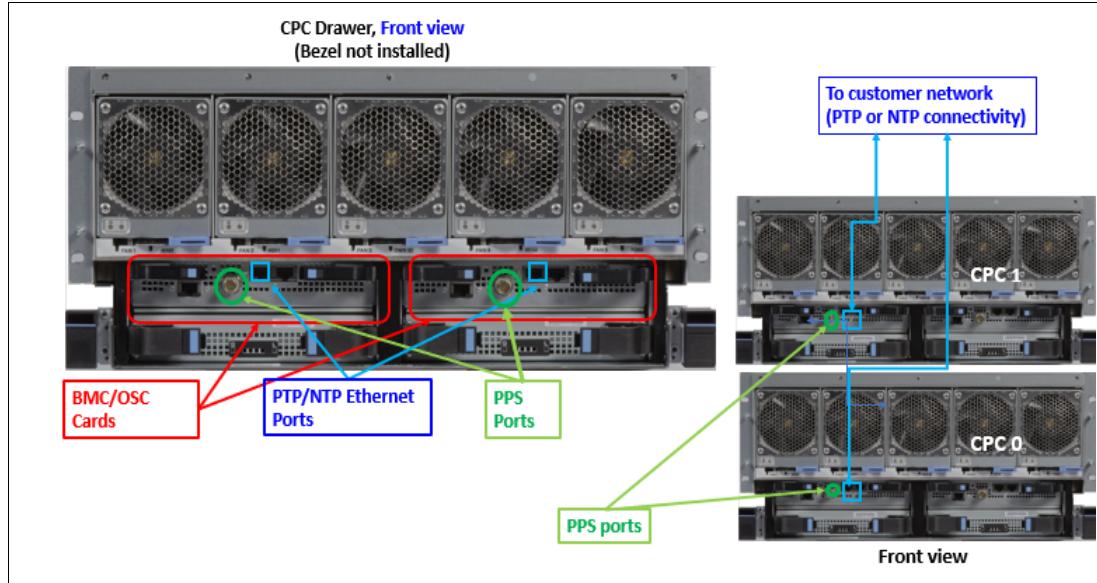


Figure 2-10 Recommended ETS cabling

Tip: STP is available as FC 1021. It is implemented in the Licensed Internal Code (LIC), and allows multiple zSystems servers to maintain time synchronization with each other and synchronization to an ETS.

For more information, see the Redbook: *IBM zSystems Server Time Protocol Guide*, SG24-8480.

2.2.3 System control

The various system elements are managed through the Baseboard Management Controllers (BMCs). The BMC is the replacement for the Flexible Support Processors (FSPs) used in previous systems.

With IBM z16 A02 and IBM z16 AGZ the CPC drawer BMC card is combined with the Oscillator card as a single Field Replaceable Unit (FRU). Two combined BMC/OSC cards are used per CPC drawer.

Also, the PCIe+ I/O drawer has a new BMC. Each BMC card has one Ethernet port that connects to the internal ethernet LANs through the internal network switches (SW1, SW2, and SW3, SW4, if configured). The BMCs communicate with the SEs and provide a subsystem interface (SSI) for controlling components.

An overview of the system control design is shown in Figure 2-11 on page 36.

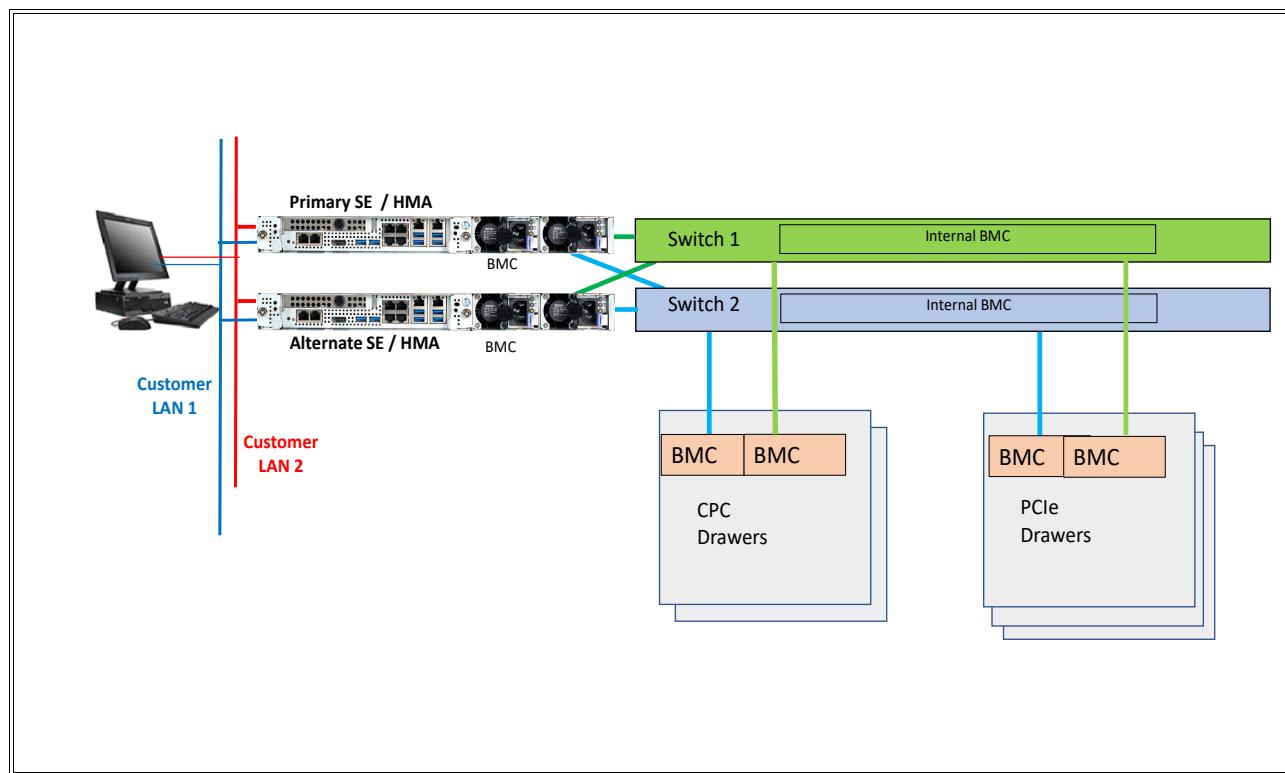


Figure 2-11 Conceptual overview of system control element - HMA and SEs/ BMCs connections

Note: The maximum IBM z16 A02 and IBM z16 AGZ configurations feature two GbE switches, two CPC drawers, and up to three PCIe I/O drawers.

A typical BMC operation is to control a power supply. An SE sends a command to the BMC to start the power supply. The BMC cycles the various components of the power supply, monitors the success of each step and the resulting voltages, and reports the status back to the SE.

SEs are duplexed (N+1), and each support element has at least one BMC and two internal Ethernet LANs for redundancy. Crossover capability between the LANs is available so both SEs can operate on both.

The Hardware Management Consoles (HMCs) and SEs/HMAs are connected directly to one or two Client Ethernet LANs. One or more HMCs can be used.

2.2.4 CPC drawer power

The power for the CPC drawer has a new design. It uses the following combinations of Power Supply Units (PSUs), Point of Load (POLs), Voltage Regulator Modules (VRMs), and Processor Power Control Cards (PPCs):

- ▶ PSUs: Provide AC conversions to 12V DC bulk/standby power and are installed at the rear of the CPC. There are four PSUs (N+1 redundancy) connected to the PDUs.
- ▶ POLs: Five N+2 redundant cards are installed next to the Memory DIMMs.
- ▶ VRMs: three modules (N+2 redundancy).

- ▶ Processor Power Control (PPC) card: Redundant processor power and control cards connect to the CPC trail board. The control function is powered from 12V standby that is provided by the PSU. The PPC cards also include pressure, temperature, and humidity sensors.

2.3 Dual chip modules

The DCM is a multi-layer metal substrate module that holds two PU chips. PU chip size is 532 mm² (23.75 mm x 22.1 mm). Each CPC drawer has two or four PU DCMs with 22.5 Billion transistors each.

The PU DCMs are shown in Figure 2-12. For both DCMs, a thermal cap is placed over the chips.

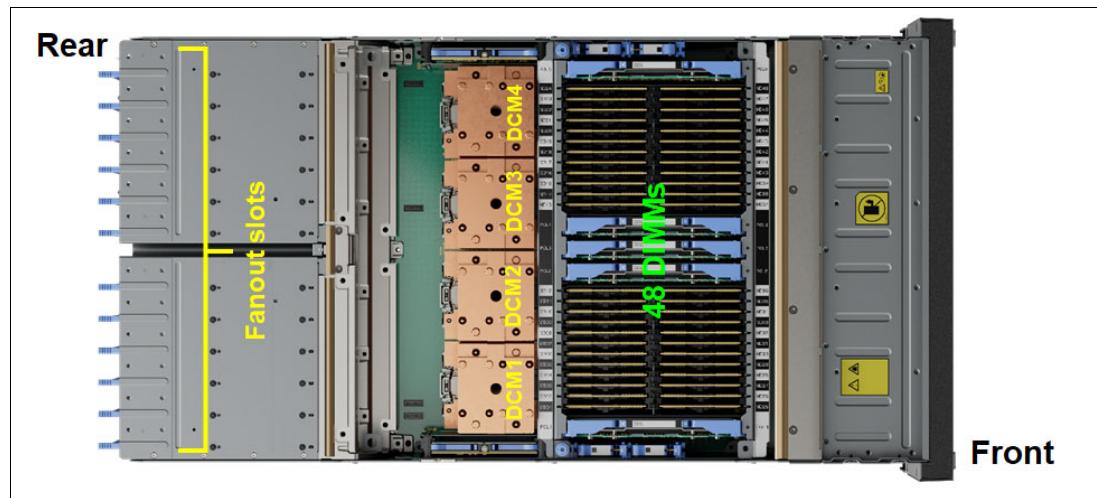


Figure 2-12 Dual chip modules (PU DCM) under the heatsink

There are two PU chips on each DCM, each DCM socket contains 4753 pins, and the module size is 71.5mm x 79mm.

The DCMs are each plugged into a socket that is part of the CPC drawer packaging. Each PU DCM is air-cooled by using the front five CPC drawer fans, and a separate heat sink for every DCM. Air flow is from the front of the drawer to the rear.

A schematic representation of the PU chip is shown in Figure 2-13.

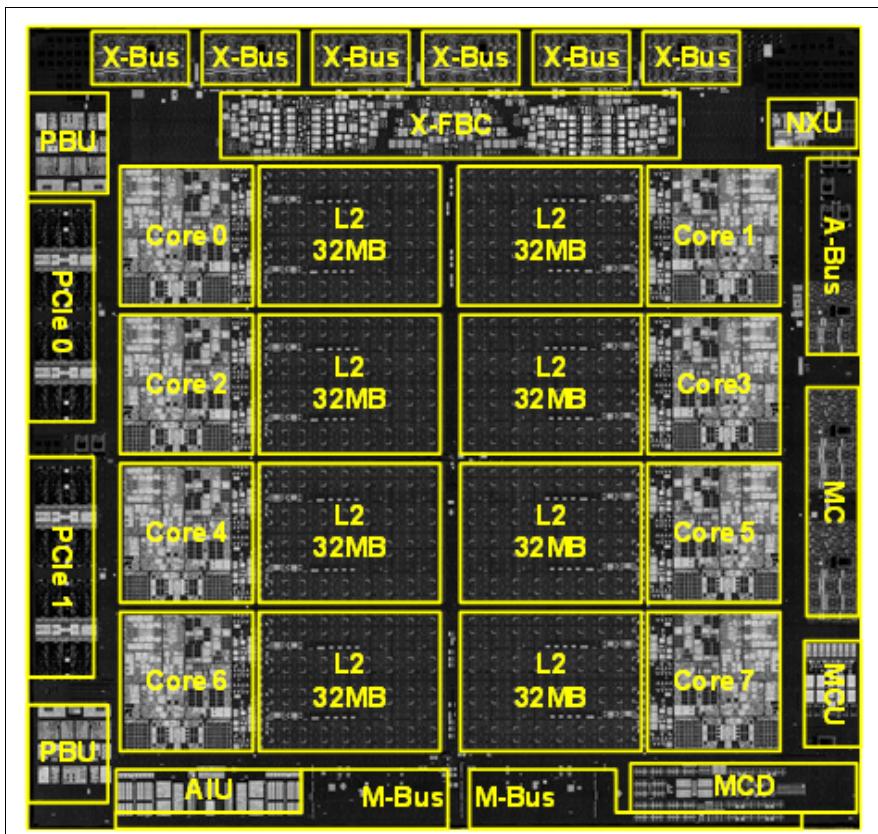


Figure 2-13 PU chip floor plan

The IBM z16 A02 and IBM z16 AGZ PU chip (two PU chips packaged on one DCM) includes the following features and improvements:

- ▶ 4.6 GHz core frequency
- ▶ eight core design (versus 10 for z15) each with 32MB L2 cache
- ▶ Two PCIe Gen5 interfaces running at Gen3/4 speeds
- ▶ DDR4 memory controller
- ▶ 2x M-Bus to support DCM internal chip to chip connectivity
- ▶ 6x X-Bus to support DCM to DCM connectivity in the CPC drawer
- ▶ 1x A-Bus to support drawer to drawer connectivity
- ▶ New cache structure design compared to z15 PU:
 - L1D(ata) and L1I(nstruction) cache - ON-core (128kB each)
 - L2 - 32 MB dense SRAM - outside the core, semi-private to the core
 - L3 (virtual) - up to 256 MB
 - L4 (virtual) - up to 2048 MB⁷
- ▶ New Core-Nest Interface
- ▶ Brand new branch prediction design using SRAM
- ▶ On chip AIU – IBM Z Accelerator for Artificial Intelligence

⁷ Max5 and Max16 L4 is limited to 1024 MB

2.3.1 Processor unit (core)

Each processor unit, or core, is a superscalar and out-of-order processor that supports 10 concurrent issues to execution units in a single CPU cycle. Figure 2-14 shows the core floor plan, which contains the following units:

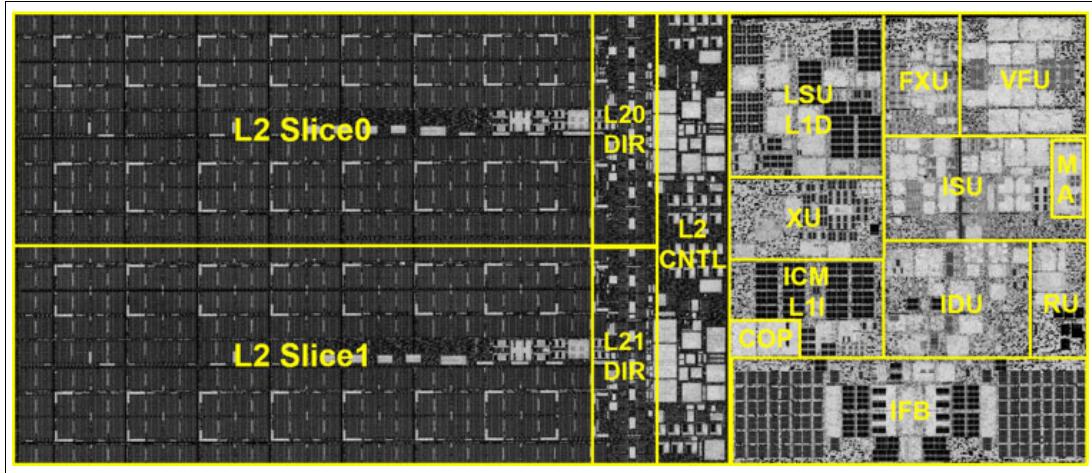


Figure 2-14 Processor core floor plan

- ▶ Fixed-point unit (FXU): The FXU handles fixed-point arithmetic.
- ▶ Load-store unit (LSU): The LSU contains the data cache. It is responsible for handling all types of operand accesses of all lengths, modes, and formats as defined in the z/Architecture.
- ▶ Instruction fetch and branch (IFB) (prediction) and Instruction cache and merge (ICM). These two sub units (IFB and ICM) contain the instruction cache, branch prediction logic, instruction fetching controls, and buffers. Its relative size is the result of the elaborate branch prediction.
- ▶ L1D and L1I are incorporated into the LSU and ICM respectively
- ▶ Instruction decode unit (IDU): The IDU is fed from the IFU buffers, and is responsible for parsing and decoding of all z/Architecture operation codes
- ▶ Translation unit (XU): The XU has a large translation lookaside buffer (TLB) and the Dynamic Address Translation (DAT) function that handles the dynamic translation of logical to physical addresses.
- ▶ Instruction sequence unit (ISU): This unit enables the Out-of-Order (OoO) pipeline. It tracks register names, Out-of-Order instruction dependency, and handling of instruction resource dispatch.
- ▶ Instruction fetching unit (IFU) (prediction): These units contain the instruction cache, branch prediction logic, instruction fetching controls, and buffers. Its relative size is the result of the elaborate branch prediction design.
- ▶ Recovery unit (RU): The RU keeps a copy of the complete state of the system that includes all registers, collects hardware fault signals, and manages the hardware recovery actions.
- ▶ Dedicated Co-Processor (CoP): The dedicated coprocessor is responsible for data compression and encryption functions for each core.
- ▶ Pervasive Core unit (PC) for instrumentation and error collection.
- ▶ Modulo arithmetic (MA) unit: Support for Elliptic Curve Cryptography.

- ▶ Vector and Floating point Units (VFU):
 - BFU: Binary floating point unit
 - DFU: Decimal floating point unit
 - DFX: Decimal fixed-point unit
 - FPd: Floating point divide unit
 - VXx: Vector fixed-point unit
 - VXs: Vector string unit
 - VXp: Vector permute unit
 - VXm: Vector multiply unit
- ▶ L2 – Level 2 cache

2.3.2 PU characterization

The PUs are characterized for client use. The characterized PUs can be used in general to run supported operating systems, such as z/OS, z/VM, and Linux on Z. They also can run specific workloads, such as Java, XML services, IPSec, and some Db2 workloads, or clustering functions, such as the Coupling Facility Control Code (CFCC).

The maximum number of characterizable PUs depends on the IBM z16 A02 and IBM z16 AGZ feature code:

- ▶ FC 0672: Max5, two PU DCMs up to 5 characterizable PUs
- ▶ FC 0673: Max16, two PU DCMs up to 16 characterizable PUs
- ▶ FC 0674: Max32, four PU DCMs up to 32 characterizable PUs
- ▶ FC 0675: Max68, eight PU DCMs up to 68 characterizable PUs

Some PUs are characterized for system use; some are characterized for client workload use. By default, one spare PU is available to assume the function of a failed PU. The maximum number of PUs that can be characterized for client use are listed in Table 2-6.

Table 2-6 PU characterization

Feature	CPs	IFLs	Unassigned IFLs	zIIPs	ICFs	IFPs	Std SAPs	Add'l SAPs	Spare PUs
Max5	0-5	0-5	0-4	0-4	0-5	2	2	0-8	2
Max16	0-6	0-16	0-15	0-15	0-16	2	2	0-8	2
Max32	0-6	0-32	0-31	0-31	0-32	2	4	0-8	2
Max68	0-6	0-68	0-67	0-67	0-68	2	8	0-8	2

At least one CP must be purchased before a zIIP can be purchased. Starting with z16, the maximum for the zIIP FCs will be one less than the feature allowed maximum PUs. For instance, an IBM z16 A02 or IBM z16 AGZ Max68 can have up to 67 zIIPs. These rules are also valid for unassigned zIIPs, and unassigned IFLs. Java and XML workloads can run on zIIPs.

Converting a PU from one type to any other type is possible by using the Dynamic Processor Unit Reassignment process. These conversions occur concurrently with the system operation.

Note: The addition of ICFs, IFLs, zIIPs, and SAPs does not change the system capacity setting or its million service units (MSU) rating, which applies to engines (processor units-PUs) characterized as Central Processor (CP).

2.3.3 Cache level structure

The cache structure comparison between CPC drawers on z16 and z15 is shown in Figure 2-15.

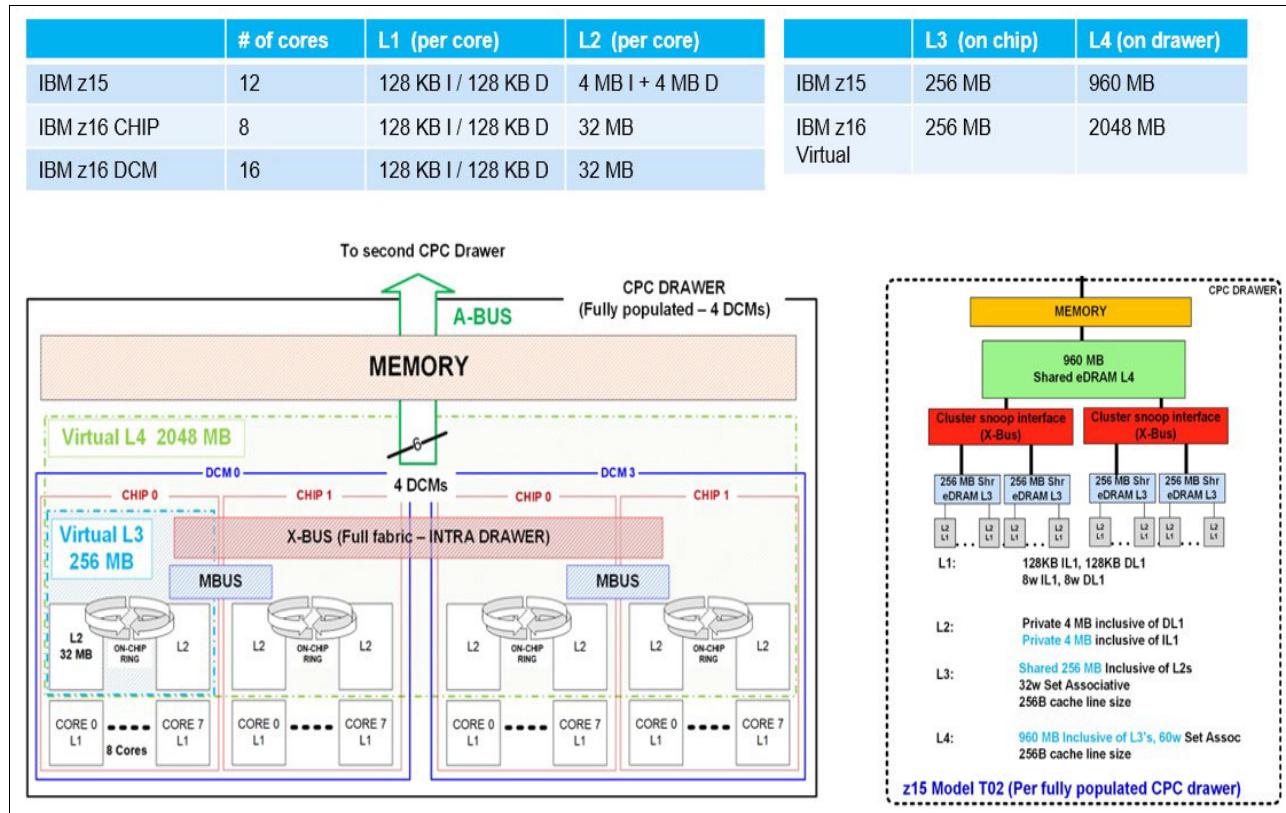


Figure 2-15 Cache structure comparison: z16 versus z15

2.4 PCIe+ I/O drawer

PCIe Generations

The z16 uses PCIe Generation 4 (Sextupled Data Rate - XDR), Generation 3 (Quadruple Data Rate - QDR), Generation 2 (Double Data Rate - DDR) and Generation 1, (Single Data Rate - SDR).

XDR is used between the CP-DCM and the fanout card. QDR is used between CEC drawer and PCIe I/O drawer with a width of 16 Lanes. Both connections operate at 16 GB/s (non full duplex). The PCIe Switch in the PCIe I/O Drawer can negotiate QDR, DDR and SDR, depending of the PCIe I/O Card. Nominal data rates in full duplex mode are: 64 GB/s for PCIe Gen-4, 32 GB/s for PCIe Gen-3, 16 GB/s for PCIe Gen-2 and 8 GB/s for PCIe Gen-1.

As shown in Figure 2-16 on page 42, each PCIe+ I/O drawer has 16 slots to support the PCIe I/O infrastructure with a bandwidth of 16 GB/s and includes the following features:

- ▶ A total of 16 I/O cards are spread over two I/O domains (0 and 1):
 - Each I/O slot reserves four PCHID numbers.
 - Left side slots are numbered LG01-LG10 and right side slots are numbered LG11-LG20 from the rear of the rack. A location and LED identifier panel is at the center of the drawer.

- With IBM z16 A02 and IBM z16 AGZ, the numbering of the PCHIDs starts with the first configured location I/O1 and continues the incremental sequence to the next configured PCIe I/O drawer(s). For more information about examples of the various configurations, see Appendix D, “Rack configurations” on page 481.
- ▶ Two PCIe+ switch cards provide connectivity to the PCIe+ Gen3 fanouts that are installed in the CPC drawers.
- ▶ Each I/O drawer domain has four dedicated support partitions (two per domain) to manage the native PCIe cards.
- ▶ Two Baseboard Management Controllers (BMC) cards are used to control the drawer functions.
- ▶ Redundant N+1 power supplies (two) are mounted on the rear and redundant blowers (six) are mounted on the front.

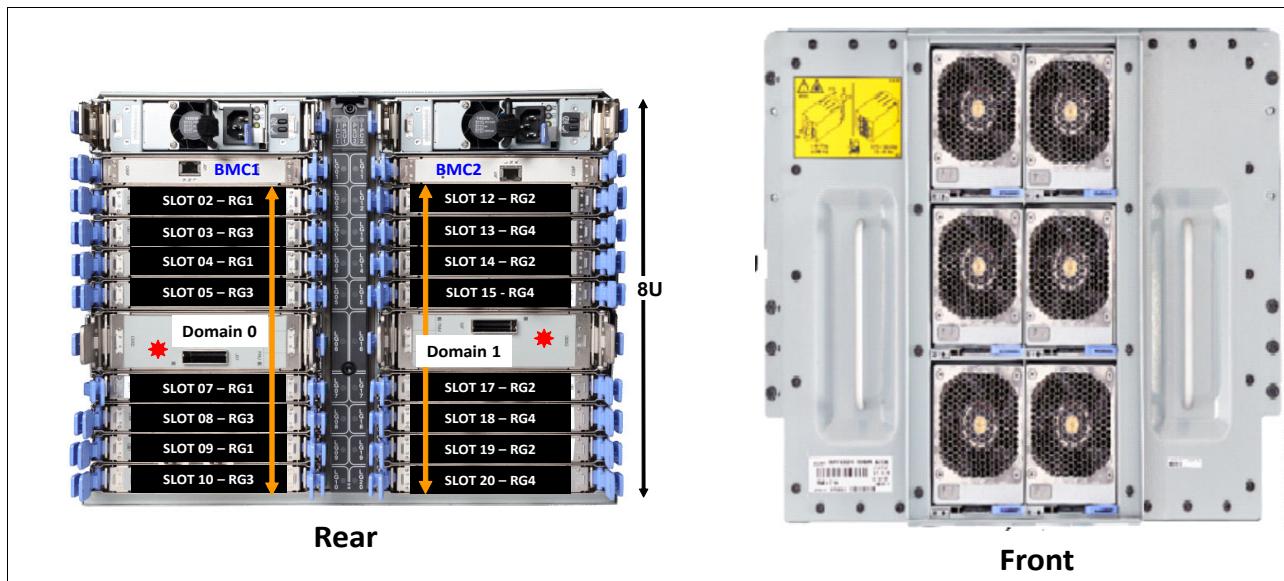


Figure 2-16 PCIe I/O drawer front and rear views

- ▶ The IBM z16 A02 and IBM z16 AGZ support up to three PCIe+ I/O drawers and use the PCHID numbering scheme (see also diagrams in Appendix D, “Rack configurations” on page 481) shown in Table 2-7:

Table 2-7 PCHID Numbering

	I/O1	I/O2	I/O3
PCHID Numbering	100 - 13F	140 - 17F	180 - 1BF

- ▶ PCHID numbering is consecutive starting from the bottom and working up.

Consideration for PCHID identification:

For IBM z16 A02 and IBM z16 AGZ, the orientation of the PCIe features is horizontal, and the top of the card is now closest to the center of the drawer for the left and right side of the drawer.

The vertical card collapsed horizontally, and the awareness of the port and PCHID layout where the top of the adapter (port D1) is closest to the location panel on both sides of the drawer are shown in Figure 2-17 on page 43.

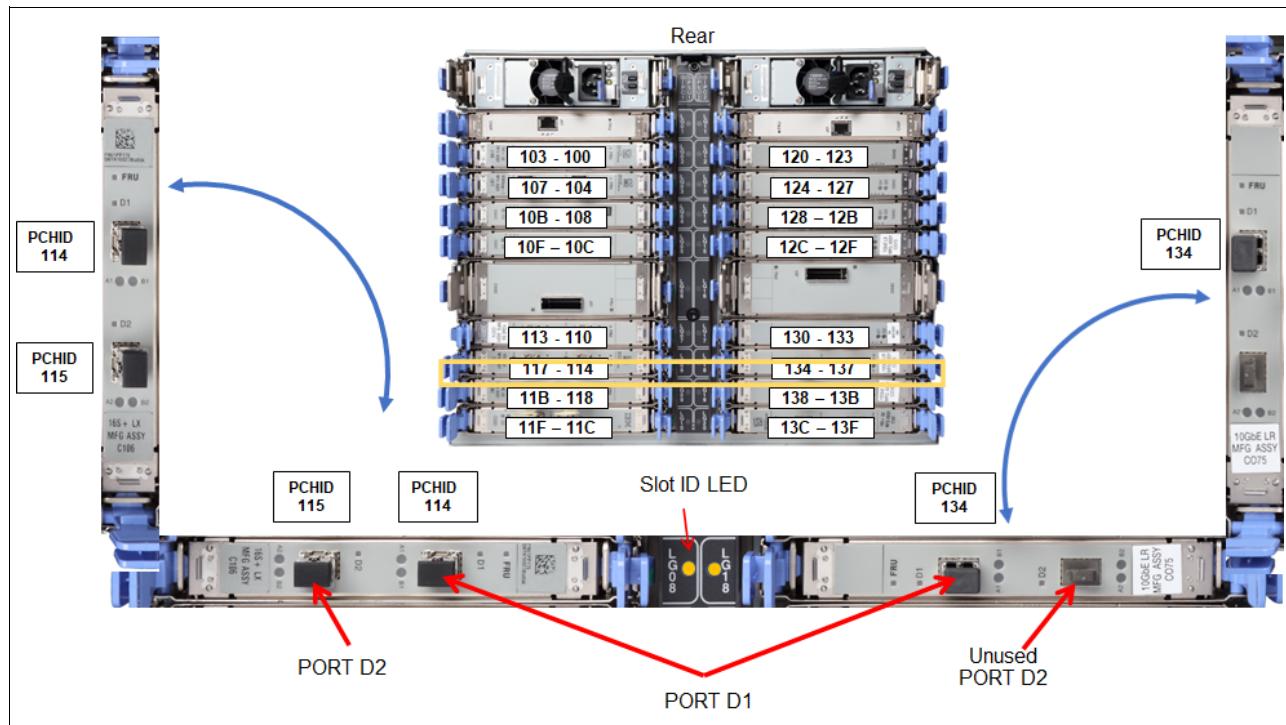


Figure 2-17 I/O feature orientation in PCIe I/O drawer (rear view)

Note: The CHPID Mapping Tool (available on ResourceLink) can be used to print a CHPID Report that displays the drawer and PCHID/CHPID layout.

2.5 Memory

The maximum physical memory size is determined by the number of CPC drawers in the system. Each CPC drawer can contain up to 8 TB of customer memory, for a total of 16 TB of memory per system.

The minimum and maximum memory sizes that you can order for each IBM z16 A02 and IBM z16 AGZ are listed in Table 2-8.

Table 2-8 Purchased Memory (Memory available for assignment to LPARs)

Feature	# of CPC drawers	Customer memory GB
Max5	1	512 - 4936
Max16	1	512 - 4936
Max32	1	512 - 8032
Max68	2	512 - 16224

The following memory types are available:

- ▶ Purchased: Memory that is available for assignment to LPARs.
- ▶ Hardware System Area (HSA): Standard 160 GB of addressable memory for system use outside of customer memory.

- Standard: Provides minimum physical memory that is required to hold customer purchase memory plus 160 GB HSA.

The memory granularity, which is based on the installed customer memory, is listed in Table 2-9.

Table 2-9 Customer offering memory increments

Memory increment (GB)	Offered memory sizes (GB)
8	64 - 96
32	128 - 544
64	608 - 736
128	864 - 2144
256	2400 - 3936
512	4448 - 16224

2.5.1 Memory subsystem topology

The IBM z16 A02 and IBM z16 AGZ memory subsystem use high-speed, differential-ended communications memory channels to link a host memory to the main memory storage devices.

The CPC drawer memory topology of an IBM z16⁸ is shown in Figure 2-18.

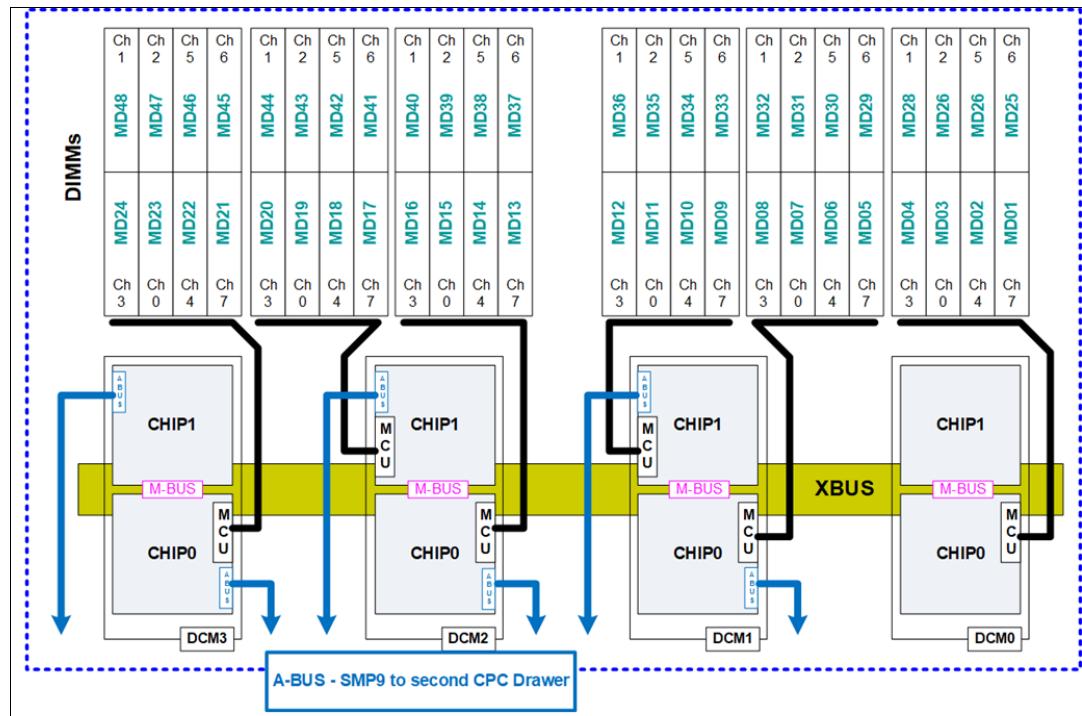


Figure 2-18 CPC drawer memory topology at maximum configuration

Consider the following points regarding the topology:

⁸ Same memory topology is used in the IBM z16 A02 and IBM z16 AGZ.

- ▶ Six memory controllers per drawer, one per PU chip
- ▶ Each memory controller supports 8 DIMM slots
- ▶ Two, four, five or six memory controllers per drawer will be populated (16, 32, 40, 48 DIMMs)
- ▶ Different memory controllers may have different size DIMMs
- ▶ Features with different DIMMs sizes can be mixed in the same drawer.
- ▶ The eight DIMMs per MCU must be the same size.
- ▶ Addressable memory is required for partitions and HSA.

2.5.2 Redundant array of independent memory

The IBM z16 A02 and IBM z16 AGZ use the new RAIM design. The new RAIM design detects and recovers from failures of dynamic random access memory (DRAM), sockets, memory channels, or DIMMs.

New RAIM Design

- Change from 4+1 RAIM (for z15) to 8 Channel R-S RAIM
- 90 -> 80 DRAMs accessed across memory channels (11% reduction, excluding unused spare)
- Staggered Memory Refresh -> Leverage RAIM to hide memory refresh penalty
- New Memory Buffer Chip Interface

2.5.3 Memory Encryption

Along with the new RAIM design, the new memory interface uses transparent memory encryption technology to protect all data leaving the processor chips before it's stored in the memory DIMMs. The encryption occurs post-RAIM encoding, the decryption occurs pre-RAIM decoding. That means the data is encrypted before being distributed according to the eight channel RAIM algorithm and written to the eight DIMMs forming a RAIM group, and the data is decrypted after being read from the DIMMs.

For encryption and decryption, an AES algorithm is used, which is executed in the Memory Control Unit (MCU). The encryption key is unique per memory channel. Since the RAIM design uses eight channels to control the eight DIMMs in a group, the data on every DIMM in a group is encrypted with another key.

The keys are generated once per IML, stored in the MCU, and locked against alteration, so the key values are Error Correction Code (ECC) protected. Since the encryption and decryption occur between RAIM error correction and memory, an encryption and decryption error is confined to a single memory channel. Encryption and decryption contain parity information to detect errors. Unauthorized physical access to the DIMMs will only deliver data which is distributed across eight DIMMs and encrypted with eight different keys.

2.5.4 Memory configurations

Memory sizes in each CPC drawer do not have to be similar. RAIM is now built into the bank of 8 DIMMs and should no longer be part of the memory equation (minus 20%).

- There are 14 drawer configurations supporting memory
- Different CPC drawers can contain different amounts of memory

- A drawer may have a mix of DIMM sizes
- Total memory includes HSA. Customer memory is remaining memory available to customer after HSA is subtracted

The 17 (40-43) and (21-33) drawer memory configurations that are supported are listed in Table 2-10. Each CPC drawer is included from manufacturing with one of these memory configurations.

Table 2-10 Drawer memory plugging configurations (all values in GB)

CFG #	DCM1 M0CP0 MD01-04 & MD25-28	DCM2 M1CP0 MD05-08 & MD29-32	DCM2 M1CP1 MD09-12 & MD33-36	DCM3 M2CP0 MD13-16 & MD37-40	DCM3 M2CP1 MD17-20 & MD41-44	DCM4 M3CP0 MD21-24 & MD45-48	Physical	INC	-HSA 160GB
40		64	64	64	128		2560		2400
41		128	64	64	128		3072	512	2912
42		128	128	64	128		3584	512	3424
43		128	128	128	128		4096	512	3936
21			32		32		512		352
22		32	32	32	32		1024	512	864
23		64	32	64	32		1536	512	1376
24		64	64	64	64		2048	512	1888
25	64	64	64	64	64		2560	512	2400
26	64	64	64	64	64	64	3072	512	2912
27	128	64	64	64	64	64	3584	512	3424
28	128	64	64	64	64	128	4096	512	3936
29	128	128	64	64	64	128	4608	512	4448
30	128	128	64	64	128	128	5120	512	4960
31	128	128	128	128	128	128	6144	1024	5984
32	256	128	128	128	128	128	7168	1024	7008
33	256	128	128	256	256		8192	1024	8032

Consider the following points:

- ▶ A CPC drawer contains a minimum of 16 (2x8) 32GB DIMMs as listed in drawer configuration number 21 in Table 2-10.
- ▶ A CPC drawer can have more memory installed than what is actually enabled for client use. The amount of memory that can be enabled by the client is the total physically installed memory minus the 160 GB of HSA memory.
- ▶ A CPC drawer can have available unused memory, which can be ordered as a memory upgrade and enabled by LICCC concurrently without DIMM changes.
- ▶ DIMM changes require a disruptive power-on reset (POR) on IBM z16 A02 and IBM z16 AGZ with a single CPC drawer. DIMM changes can be done concurrently on configurations with two CPC drawers using Enhanced Drawer Availability (EDA).

DIMM plugging for the configurations in each CPC drawer do not have to be similar. Each memory 8 slot DIMM bank must have the same DIMM size; however, a drawer can have a mix of DIMM banks.

The support element ***View Hardware Configuration*** task can be used to determine the size and quantity of the memory plugged in each drawer. Figure 2-19 shows an example of an IBM z16 AGZ using drawer configuration number 30 from the previous tables, and displays some of the locations and descriptions of the installed memory modules.

View Hardware Configuration - VELA			
Machine Type - Model: 3932 - AGZ			
Machine serial number: 000020087F28			
Processor location: ASYS			
Select	Location	Identifier	Description
<input type="radio"/>	ACP0MD01	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD02	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD03	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD04	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD05	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD06	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD07	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD08	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD09	32CC	Memory DIMM 64 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD10	32CC	Memory DIMM 64 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD11	32CC	Memory DIMM 64 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD12	32CC	Memory DIMM 64 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD13	32CC	Memory DIMM 64 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD14	32CC	Memory DIMM 64 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD15	32CC	Memory DIMM 64 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD16	32CC	Memory DIMM 64 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD17	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD18	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD19	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)
<input type="radio"/>	ACP0MD20	32CD	Memory DIMM 128 GB DDR4 (16 Gb DRAM)

Figure 2-19 View Hardware Configuration task on the Support Element

Figure 2-20 shows the CPC drawer and DIMM locations for an IBM z16 A02 and IBM z16 AGZ.

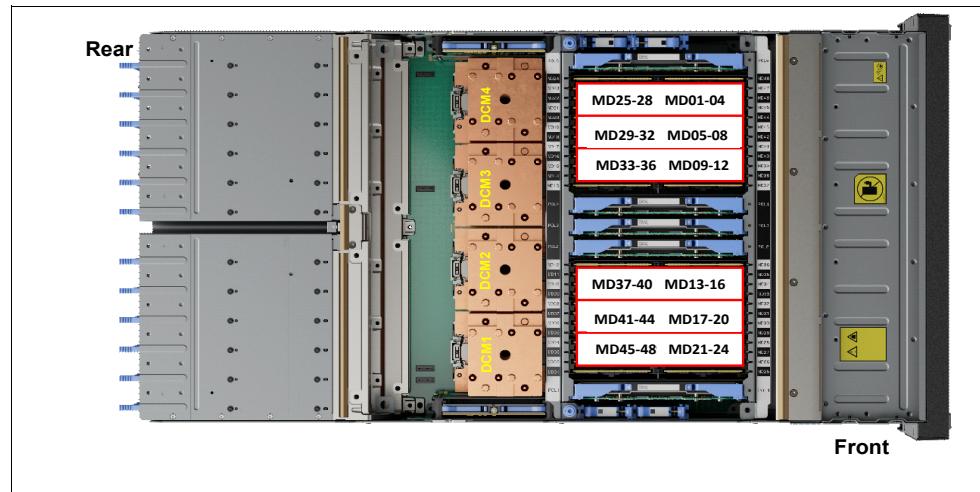


Figure 2-20 CPC Drawer DIMM locations

Table 11 lists the physical memory plugging configurations by feature code from manufacturing when the system is ordered. Consider the following points:

- ▶ The CPC drawer columns for the specific feature contain the Memory Plug Drawer Configuration number that is referenced in Table 2-10 and the Population by DIMM Bank that is listed in Table .
- ▶ Dial Max indicates the maximum memory that can be enabled by way of the LICC concurrent upgrade.

If more storage is ordered by using other feature codes, such as Virtual Flash Memory, the extra storage is installed and plugged as necessary.

Example1: A customer orders FC 3146 that provides 4448 GB customer memory, and FC 0675 (Max68). The two drawer configurations include the following components: CPC0 (2560 GB) and CPC1 (2560 GB) - both CPC drawers use config #25 (See Table 2-10). Total Installed memory: 5120GB - 160GB HSA = 4960 (Dial Max).

Example 2: Customer orders FC 3145 that features 3936 GB customer memory, and FC 0674 Max32 (1 CPC drawer). The drawer configuration include the following components: CPC0 (4096 GB Physical) - Configuration #43 (from Table 2-10). Total 4096 GB - 160GB HSA= 3936 GB (Dial Max).

Example 3: Customer orders FC 3153 that features 8032 GB customer memory, and FC 0674 Max68 (2 CPC drawers). The drawer configuration include the following components: CPC0 (4096GB Physical) - Configuration #28 and CPC1 (4096 GB Physical) - Configuration #28 (from Table 2-10). Total 8192 GB - 160GB HSA= 8032 GB (Dial Max).

The previous examples do not consider the amount of Virtual Flash Memory (VFM). If VFM is part of the initial order, you need to subtract the amount of VFM (FC 0644) from the Dial Max and ensure that the reminder satisfies customer memory requirements.

2.5.5 Memory Offerings (initial order)

Table 11 provides the list of memory features orderable and the associates memory configurations.

Table 11 Memory Offerings

FC	Increments	Customer Memory Increments	Max5 (CPC0)			Max16 (CPC0)			Max32 (CPC0)			Max68 (CPC0 + CPC1)				
			Cfg. #	Dial Max	Extra Mem	Cfg. #	Dial Max	Extra Mem	Cfg. #	Dial Max	Extra Mem	Cfg. #	CPC0	Cfg. #	CPC1	Dial Max
3106	8	64	21	352	288	21	352	288	21	352	288	21	21	21	864	800
3107		72	21	352	280	21	352	280	21	352	280	21	21	21	864	792
3108		80	21	352	272	21	352	272	21	352	272	21	21	21	864	784
3109		88	21	352	264	21	352	264	21	352	264	21	21	21	864	776
3110		96	21	352	256	21	352	256	21	352	256	21	21	21	864	768

FC	Increments	Max5 (CPC0)			Max16 (CPC0)			Max32 (CPC0)			Max68 (CPC0 + CPC1)				
		Cfg. #	Dial Max	Extra Mem	Cfg. #	Dial Max	Extra Mem	Cfg. #	Dial Max	Extra Mem	Cfg. #	CPC0	Cfg. #	CPC1	Dial Max
3111	32	128	21	352	224	21	352	224	21	352	224	21	21	864	736
3112		160	21	352	192	21	352	192	21	352	192	21	21	864	704
3113		192	21	352	160	21	352	160	21	352	160	21	21	864	672
3114		224	21	352	128	21	352	128	21	352	128	21	21	864	640
3115		256	21	352	96	21	352	96	21	352	96	21	21	864	608
3116		288	21	352	64	21	352	64	21	352	64	21	21	864	576
3117		320	21	352	32	21	352	32	21	352	32	21	21	864	544
3118		352	21	352	0	21	352	0	21	352	0	21	21	864	512
3119		384	22	864	480	22	864	480	22	864	480	21	21	864	480
3120		416	22	864	448	22	864	448	22	864	448	21	21	864	448
3121		448	22	864	416	22	864	416	22	864	416	21	21	864	416
3122		480	22	864	384	22	864	384	22	864	384	21	21	864	384
3123		512	22	864	352	22	864	352	22	864	352	21	21	864	352
3124		544	22	864	320	22	864	320	22	864	320	21	21	864	320
3125	64	608	22	864	256	22	864	256	22	864	256	21	21	864	256
3126		672	22	864	192	22	864	192	22	864	192	21	21	864	192
3127		736	22	864	128	22	864	128	22	864	128	21	21	864	128
3128	128	864	22	864	0	22	864	0	22	864	0	21	21	864	0
3129		992	23	137 6	384	23	1376	384	23	1376	384	22	21	1376	384
3130		1120	23	137 6	256	23	1376	256	23	1376	256	22	21	1376	256
3131		1248	23	137 6	128	23	1376	128	23	1376	128	22	21	1376	128
3132		1376	23	137 6	0	23	1376	0	23	1376	0	22	21	1376	0
3133		1504	24	188 8	384	24	1888	384	24	1888	384	22	22	1888	384
3134		1632	24	188 8	256	24	1888	256	24	1888	256	22	22	1888	256
3135		1760	24	188 8	128	24	1888	128	24	1888	128	22	22	1888	128
3136		1888	24	188 8	0	24	1888	0	24	1888	0	22	22	1888	0
3137		2016	40	240 0	384	40	2400	384	25	2400	384	23	22	2400	384
3138		2144	40	240 0	256	40	2400	256	25	2400	256	23	22	2400	256

FC	Customer Memory Increments	Max5 (CPC0)			Max16 (CPC0)			Max32 (CPC0)			Max68 (CPC0 + CPC1)				
		Cfg. #	Dial Max	Extra Mem	Cfg. #	Dial Max	Extra Mem	Cfg. #	Dial Max	Extra Mem	Cfg. #	CPC0	CPC1	Dial Max	Extra Mem
3139	256	2400	40	2400	0	40	2400	0	25	2400	0	23	22	2400	0
3140		2656	41	2912	256	41	2912	256	26	2912	256	23	23	2912	256
3141		2912	41	2912	0	41	2912	0	26	2912	0	23	23	2912	0
3142		3168	42	3424	256	42	3424	256	27	3424	256	24	23	3424	256
3143		3424	42	3424	0	42	3424	0	27	3424	0	24	23	3424	0
3144		3680	43	3936	256	43	3936	256	28	3936	256	24	24	3936	256
3145		3936	43	3936	0	43	3936	0	28	3936	0	24	24	3936	0

FC	Increments	Max5 (CPC0)			Max16 (CPC0)			Max32 (CPC0)			Max68 (CPC0 + CPC1)					
		Cfg. #	Dial Max	Extra Mem	Cfg. #	Dial Max	Extra Mem	Cfg. #	Dial Max	Extra Mem	Cfg. #	CPC0	Cfg. #	CPC1	Dial Max	Extra Mem
3146	512	4448						30	4960	512	25	25	4960	512		
3147		4960						30	4960	0	25	25	4960	0		
3148		5472						31	5984	512	26	26	5984	512		
3149		5984						31	5984	0	26	26	5984	0		
3150		6496						32	7008	512	27	27	7008	512		
3151		7008						32	7008	0	27	27	7008	0		
3152		7520						33	8032	512	28	28	8032	512		
3153		8032						33	8032	0	28	28	8032	0		
3154		8544									29	29	9056	512		
3155		9056									29	29	9056	0		
3156		9568									30	30	10080	512		
3157		10080									30	30	10080	0		
3158		10592									31	30	11104	512		
3159		11104									31	30	11104	0		
3160		11616									31	31	12128	512		
3161		12128									31	31	12128	0		
3162		12640									32	31	13152	512		
3163		13152									32	31	13152	0		
3164		13664									32	32	14176	512		
3165		14176									32	32	14176	0		
3166		14688									33	32	15200	512		
3167		15200									33	32	15200	0		
3168		15712									33	33	16224	512		
3169		16224									33	33	16224	0		

Memory upgrades can be ordered and enabled by LIC, upgrading the DIMM cards, adding DIMM cards, or adding a CPC drawer.

For an upgrade that results in the addition of a CPC drawer, the minimum memory increment is added to the system. Each CPC drawer has a minimum physical memory size of 1024 GB.

During an upgrade, adding a CPC drawer is a concurrent operation. Adding physical memory to the added drawer is also concurrent. If all or part of the added memory is enabled for use, it might become available to an active LPAR if the partition includes defined reserved storage. (For more information, see 3.7.3, “Reserved storage” on page 126.) Alternatively, the added memory can be used by a defined LPAR that is activated after the memory is added.

Note: Memory downgrades within an IBM z16 A02 and IBM z16 AGZ are not supported. Feature downgrades (removal of a CPC quantity feature) are also not supported.

2.5.6 Drawer replacement and memory

With Enhanced Drawer Availability (EDA), which is supported for IBM z16 A02 and IBM z16 AGZ when two CPC drawers are installed, sufficient resources must be available to accommodate the ones that are rendered unavailable when a CPC drawer is removed for upgrade or repair. For more information, see 2.7.1, “Redundant I/O interconnect” on page 54.

Note: Removing a CPC drawer during EDA often results in reducing total active memory.

2.5.7 Virtual Flash Memory

IBM Virtual Flash Memory (VFM) FC 0644 replaces the Flash Express features (0402 and 0403) that were available on previous IBM zSystems. It offers up to 2.0 TB of virtual flash memory in 512 GB (0.5 TB) increments for improved application availability and to handle paging workload spikes.

No application changes are required to change from IBM Flash Express to VFM. Consider the following points:

- ▶ Dialed memory + VFM = total hardware plugged
- ▶ VFM is offered in 0.5 TB increment size; VFM for z16 is FC 0644 - Min=0, Max=4

VFM is designed to help improve availability and handling of paging workload spikes when z/OS V2.1, V2.2, V2.3, V2.4 or V2.5 is run. With this support, z/OS is designed to help improve system availability and responsiveness by using VFM across transitional workload events. z/OS is also designed to help improve processor performance by supporting middleware use of pageable large (1 MB) pages.

VFM can also be used by coupling facility images to provide extended capacity and availability for workloads that use IBM WebSphere MQ Shared Queues structures. The use of VFM can improve availability by reducing latency from paging delays that can occur at the start of the workday or during other transitional periods. It is also designed to help eliminate delays that can occur when collecting diagnostic data.

VFM can help organizations meet their most demanding service level agreements and compete more effectively. VFM is easy to configure in the LPAR Image Profile and provides rapid time to value.

Once VFM is installed, the increments can be decremented. The reduction of VFM increments is treated like a memory downgrade and therefore is disruptive to the customer.

2.6 Reliability, availability, and serviceability

IBM zSystems continue to deliver enterprise class RAS with IBM z16 A02 and IBM z16 AGZ. The main philosophy behind RAS is about preventing or tolerating (masking) outages. It is also about providing the necessary instrumentation (in hardware, LIC and microcode, and software) to capture or collect the relevant failure information to help identify an issue without requiring a re-creation of the event. These outages can be planned or unplanned. Planned

and unplanned outages can include the following situations (examples are not related to the RAS features of IBM zSystems):

- A planned outage because of the addition of physical processor capacity or memory
- A planned outage because of the addition of I/O cards
- An unplanned outage because of a failure of a power supply
- An unplanned outage because of a memory failure

The IBM Systems hardware has decades of intense engineering behind it, which results in a robust and reliable platform. The hardware has many RAS features that are built into it. For more information, see Chapter 9, “Reliability, availability, and serviceability” on page 363.

2.7 Connectivity

Connections to PCIe+ I/O drawers and Integrated Coupling Adapters are driven from the CPC drawer fanout cards. (see 2.4, “PCIe+ I/O drawer” on page 41). These fanouts are installed in the rear of the CPC drawer.

Figure 2-21 shows the location of the fanout slots. Each slot is identified with a location code (label) of LGxx.

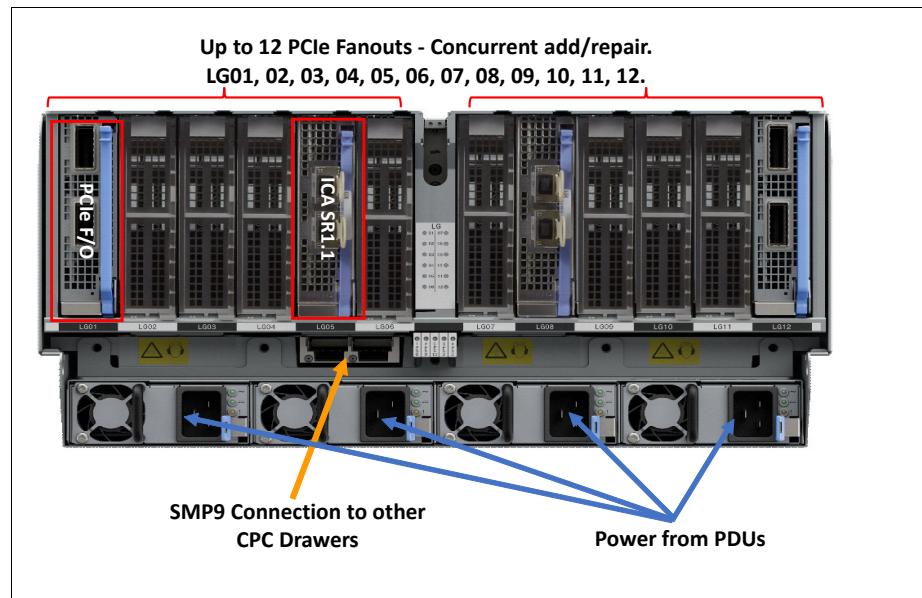


Figure 2-21 Fanout locations in the CPC drawer

Up to 12 PCIe fanouts (LG01 - LG12) can be installed in each CPC drawer.

A fanout can be repaired concurrently with the use of redundant I/O interconnect. For more information, see 2.7.1, “Redundant I/O interconnect” on page 54.

The following types of fanouts are available:

- ▶ A new PCIe+ Generation 3 dual port fanout card: This fanout provides connectivity to the PCIe switch cards in the PCIe+ I/O drawer.
- ▶ A new Integrated Coupling Adapter (ICA SR1.1): This adapter provides coupling connectivity to z16, z15, and z14.

- One, two or three pairs of redundant SMP9 connectors providing connectivity to the other one, two, or three CPC Drawers in the configuration⁹.

When configured for availability, the channels and coupling links are balanced across CPC drawers. In a system that is configured for maximum availability, alternative paths maintain access to critical I/O devices, such as disks and networks. The CHPID Mapping Tool can be used to assist with configuring a system for high availability.

Enhanced (CPC) drawer availability (EDA) allows a single CPC drawer in a multidrawer CPC to be removed and reinstalled (serviced) concurrently for an upgrade or a repair. Removing a CPC drawer means that the connectivity to the I/O devices that are connected to that CPC drawer is lost. To prevent connectivity loss, the redundant I/O interconnect feature allows you to maintain connection to critical I/O devices (except for ICA SR1.1) when a CPC drawer is removed.

2.7.1 Redundant I/O interconnect

Redundancy is provided for PCIe I/O interconnects.

The PCIe+ I/O drawer supports up to 16 PCIe features, which are organized in two hardware domains (for each drawer). The infrastructure for the fanout to I/O drawers and external coupling is shown in Figure 2-22 on page 54.

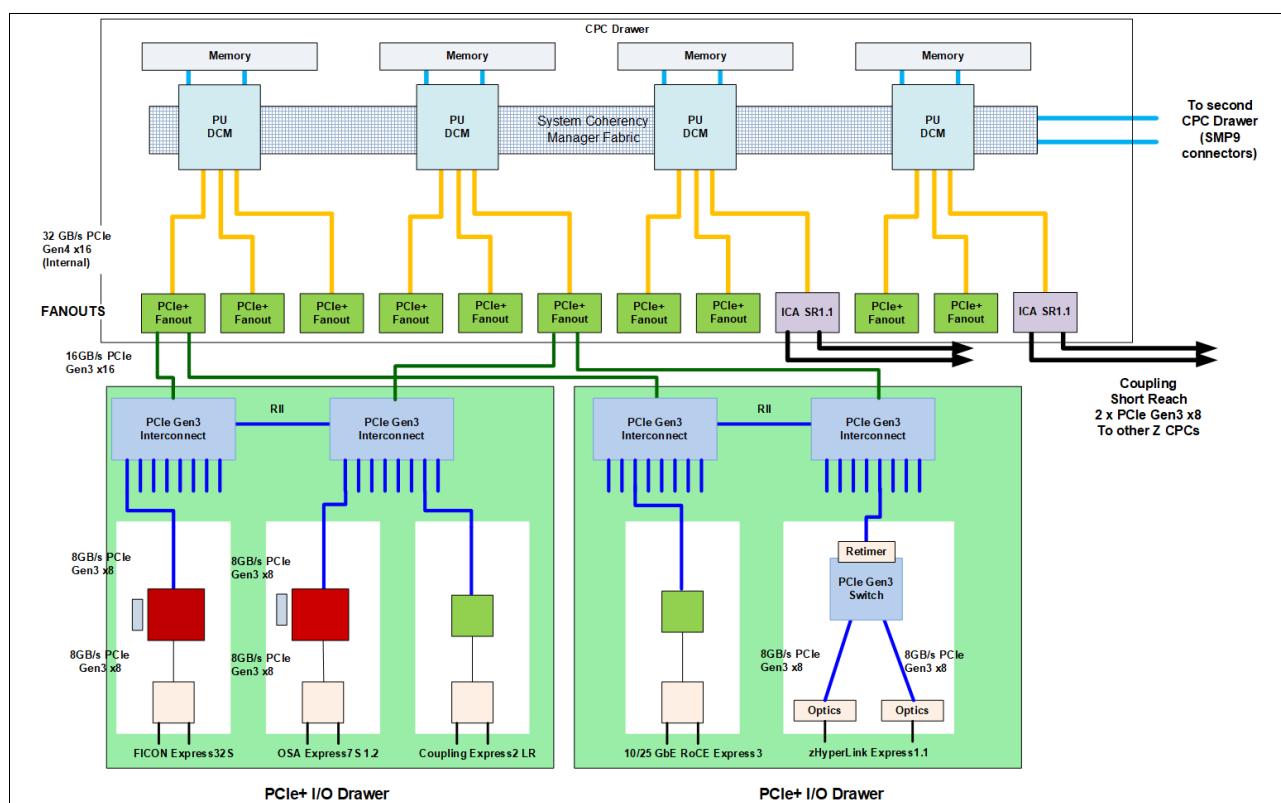


Figure 2-22 Infrastructure for PCIe+ I/O drawer (system with two PCIe+ I/O drawers)

The new PCIe+ Gen3 fanout cards are used to provide the connection from the PU DCM PCIe Bridge Unit (PBU), which uses split the PCIe Gen4 (@32GBps) processor busses into two PCIe Gen3 x16 (@16 GBps) interfaces to the PCIe switch card in the PCIe+ I/O drawer.

⁹ The IBM z16 A02 and IBM z16 AGZ with two CPC drawers use only two redundant SMP9 connectors.

The PCIe switch card spreads the x16 PCIe bus to the PCIe I/O slots in the domain.

In the PCIe+ I/O drawer, the two PCIe switch cards (LG06 and LG16, see Figure 2-23) provide a backup path (Redundant I/O Interconnect - RII) for each other through the passive connection in the PCIe+ I/O drawer backplane.

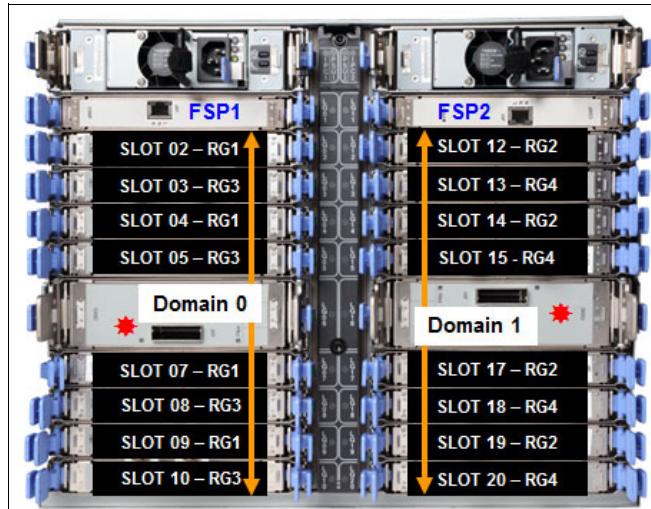


Figure 2-23 PCIe+ I/O drawer locations

To support Redundant I/O Interconnect (RII) between domain pair 0 and 1, the two interconnects to each pair must be driven from two different PCIe fanouts. Normally, each PCIe interconnect in a pair supports the eight features in its domain. In backup operation mode, one PCIe interconnect supports all 16 features in the domain pair.

Note: The PCIe Interconnect (switch) adapter *must* be installed in the PCIe+ I/O drawer to maintain the interconnect across I/O domains. If the adapter is removed (for a service operation), the I/O cards in that domain (up to eight) become unavailable.

During a CPC Drawer PCIe+ Gen3 fanout or cable failure, all 16 PCIe cards in the two domains can be driven through a single PCIe switch card (Figure 2-24).

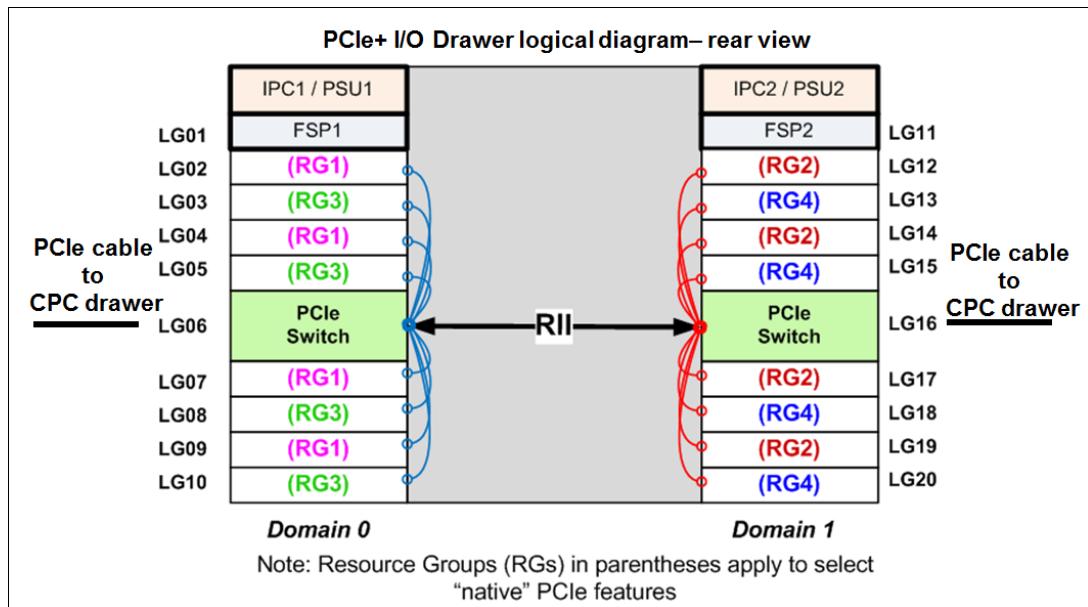


Figure 2-24 Redundant I/O Interconnect

2.7.2 Enhanced drawer availability (EDA)

With EDA, the effect of CPC drawer replacement is minimized. In a dual CPC drawer system, a single CPC drawer can be concurrently removed and reinstalled for an upgrade or repair. Removing a CPC drawer without affecting the workload requires sufficient resources in the remaining CPC drawer.

Before removing the CPC drawer, the contents of the PUs and memory of the drawer must be relocated. PUs must be available on the remaining CPC drawers to replace the deactivated drawer. Also, sufficient redundant memory must be available if no degradation of applications is allowed. Any CPC drawer can be replaced, including the first CPC drawer that initially contains the HSA.

Removal of a CPC drawer also removes the CPC drawer connectivity to the I/O drawers, PCIe I/O drawers, and coupling links. The effect of the removal of the CPC drawer on the system is limited by the use of redundant I/O interconnect. (For more information, see 2.7.1, “Redundant I/O interconnect” on page 54.) However, all ICA SR1.1 links that are installed in the removed CPC drawer must be configured offline.

If the enhanced drawer availability is *not* used when a CPC drawer must be replaced, the memory in the failing drawer is also removed. This process might be necessary during an upgrade or a repair action. Until the removed CPC drawer is replaced, a power-on reset of the system with the working CPC drawer is supported. The CPC drawer can then be replaced and added back into the configuration concurrently.

2.7.3 CPC drawer upgrade

All fanouts that are used for I/O and coupling links are rebalanced concurrently as part of a CPC drawer addition to support better RAS characteristics.

2.8 Processor configurations

When a z16 is ordered, the PUs are characterized according to their intended usage. The PUs can be ordered as any of the following items:

CP	The processor is purchased and activated. PU supports running the z/OS, z/VSE, z/VM, z/TPF, and Linux on IBM Z ¹⁰ operating systems. It can also run Coupling Facility Control Code.
IFL	The Integrated Facility for Linux (IFL) is a processor that is purchased and activated for use by z/VM for Linux guests and Linux on Z ¹⁰ operating systems.
Unassigned IFL	A processor that is purchased for future use as an IFL. It is offline and cannot be used until an upgrade for the IFL is installed. It does not affect software licenses or maintenance charges.
ICF	An internal coupling facility (ICF) processor that is purchased and activated for use by the Coupling Facility Control Code.
zIIP	An “Off Load Processor” for workload that supports applications such as DB2, Java, XML and z/OS Container Extensions. It can also be used for System recovery Boost (SRB). For more information, see <i>IBM Z System Recovery Boost</i> , REDP-5563-02.
Unassigned zIIP	A processor that is purchased for future use as a zIIP. It is offline and cannot be used until an upgrade for the zIIP is installed. It does not affect software licenses or maintenance charges.
Additional SAP	An optional processor that is purchased and activated for use as SAP (System Assist Processor).

A minimum of one PU that is characterized as a CP, IFL, or ICF is required per system. The maximum number of characterizable PUs is 68. At least one CP must be purchased before a zIIP can be purchased. Starting with z16, the maximum for the zIIP FCs will be one less than the feature allowed maximum PUs. For instance, an IBM z16 A02 and IBM z16 AGZ Max68 can have up to 67 zIIPS. These rules are also valid for unassigned zIIPs and unassigned IFLs.

The following components are present in the z16 configurations, but they are not part of the PUs that clients purchase and require no characterization:

- ▶ SAP to be used by the channel subsystem. The number of predefined SAPs depends on the z16 feature and is fixed for the specified configuration.
- ▶ Two IFP, which are used in the support of designated features and functions, such as RoCE (all features), Coupling Express LR, zHyperlink Express 1.1, Internal Shared Memory (ISM) SMC-D, and other management functions.
- ▶ Two spare PUs, which can transparently assume any characterization during a permanent failure of another PU.

The z16 uses features to define the number of PUs that are available for client use in each configuration. The features are listed in Figure 2-12.

¹⁰ The KVM hypervisor is part of supported IBM Linux on Z distributions.

Table 2-12 IBM z16 Az16 AGZ Processor Configurations

Feature	CPC Drawers	PUs per drawer	Active PUs				zIIP	IFP	STD SAPs	Spares	MAX Memory TB
			CPs	IFLs	ICFs	ulIFLs					
Max5	1	32	0-5	0-5	0-5	0-4	0-4	2	2	2	4
Max16	1	32	0-6	0-16	0-16	0-15	0-15	2	2	2	4
Max32	1	32	0-6	0-32	0-32	0-31	0-31	2	4	2	8
Max68	2	68	0-6	0-68	0-68	0-67	0-67	2	8	2	16

- ▶ Not all PUs available on a feature are required to be characterized with a feature code. Only the PUs purchased by a client are identified with a feature code.
- ▶ All PU conversions can be performed concurrently.

A *capacity marker* identifies the number of CPs that were purchased. This number of purchased CPs is higher than or equal to the number of CPs that is actively used. The capacity marker marks the availability of purchased but unused capacity that is intended to be used as CPs in the future. This status often is present for software-charging reasons.

Unused CPs are not a factor when establishing the millions of service units (MSU) value that is used for charging monthly license charge (MLC) software, or when charged on a per-processor basis.

2.8.1 Upgrades

Concurrent upgrades of CPs, IFLs, ICFs, zIIPs are available for the IBM z16 A02 and IBM z16 AGZ. However, concurrent PU upgrades require that more PUs are installed but not activated.

Spare PUs are used to replace defective PUs. Two spare PUs always are on an IBM z16 A02 and IBM z16 AGZ. In the rare event of a PU failure, a spare PU is activated concurrently and transparently and is assigned the characteristics of the failing PU.

If an upgrade request cannot be accomplished within the configuration, a hardware upgrade is required.

The upgrade paths for IBM z16 A02 and IBM z16 AGZ are shown in Figure 2-25 on page 59:

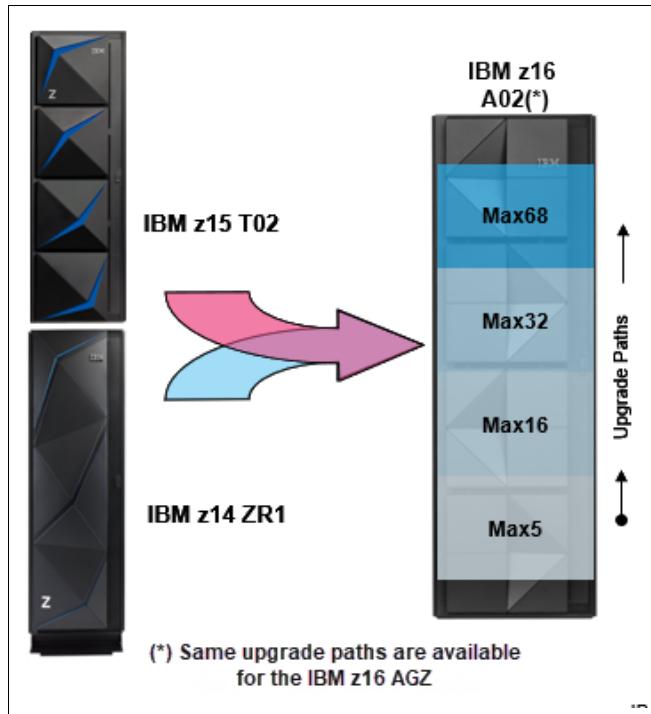


Figure 2-25 IBM z16 A02 upgrade paths

- ▶ IBM z16 A02 and IBM z16 AGZ upgrades:
 - Max5 to Max16 is concurrent
 - Max5 to Max32 or Max68 is disruptive
 - Max16 to Max32 or Max68 is disruptive
 - More I/O drawers can be added concurrently based on available space in current frames for IBM z16 A02 and IBM z16 AGZ or with plan ahead for the IBM z16 A02 and IBM z16 AGZ.
- ▶ IBM z14 (M/T 3907-ZR1) to IBM z16 A02 and IBM z16 AGZ:
 - Feature conversion of installed zAAPs to zIIPs (default) or another processor type
 - For installed OnDemand Records, stage the records
- ▶ IBM z15 (M/T 8562-T02) to IBM z16 A02 or IBM z16 AGZ:

Supported feature MaxYY upgrade paths are summarized in Table 2-13.

Table 2-13 Supported MaxYY upgrade paths

Feature To / From	Max5	Max16	Max32	Max68
Max5	N/A	Concurrent	Disruptive	Disruptive
Max16	N/A	NA	Disruptive	Disruptive
Max32	N/A	NA	NA	Concurrent w/plan ahead

2.8.2 Model capacity identifier

To recognize how many PUs are characterized as CPs, the Store System Information (STSI) instruction returns a Model Capacity Identifier (MCI). The MCI determines the number and speed of characterized CPs. Characterization of a PU as an IFL, ICF, or zIIP is not reflected in the output of the STSI instruction because characterization has no effect on software charging. For more information about STSI output, see “Processor identification” on page 357.

The following distinct model capacity identifier ranges are recognized (one for full capacity and three for granular capacity):

- ▶ For full-capacity engines, model capacity identifiers Z01 - Z06 are used. They express capacity settings for 1 - 6 characterized CPs.
- ▶ Three model capacity identifier ranges offer a unique level of granular capacity at the low end. They are available as A to Z capacity setting for 1 - 6 CPs characterized, which offers 156 subcapacity settings. For more information, see “Granular capacity”.

Granular capacity

The z16 A02 and the IBM z16 AGZ configurations offer 156 capacity settings for 1 - 6 CPs. The subcapacity settings are defined as models A01 - Z06. Up to 6 CPs can be characterized for customer use. The reminder PUs (feature dependent) can be characterized as IFLs, ICFs, unassigned IFLs, zIIPs (if CPs are available) and extra SAPs.

The 156 defined ranges of subcapacity settings feature model capacity identifiers numbered A01- Z01, A02 - Z02, A03 - Z03, A04 - Z04, A05 - Z05, and A06 - Z06¹¹.

Consideration: All CPs have the same capacity identifier. Specialty engines (IFLs, zIIPs, and ICFs) operate at full speed.

List of model capacity identifiers (MCI)

Regardless of the number of CPC drawers, a configuration with one characterized CP is possible, as listed in Figure 2-26 on page 61.

Note: A00 MCI has ICFs or IFLs only (zero CPs).

¹¹ Max5 max subcapacities are A05 to z05.

Z01	Z02	Z03	Z04	Z05	Z06
Y01	Y02	Y03	Y04	Y05	Y06
X01	X02	X03	X04	X05	X06
W01	W02	W03	W04	W05	W06
V01	V02	V03	V04	V05	V06
U01	U02	U03	U04	U05	U06
T01	T02	T03	T04	T05	T06
S01	S02	S03	S04	S05	S06
R01	R02	R03	R04	R05	R06
Q01	Q02	Q03	Q04	Q05	Q06
P01	P02	P03	P04	P05	P06
O01	O02	O03	O04	O05	O06
N01	N02	N03	N04	N05	N06
M01	M02	M03	M04	M05	M06
L01	L02	L03	L04	L05	L06
K01	K02	K03	K04	K05	K06
J01	J02	J03	J04	J05	J06
I01	I02	I03	I04	I05	I06
H01	H02	H03	H04	H05	H06
G01	G02	G03	G04	G05	G06
F01	F02	F03	F04	F05	F06
E01	E02	E03	E04	E05	E06
D01	D02	D03	D04	D05	D06
C01	C02	C03	C04	C05	C06
B01	B02	B03	B04	B05	B06
A01	A02	A03	A04	A05	A06
1-way	2-way	3-way	4-way	5-way	6-way
Specialty Engine	Specialty Engine	Specialty Engine	Specialty Engine	Specialty Engine	Specialty Engine
→ 62 more specialty engines (w/ FC 0675 – Max68)					

Figure 2-26 IBM z16 A02 and IBM z16 AGZ Model capacity identifiers

For more information about temporary capacity increases, see Chapter 8, “System upgrades” on page 317.

2.8.3 Capacity Backup Upgrade

Capacity Backup Upgrade (CBU) delivers temporary backup capacity in addition to the capacity that an installation might have available in numbers of assigned CPs, IFLs, ICFs, and zIIPs. CBU has the following types:

- ▶ CBU for CP
- ▶ CBU for IFL
- ▶ CBU for ICF
- ▶ CBU for zIIP

When CBU for CP is added within the same capacity setting range (indicated by the model capacity identifier) as the currently assigned PUs, the total number of active PUs (the sum of all assigned CPs, IFLs, ICFs, and zIIPs) plus the number of CBUs cannot exceed the total number of PUs that are available in the system.

When CBU for CP capacity is acquired by switching from one capacity setting to another, no more CBUs can be requested than the total number of PUs available for that capacity setting.

CBU and granular capacity

When CBU for CP is ordered, it replaces lost capacity for disaster recovery. Specialty engines (ICFs, IFLs, and zIIPs) always run at full capacity, and when running as a CBU to replace lost capacity for disaster recovery.

When you order CBU, specify the maximum number of CPs, ICFs, IFLs, and zIIPs to be activated for disaster recovery. If a disaster occurs, you decide how many of each of the contracted CBUs of any type to activate. The CBU rights are registered in one or more records in the CPC. Up to eight records can be active, which can contain various CBU activation variations that apply to the installation.

The number of CBU test activations that you can run for no extra fee in each CBU record is now determined by the number of years that are purchased with the CBU record. For example, a three-year CBU record includes three test activations, as compared to a one-year CBU record that has one test activation.

You can increase the number of tests up to a maximum of 15 for each CBU record. The real activation of CBU lasts up to 90 days with a grace period of two days to prevent sudden deactivation when the 90-day period expires. The contract duration can be set 1 - 5 years.

The CBU record describes the following properties that are related to the CBU:

- ▶ Number of CP CBUs that are allowed to be activated
- ▶ Number of IFL CBUs that are allowed to be activated
- ▶ Number of ICF CBUs that are allowed to be activated
- ▶ Number of zIIP CBUs that are allowed to be activated
- ▶ Number of SAP CBUs that are allowed to be activated
- ▶ Number of extra CBU tests that are allowed for this CBU record
- ▶ Number of total CBU years ordered (duration of the contract)
- ▶ Expiration date of the CBU contract

The record content of the CBU configuration is documented in IBM configurator output, which is shown in Example 2-1. In this example, one CBU record is made for a five-year CBU contract without more CBU tests for the activation of one CP CBU.

Example 2-1 Simple CBU record and related configuration features

On-Demand Capacity Selections:

NEW00001 - CBU - CP(1) - Years(5) - Tests(5)

Resulting feature numbers in configuration:

6817 Total CBU Years Ordered	5
6818 CBU Records Ordered	1
6820 Single CBU CP-Year	5

In Example 2-2, a second CBU record is added to the configuration for two CP CBUs, two IFL CBUs, and two zIIP CBUs, with five more tests and a five-year CBU contract. The result is that a total number of 10 years of CBU ordered: Five years in the first record and five years in the second record. The two CBU records are independent and can be activated individually. Five more CBU tests were requested. Because a total of five years are contracted for a total of three CP CBUs (two IFL CBUs and two zIIP CBUs), they are shown as 15, 10, 10, and 10 CBU years for their respective types.

Example 2-2 Second CBU record and resulting configuration features

NEW00001 - CBU - Replenishment is required to reactivate
Expiration(06/21/2017)
NEW00002 - CBU - CP(2) - IFL(2) - zIIP(2)
Total Tests(5) - Years(5)

Resulting cumulative feature numbers in configuration:

6817	Total CBU Years Ordered	10
6818	CBU Records Ordered	2
6819	5 Additional CBU Tests	1
6820	Single CBU CP-Year	15
6822	Single CBU IFL-Year	10
6828	Single CBU zIIP-Year	10

CBU for CP rules

Consider the following guidelines when you are planning for CBU for CP capacity:

- ▶ The total CBU CP capacity features are equal to the number of added CPs plus the number of permanent CPs that change the capacity level. For example, if two CBU CPs are added to the current model D03, and the capacity level does not change, the D03 becomes D05, as shown in the following example:

$$(D03 + 2 = D05)$$

If the capacity level changes to a E06, the number of extra CPs (three) is added to the three CPs of the D03, which results in a total number of CBU CP capacity features of six:

$$(3 + 3 = 6)$$

- ▶ The CBU cannot decrease the number of CPs.
- ▶ The CBU cannot lower the capacity setting.

Remember: CBU for CPs, IFLs, ICFs, and zIIPs can be activated together with On/Off Capacity on-Demand (CoD) temporary upgrades. Both facilities can be on a single system, and can be activated simultaneously.

CBU for specialty engines

Specialty engines (ICFs, IFLs, and zIIPs) run at full capacity for all capacity settings. This fact also applies to CBU for specialty engines. The minimum and maximum (min-max) numbers of all types of CBUs that can be activated on each of the features are listed in Table 2-14. The CBU record can contain larger numbers of CBUs than can fit in the current feature.

Table 2-14 Capacity Backup matrix

IBM z16 A02 and IBM z16 AGZ feature	Total PUs available	CBU CPs min - max	CBU IFLs min - max	CBU ICFs min - max	CBU ^a zIIPs min - max
Max68	68	0-6	0 - 68	0 - 68	0 - 67
Max32	31	0-6	0 - 32	0 - 32	0 - 31
Max16	21	0-6	0 - 16	0 - 16	0 - 15
Max5	5	0-5	0 - 5	0 - 5	0 - 4

a. At least one CP is needed to achieve the maximum number of zIIPs.

2.8.4 On/Off Capacity on Demand and CPs

On/Off CoD provides temporary capacity for all types of characterized PUs. Relative to granular capacity, On/Off CoD for CPs is treated similarly to the way that CBU is handled.

On/Off CoD and granular capacity

When temporary capacity that is requested by On/Off CoD for CPs matches the model capacity identifier range of the permanent CP feature, the total number of active CPs equals the sum of the number of permanent CPs plus the number of temporary CPs ordered. For example, when a model capacity identifier D03 has two CPs added temporarily, it becomes a model capacity identifier D05.

When the addition of temporary capacity that is requested by On/Off CoD for CPs results in a cross-over from one capacity identifier range to another, the total number of CPs active when the temporary CPs are activated is equal to the number of temporary CPs ordered. For example, when a configuration with model capacity identifier D03 specifies four temporary CPs through On/Off CoD, the result is a configuration with model capacity identifier E05.

A cross-over does not necessarily mean that the CP count for the extra temporary capacity increases. The same D03 can temporarily be upgraded to a configuration with model capacity identifier F03. In this case, the number of CPs does not increase, but more temporary capacity is achieved.

On/Off CoD guidelines

When you request temporary capacity, consider the following guidelines:

- ▶ Temporary capacity must be greater than permanent capacity.
- ▶ Temporary capacity cannot be more than double the purchased capacity.
- ▶ On/Off CoD cannot decrease the number of engines on the CPC.
- ▶ The number of engines cannot be increased to more than what is installed.

For more information about temporary capacity increases, see Chapter 8, “System upgrades” on page 317.

Flexible Capacity for Cyber Resiliency

IBM Z Flexible Capacity for Cyber Resiliency can be used to remotely shift capacity and production workloads between IBM z16 A01, IBM z16 A02 and IBM z16 AGZ systems at different sites on demand, and stay at the alternate site for up to one year. This capability can help demonstrate compliance with regulations that require organizations to be able to dynamically shift production to an alternate site and remain there for an extended period. This capability is also designed to help you proactively avoid disruptions from unplanned events. For example, it enables you to move production workload to avoid disruptions from an impending hurricane, flood, or wildfire, as well from planned scenarios such as site facility maintenance. For additional information, refer to this Redpaper publication: [IBM Z Flexible Capacity for Cyber Resiliency, REDP-5702-00](#)

2.9 Power and cooling

The IBM z16 A02 power and cooling is similar to IBM z15 T02 and uses intelligent power distribution units (iPDUs) to supply power to the system components. Consider the following points:

- ▶ The power subsystem is based on the following offerings:
 - Power Distribution Units (iPDUs) - single phase. Single phase power is available for systems with a single CPC drawer and w/o FC 2271
 - Power Distribution Units (iPDUs) - three phase - available for all configurations
- ▶ The three-phase power iPDUs can be:
 - Low voltage 4 wire “Delta”
 - High voltage 5 wire “Wye”
- ▶ No EPO (emergency power off) switch is used.
IBM z16 A02 has a support element task to simulate the EPO function (only used when necessary to do a System Reset Function).
- ▶ No DC input feature is available. Internal Battery Feature *is not available* for z16 A02.
- ▶ The system is air-cooled and has redundant design for the blowers (fans) and power supplies.
- ▶ No Top Exit Power feature is available because the 19-inch frame is capable of top or bottom exit of power. All line cords are 4.26 meters (14 feet). Combined with the Top Exit I/O Cabling feature, more options are available when you are planning your computer room cabling.
- ▶ The new PSCN¹² structure uses industry standard Ethernet switches (redundant, 1+1).

2.9.1 Power considerations- IBM z16 A02 (factory frame)

The IBM zSystems operate with redundant power infrastructure.

The IBM z16 A02 is designed with a power infrastructure that is based on intelligent Power Distribution Units (iPDUs) that are mounted vertically on the rear side of the 19-inch frame and Power Supply Units for the internal components.

The iPDUs are controlled by using an Ethernet port and support the following inputs:

- ▶ 3-phase 200 - 240 V AC (wired as “Delta”)

¹² Power System Control Network - PSCN

- ▶ 3-phase 380 - 415 V AC (wired as “Wye”)
- ▶ Single phase 200 - 240 V AC.

The power supply units convert the AC power to DC power that is used as input for the Points of Load (POLs) in the CPC drawer and the PCIe+ I/O drawers.

The power requirements depend on the number of CPC drawers (1 or 2), number of PCIe I/O drawers (0 - 3) and I/O features that are installed in the PCIe I/O drawers.

iPDUs are installed and serviced from the rear of the frame. Unused power ports are never used by any external device.

Each iPDU installed requires a customer supplied power feed. The number of power cords that are required depends on the system configuration.

Note: For initial installation, all power sources are required to run the system checkout diagnostics successfully.

iPDUs are installed in pairs. A system can have two or four iPDUs, depending on the configuration. Consider the following points:

- ▶ Paired iPDUs are A1/A2 and A3/A4.
- ▶ From the rear of the system, the odd-numbered PDUs are on the left side of the rack, and the even-numbered iPDUs are on the right side of the rack.
- ▶ The total loss of one iPDU in a pair has no effect on the system operation.

Components that plug into the PDUs for redundancy (using two power cords) include the following features:

- ▶ CPC Drawers, PCIe+ I/O drawers, Radiators, and Support Elements
- ▶ The redundancy for each component is achieved by plugging the power cables into the paired iPDUs.

For example, the top Support Element (1), has one power supply plugged into iPDU A1 and the second power supply plugged into the paired iPDU A2 for redundancy.

Note: Customer power sources should always maintain redundancy across PDU pairs; that is, one power source or distribution panel supplies power for iPDU A1 and the separate power source or distribution panel supplies power for iPDU A2.

As a best practice, connect the odd-numbered iPDUs (A1, B1) to one power source or distribution panel, and the even-numbered iPDUs (A2, B2) to a separate power source or distribution panel.

2.9.2 Power considerations - IBM z16 AGZ

The IBM z16 AGZ Power Distribution Units (PDUs) are supplied by the client and are installed on the rear side of the 19-inch rack. The system order comes with a set of power cables for connecting the system components to the PDUs.

The PDUs must be installed in the same rack with the system components to provide connectivity from the rear side of the rack.

For additional details and planning, see the *IBM z16 and LinuxONE Rockhopper 4 Rack Mount Bundle Installation Manual (Models AGZ/AGL)*, GC28-7036.

The PDUs must support the following inputs (depending on system configuration):

- ▶ 3-phase 200 - 240 V AC (wired as “Delta”)
- ▶ 3-phase 380 - 415 V AC (wired as “Wye”)
- ▶ Single phase 200 - 240 V AC.

The power supply units convert the AC power to DC power that is used as input for the Points of Load (POLs) in the CPC drawer and the PCIe+ I/O drawers.

The power requirements depend on the number of CPC drawers (1 or 2), number of PCIe I/O drawers (0 - 3) and I/O features that are installed in the PCIe I/O drawers.

Each PDU installed requires a customer supplied power feed. The number of power cords that are required depends on the system configuration.

System components must be plugged into the PDUs for redundancy (using two power cords).

2.9.3 Power estimation tool

The power estimation tool for the z16 A02 allows you to enter your precise and detailed configuration to obtain an *estimate* of power consumption. Log in to the [Resource link](#) with your user ID. Click **Planning** → **Tools** → **Power and weight estimation**. Specify the quantity for the features that are installed in your system.

This tool estimates the power consumption for the specified configuration. The tool does *not* verify that the specified configuration can be physically built.

Tip: The exact power consumption for your system varies. The object of the tool is to estimate the power requirements to aid you in planning for your system installation. Actual power consumption after installation can be confirmed by using the HMC Monitors Dashboard task.

2.9.4 Cooling

The PU DCMs are air-cooled. For IBM z16 A02 and IBM z16 AGZ, the CPC drawer components and the PCIe+ I/O drawers are air cooled by redundant fans. Airflow of the system is directed from front (cool air) to the back of the system (hot air).

2.10 Summary

All aspects of the IBM z16 A02 and IBM z16 AGZ structure are listed in Table 2-15.

Table 2-15 System structure summary

Description	Max5	Max16	Max32	Max68
Maximum number of characterized PUs	5	16	32	68
Number of CPC Drawers	1	1	1	2
Number of CP chips / DCMs	4 / 2	4 / 2	8 / 4	16 / 8

Description	Max5	Max16	Max32	Max68
Number of CPs	0 - 5	0 - 6	0 - 6	0 - 6
Number of IFLs	0 - 5	0 - 16	0 - 32	0 - 68
Number of Unassigned IFLs	0 - 4	0 - 15	0 - 31	0 - 67
Number of ICFs	0 - 5	0 - 16	0 - 32	0 - 68
Number of Unassigned ICFs	0 - 4	0 - 15	0 - 31	0 - 67
Number of zIIPs	0 - 4	0 - 15	0 - 31	0 - 67
Number of Unassigned zIIPs	0 - 3	0 - 114	0 - 30	0 - 66
Standard SAPs	2	2	4	8
Additional SAPs	0	0	0	0
Number of IFP	2	2	2	2
Standard spare PUs	2	2	2	2
Enabled Memory sizes GB	64 - 3936	64 - 3936	64 - 8032	64 - 16128
L1 cache per PU (I/D)	128/128 KB	128/128 KB	128/128 KB	128/128 KB
L2 private unified cache per PU	32 MB	32 MB	32 MB	32 MB
L3 virtual cache on PU chip	32 x 7 =256MB	32 x 7 =256MB	32 x 7 =256MB	32 x 7 =256MB
L4 virtual cache on chips in drawer	32 x 8 x 7 =1024GB	32 x 8 x 7 =1024GB	32 x 8 x 7 =2048GB	32 x 8 x 7 =2048GB
Cycle time (ns)	0.217	0.217	0.217	0.217
Clock frequency	4.6GHz	4.6 GHz	4.6 GHz	4.6 GHz
Maximum number of PCIe fanouts	6	6	12	24
PCIe Bandwidth	16 GBps	16 GBps	16 GBps	16 GBps
Number of support elements or HMA	2	2	2	2
External AC power	1- or 3-phase	1- or 3-phase	1- or 3-phase	3-phase



Central processor complex design

This chapter describes the design of the IBM z16 A02 and IBM z16 AGZ processor unit. By understanding this design, users become familiar with the functions that make the IBM z16 A02 and IBM z16 AGZ a system that accommodates a broad mix of workloads for the enterprise.

This chapter includes the following topics:

- ▶ 3.1, “Overview” on page 70
- ▶ 3.2, “Design highlights” on page 71
- ▶ 3.3, “CPC drawer design” on page 73
- ▶ 3.4, “Processor unit design” on page 78
- ▶ 3.5, “Processor unit functions” on page 98
- ▶ 3.6, “Memory design” on page 113
- ▶ 3.7, “Logical partitioning” on page 117
- ▶ 3.8, “Intelligent Resource Director” on page 128
- ▶ 3.9, “Clustering technology” on page 129
- ▶ 3.10, “Virtual Flash Memory” on page 132
- ▶ 3.11, “Secure Service Container” on page 133

3.1 Overview

The IBM z16 A02 and IBM z16 AGZ symmetric multiprocessor (SMP) system is the next step in an evolutionary journey that began with the introduction of the IBM System/360 in 1964. Over time, the design was adapted to the changing requirements that were dictated by the shift toward new types of applications on which clients depend.

IBM zSystems offer high levels of reliability, availability, serviceability (RAS), resilience, and security. The IBM z16 A02 and IBM z16 AGZ fits into the IBM strategy in which mainframes play a central role in creating an infrastructure for cloud, artificial intelligence, and analytics, which is underpinned by security. The IBM z16 A02 and IBM z16 AGZ ares designed so that everything around it, such as operating systems, middleware, storage, security, and network technologies that support open standards, helps you achieve your business goals.

The IBM z16 A02 and IBM z16 AGZ extend the platform's capabilities and adds value with breakthrough technologies, such as the following examples:

- ▶ On-chip Artificial Intelligence (AI) at speed and scale that is designed to leave no transaction behind.
- ▶ An industry-first system that uses quantum-safe technologies, cryptographic discovery tools, and end-to-end data encryption to protect against future attacks now.
- ▶ A continuous compliance solution to help keep up with changing regulations, which reduces cost and risk exposure.
- ▶ A consistent cloud experience to enable accelerated modernization, rapid delivery of new services, and end-to-end automation.
- ▶ New options in flexible and responsible consumption to manage system resources across geographical locations, with sustainability that is built in across its lifecycle.

The modular CPC drawer design aims to reduce (or in some cases even eliminate) planned and unplanned outages. The design does so by offering concurrent repair, replace, and upgrade functions for processors, memory, and I/O.¹

For more information about the IBM z16 A02 and IBM z16 AGZ RAS features, see Chapter 9, “Reliability, availability, and serviceability” on page 363.

IBM z16 A02 and IBM z16 AGZ configurations include the following features:

- ▶ Ultra-high frequency, large, high-speed buffers (caches) and memory
- ▶ Superscalar processor design
- ▶ Improved out-of-order core execution
- ▶ Simultaneous multithreading (SMT)
- ▶ Single-instruction multiple-data (SIMD)
- ▶ On-core integrated accelerator for Z SORT, one per PU core
- ▶ On-chip integrated accelerator for IBM zEnterprise® Data Compression (zEDC), one per PU chip
- ▶ On-chip integrated accelerator for AI (AI unit or AIU), one per PU chip
- ▶ Quantum-safe cryptography support
- ▶ Flexible configuration options

IBM z16 A02 and IBM z16 AGZ are the next implementation of IBM zSystems to address the ever-changing IT environment.

¹ Some concurrent actions are only available on the Max68 feature.

For more information about frames and configurations, see Chapter 2, “Central processor complex hardware components” on page 21.

3.2 Design highlights

The physical packaging of IBM z16 A02 and IBM z16 AGZ CPC drawer is a continuation and evolution of the previous generations of IBM zSystems. Its modular CPC drawer and new dual chip module (DCM) design address the augmenting costs that are related to building systems with ever-increasing capacities.

The modular CPC drawer design is flexible and expandable. It offers unprecedented capacity and security features to meet consolidation needs.

IBM z16 A02 and IBM z16 AGZCPC continues the line of mainframe processors that are compatible with an earlier version. The IBM z16 A02 and IBM z16 AGZ brings the following processor design enhancements:

- ▶ 7 nm EUV lithography using FinFET silicon process
- ▶ Eight cores per PU chip design with 22.5 billion transistors per PU chip
- ▶ Redesigned cache structure that is implemented in dense SRAM
- ▶ Four PU Dual Chip Modules per CPC Drawer
- ▶ Each PU chip features:
 - Two PCIe Generation 4 interfaces (x16 @ 32 GBps)
 - IBM integrated accelerator for AI (on-chip AI accelerator)
 - Transparent memory encryption.
 - Optimized pipeline
 - Improved SMT and SIMD
 - Improved branch prediction
 - Improved co-processor functions (CPACF)
 - IBM integrated accelerator for zEnterprise Data Compression (zEDC) (on-chip compression accelerator)
 - IBM integrated accelerator for Z Sort (on-core sort accelerator)

The processor architecture uses 24-, 31-, and 64-bit addressing modes, multiple arithmetic formats, and multiple address spaces for robust interprocess security.

The IBM z16 A02 and IBM z16 AGZ system design features the following main objectives:

- ▶ Offer a data-centric approach to information (data) security that is simple, transparent, and consumable (extensive data encryption from inception to archive, in-flight, and at-rest).
- ▶ Offer a flexible infrastructure to concurrently accommodate a wide range of operating systems and applications, from the traditional systems (for example, z/OS and z/VM) to the world of Linux, cloud, analytics, and mobile computing.
- ▶ Offer state-of-the-art integration capability for server consolidation by using virtualization capabilities in a highly secure environment:
 - Logical partitioning, which allows up to 40 independent logical servers.
 - z/VM, which can virtualize hundreds to thousands of servers as independently running virtual machines (guests).
 - HiperSockets, which implement virtual LANs between logical partitions (LPARs) within the system.
 - Efficient data transfer that uses direct memory access (SMC-D), Remote Direct Memory Access (SMC-R), and reduced storage access latency for transactional environments.

- The IBM zSystems Processor Resource/System Manager (PR/SM) is designed for Common Criteria Evaluation Assurance Level 5+ (EAL 5+) certification for security; therefore, an application that is running on one partition (LPAR) cannot access another application on a different partition, which provides essentially the same security as an air-gapped system.
- The Secure Execution feature securely separates second-level guest operating systems running under KVM for IBM Z from each other and securely separates access to second-level guests from the hypervisor.

This configuration allows for a logical and virtual server coexistence and maximizes system utilization and efficiency by sharing hardware resources.

- ▶ Offer high-performance computing to achieve the outstanding response times that are required by new workload-type applications. This performance is achieved by high-frequency, enhanced superscalar processor technology, out-of-order core execution, large high-speed buffers (cache) and memory, an architecture with multiple complex instructions, and high-bandwidth channels.
- ▶ Offer the high capacity and scalability that are required by the most demanding applications, from the single-system and clustered-systems points of view.
- ▶ Offer the capability of concurrent upgrades for processors, memory, and I/O connectivity, which prevents system outages in planned situations.
- ▶ Implement a system with high availability and reliability. These goals are achieved with redundancy of critical elements and sparing components of a single system, and the clustering technology of the Parallel Sysplex environment.
- ▶ Have internal and external connectivity offerings, supporting open standards, such as Gigabit Ethernet (GbE) and Fibre Channel Protocol (FCP).
- ▶ Provide leading cryptographic performance. Every processor unit (PU) includes a dedicated and optimized CP Assist for Cryptographic Function (CPACF). Optional Crypto Express features with cryptographic coprocessors provide the highest standardized security certification.² These optional features also can be configured as Cryptographic Accelerators to enhance the performance of Secure Sockets Layer/Transport Layer Security (SSL/TLS) transactions.
- ▶ Provide on-chip compression. Every PU chip design incorporates a compression unit, which is the IBM Integrated Accelerator for z Enterprise Data Compression (zEDC). This configuration is different from the CMPSC (Compression Coprocessor) that is implemented in each core.
- ▶ Provide a new dedicated on-chip integrated AI Accelerator for high-speed inference to enable real-time AI embedded directly in transactional workloads, and improvements for performance, security, and availability.
- ▶ Be self-managing and self-optimizing, adjusting itself when the workload changes to achieve the best system throughput. This process can be done by using the Intelligent Resource Director or the Workload Manager functions, which are assisted by HiperDispatch.
- ▶ Have a balanced system design with pervasive encryption, which provides large data rate bandwidths for high-performance connectivity along with processor and system capacity, while having the capability of protecting every byte that enters and exits the IBM z16 A02 and IBM z16 AGZ.

² Federal Information Processing Standard (FIPS) 140-2 Security Requirements for Cryptographic Modules.

The remaining sections in this chapter describe the IBM z16 A02 and IBM z16 AGZ system structure. It shows a logical representation of the data flow from PUs, caches, memory cards, and various interconnect capabilities.

3.3 CPC drawer design

An IBM z16 A02 and IBM z16 AGZ can have up to two CPC drawers in a full configuration, with up to 68 PUs that can be characterized for customer use, and up to 16 TB of customer usable memory.

The following features for CPC drawer configurations are available for the IBM z16 A02 and IBM z16 AGZ:

- ▶ One drawer, two PU dual chip modules (DCM): Max5
- ▶ One drawer, two PU DCMs: Max16
- ▶ One drawer, four PU DCMs: Max32
- ▶ Two drawers, eight PU DCMs: Max68

The IBM z16 A02 and IBM z16 AGZ have 12 memory controller units (MCUs) for a Max68 feature (one MCU per PU chip, and up to six MCUs populated per CPC drawer). The MCU configuration uses eight-channel Reed-Solomon redundant array of independent memory (RAIM).

The RAIM design is new compared to the IBM z15 T02, moving from a 4+1 DIMM structure that is based on 5-channel RAIM design on IBM z15 to an 8-channel R-S RAIM design on the IBM z16 A02 and IBM z16 AGZ. The new memory architecture provides approximately 15% DRAM reduction for a similar RAS, but a higher memory bandwidth at drawer level.

The DIMM sizes (32, 64, 128, or 256 GB) include RAIM overhead. An IBM z16 A02 and IBM z16 AGZ CPC drawer can have up to 48 memory DDR4 DIMMs (populated with 16, 32, 40, or 48 DIMMs).

The IBM z16 A02 and IBM z16 AGZ microprocessor chip (called IBM Telum) integrates a new cache hierarchy design with only two levels of physical cache (L1 and L2). The cache hierarchy (L1, L2) is implemented with dense static random access memory (SRAM).

Unlike the IBM z15, eDRAM cache is no longer used in the IBM Telum processor. On an IBM z16 A02 and IBM z16 AGZ, L2 cache (32 MB) is semi-private with 16 MB dedicated to the associated core, and 16 MB shared with the system (the 50/50 split is adjustable) implemented in SRAM³. Level 3 (L3) and Level 4 (L4) caches are now virtual caches and are allocated on L2.

Two processor chips (up to eight active cores per PU chip) are combined in a Dual Chip Module (DCM) and up to four DCMs are assembled in a CPC drawer. An IBM z16 A02 and IBM z16 AGZ can have either one CPC drawer (Max5, Max16, and Max32) or two CPC drawers (Max68).

Figure 3-1 on page 74 shows the new IBM Telum processor.

³ SRAM - Static Random Access Memory. Unlike eDRAM, SRAM does not require refresh.

- 7nm silicon technology (FinFET)
- 530 mm² chip size
- 22.5 Billion transistors
- 4.6 GHz clock frequency
- New cache structure
 - L1 cache - ON-core
 - L1D(data) and L1I(instruction) caches - 128K each
 - L2 - dense SRAM
 - outside the core, semi-private to the core – 32 MB
 - L3 (virtual) == up to 256 MB
 - L4 (virtual) == up to 2048 MB
- New Core-Nest Interface
- Brand new branch prediction design using SRAM
- Significant architecture changes – COBOL compiler & more
- On chip AI – deep learning focus for inference

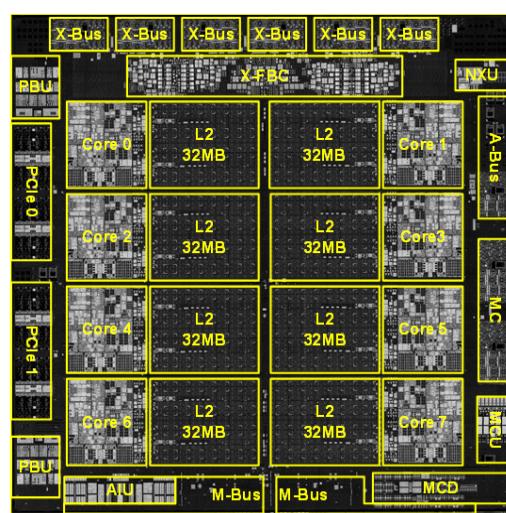


Figure 3-1 IBM Telum processor

The new IBM z16 A02 and IBM z16 AGZ Dual Chip Module (DCM) is shown in Figure 3-2.

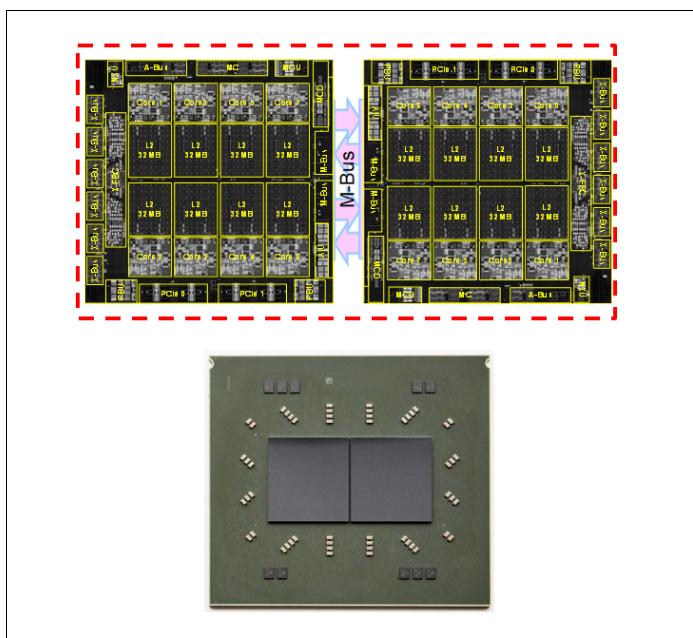


Figure 3-2 IBM z16 A02 and IBM z16 AGZ Dual Chip Module (DCM)

Concurrent maintenance allows dynamic central processing complex (CPC) drawer add and repair.⁴

IBM z16 A02 and IBM z16 AGZ processors are manufactured using 7 nm extreme ultraviolet (EUV) FinFET silicon technology with advanced low latency pipeline design, which creates high-speed yet power-efficient circuit designs. The PU DCMs are air-cooled. For more information, see 2.9, “Power and cooling” on page 65.

⁴ Repair only for configurations with two CPC drawers installed.

3.3.1 Cache levels and memory structure

The IBM z16 A02 and IBM z16 AGZ include a new optimized memory subsystem design that focuses on keeping data closer to the PU core. With the current processor configuration, all on-chip cache levels increased.

The cache hierarchy is shown in Figure 3-3.

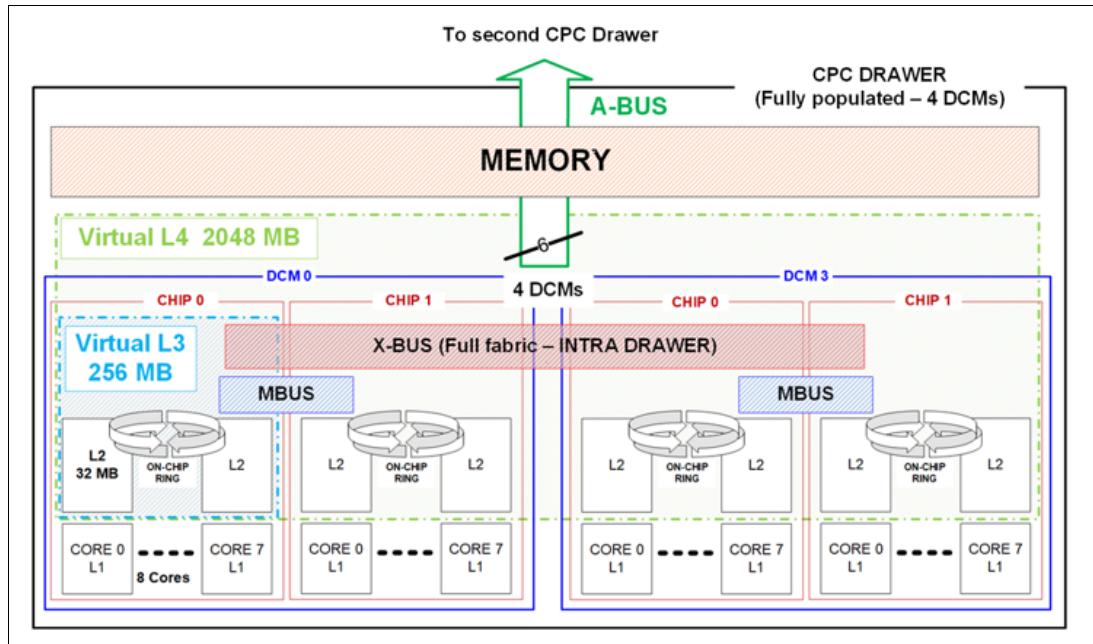


Figure 3-3 IBM z16 A02 and IBM z16 AGZ cache levels and memory hierarchy

The cache structure of the IBM z16 A02 and IBM z16 AGZ features the following characteristics:

- ▶ Large L1, L2 caches (more data closer to the core).
- ▶ L1 cache is implemented as SRAM⁵ and has the same size as on IBM z15 T02 (128 KB for instructions and 128 KB for data).
- ▶ L2 cache (32 MB in total) also uses SRAM technology, and is semi-private to each PU core with 16 MB dedicated to the associated core, and 16 MB shared with the system (the 50/50 split is adjustable).
- ▶ L3 cache (up to 256 MB per chip) now becomes a virtual cache and can be allocated on any of the share part of a L2 cache.
- ▶ L4 cache (per drawer - up to 1024 MB for the Max5 and Max16, up to 2048 MB for the Max32 and Max68) is also a virtual cache and can be allocated on any of the share part of a L2 cache.

Figure 3-4 shows the new cache structure that is implemented in an fully populated IBM z16 A02 and IBM z16 AGZ CPC drawer.

⁵ SRAM - Static Random Access Memory (not refresh required)

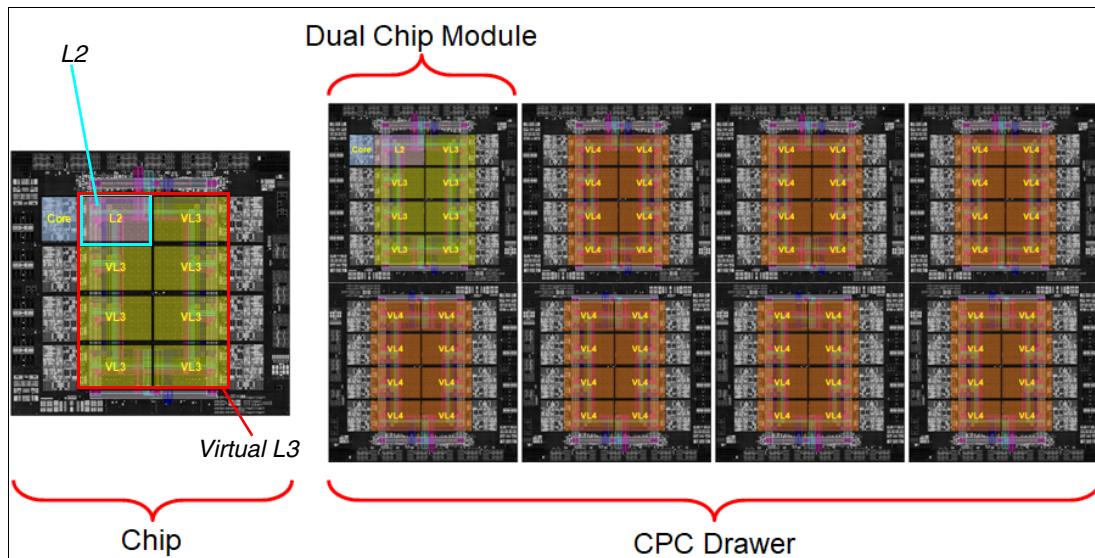


Figure 3-4 IBM z16 A02 and IBM z16 AGZ cache structure at CPC drawer level

Main storage has up to 8 TB addressable memory per CPC drawer, which uses up to 48 DDR4 DIMMs. A system with two CPC drawers can have up to 16 TB of main storage.

Considerations

Cache sizes are limited by ever-shrinking cycle times because they must respond quickly without creating bottlenecks. Access to large caches costs more cycles. Instruction and data cache (L1) sizes must be limited because larger distances must be traveled to reach long cache lines. This L1 access time generally occurs in one cycle, which prevents increased latency.

Also, the distance to remote caches as seen from the microprocessor becomes a significant factor. For example, on an IBM z15, access to L4 physical cache (on the SC chip and which might not even be in the same CPC drawer) requires several cycles to travel the distance to the cache. On an IBM z16 A02 and IBM z16 AGZ, having an L4 virtual, physically allocated on the shared L2 requires fewer processor cycles in many instances.

Although large caches mean increased access latency, the new technology 7 nm EUV chip lithography and the lower cycle time allows IBM z16 A02 and IBM z16 AGZ to increase the size of L2 cache level within the PU chip.

To overcome the inherent delays of the SMP CPC drawer design and save cycles to access the remote virtual L4 content, the system keeps instructions and data as close to the processors as possible. This configuration can be managed by directing as much work of a specific LPAR workload to the processors in the same CPC drawer as the L4 virtual cache.

This configuration is achieved by having the IBM Processor Resource/Systems Manager (PR/SM) scheduler and the z/OS WLM and dispatcher work together. Have them keep as much work as possible within the boundaries of as few processors and L4 virtual cache space (which is best within a CPC drawer boundary) without affecting throughput and response times.

The cache structures of IBM z16 A02 and IBM z16 AGZ systems are compared to the previous generation (IBM z15 T02) in Figure 3-5 on page 77. Logical cache hierarchy is explained in more detail in Figure 9-4 on page 367.

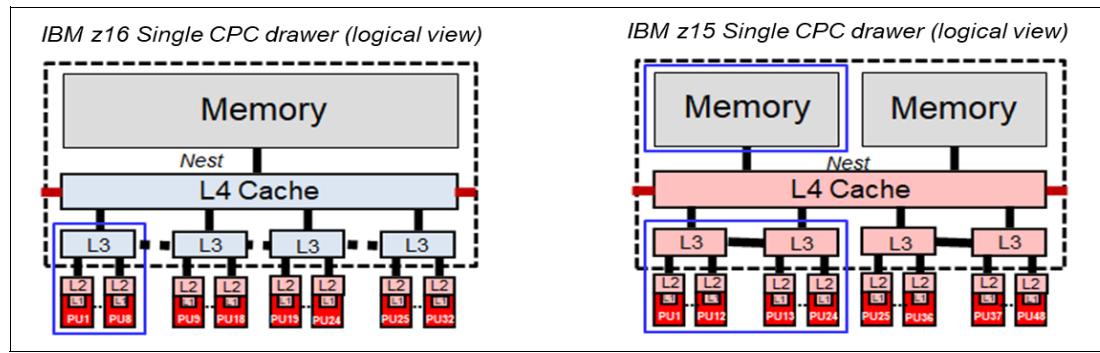


Figure 3-5 IBM z16 A02 and IBM z16 AGZ and IBM z15 cache level comparison

Compared to IBM z15, the IBM z16 A02 and IBM z16 AGZ cache design have larger L2 cache size, while L3 and L4 are now virtual caches. More affinity exists between the memory of a partition, the L4 virtual cache in a drawer, and the cores in the PU chips. As in IBM z15, the IBM z16 A02 and IBM z16 AGZ cache level structure is focused on keeping more data closer to the PU. This design can improve system performance on many production workloads.

HiperDispatch

To help avoid latency in a high-frequency processor design, PR/SM and the dispatcher must be prevented from scheduling and dispatching a workload on any processor available, which keeps the workload in as small a portion of the system as possible. The cooperation between z/OS and PR/SM is bundled in a function that is called HiperDispatch. HiperDispatch uses the IBM z16 A02 and IBM z16 AGZ cache topology, which features reduced cross-cluster “help” and better locality for multi-task address spaces.

PR/SM can use dynamic PU reassignment to move processors (CPs, ZIIPs, IFLs, ICFs, SAPs, and spares) to a different chip and drawer to improve the reuse of shared caches by processors of the same partition. It can use dynamic memory relocation (DMR) to move a running partition’s memory to different physical memory to improve the affinity and reduce the distance between the memory and processors of a partition.

For more information about HiperDispatch, see 3.7, “Logical partitioning” on page 117.

3.3.2 CPC drawer interconnect topology

In a configuration with two CPC drawers (Max68), the drawers are interconnected in a point-to-point topology that allows CPC drawers to communicate with each other to make the system appear as a single large SMP structure.

The IBM z16 A02 and IBM z16 AGZ intra-CPC drawer SMP communication structure for CPC drawers with 4 DCMs is shown in Figure 3-6 on page 78.

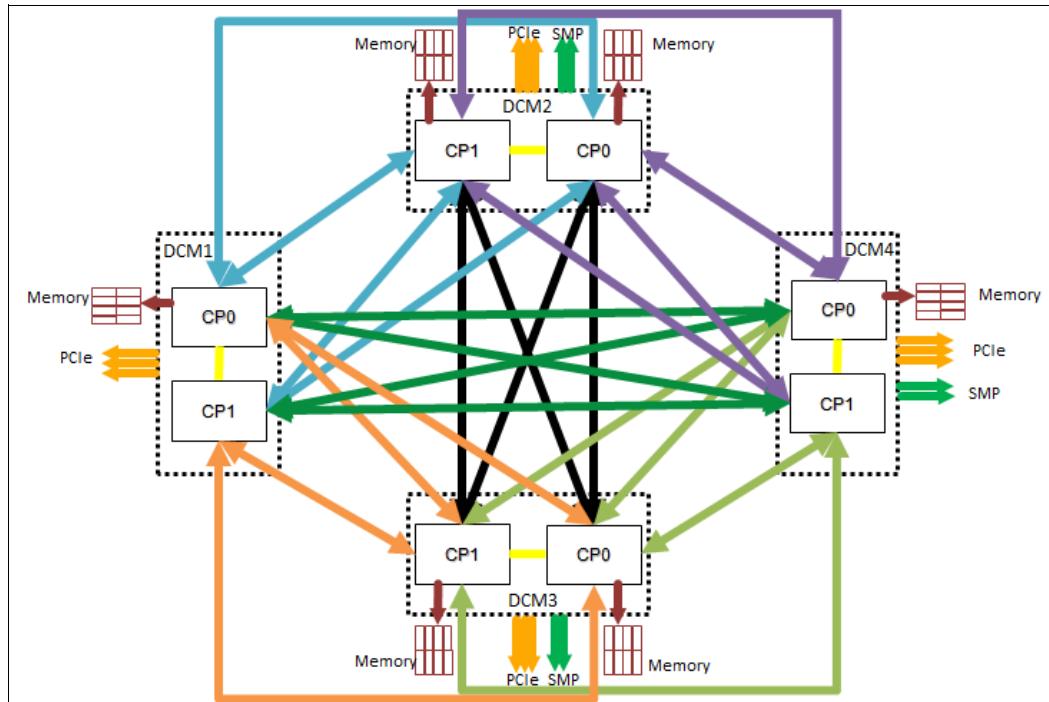


Figure 3-6 IBM z16 A02 and IBM z16 AGZ CPC drawer communication topology for a 4 DCM configuration (Max32 & Max68)

A simplified topology of a two-CPC drawer (Max68) system is shown in Figure 3-7.

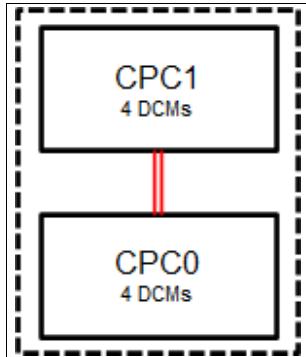


Figure 3-7 Point-to-point topology with a systems with two CPC drawers

Inter-CPC drawer communication occurs at the Level 4 virtual cache level, which is implemented on the semi-private part of one of the Level 2 caches in a chip module. The Level 4 cache function regulates coherent drawer-to-drawer traffic.

3.4 Processor unit design

Processor cycle time is especially important for processor-intensive applications. Current systems design is driven by processor cycle time, although improved cycle time does not automatically mean that the performance characteristics of the system improve.

IBM z16 A02 and IBM z16 AGZ core frequency is 4.6 GHz (compared to 4.5 GHz for the IBM z15 T02), and with increased number of processors that share larger caches to have shorter access times and improved capacity and performance.

Through innovative processor design (significant architecture changes, new cache structure, new Core-Nest interface, new branch prediction design that uses dense SRAM, and new on-chip AI accelerator for inference), the IBM zSystems processor performance continues to evolve.

Enhancements were made on the processor unit design, including the following examples:

- ▶ Cache structure
- ▶ Branch prediction mechanism
- ▶ Floating point unit
- ▶ Divide engine scheduler
- ▶ Load/Store Unit and Operand Store Compare (OSC)
- ▶ Simultaneous multi-threading
- ▶ Relative nest intensity (RNI) redesigns

For more information about RNI, see 12.4, “Relative Nest Intensity” on page 459.

The processing performance was enhanced through the following changes to the IBM z16 A02 and IBM z16 AGZ processor design:

- ▶ Core optimization to enable performance and capacity growth.
- ▶ New cache structure design, including a larger cache Level 2 (SRAM) and virtual Level 3 and Level 4 cache to reduce latency.
- ▶ On-chip IBM Integrated Accelerator for zEnterprise Data Compression (Nest compression accelerator, or NXU. For more information, see Figure 2-14 on page 31.)
- ▶ Enhancement of nest-core staging.
- ▶ On-chip IBM Integrated Accelerator for AI. For more information, see 3.4.6, “IBM Integrated Accelerator for Artificial Intelligence (on-chip)” on page 90, and Appendix B, “IBM Z Integrated Accelerator for AI” on page 471.

Because of these enhancements, the IBM z16 A02 and IBM z16 AGZ processor full speed z/OS single-thread performance is on average 1.14 times faster than the IBM z15 T02 at equal N-way. For more information about performance, see Chapter 12, “Performance” on page 453.

IBM z13 introduced architectural extensions with instructions that reduce processor quiesce effects, cache misses, and pipeline disruption, and increase parallelism with instructions that process several operands in a single instruction (SIMD). The processor architecture was further developed for IBM z14 and IBM z15 generations.

IBM z16 A02 and IBM z16 AGZ include the following enhancements:

- ▶ Optimized third-generation SMT
- ▶ Improved Out-of-Order core execution
- ▶ Improvements in branch prediction and handling
- ▶ Pipeline optimization
- ▶ Secure Execution⁶
- ▶ Co-processor compression enhancements

The IBM z16 A02 and IBM z16 AGZ enhanced Instruction Set Architecture (ISA) includes a set of instructions that are added to improve compiled code efficiency. These instructions

⁶ Secure execution requires operating system support.

optimize PUs to meet the demands of various business and analytics workload types without compromising the performance characteristics of traditional workloads.

3.4.1 Simultaneous multithreading

Aligned with industry directions, the IBM z16 A02 and IBM z16 AGZ can process up to two simultaneous threads in a single core while sharing certain resources of the processor, such as execution units, translation lookaside buffers (TLBs), and caches. When one thread in the core is waiting for other hardware resources, the second thread in the core can use the shared resources rather than remaining idle. This capability is known as *simultaneous multithreading (SMT)*.

An operating system with SMT support can be configured to dispatch work to a thread on a zIIP (for eligible workloads in z/OS) or an IFL (for z/VM and Linux on IBM Z) core in single thread or SMT mode so that HiperDispatch cache optimization can be considered. For more information about operating system support, see Chapter 7, “Operating system support” on page 241.

SMT technology allows instructions from more than one thread to run in any pipeline stage at a time. SMT can handle up to four pending translations.

Each thread has its own unique state information, such as Program Status Word (PSW) and registers. The simultaneous threads cannot necessarily run instructions instantly and must at times compete to use certain core resources that are shared between the threads. In some cases, threads can use shared resources that are not experiencing competition.

Two threads (A and B) that are running on the same processor core on different pipeline stages and sharing the core resources is shown in Figure 3-8.

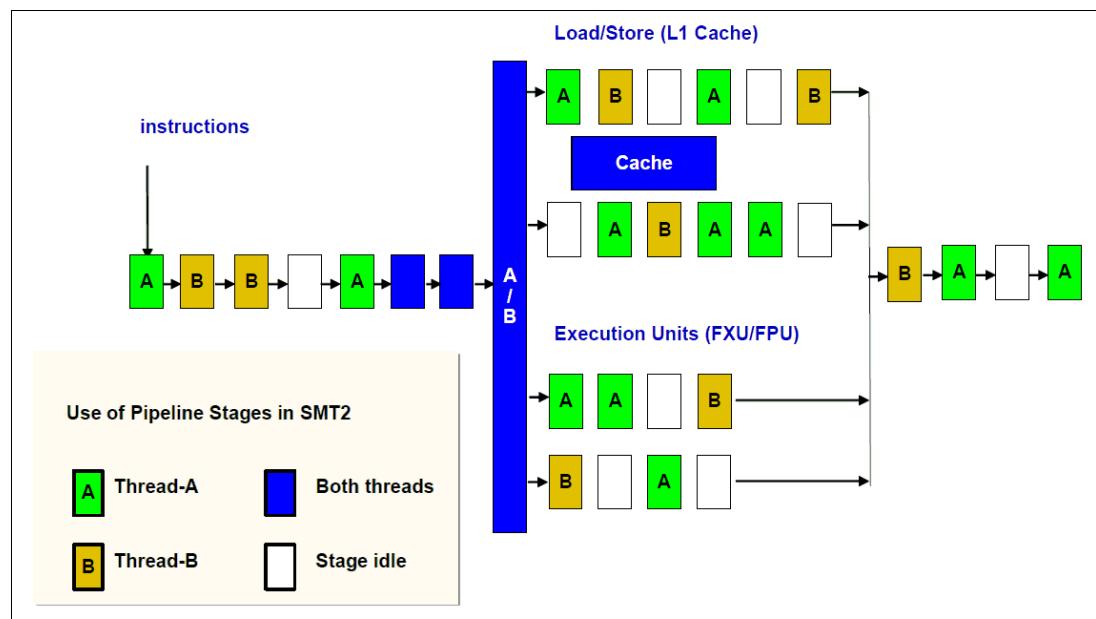


Figure 3-8 Two threads running simultaneously on the same processor core

The use of SMT provides more efficient use of the processors' resources and helps address memory latency, which results in overall throughput gains. The active thread shares core resources in space, such as data and instruction caches, TLBs, branch history tables, and, in time, pipeline slots, execution units, and address translators.

Although SMT increases the processing capacity, the performance in some cases might be superior if a single thread is used. Enhanced hardware monitoring supports measurement through CPUMF for thread usage and capacity.

For workloads that need maximum thread speed, the partition's SMT mode can be turned off. For workloads that need more throughput to decrease the dispatch queue size, the partition's SMT mode can be turned on.

SMT use is functionally transparent to middleware and applications, and no changes are required to run them in an SMT-enabled partition.

3.4.2 Single-instruction multi-data

The IBM z16 A02 and IBM z16 AGZ superscalar processor have 32 vector registers and an instruction set architecture that includes a subset of instructions (known as SIMD) that were added to improve the efficiency of complex mathematical models and vector processing. These new instructions allow a larger number of operands to be processed with a single instruction. The SIMD instructions use the superscalar core to process operands in parallel.

SIMD provides the next phase of enhancements of IBM zSystems analytics capability. The set of SIMD instructions is a type of data parallel computing and vector processing that can decrease the amount of code and accelerate code that handles integer, string, character, and floating point data types. The SIMD instructions improve performance of complex mathematical models and allow integration of business transactions and analytic workloads on IBM zSystems servers.

The 32 vector registers feature 128 bits. The instructions include string operations, vector integer, and vector floating point operations. Each register contains multiple data elements of a fixed size. The following instructions code specifies which data format to use and the size of the elements:

- ▶ Byte (16 8-bit operands)
- ▶ Halfword (eight 16-bit operands)
- ▶ Word (four 32-bit operands)
- ▶ Doubleword (two 64-bit operands)
- ▶ Quadword (one 128-bit operand)

The collection of elements in a register is called a vector. A single instruction operates on all of the elements in the register. Instructions include a nondestructive operand encoding that allows the addition of the register vector A and register vector B and stores the result in the register vector A ($A = A + B$).

A schematic representation of a SIMD instruction with 16-byte size elements in each vector operand is shown in Figure 3-9.

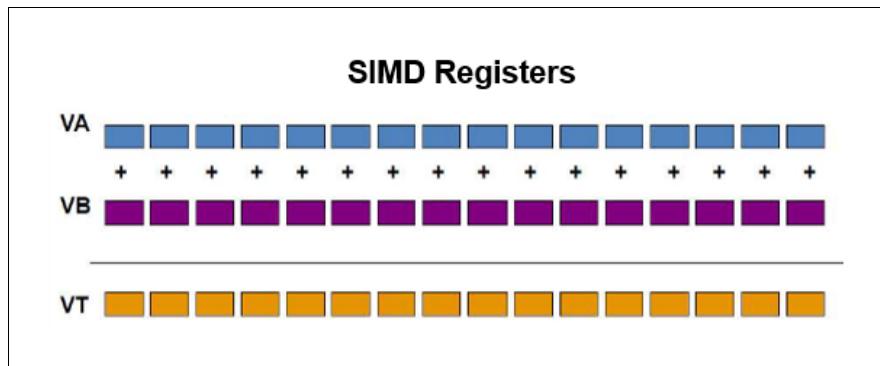


Figure 3-9 SIMD operation logic

The vector register file overlays the floating-point registers (FPRs), as shown in Figure 3-10. The FPRs use the first 64 bits of the first 16 vector registers, which saves hardware area and power, and makes it easier to mix scalar and SIMD codes. Effectively, the core gets 64 FPRs, which can further improve FP code efficiency.

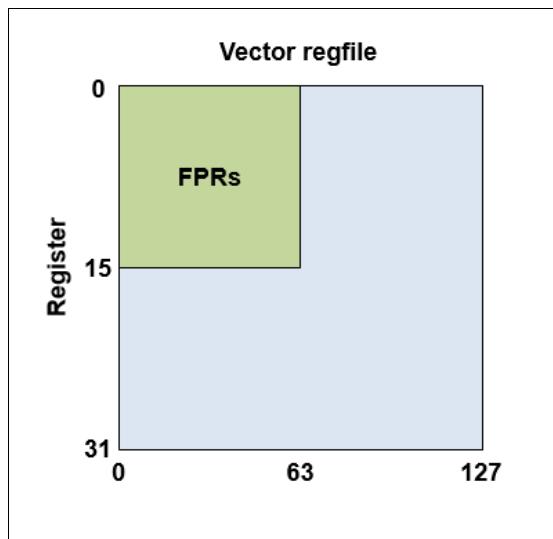


Figure 3-10 Floating point registers overlaid by vector registers

SIMD instructions include the following examples:

- ▶ Integer byte to quadword add, sub, and compare
- ▶ Integer byte to doubleword min, max, and average
- ▶ Integer byte to word multiply
- ▶ String find 8-bit, 16-bit, and 32-bit
- ▶ String range compare
- ▶ String find any equal
- ▶ String load to block boundaries and load/store with length

For most operations, the condition code is not set. A summary condition code is used only for a few instructions.

3.4.3 Out-of-Order execution

IBM z16 A02 and IBM z16 AGZ have an Out-of-Order core, much like the IBM z15 and IBM z14. This optimized Out-of-Order feature yields significant performance benefits for

compute-intensive applications. It does so by reordering instruction execution, which allows later (younger) instructions to be run ahead of a stalled instruction, and reordering storage accesses and parallel storage accesses. Out-of-Order maintains good performance growth for traditional applications.

Out-of-Order execution can improve performance in the following ways:

- ▶ Reordering instruction execution

Instructions stall in a pipeline because they are waiting for results from a previous instruction or the execution resource that they require is busy. In an in-order core, this stalled instruction stalls all later instructions in the code stream. In an out-of-order core, later instructions are allowed to run ahead of the stalled instruction.

- ▶ Reordering storage accesses

Instructions that access storage can stall because they are waiting on results that are needed to compute the storage address. In an in-order core, later instructions are stalled. In an out-of-order core, later storage-accessing instructions that can compute their storage address are allowed to run.

- ▶ Hiding storage access latency

Many instructions access data from storage. Storage accesses can miss the L1 and require 7 - 50 more clock cycles to retrieve the storage data. In an in-order core, later instructions in the code stream are stalled. In an out-of-order core, later instructions that are not dependent on this storage data are allowed to run.

The IBM z16 A02 and IBM z16 AGZ processor includes pipeline enhancements that benefit Out-of-Order execution. The processor design features advanced micro-architectural innovations that provide the following benefits:

- ▶ Maximized instruction-level parallelism (ILP) for a better cycles per instruction (CPI) design.
- ▶ Maximized performance per watt.
- ▶ Enhanced instruction dispatch and grouping efficiency.
- ▶ Increased Out-of-Order resources, such as Global Completion Table entries, physical GPR entries, and physical FPR entries.
- ▶ Improved completion rate.
- ▶ Reduced cache/TLB miss penalty.
- ▶ Improved execution of D-Cache store and reload and new Fixed-point divide.
- ▶ New Operand Store Compare (OSC) (load-hit-store conflict) avoidance scheme.
- ▶ Enhanced branch prediction structure and sequential instruction fetching.

Program results

The Out-of-Order execution does not change any program results. Execution can occur out of (program) order, but all program dependencies are accepted, and the same results are seen as in-order (program) execution. The design was optimized by increasing the Global Completion Table (GCT) from 48x3 to 60x3, which increased the issue queue size from 2x30 to 2x36 and designed a new Mapper.

This implementation requires special circuitry to make execution and memory accesses display in order to the software. The logical diagram of an IBM z16 A02 and IBM z16 AGZ core is shown in Figure 3-11 on page 84.

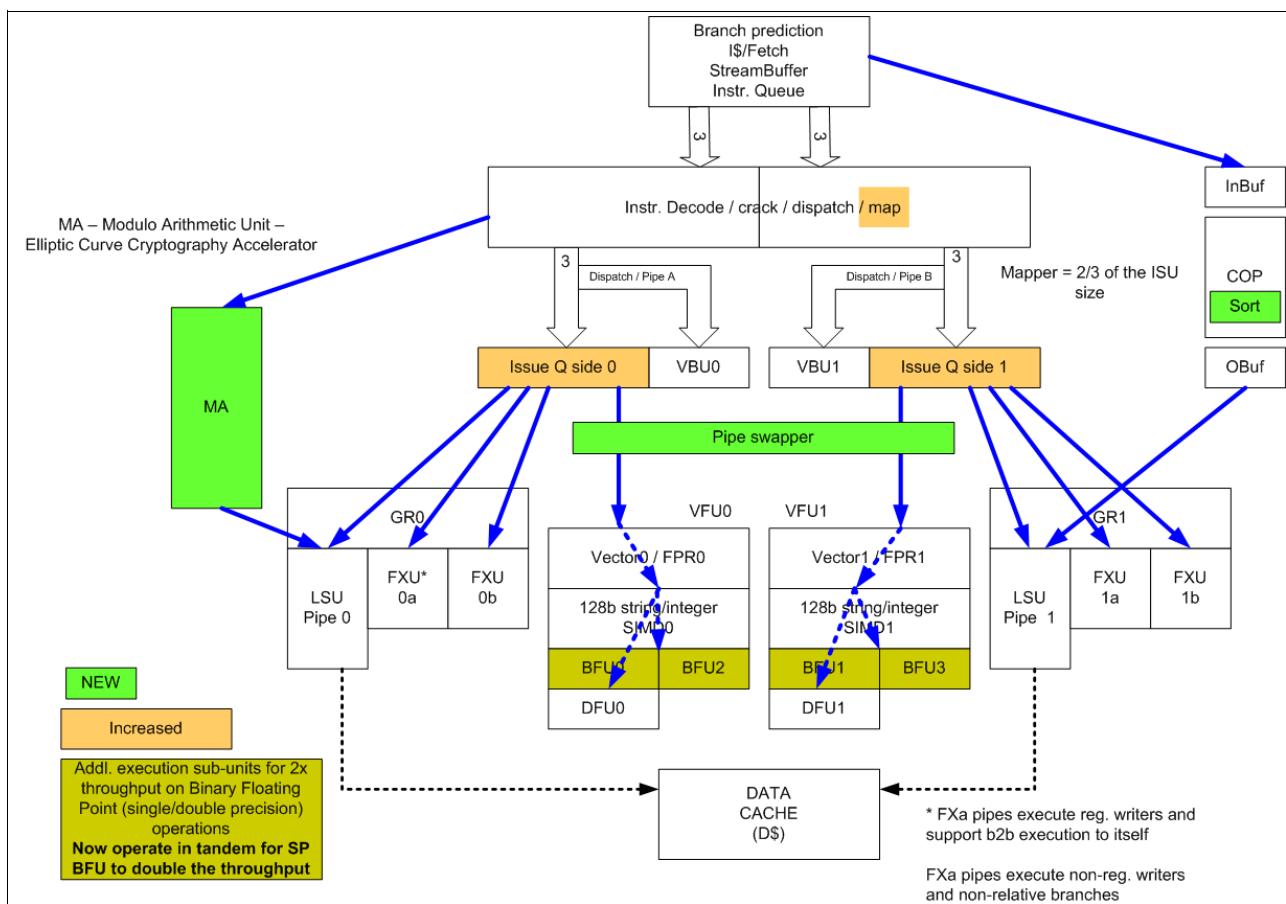


Figure 3-11 IBM z16 A02 and IBM z16 AGZ PU core logical diagram

Memory address generation and memory accesses can occur out of (program) order. This capability can provide a greater use of the IBM z16 A02 and IBM z16 AGZ superscalar core, and improve system performance.

The IBM z16 A02 and IBM z16 AGZ processor unit core is a superscalar, out-of-order, SMT processor with eight execution units. Up to six instructions can be decoded per cycle, and up to 12 instructions or operations can be started to run per clock cycle (0.192 ns). The execution of the instructions can occur out of program order and memory address generation and memory accesses can also occur out of program order. Each core includes special circuitry to display execution and memory accesses in order to the software.

The IBM z16 A02 and IBM z16 AGZ superscalar PU core can have up to 10 instructions or operations that are running per cycle. This technology results in shorter workload runtime.

Enhanced branch prediction

If the branch prediction logic of the microprocessor makes the wrong prediction, all instructions in the parallel pipelines are removed. The wrong branch prediction is expensive in a high-frequency processor design. Therefore, the branch prediction techniques that are used are important to prevent as many wrong branches as possible.

For this reason, various history-based branch prediction mechanisms are used, as shown in the in-order part of the IBM z16 A02 and IBM z16 AGZ PU core logical diagram in Figure 3-11. The branch target buffer (BTB) runs ahead of instruction cache prefetches to prevent branch misses in an early stage. Furthermore, a branch history table (BHT) offers a high

branch prediction success rate in combination with a pattern history table (PHT) and the use of tagged multi-target prediction technology branch prediction.

Branch prediction is now implemented as a two level BTB, BTB1 (“small” and “fast”), and BTB2 (large, dense-SRAM). Now, BTB1 and BTB2 feature dynamic (variable) capacity:

- ▶ BTB1: First Level Branch Target Buffer, smaller than IBM z15, dynamic director, variable capacity:
 - Minimum total branches in all parents (all large branches) = 8 K
 - Maximum total branches in all parents (all medium branches) = 12 K
- ▶ BTB2: Second Level Branch Target Buffer, also variable capacity (variable directory), up to 260 k branches

Branch prediction also implements auxiliary predictors for:

- ▶ Direction:
 - Two table TAGE⁷ Pattern History Table (PHT): A two-level table (with different history lengths). Branch direction is predicted based on history.
 - Perceptron: Called a Perceptron because this is a neural network algorithm that learns to correlate branch history over time and predicts direction of branches that the other mechanisms cannot catch with sufficient accuracy
- ▶ Target:
 - Two table TAGE Changing Target Buffer (CTB): A two-level table (with different history lengths). Branches are remembered that have different targets depending on history.
 - Return Address Table Call/Return Stack (RAT CRS): Multi-level CRS that is implemented as a table lookup

3.4.4 Superscalar processor

A *scalar processor* is a processor that is based on a single-issue architecture, which means that only a single instruction is run at a time. A *superscalar processor* allows concurrent (parallel) execution of instructions by adding resources to the microprocessor in multiple pipelines, each working on its own set of instructions to create parallelism.

A superscalar processor is based on a multi-issue architecture. However, when multiple instructions can be run during each cycle, the level of complexity is increased because an operation in one pipeline stage might depend on data in another pipeline stage. Therefore, a superscalar design demands careful consideration of which instruction sequences can successfully operate in a long pipeline environment.

IBM z16 A02 and IBM z16 AGZ are superscalar processors. Each processor unit, or core, is a superscalar and out-of-order processor that supports 10 concurrent issues to execution units in a single CPU cycle:

- ▶ Fixed-point unit (FXU): The FXU handles fixed-point arithmetic.
- ▶ Load-store unit (LSU): The LSU contains the data cache. It is responsible for handling all types of operand accesses of all lengths, modes, and formats as defined in the z/Architecture.
- ▶ Instruction fetch and branch (IFB) (prediction) and Instruction cache and merge (ICM). These two sub units (IFB and ICM) contain the instruction cache, branch prediction logic,

⁷ TAgged GEometric predictor.

instruction fetching controls, and buffers. Its relative size is the result of the elaborate branch prediction.

- ▶ L1 data and L1 instruction are incorporated into the LSU and ICM, respectively.

COBOL enhancements

IBM z16 A02 and IBM z16 AGZ core implement new instructions for the compiler to accelerate numeric formatting, and hardware support for new numeric conversion instructions (exponents and arithmetic common in financial applications).

3.4.5 On-chip coprocessors and accelerators

This section introduces the CPACF enhancements for IBM z16 A02 and IBM z16 AGZ and the IBM Integrated Accelerator for zEnterprise Data Compression (zEDC).

IBM integrated Accelerator for zEDC (on-chip)

Introduced in IBM z15, the On-Chip data compression accelerator (Nest Accelerator Unit - NXU, see Figure 3-12 on page 88) provides real value for existing and new data compression use cases.

IBM z16 A02 and IBM z16 AGZ Compression/Decompression accelerator is implemented in the Nest Accelerator Unit (NXU) on each processor chip of the IBM Telum microprocessor. IBM z16 A02 and IBM z16 AGZ On-Chip Compression delivers industry-leading throughput and replaces the zEDC Express PCIe adapter available on the IBM z14.

One Nest Accelerator Unit (NXU) is used per processor chip, which is shared by all cores on the chip and features the following benefits:

- ▶ Brand new concept of sharing and operating an accelerator function in the nest
- ▶ Supports DEFLATE compliant compression/decompression and GZIP CRC/ZLIB Adler
- ▶ Low latency
- ▶ High bandwidth
- ▶ Problem state execution
- ▶ Hardware/Firmware interlocks to ensure system responsiveness
- ▶ Designed instruction
- ▶ Run in millicode

The On-Chip Compression Accelerator removes this virtualization constraint because it is shared by all PUs on the processors chip; therefore, it is available to all LPARs and guests.

Moving the compression function from the I/O drawer to the processor chip means that compression can operate directly on L2 cache and data does not need to be passed by using I/O.

Data compression is running in one of the two execution modes available: Synchronous mode or Asynchronous mode:

- ▶ Synchronous execution occurs in problem states where the user application starts the instruction in its virtual address space.
- ▶ Asynchronous execution is optimized for Large Operations under z/OS for authorized applications (for example, BSAM/QSAM) and issues I/O by using EADMF for asynchronous execution.
Asynchronous execution maintains the current user experience and provides a transparent implementation for existing authorized users of zEDC.

The On-Chip data compression implements compression as defined by RFC1951 (DEFLATE).

Figure 3-12 shows the nest compression accelerator (NXU) for On-Chip Compression acceleration.

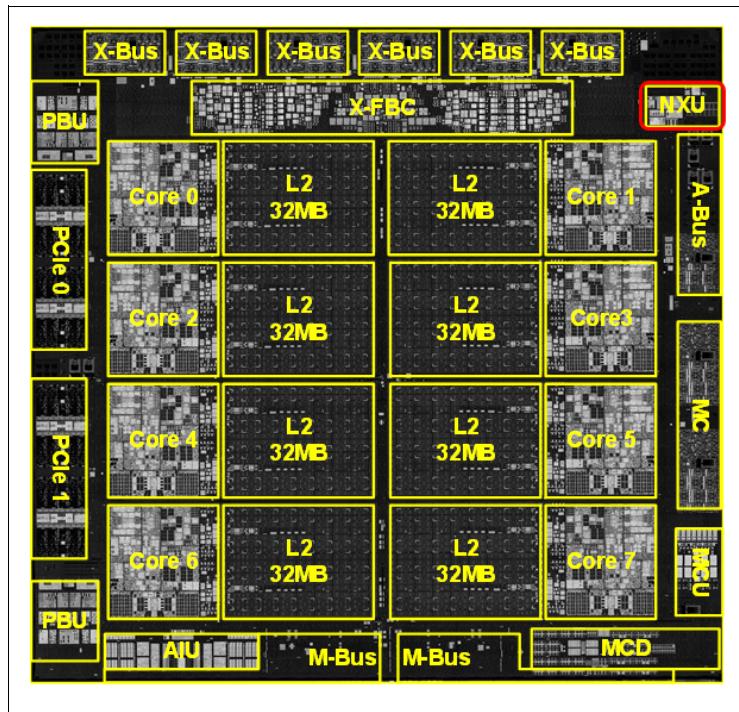


Figure 3-12 Integrated Accelerator for zEDC (NXU) on the IBM z16 A02 and IBM z16 AGZ PU chip

Coprocessor units (on-core)

A data compression coprocessor and a cryptography coprocessor unit is available on each core in the IBM Telum chip.

The compression engine uses static dictionary compression and expansion that is based on CMPSC instruction. The compression dictionary uses the level 1 (L1) cache (instruction cache).

The cryptography coprocessor is used for CPACF, which offers a set of symmetric cryptographic functions for encrypting and decrypting of clear key operations.

The compression and cryptography coprocessors feature the following characteristics:

- ▶ Each core has an independent compression and cryptographic engine.
- ▶ The coprocessor was redesigned to support SMT operation and for throughput increase.
- ▶ It is available to any processor type (regardless of the processor characterization).
- ▶ The owning processor is busy when its coprocessor is busy.

The location of the coprocessor on the IBM z16 A02 and IBM z16 AGZ chip is shown in Figure 3-13 on page 88.

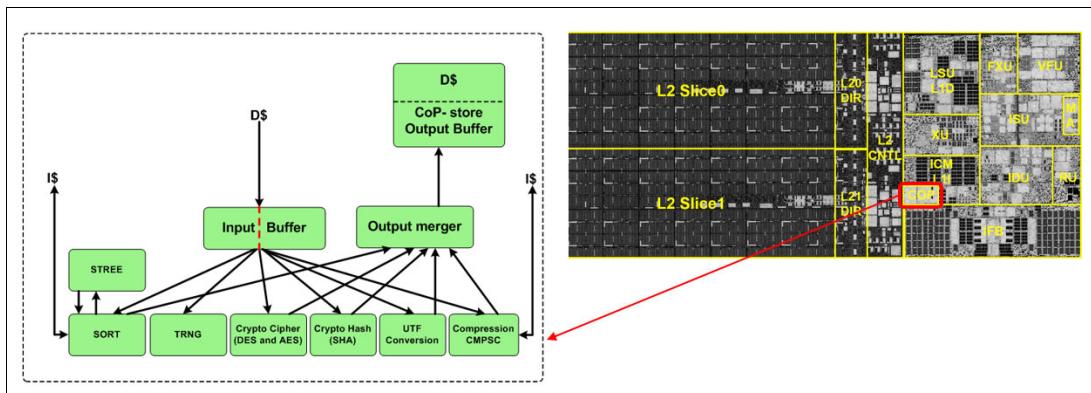


Figure 3-13 IBM z16 A02 and IBM z16 AGZ Core co-processor

On-core compression (CMPSC) on IBM z16 A02 and IBM z16 AGZ

The compression coprocessor on IBM z16 A02 and IBM z16 AGZ provides the same functions that are available on IBM z15.

On-core cryptography coprocessor (CPACF)

CPACF accelerates the encrypting and decrypting of SSL/TLS transactions, virtual private network (VPN)-encrypted data transfers, and data-storing applications that do not require FIPS 140-2 level 4 security. The assist function uses a special instruction set for symmetrical clear key cryptographic encryption and decryption, and for hash operations. This group of instructions is known as the *Message-Security Assist (MSA)*.

For more information about these instructions, see *z/Architecture Principles of Operation*, SA22-7832.

Crypto functions enhancements

The IBM z16 A02 and IBM z16 AGZ microprocessor structure was optimized and aligned to the new cache hierarchy. Co-processor results (data) are stored by way of level 1 (L1) cache.

The crypto/hashing/UTF-conversion/compression engines were redesigned for increased throughput.

CPACF accelerator that is built into every core supports Pervasive Encryption by providing fast synchronous cryptographic services:

- ▶ Encryption (DES, TDES, and AES)
- ▶ Hashing (SHA-1, SHA-2, SHA-3, and SHAKE)
- ▶ Random Number Generation (PRNG, DRNG, and TRNG)
- ▶ Elliptic Curve supported operations:
 - ECDH[E]:
 - P256, P384, and P521
 - X25519 and X448
 - ECDSA:
 - Keygen, sign, and verify
 - P256, P384, and P521
 - EdDSA:
 - Keygen, sign, and verify
 - Ed25519 and Ed448

For more information about cryptographic functions on IBM z16 A02 and IBM z16 AGZ, see Chapter 6, “Cryptographic features” on page 197.

IBM Integrated Accelerator for Z Sort (on-core)

Sorting data is a significant part of IBM zSystems workloads including batch workloads, database query processing, and utility processing. The amount of data that is stored and processed on IBM zSystems continues to grow at a high rate, which drives an ever-increasing sort workload.

Introduced on IBM z15 was the sort accelerator that is known as the IBM Integrated Accelerator for Z Sort (see Figure 3-13 on page 89). The SORTL hardware instruction that is implemented on each core is used by DFSORT and the Db2 utilities for z/OS Suite to allow the use of a hardware-accelerated approach to sorting.

The IBM Integrated Accelerator for Z Sort feature termed as “ZSORT” helps to reduce the CPU costs and improve the elapsed time for eligible workloads. One of the primary requirements for ZSORT is providing enough virtual, real, and auxiliary storage.

Sort jobs that run in memory-constrained environments in which the amount of memory that is available to be used by DFSORT jobs is restricted might not achieve optimal performance results or might not be able to use ZSORT.

The 64-bit memory objects (above-the-bar-storage) can use the ZSORT accelerator for sort workloads for optimal results. Because ZSORT is part of the CPU and memory latency is much less than disk latency, sorting in memory is more efficient than sorting with memory and disk workspace. By allowing ZSORT to process the input completely in memory, it can achieve the best results in elapsed time and CPU time.

Because the goal of ZSORT is to reduce CPU time and elapsed time, it can require more storage than a DFSORT application that does not use ZSORT.

Note: Not all sorts are eligible to use ZSORT. IBM’s zBNA tool provides modeling support for identifying potential ZSORT-eligible candidate jobs and estimates the benefits of ZSORT. The tool uses information in the SMF type 16 records.

The following restrictions disable ZSORT and revert to the use of traditional sorting technique:

- ▶ SORTL facility is not enabled/unavailable on the processor
- ▶ ZSORT is not enabled
- ▶ OPTION COPY or SORT FIELDS=COPY is specified
- ▶ Use of:
 - INREC
 - JOINKEYS
 - MERGE FIELDS
 - MODS(EXITS) statements
 - OUTREC
 - OUTFIL
 - SUM FIELDS
- ▶ Program started sorts
- ▶ Memory objects cannot be created
- ▶ Insufficient memory object storage available (required more than currently available)
- ▶ Unsupported sort fields specified (examples Unicode, Locale, and ALTSEQ)

- ▶ Unknown file size or file size=0.
- ▶ SIZE/FILSZ=Uxxxxxx is specified
- ▶ SORTIN/SORTOUT is a VSAM Cluster
- ▶ Sort control field positions are beyond 4092 and VLSHRT is specified
- ▶ Use of EXCP access method was requested
- ▶ Insufficient storage (for example, above or below the line)
- ▶ Sorting key greater than 4088 bytes or greater than 4080 bytes if EQUALS is specified
- ▶ For variable records, the record length (LRECL) must be greater than 24
- ▶ zHPF is unavailable for a sort that cannot be performed entirely in memory
- ▶ Insufficient amount of sort workspace

3.4.6 IBM Integrated Accelerator for Artificial Intelligence (on-chip)

The IBM zSystems processor chip was enhanced from one generation to another. This enhancement enables various data manipulations (such as compression, sorting, cryptography) directly in hardware, on the processor chip by way of purpose-built accelerators. It also provides eligible workloads with low latency time, high performance, and high throughput.

The new IBM z16 A02 and IBM z16 AGZ microprocessor chip, also called the IBM Telum processor, integrates a new AI accelerator. This innovation brings incredible value to applications and workloads that are running on IBM zSystems platform.

Customers can benefit from the integrated AI accelerator by adding AI operations that are used to perform fraud prevention and fraud detection, customer behavior predictions, and supply chain operations. All of these operations are done in real time and fully integrated in transactional workloads. As a result, valuable insights are gained from their data instantly.

The integrated accelerator for AI delivers AI inference in real time, at large scale, and high throughput rate, with no transaction left behind. The AI capability applies directly to the running transaction. It shifts the traditional paradigm of applying AI to the transactions that were completed. This innovative technology also can be used for intelligent IT workloads placement algorithms, which contributes to the better overall system performance.

The Telum processor also integrates powerful mechanisms of data prefetch, fast and high capacity level 1 (L1) and level 2 (L2) caches, enhanced branch prediction, and other improvements and innovations that streamlines the data processing by the AI accelerator. The hardware, firmware, and software are vertically integrated to deliver the new AI for inference functions seamless to the applications.

The location of the integrated accelerator for AI on the Telum chip is shown in Figure 3-14 on page 91.

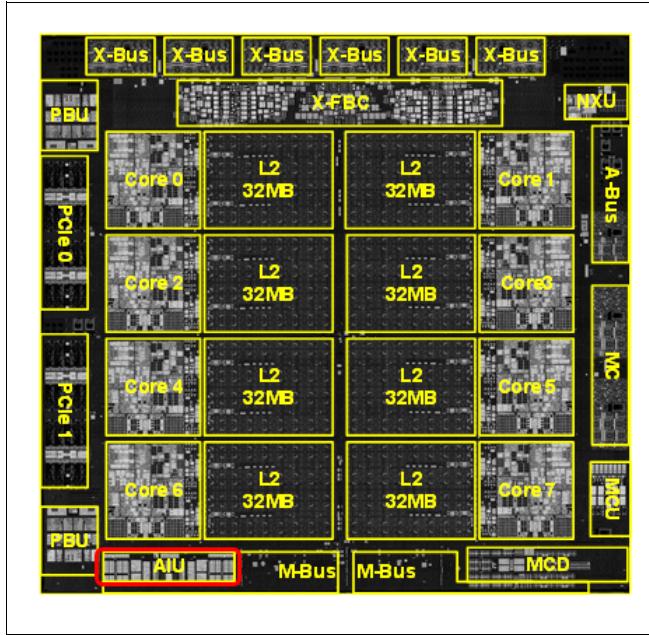


Figure 3-14 Integrated Accelerator for AI on the IBM Telum processor

The AI accelerator is driven by the new Neural Networks Processing Assist (NNPA) instruction.

NNPA is a new non privileged Complex Instruction Set Computer (CISC) memory-to-memory instruction that operates on tensor objects that are in user program's memory. AI functions and macros are abstracted by NNPA.

Figure 3-15 on page 92 shows the AI accelerator and its components:

- ▶ Data movers surround the compute arrays that consist of the Processor Tiles (PT)
- ▶ Processing Elements (PE)
- ▶ Special Function Processors (SFP)

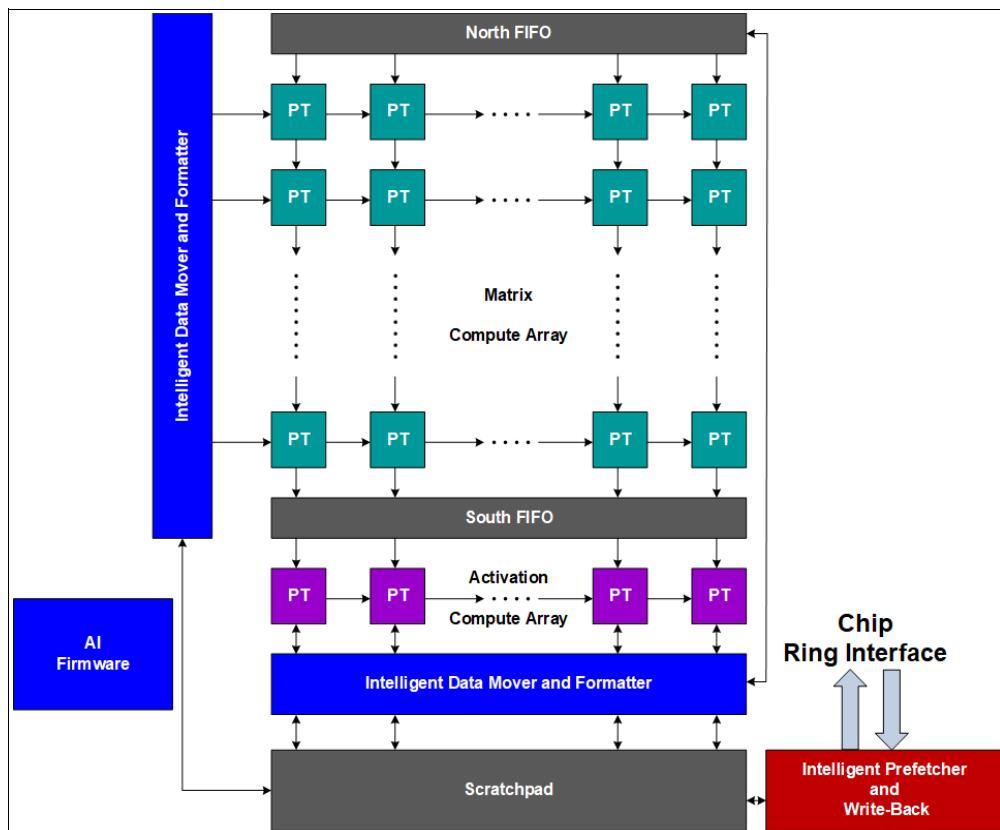


Figure 3-15 IBM z16 A02 and IBM z16 AGZ Integrated accelerator for AI logical diagram

Intelligent data movers and prefetchers are connected to the chip by way of ring interface for high-speed, low-latency, read/write cache operations at 200+ GBps read/store bandwidth, and 600+ GBps bandwidth between engines.

Compute Arrays consist of 128 processor tiles with 8-way FP-16 FMA SIMD, which are optimized for matrix multiplication and convolution, and 32 processor tiles with 8-way FP-16/FP-32 SIMD, which are optimized for activation functions and complex functions.

The AI accelerator is shared by all cores on the chip. The firmware, running on the cores and accelerator, orchestrates and synchronizes the execution on the accelerator.

Using IBM Integrated AI Accelerator in your enterprise

Figure 3-16 on page 93 shows the software ecosystem and high-level integration of the AI accelerator into enterprise AI/Machine Learning solution stack. Great flexibility and interoperability are available for training and building models.

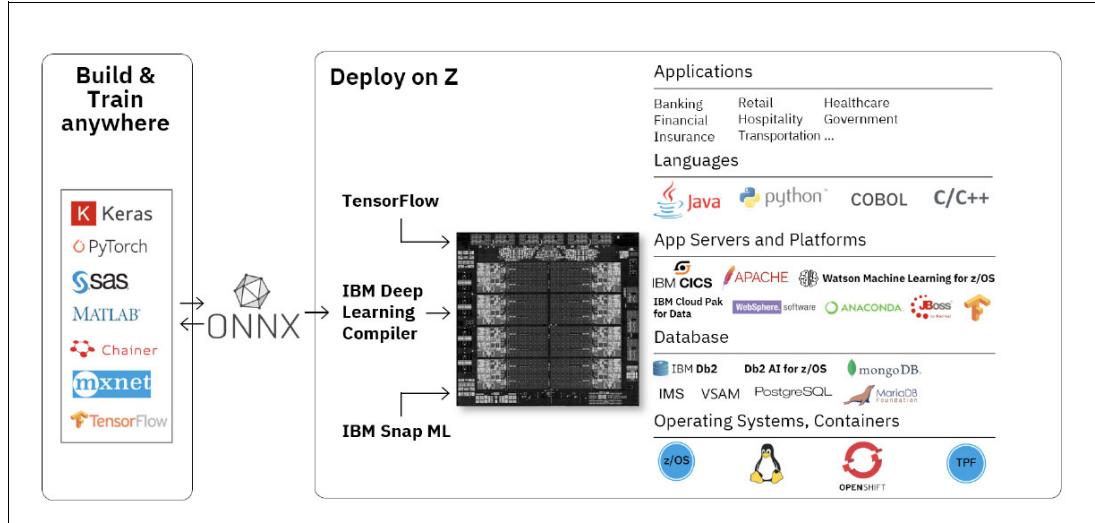


Figure 3-16 Software ecosystem for the AI accelerator

Acknowledging the diverse AI training frameworks, customers can train their models on platforms of their choice, including IBM zSystems (on-premises and in hybrid cloud) and then, deploy it efficiently on IBM zSystems in collocation with the transactional workloads. No other development effort is needed to enable this strategy.

IBM has invested into Open Neural Network Exchange (ONNX), which is a standard format for representing AI models that allows a data scientist to build and train a model in the framework of choice without worrying about the downstream inference implications.

To enable deployment of ONNX models, IBM provides an ONNX model compiler that is optimized for IBM zSystems. IBM also optimized key Open Source frameworks, such as TensorFlow and TensorFlow Serving, for use on IBM zSystems platform.

IBM open-sourced zDNN library provides common APIs for the functions that allow to convert tensor format to the accelerator required one. Customers can run zDNN under z/OS (in zCX) and Linux on IBM zSystems.

A Deep Learning Compiler (DLC) for z/OS and for Linux on IBM zSystems provides the AI functions to the applications.

3.4.7 Decimal floating point accelerator

The decimal floating point (DFP) accelerator function is on each of the microprocessors (cores) on the 8-core chip. Its implementation meets business application requirements for better performance, precision, and function.

Base 10 arithmetic is used for most business and financial computation. Floating point computation that is used for work that is typically done in decimal arithmetic involves frequent data conversions and approximation to represent decimal numbers. This process makes floating point arithmetic complex and error-prone for programmers who use it for applications in which the data is typically decimal.

Hardware DFP computational instructions provide the following features:

- ▶ Data formats of 4, 8, and 16 bytes
- ▶ An encoded decimal (base 10) representation for data

- ▶ Instructions for running decimal floating point computations
- ▶ An instruction that runs data conversions to and from the decimal floating point representation

Benefits of the DFP accelerator

The DFP accelerator offers the following benefits:

- ▶ Avoids rounding issues, such as those issues that occur with binary-to-decimal conversions.
- ▶ Controls binary-coded decimal (BCD) operations better.
- ▶ Follows the standardization of the dominant decimal data and decimal operations in commercial computing, which supports the industry standardization (IEEE 745R) of decimal floating point operations. Instructions are added in support of the Draft Standard for Floating-Point Arithmetic - IEEE 754-2008, which is intended to supersede the ANSI/IEEE Standard 754-1985.
- ▶ Allows COBOL programs that use zoned-decimal operations to take advantage of the z/Architecture DFP instructions.

IBM z16 A02 and IBM z16 AGZ have two DFP accelerator units per core, which improve the decimal floating point execution bandwidth. The floating point instructions operate on newly designed vector registers (32 new 128-bit registers).

IBM z16 A02 and IBM z16 AGZ include decimal floating point packed conversion facility support with the following benefits:

- ▶ Reduces code path length because extra instructions to format conversion are no longer needed.
- ▶ Packed data is operated in memory by all decimal instructions without general-purpose registers, which were required only to prepare for decimal floating point packed conversion instruction.
- ▶ Converting from packed can now force the input packed value to positive instead of requiring a separate OI, OILL, or load positive instruction.
- ▶ Converting to packed can now force a positive zero result instead of requiring ZAP instruction.

Cobol and PL/I compilers were updated to support the new IBM z16 A02 and IBM z16 AGZ enhancements:

- ▶ BCD to HFP conversions
- ▶ Numeric editing operation
- ▶ Zoned decimal operations

Software support

DFP is supported in the following programming languages and products:

- ▶ Release 4 and later of the High Level Assembler
- ▶ C/C++, which requires supported z/OS version
- ▶ Enterprise PL/I Release 3.7 and Debug Tool Release 8.1 or later
- ▶ Java Applications that use the BigDecimal Class Library
- ▶ SQL support as of Db2 Version 9 and later

3.4.8 IEEE floating point

Binary and hexadecimal floating-point instructions are implemented in IBM z16 A02 and IBM z16 AGZ. They incorporate IEEE standards into the system.

The IBM z16 A02 and IBM z16 AGZ core implements two other execution subunits for 2x throughput on BFP (single/double precision) operations (see Figure 3-11 on page 84).

The key point is that Java and C/C++ applications tend to use IEEE BFP operations more frequently than earlier applications. Therefore, the better the hardware implementation of this set of instructions, the better the performance of applications.

3.4.9 Processor error detection and recovery

The PU uses a process called transient recovery as an error recovery mechanism. When an error is detected, the instruction unit tries the instruction again, and attempts to recover the error. If the second attempt is unsuccessful (that is, a permanent fault exists), a relocation process is started that restores the full capacity by moving work to another PU.

Relocation under hardware control is possible because the R-unit has the full designed state in its buffer. PU error detection and recovery are shown in Figure 3-17.

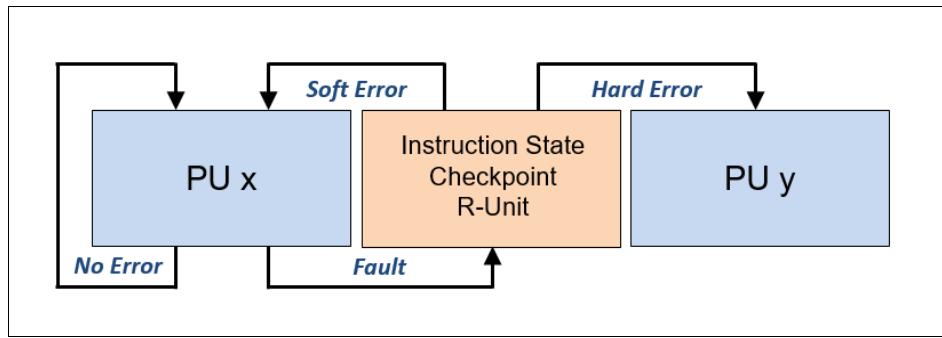


Figure 3-17 PU error detection and recovery

3.4.10 Branch prediction

Because of the ultra-high frequency of the PUs, the penalty for a wrongly predicted branch is high. Therefore, a multi-pronged strategy for branch prediction is implemented on each core based on gathered branch history that is combined with other prediction mechanisms.

The BHT (Branch History Table) implementation on processors provides a large performance improvement. Originally introduced on the IBM ES/9000 9021 in 1990, the BHT is continuously improved.

It offers significant branch performance benefits. The BHT allows each PU to take instruction branches that are based on a stored BHT, which improves processing times for calculation routines.

In addition to the BHT, IBM z16 A02 and IBM z16 AGZ use the following techniques to improve the prediction of the correct branch to be run:

- ▶ BTB
- ▶ PHT
- ▶ CTB

The success rate of branch prediction contributes significantly to the superscalar aspects of IBM z16 A02 and IBM z16 AGZ processor. This success is because the architecture rules prescribe that the correctly predicted result of the branch is essential for successful parallel execution of an instruction stream.

IBM z16 A02 and IBM z16 AGZ integrate a new branch prediction design that uses SRAM and supports the following enhancements over IBM z15:

- ▶ BTB1: 8 K - 12 K
- ▶ BTB2: up to 260 K
- ▶ TAGE PHT: 4 k x 2
- ▶ TAGE CTB: 1 k x 2

3.4.11 Wild branch

When a bad pointer is used or when code overlays a data area that contains a pointer to code, a random branch is the result. This process causes a 0C1 or 0C4 abend. Random branches are difficult to diagnose because clues about how the system got to that point are not evident.

With the wild branch hardware facility, the last address from which a successful branch instruction was run is kept. z/OS uses this information with debugging aids, such as the **SLIP** command, to determine from where a wild branch came.

It also can collect data from that storage location. This approach decreases the number of debugging steps that are necessary when you want to know from where the branch came.

3.4.12 Translation lookaside buffer

The TLB in the instruction and data L1 caches use a secondary TLB to enhance performance.

The size of the TLB is kept as small as possible because of its short access time requirements and hardware space limitations. Because memory sizes recently increased significantly as a result of the introduction of 64-bit addressing, a smaller working set is represented by the TLB.

To increase the working set representation in the TLB without enlarging the TLB, large (1 MB) page and giant page (2 GB) support is available and can be used when suitable. For more information, see “Large page support” on page 116.

With the enhanced DAT-2 (EDAT-2) improvements, the IBM zSystems support 2 GB page frames.

IBM z16 A02 and IBM z16 AGZ TLB

IBM z16 A02 and IBM z16 AGZ switch to a logical-tagged L1 directory and inline TLB2. Each L1 cache directory entry contains the virtual address and Address Space Control Element (ASCE) because it no longer must access TLB for L1 cache hit. TLB2 is accessed in parallel to L2, which saves significant latency compared to TLB1-miss.

The new translation engine allows up to four translations pending concurrently. Each translation step is ~2x faster, which helps second level guests.

3.4.13 Instruction fetching, decoding, and grouping

The superscalar design of the microprocessor allows for the decoding of up to six instructions per cycle and the execution of up to 12 instructions per cycle. Both execution and storage accesses for instruction and operand fetching can occur out of sequence.

Instruction fetching

Instruction fetching normally tries to get as far ahead of instruction decoding and execution as possible because of the relatively large instruction buffers that are available. In the microprocessor, smaller instruction buffers are used. The operation code is fetched from the I-cache and put in instruction buffers that hold prefetched data that is awaiting decoding.

Instruction decoding

The processor can decode up to six instructions per cycle. The result of the decoding process is queued and later used to form a group.

Instruction grouping

From the instruction queue, up to 12 instructions can be completed on every cycle. A complete description of the rules is beyond the scope of this publication.

The compilers and JVMs are responsible for selecting instructions that best fit with the superscalar microprocessor. They abide by the rules to create code that best uses the superscalar implementation. All IBM zSystems compilers and JVMs are constantly updated to benefit from new instructions and advances in microprocessor designs.

3.4.14 Extended Translation Facility

The z/Architecture instruction set includes instructions in support of the Extended Translation Facility. They are used in data conversion operations for Unicode data, which causes applications that are enabled for Unicode or globalization to be more efficient. These data-encoding formats are used in web services, grid, and on-demand environments in which XML and SOAP technologies are used. The High Level Assembler supports the Extended Translation Facility instructions.

3.4.15 Instruction set extensions

Thirty new instructions were added to the IBM z16 A02 and IBM z16 AGZ processor. The following new mnemonics were added to the IBM z/Architecture:

- ▶ LBEAR, LFI, LLGFI, and LPSWEY
- ▶ NNPA
- ▶ QPACI
- ▶ RDP
- ▶ SLLHH, SLLHL, SLLLH, SRLHH, SRLHL, SRLLH, and STBEAR
- ▶ VCFN, VCLFNH, VCLFNL, VCLZDP, VCNF, VCRNF, VCSPH, VPKZR, VSCHDP, VSCHP, VSCHSP, VSCHXP, VSCSHP, VSRPR, VUPKZH, and VUPKZL

A new procedure is available to run against customer macros in assembler libraries to verify that no conflicts exist between some older IBM Macro names and new IBM z16 A02 and IBM z16 AGZ (M/T 3932) hardware instruction mnemonics.

3.4.16 Transactional Execution

The Transactional Execution (TX) capability, which is known in the industry as *hardware transactional memory*, runs a group of instructions atomically; that is, all of their results are committed or no result is committed. The execution is optimistic. The instructions are run, 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.

The Transaction Execution Facility provides instructions, including declaring the beginning and end of a transaction, and canceling the transaction. TX is expected to provide significant performance benefits and scalability by avoiding most locks. This benefit is especially important for heavily threaded applications, such as Java.

Removal of support of the transactional execution and constrained transactional execution facility ^a: In a future IBM zSystems hardware system family, the transactional execution and constrained transactional execution facility will no longer be supported. Users of the facility on current zSystems servers should always check the facility indications before using it.

- a. Statements by IBM regarding its plans, directions, and intent are subject to change or withdrawal without notice at the sole discretion of IBM. Information regarding potential future products is intended to outline general product direction and should not be relied on in making a purchasing decision.

3.4.17 Runtime Instrumentation

Runtime Instrumentation (RI) is a hardware facility for managed run times, such as the Java Runtime Environment (JRE). RI allows dynamic optimization of code generation as it is being run. It requires fewer system resources than the current software-only profiling, and provides information about hardware and program characteristics. RI also enhances JRE in making the correct decision by providing real-time feedback.

3.5 Processor unit functions

The PU functions are described in this section.

3.5.1 Overview

All PUs on an IBM z16 A02 and IBM z16 AGZ are physically identical. When the system is initialized, two integrated firmware processors (IFP) are allocated from the pool of PUs that is available for the entire system. The other PUs can be characterized to specific functions (CP, IFL, ICF, zIIP, or SAP).

The function that is assigned to a PU is set by the Licensed Internal Code (LIC). The LIC is loaded when the system is initialized at power-on reset (POR) and the PUs are characterized.

Only characterized PUs include a designated function. Non-characterized PUs are considered spares. You must order at least one CP, IFL, or ICF on IBM z16 A02 and IBM z16 AGZ.

This design brings outstanding flexibility to IBM z16 A02 and IBM z16 AGZ because any PU can assume any available characterization. The design also plays an essential role in system availability because PU characterization can be done dynamically, with no system outage.

For more information about software level support of functions and features, see Chapter 7, “Operating system support” on page 241.

Concurrent upgrades

For all IBM z16 A02 and IBM z16 AGZ features that have more processor units (PUs) installed (non-characterized) than activated, concurrent upgrades can be done by using LIC activation.

This activation assigns a PU function to a previously non-characterized PU. No hardware changes are required.

The upgrade can be done concurrently through the following facilities:

- ▶ Customer Initiated Upgrade (CIU) for permanent upgrades
- ▶ On/Off Capacity on Demand (On/Off CoD) for temporary upgrades
- ▶ Capacity BackUp (CBU) for temporary upgrades
- ▶ Capacity for Planned Event (CPE) for temporary upgrades (available only if the CPE feature code was carried forward from a previous IBM z14 or IBM z15 machine)
- ▶ Flexible Capacity for Cyber Resilience upgrades

If the PU chips in the installed in the first CPC drawer have no available remaining PUs, an upgrade results in a feature upgrade and potential installation of extra PU chips (e.g. from Max5 or Max16 to a Max32) or installation of the second CPC drawer (upgrades to a Max68).

The mentioned addition of PU chips in the first CPC drawer is a disruptive upgrade, the addition of the second CPC drawer is always non-disruptive for the IBM z16 A02, but for the IBM z16 AGZ only if planned ahead when the original machine was ordered.

For more information about Capacity on Demand, see Chapter 8, “System upgrades” on page 317.

PU sparing

If a PU failure occurs, the failed PU’s characterization is dynamically and transparently reassigned to a spare PU. IBM z16 A02 and IBM z16 AGZ have two spare PUs. PUs that are not characterized on a CPC configuration can also be used as extra spare PUs. For more information about PU sparing, see 3.5.10, “Sparing rules” on page 112.

PU pools

PUs that are defined as CPs, IFLs, ICFs, and zIIPs are grouped in their own pools from where they can be managed separately. This configuration significantly simplifies capacity planning and management for LPARs. The separation also affects weight management because CP and zIIP weights can be managed separately.

For more information, see “PU weighting” on page 100.

All assigned PUs are grouped in the PU pool. These PUs are dispatched to online logical PUs. As an example, consider a z16 A02 Max32 with 6 CPs, 6 IFLs, 5 zIIPs, and 1 ICF. This system has a PU pool of 18 PUs, called the *pool width*. Subdivision defines the following pools:

- ▶ A CP pool of six CPs
- ▶ An ICF pool of one ICF
- ▶ An IFL pool of six IFLs
- ▶ A zIIP pool of five zIIPs

PUs are placed in the pools in the following circumstances:

- ▶ When the system is PORed (Power on Reset / IML)
- ▶ At the time of a concurrent upgrade
- ▶ As a result of adding PUs during a CBU
- ▶ Following a capacity on-demand upgrade through On/Off CoD or CIU

PUs are removed from their pools when a concurrent downgrade occurs as the result of the removal of a CBU. They are also removed through the On/Off CoD process and the conversion of a PU. When a dedicated LPAR is activated, its PUs are taken from the correct pools. This process is also the case when an LPAR logically configures a PU as on, if the width of the pool allows for it.

For an LPAR, logical PUs are dispatched from the supporting pool only. The logical CPs are dispatched from the CP pool, logical zIIPs from the zIIP pool, logical IFLs from the IFL pool, and the logical ICFs from the ICF pool.

PU weighting

Because CPs, zIIPs, IFLs, and ICFs have their own pools from where they are dispatched, they can be given their own weights. For more information about PU pools and processing weights, see the *Processor Resource/Systems Manager Planning Guide*, SB10-7178.

3.5.2 Central processors

A central processor (CP) is a PU that uses the full z/Architecture instruction set. It can run z/Architecture-based operating systems (z/OS, z/VM, TPF, z/TPF, z/VSE, and Linux on IBM Z) and the Coupling Facility Control Code (CFCC). Up to six PUs can be characterized as CPs, depending on the IBM z16 A02 and IBM z16 AGZ configuration.

The IBM z16 A02 and IBM z16 AGZ can be initialized in LPAR (PR/SM) mode or in Dynamic Partition Manger (DPM) mode.

CPs are defined as dedicated or shared. Reserved CPs can be defined to an LPAR to allow for nondisruptive image upgrades. If the operating system in the LPAR supports the logical processor add function, reserved processors are no longer needed.

All PUs that are characterized as CPs within a configuration are grouped into the CP pool. The CP pool can be seen on the Hardware Management Console (HMC) workplace. Any z/Architecture operating systems and CFCCs can run on CPs that are assigned from the CP pool.

The IBM z16 A02 and IBM z16 AGZ can be ordered with 26 distinct capacity settings for CPs. Full-capacity CPs are identified as "Z". In addition to full-capacity CPs, three subcapacity settings (A to Y, each for up to 6 PUs⁸, are offered. Table 3-1 lists the capacity settings that appear in hardware descriptions.

Table 3-1 CP capacity settings for one CP

CP Capacity	Feature Code
CP-A	6156
CP-B	6157
CP-C	6158

⁸ Limited to five PUs on an IBM z16 A02 and IBM z16 AGZ Max5

CP Capacity	Feature Code
CP-D	6159
CP-E	6160
CP-F	6161
CP-G	6162
CP-H	6163
CP-I	6164
CP-J	6165
CP-K	6166
CP-L	6167
CP-M	6168
CP-N	6169
CP-O	6170
CP-P	6171
CP-Q	6172
CP-R	6173
CP-S	6174
CP-T	6175
CP-U	6176
CP-V	6177
CP-W	6178
CP-X	6179
CP-Y	6180
CP-Z	6181

Granular capacity provides 156 subcapacity settings for 6 CPs capacity. Information about CPs in the remainder of this chapter applies to all CP capacity settings, unless indicated otherwise.

Note: Information about CPs in the remainder of this chapter applies to all CP capacity settings, unless indicated otherwise. For more information about granular capacity, see 2.3.2, “PU characterization” on page 40.

3.5.3 Integrated Facility for Linux (FC 1959)

An IFL is a PU that can be used to run Linux, Linux guests on z/VM operating systems, and Secure Service Container (SSC). Up to 68 PUs can be characterized as IFLs, depending on the configuration.

Note: IFLs can be dedicated to a Linux, a z/VM, or LPAR, or can be shared by multiple Linux guests, z/VM LPARs, or SSC that are running on the same IBM z16 A02 or IBM z16 AGZ. Only z/VM, Linux on Z operating systems, SSC, and designated software products can run on IFLs. IFLs are orderable by using FC 1959.

IFL pool

All PUs that are characterized as IFLs within a configuration are grouped into the IFL pool. The IFL pool can be seen on the HMC workplace.

IFLs do not change the model capacity identifier of the IBM z16 A02 or IBM z16 AGZ. Software product license charges that are based on the model capacity identifier are not affected by the addition of IFLs.

Unassigned IFLs

An IFL that is purchased but not activated is registered as an Unassigned IFL (FC 1962). When the system is later upgraded with another IFL, the system recognizes that an IFL was purchased and is present.

The allowable number of IFLs and Unassigned IFLs numbers per feature is listed in Table 3-2.

Table 3-2 IFLs and Unassigned IFLs per feature

Features	Max5	Max16	Max32	Max68
Maximum of IFLs FC 1959	5	16	32	68
Maximum of Unassigned IFLs FC 1962	4	15	31	67

Unassigned zIIPs

A zIIP that is purchased but not activated is registered as an Unassigned zIIP (FC 1975). When the system is later upgraded with another zIIP, the system recognizes that a zIIP was purchased and is present.

The allowable number of zIIPs and Unassigned zIIPs numbers per feature is listed in Table 3-3

Table 3-3 zIIPs and unassigned zIIPs per feature

Features	Max5	Max16	Max32	Max68
Maximum of zIIPs FC 1961	5	16	32	68
Maximum of Unassigned zIIPs FC 1975	4	15	31	67

3.5.4 Internal Coupling Facility (FC 1960)

An Internal Coupling Facility (ICF) is a PU that is used to run the CFCC for Parallel Sysplex environments. Within the sum of all unassigned PUs in up to five CPC drawers, up to 68 ICFs can be characterized, depending on the feature. However, the maximum number of ICFs that can be defined on a coupling facility LPAR is limited to 16. ICFs are orderable by using FC 1960.

Unassigned ICFs

New on IBM z16 A02 and IBM z16 AGZ, an ICF that is purchased but not activated is registered as an unassigned ICF (FC 1974). When the system is later upgraded with another ICF, the system recognizes that an ICF was purchased and is present.

The allowable number of ICFs and Unassigned ICFs for each feature is listed in Table 3-4.

Table 3-4 ICFs per feature

Features	Max5	Max16	Max32	Max68
Maximum of ICFs FC 1960	5	16	32	68
Maximum of Unassigned ICFs FC 1974	4	15	31	67

ICFs exclusively run CFCC. ICFs do not change the model capacity identifier of the z16 A02 and IBM z16 AGZ. Software product license charges that are based on the model capacity identifier are not affected by the addition of ICFs.

All ICFs within a configuration are grouped into the ICF pool. The ICF pool can be seen on the HMC workplace.

The ICFs can be used by coupling facility LPARs only. ICFs are dedicated or shared. ICFs can be dedicated to a CF LPAR, or shared by multiple CF LPARs that run on the same system. However, having an LPAR with dedicated and shared ICFs at the same time is not possible.

Coupling Thin Interrupts

With the introduction of Driver 15F (zEC12 and zBC12), the IBM z/Architecture provides a thin interrupt class called *Coupling Thin Interrupts*⁹. The capabilities that are provided by hardware, firmware, and software support the generation of coupling-related “Thin Interrupts” when the following situations occur:

- ▶ On the coupling facility (CF) side:
 - A CF command or a CF signal (arrival of a CF-to-CF duplexing signal) is received by a shared-engine CF image.
 - The completion of a CF signal that was previously sent by the CF occurs (completion of a CF-to-CF duplexing signal).
- ▶ On the z/OS side:
 - CF signal is received by a shared-engine z/OS image (arrival of a List Notification signal).
 - An asynchronous CF operation completes.

The interrupt causes the receiving partition to be dispatched by an LPAR if it is not dispatched. This process allows the request, signal, or request completion to be recognized and processed in a more timely manner.

After the image is dispatched, “poll for work” logic in CFCC and z/OS can be used largely as-is to locate and process the work. The new interrupt expedites the redispatching of the partition.

⁹ It is the only option for shared processors in a CF image (whether they be ICFs or CPs) on IBM z16 A02 and IBM z16 AGZ.

LPAR presents these Coupling Thin Interrupts to the guest partition, so CFCC and z/OS both require interrupt handler support that can deal with them. CFCC also changes to relinquish control of the processor when all available pending work is exhausted, or when the LPAR undispatches it off the shared processor, whichever comes first.

CF processor combinations

A CF image can have one of the following combinations that are defined in the image profile:

- ▶ Dedicated ICFs
- ▶ Shared ICFs
- ▶ Dedicated CPs
- ▶ Shared CPs

Shared ICFs add flexibility. However, running only with shared coupling facility PUs (ICFs or CPs) is not a preferable production configuration. It is preferable for a production CF to operate by using dedicated ICFs.

In Figure 3-18, the CPC on the left has two environments that are defined (production and test), and each has one z/OS and one coupling facility image. The coupling facility images share an ICF.

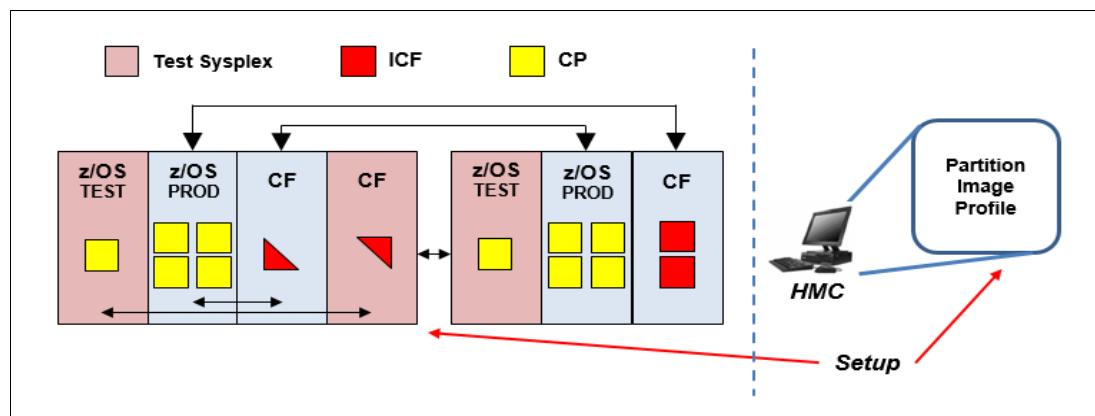


Figure 3-18 ICF options - shared ICFs

The LPAR processing weights are used to define how much processor capacity each CF image can include. The capped option also can be set for a test CF image to protect the production environment.

Connections between these z/OS and CF images can use internal coupling links to avoid the use of real (external) coupling links, and get the best link bandwidth available.

Dynamic CF dispatching

The *dynamic coupling facility dispatching* (DYNDISP) function features a dispatching algorithm that you can use to define a backup CF in an LPAR on the system. When this LPAR is in backup mode, it uses few processor resources.

DYNDISP allows more environments with multiple CF images to coexist in a server, and to share CF engines with reasonable performance. DYNDISP THIN is the only option for CF images that use shared processors on IBM z16 A02 and IBM z16 AGZ. For more information, see 3.9.3, “Dynamic CF dispatching” on page 131.

Coupling Facility Processor scalability

CF work management and dispatcher changed to improve efficiency as processors are added to scale up the capacity of a CF image.

CF images support up to 16 processors. To obtain sufficient CF capacity, customers might be forced to split the CF workload across more CF images. However, this change brings more configuration complexity and granularity (more, smaller CF images, more coupling links, and logical CHPIs to define and manage for connectivity, and so on).

To improve CF processor scaling for the customer's CF images and to make effective use of more processors as the sysplex workload increases, CF work management and dispatcher provide the following improvements IBM z16 A02 and IBM z16 AGZ:

- ▶ IBM z16 A02 and IBM z16 AGZ provide improved CF processor scalability for CF images.
- ▶ Increased number of CF tasks
- ▶ ICA SR latency improvements that improve coupling efficiency in a parallel sysplex.

Coupling Facility Enhancements with CFCC level 25

CFCC level 25 is available on IBM z16 A02 and IBM z16 AGZ with driver 51. For more information about CFCC Level 25 enhancements, see 7.4.3, "Coupling and clustering features and functions" on page 274.

3.5.5 IBM Z Integrated Information Processor (FC 1961)

A zIIP¹⁰ reduces the standard processor (CP) capacity requirements for z/OS Java, XML system services applications, and a portion of work of z/OS Communications Server and Db2 UDB for z/OS Version 8 or later, which frees up capacity for other workload requirements.

Tip: Starting with IBM z16 A02 and IBM z16 AGZ announcement, the 2:1 zIIP:CP ratio restriction has been removed. With one CP ordered, the number of zIIPs orderable is now (MaxYY-1). The restriction has been lifted for IBM z16 A02, IBM16 AGZ and IBM z16 A01 as well.

A zIIP enables eligible z/OS workloads to have a portion of them directed for execution to a processor that is characterized as a zIIP. Because the zIIPs do not increase the MSU value of the processor, they do not affect the IBM software license changes.

IBM z16 A02 and IBM z16 AGZ are the fourth generation of IBM zSystems processors to support SMT. IBM z16 A02 and IBM z16 AGZ implement two threads per core on IFLs and zIIPs. SMT must be enabled at the LPAR level and supported by the z/OS operating system. SMT was enhanced for IBM z16 A02 and IBM z16 AGZ and it is enabled for SAPs by default (no customer intervention required).

Introduced in z/OS V2R4, the z/OS Container Extensions¹¹ allows deployment of Linux on IBM zSystems software components, such as Docker Containers in a z/OS system, in direct support of z/OS workloads without requiring a separately provisioned Linux server. It also maintains overall solution operational control within z/OS and with z/OS qualities of service. Workload deployed in z/OS Container Extensions is zIIP eligible.

¹⁰ IBM zSystems Application Assist Processors (zAAPs) are not available since IBM z14. A zAAP workload is dispatched to available zIIPs (zAAP on zIIP capability).

¹¹ z/OS Container Extensions that are running on IBM z16 A02 and IBM z16 AGZ require "IBM Container Hosting Foundation for z/OS" software product (5655-HZ1) and/or the "IBM zCX Foundation for Red Hat OpenShift" software product (5655-ZCX) when running OCP in zCX.

How zIIPs work

zIIPs are designed for supporting designated z/OS workloads. One of the workloads is Java code execution. When Java code must be run (for example, under control of IBM WebSphere), the z/OS JVM calls the function of the zIIP. The z/OS dispatcher then suspends the JVM task on the CP that it is running on and dispatches it on an available zIIP. After the Java application code execution is finished, z/OS redispaches the JVM task on an available CP. After this process occurs, normal processing is resumed.

This process reduces the CP time that is needed to run Java WebSphere applications, which frees that capacity for other workloads.

The logical flow of Java code running on an IBM z16 A02 or IBM z16 AGZ with a zIIP available is shown in Figure 3-19. When JVM starts the execution of a Java program, it passes control to the z/OS dispatcher that verifies the availability of a zIIP.

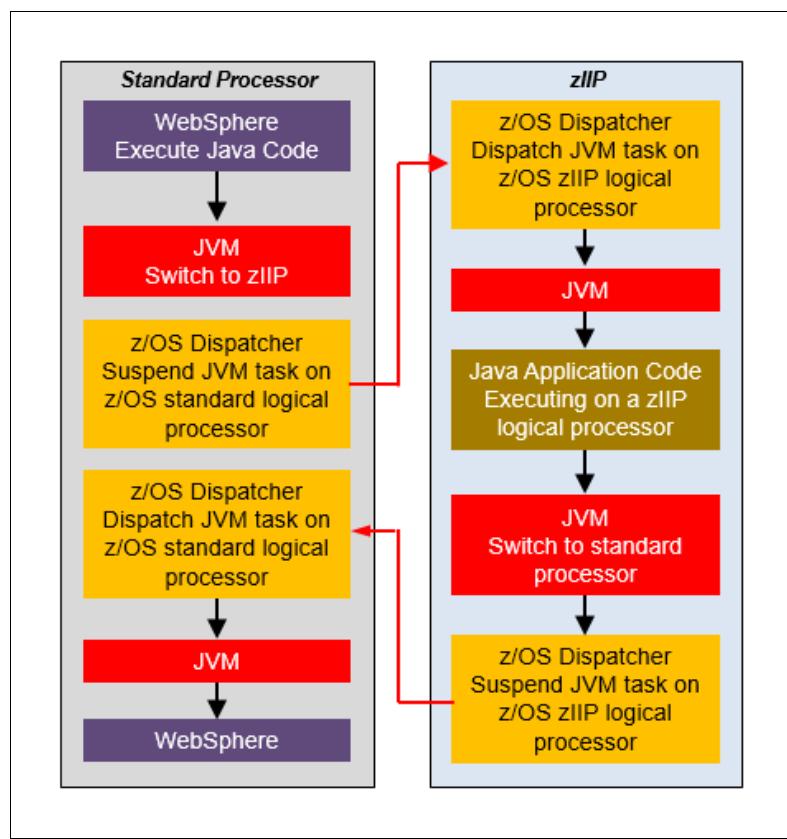


Figure 3-19 Logical flow for Java code execution on a zIIP

The availability is treated in the following manner:

- ▶ If a zIIP is available (not busy), the dispatcher suspends the JVM task on the CP and assigns the Java task to the zIIP. When the task returns control to the JVM, it passes control back to the dispatcher. The dispatcher then reassigns the JVM code execution to a CP.
- ▶ If no zIIP is available (all busy), the z/OS dispatcher allows the Java task to run on a standard CP. This process depends on the option that is used in the OPT statement in the IEAOPTxx member of SYS1.PARMLIB.

A zIIP runs IBM authorized code only. This IBM authorized code includes the z/OS JVM in association with parts of system code, such as the z/OS dispatcher and supervisor services.

A zIIP cannot process I/O or clock comparator interruptions. It also does not support operator controls, such as IPL.

Java application code can run on a CP or a zIIP. The installation can manage the use of CPs so that Java application code runs only on CPs or zIIPs, or on both.

The following execution options for zIIP-eligible code execution are available and supported for z/OS¹². These options are user-specified in IEAOPTxx and can be dynamically altered by using the **SET OPT** command:

- ▶ Option 1: Java dispatching by priority (IIPHONORPRIORITY=YES)

This option is the default option and specifies that CPs must not automatically consider zIIP-eligible work for dispatching on them. The zIIP-eligible work is dispatched on the zIIP engines until Workload Manager (WLM) determines that the zIIPs are overcommitted.

WLM then requests help from the CPs. When help is requested, the CPs consider dispatching zIIP-eligible work on the CPs based on the dispatching priority relative to other workloads. When the zIIP engines are no longer overcommitted, the CPs stop considering zIIP-eligible work for dispatch.

This option runs as much zIIP-eligible work on zIIPs as possible. It also allows it to spill over onto the CPs only when the zIIPs are overcommitted.

- ▶ Option 2: Java dispatching by priority (IIPHONORPRIORITY=NO)

zIIP-eligible work runs on zIIPs only while at least one zIIP engine is online. zIIP-eligible work is not normally dispatched on a CP, even if the zIIPs are overcommitted and CPs are unused. The exception is that zIIP-eligible work can sometimes run on a CP to resolve resource conflicts.

Therefore, zIIP-eligible work does not affect the CP utilization that is used for reporting through the subcapacity reporting tool (SCRT), no matter how busy the zIIPs are.

If zIIPs are defined to the LPAR but are not online, the zIIP-eligible work units are processed by CPs in order of priority. The system ignores the IIPHONORPRIORITY parameter in this case and handles the work as though it had no eligibility to zIIPs.

zIIPs provide the following benefits:

- ▶ Potential software cost savings.
- ▶ Simplification of infrastructure as a result of the collocation and integration of new applications with their associated database systems and transaction middleware, such as Db2, IMS, or CICS. Simplification can happen, for example, by introducing a uniform security environment, and by reducing the number of TCP/IP programming stacks and system interconnect links.
- ▶ Prevention of processing latencies that occur if Java application servers and their database servers are deployed on separate server platforms.

The following Db2 UDB for z/OS V8 or later workloads can run in Service Request Block (SRB) mode:

- ▶ Query processing of network-connected applications that access the Db2 database over a TCP/IP connection by using IBM Distributed Relational Database Architecture (DRDA).

DRDA enables relational data to be distributed among multiple systems. It is native to Db2 for z/OS, which reduces the need for more gateway products that can affect performance and availability. The application uses the DRDA requester or server to access a remote database. IBM Db2 Connect is an example of a DRDA application requester.

¹² z/OS V2R2 and later (older z/OS versions are out of support)

- ▶ Star schema query processing, which is mostly used in business intelligence work.
A *star schema* is a relational database schema for representing multidimensional data. It stores data in a central fact table and is surrounded by more dimension tables that hold information about each perspective of the data. For example, a star schema query joins various dimensions of a star schema data set.
- ▶ Db2 utilities that are used for index maintenance, such as LOAD, REORG, and REBUILD.
Indexes allow quick access to table rows. However, the databases become less efficient over time and must be maintained as data in large databases is manipulated.

The zIIP runs portions of eligible database workloads, which helps to free computer capacity and lower software costs. Not all Db2 workloads are eligible for zIIP processing. Db2 UDB for z/OS V8 and later gives z/OS the information to direct portions of the work to the zIIP. The result is that in every user situation, different variables determine how much work is redirected to the zIIP.

On an IBM z16 A02 and IBM z16 AGZ, the following workloads also can benefit from zIIPs:

- ▶ z/OS Communications Server uses the zIIP for eligible Internet Protocol Security (IPSec) network encryption workloads. Portions of IPSec processing take advantage of the zIIPs, specifically end-to-end encryption with IPSec. The IPSec function moves a portion of the processing from the general-purpose processors to the zIIPs. In addition, to run the encryption processing, the zIIP also handles the cryptographic validation of message integrity and IPSec header processing.
- ▶ z/OS Global Mirror, formerly known as *Extended Remote Copy* (XRC), also uses the zIIP. Most z/OS Data Facility Storage Management Subsystem (DFSMS) system data mover (SDM) processing that is associated with z/OS Global Mirror can run on the zIIP.
- ▶ The first IBM user of z/OS XML system services is Db2 V9. For Db2 V9 before the z/OS XML System Services enhancement, z/OS XML System Services non-validating parsing was partially directed to zIIPs when used as part of a distributed Db2 request through DRDA. This enhancement benefits Db2 by making all z/OS XML System Services non-validating parsing eligible to zIIPs. This configuration is possible when processing is used as part of any workload that is running in enclave SRB mode.
- ▶ z/OS Communications Server also allows the HiperSockets Multiple Write operation for outbound large messages (originating from z/OS) to be run by a zIIP. Application workloads that are based on XML, HTTP, SOAP, and Java, and traditional file transfer can benefit.
- ▶ During the SRB boost period, ANY work in a boosting image is eligible to run on a zIIP processor associated with the image (LPAR).

Many more workloads and software can use zIIP processors, such as the following examples:

- ▶ IBM z/OS Container Extensions (zCX)
- ▶ IBM z/OS CIM monitoring
- ▶ IBM z/OS Management Facility (z/OSMF)
- ▶ System Display and Search Facility (SDSF)
- ▶ IBM z/OS Connect EE components
- ▶ IBM Sterling™ Connect:Direct®
- ▶ IBM Z System Automation:
- ▶ Java components of IBM Z SMS and SAS
- ▶ IBM Z NetView RESTful API server
- ▶ IBM Z Workload Scheduler & Dynamic Workload Console (under WebSphere Liberty)
- ▶ IMS workloads (DRDA, SOAP, MSC, ISC)
- ▶ IBM Python V3.11
- ▶ Db2 for z/OS Data Gate

- ▶ Db2 Sort for z/OS
- ▶ Db2 Analytics Accelerator Loader for z/OS
- ▶ Db2 Utilities Suite for z/OS
- ▶ Db2 Log Analysis Tool for z/OS
- ▶ Data Virtualization Manager for z/OS (DVM)
- ▶ IzODA (Apache Spark workloads)
- ▶ Watson Machine Learning for z/OS (WMLz) for Mleap and Spark workloads
- ▶ IBM Z Common Data Provider (CDP)
- ▶ IBM Omegamon Portfolio components
- ▶ IBM RMF (Monitor III work)
- ▶ IBM Developer for z/OS Enterprise Edition components.

For more information about zIIP and eligible workloads, see the [IBM zIIP web page](#).

zIIP installation

One CP must be installed with or before any zIIP is installed.

Unassigned zIIPs

New on IBM z16 A02 and IBM z16 AGZ, a zIIP that is purchased but not activated is registered as an Unassigned zIIP (FC 1975). When the system is later upgraded with another zIIP, the system recognizes that a zIIP was purchased and is present. zIIPs are orderable by using FC 1961.

The allowable number of zIIPs for each feature is listed in Table 3-5.

Table 3-5 Number of zIIPs per feature

Features	Max5	Max16	Max32	Max68
Maximum of zIIPs FC 1961	4	15	31	67
Maximum of Unassigned zIIPs FC 1975 ^a	3	15	31	67

a. The numbers for FC 1961 and 1975 are based on one active CP in the configuration.

The maximum number of Unassigned zIIPs decreases by the number of active zIIPs and active CPs.

Note: Starting with IBM z16 A02 and IBM z16 AGZ announcement, the 2:1 zIIP:CP ratio restriction has been removed. With one CP ordered, the number of zIIPs orderable is now (MaxYY-1). The restriction has been lifted for IBM z16 A02, IBM16 AGZ and IBM z16 A01 as well.

If the installed CPC drawer has no remaining unassigned PUs, the assignment of the next zIIP might require the installation of another CPC drawer.

PUs that are characterized as zIIPs within a configuration are grouped into the zIIP pool. This configuration allows zIIPs to have their own processing weights, independent of the weight of parent CPs. The zIIP pool can be seen on the hardware console.

The number of permanent zIIPs plus temporary zIIPs cannot exceed twice the number of purchased CPs plus temporary CPs. Also, the number of temporary zIIPs cannot exceed the number of permanent zIIPs.

LPAR: In an LPAR, as many zIIPs as are available can be defined together with at least one CP.

3.5.6 System assist processors

A system assist processor (SAP) is a PU that runs the channel subsystem LIC to control I/O operations. All SAPs run I/O operations for all LPARs. As with IBM z14, z15 and z16, in IBM z16 A02 and IBM z16 AGZ, SMT is enabled¹³ for SAPs. All features include standard SAPs configured. SAPs increase the capability of the channel subsystem to run I/O operations.

The number of standard SAPs depends on the IBM z16 A02 and IBM z16 AGZ feature, as listed in Table 3-6.

Table 3-6 Standard SAPs per feature

Features	Max5	Max16	Max32	Max68
Standard SAPs	2	2	4	8

Note: On the IBM z16 A02 and IBM z16 AGZ configurations it is no longer possible to order additional Optional SAPs.

3.5.7 Reserved processors

Reserved processors are defined by PR/SM to allow for a nondestructive capacity upgrade. Reserved processors are similar to spare logical processors and can be shared or dedicated. Reserved CPs can be defined to an LPAR dynamically to allow for nondisruptive image upgrades.

Reserved processors can be dynamically configured online by an operating system that supports this function if enough unassigned PUs are available to satisfy the request. The PR/SM rules that govern logical processor activation remain unchanged.

By using reserved processors, you can define more logical processors than the number of available CPs, IFLs, ICFs, and zIIPs in the configuration to an LPAR. This process makes it possible to nondisruptively configure online more logical processors after more CPs, IFLs, ICFs, and zIIPs are made available concurrently. They can be made available with one of the capacity on-demand options.

3.5.8 Integrated Firmware Processors

Integrated Firmware Processors (IFP) are allocated from the pool of PUs and are available for the entire system. Unlike other characterized PUs, IFPs are standard on IBM z16 A02 and IBM z16 AGZ and not defined by the client.

The two PUs that are characterized as IFP are dedicated to supporting firmware functions that are implemented in Licensed Internal Code (LIC); for example, the resource groups (RGs) that are used for managing the following *native* Peripheral Component Interconnect Express (PCIe) features:

- ▶ 10GbE and 25GbE RoCE Express3 Short Reach (SR) and Long Reach (LR)

¹³ Enabled by default, cannot be changed or altered by user

- ▶ 10GbE and 25GbE RoCE Express2.1
- ▶ 10GbE and 25GbE RoCE Express2
- ▶ Coupling Express2 Long Reach

IFPs are initialized at POR. They support various firmware functions such as Resource Group (RG) LIC¹⁴ to provide native PCIe I/O feature management and virtualization functions.

3.5.9 Processor unit assignment

The processor unit assignment of characterized PUs is done at POR time, when the system is initialized. The initial assignment rules keep PUs of the same characterization type grouped as much as possible in relation to PU chips and CPC drawer boundaries to optimize shared cache usage.

The PU assignment is based on CPC drawer plug order (not “ordering”). Feature upgrade provides additional dual-chip modules (e.g. Max16 to Max32) or a fully populated CPC drawer (e.g. Max32 to Max68).

The CPC drawers are populated from the bottom up. This process defines following the low-order and the high-order CPC drawers:

- ▶ CPC drawer 1 (CPC 0 at position A10)¹⁵: Plug order 1 (low-order CPC drawer)
- ▶ CPC drawer 2 (CPC 1 at position A15)¹⁵: Plug order 2

The assignment rules comply with the following order:

- ▶ Spare: CPC drawers 0 and 1 are assigned one spare each on the high PU chip. In the features Max5, Max16, and Max32, both spares are assigned to CPC drawer 0.
- ▶ IFP: Two IFP's are assigned to CPC drawer 0.
- ▶ SAPs: Spread across CPC drawers and high PU chips. Each CPC drawer includes at least five standard SAPs. Start with the highest PU chip high core, then the next highest PU chip high core. This process prevents all the SAPs from being assigned on one PU chip.
- ▶ IFLs and ICFs: Assign IFLs and ICFs to cores on chips in the higher order CPC drawer working downward.
- ▶ CPs and zIIPs: Assign CPs and zIIPs to cores on chips in lower CPC drawers working upward.

These rules are intended to isolate processors that are used by different operating systems as much as possible on different CPC drawers and even on different PU chips. This configuration ensures that different operating systems do not use the same shared caches. For example, CPs and zIIPs are all used by z/OS, and can benefit by using the same shared caches. However, IFLs are used by z/VM and Linux, and ICFs are used by CFCC.

This initial PU assignment, which is done at POR, can be dynamically rearranged by an LPAR by swapping an active core to a core in a different PU chip in a different CPC drawer to improve system performance. For more information, see “LPAR dynamic PU reassignment” on page 123.

When a CPC drawer is added concurrently after POR and new LPARs are activated, or processor capacity for active partitions is dynamically expanded, the extra PU capacity can

¹⁴ IBM zHyperLink Express1.1 and IBM zHyperLink Express are not managed by Resource Groups LIC

¹⁵ A10 & A15 for the IBM z16 A02. For the IBM z16 AGZ configurations, rack position is decided by client.

be assigned from the new CPC drawer. The processor unit assignment rules consider the newly installed CPC drawer dynamically.

3.5.10 Sparing rules

On a IBM z16 A02 and IBM z16 AGZ configuration, all features have two (2) spares. These spare PUs are available to replace any two characterized PUs, whether they are CP, IFL, ICF, zIIP, SAP, or IFP.

Systems with a failed PU for which no spare is available *call home* for a replacement. A system with a failed PU that is spared and requires an DCM to be replaced (referred to as a *pending repair*) can still be upgraded when sufficient PUs are available.

Transparent CP, IFL, ICF, zIIP, SAP, and IFP sparing

Depending on the feature, sparing of CP, IFL, ICF, zIIP, SAP, and IFP is transparent and does not require operating system or operator intervention.

With *transparent sparing*, the status of the application that was running on the failed processor is preserved. The application continues processing on a newly assigned CP, IFL, ICF, zIIP, SAP, or IFP (allocated to one of the spare PUs) without client intervention.

Application preservation

If no spare PU is available, *application preservation* (z/OS only) is started. The state of the failing processor is passed to another active processor that is used by the operating system. Through operating system recovery services, the task is resumed successfully (in most cases, without client intervention).

Dynamic SAP and IFP sparing and reassignment

Dynamic recovery is provided if a failure of the SAP or IFP occurs. If the SAP or IFP fails, and if a spare PU is available, the spare PU is dynamically assigned as a new SAP or IFP. If no spare PU is available, and more than one CP is characterized, a characterized CP is reassigned as an SAP or IFP. In either case, client intervention is not required. This capability eliminates an unplanned outage and allows a service action to be deferred to a more convenient time.

3.5.11 CPC drawer numbering¹⁶

For the IBM z16 A02 and IBM z16 AGZ configurations, CPC drawer numbering starts with CPC0. The first CPC drawer is installed in the frame at location A10 (IBM z16 A02) or ACP0 (IBM z16 AGZ). The second CPC drawer (CPC 1) is installed at location at A15 (IBM z16 A02) or ACP1 (IBM z16 AGZ). For additional details see Appendix D, “Rack configurations” on page 481.

Figure 3-20 on page 113 shows CPC drawer numbering.

¹⁶ For IBM z16 A02 (IBM factory frame installed) position of all components are fixed. IBM z16 AGZ does not have any “Reserved” space features (also, no carry forward of any Reserve feature from previous systems).

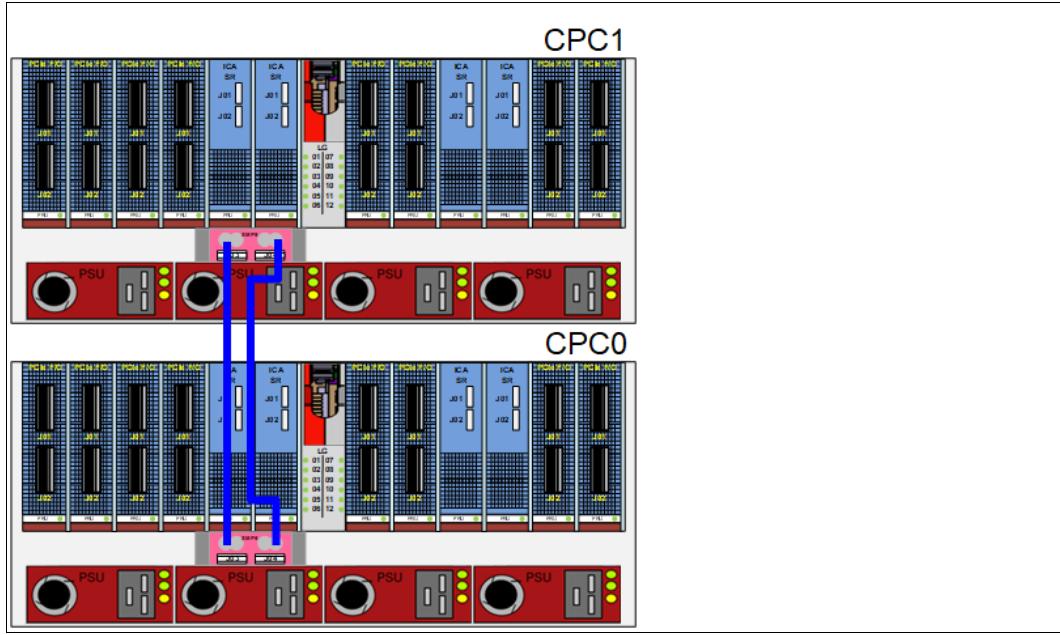


Figure 3-20 CPC drawer number

3.6 Memory design

Various considerations of the IBM z16 A02 and IBM z16 AGZ memory design are described in this section.

3.6.1 Overview

The IBM z16 A02 and IBM z16 AGZ memory design also provides flexibility, high availability, and the following upgrades:

- ▶ Concurrent memory upgrades if the physically installed capacity is not yet reached. IBM z16 A02 and IBM z16 AGZ can have more physically installed memory than the initial available capacity. Memory upgrades within the physically installed capacity can be done concurrently by LIC, and no hardware changes are required. However, memory upgrades *cannot* be done through CBU or On/Off CoD.
- ▶ Concurrent memory upgrades if the physically installed capacity is reached are available only for the Max68 system, or very specific upgrades when combining the memory upgrade with a Max32 to Max68 upgrade.

For more information, see 2.5.6, “Drawer replacement and memory” on page 52. Memory upgrades which require the addition or replacement for existing DIMMs in a single CPC drawer system require the system to be powered off during this operation.

When the total capacity that is installed has more usable memory than required for a configuration, the Licensed Internal Code Configuration Control (LICCC) determines how much memory is used from each processor drawer. The sum of the LICCC provided memory from each CPC drawer is the amount that is available for use in the system.

Memory allocation

When the system is activated by using a POR, PR/SM determines the total installed memory and the customer enabled memory. Later in the process, during LPAR activation, PR/SM assigns and allocates each partition memory according to their image profile.

PR/SM controls all physical memory, and can make physical memory available to the configuration when a CPC drawer is added.

In older IBM zSystems processors, memory allocation was striped across the available CPC drawers because relatively fast connectivity (that is, relatively fast to the processor clock frequency) existed between the drawers. Splitting the work between all of the memory controllers allowed a smooth performance variability.

The memory allocation algorithm changed starting with IBM z13®. For IBM z16 A02 and IBM z16 AGZ, PR/SM tries to allocate memory into a single CPC drawer. If memory does not fit into a single drawer, PR/SM tries to allocate the memory into the CPC drawer with the most processor entitlement.

The PR/SM memory and logical processor resources allocation goal is to place all partition resources on a single CPC drawer, if possible. The resources, such as memory and logical processors, are assigned to the logical partitions at the time of their activation. Later on, when all partitions are activated, PR/SM can move memory between CPC drawers to benefit the performance of each LPAR, without operating system knowledge. This process was done on the previous families of IBM zSystems servers only for PUs that use PR/SM dynamic PU reallocation.

With IBM z16 A02 and IBM z16 AGZ, this process occurs whenever the configuration changes, such as in the following circumstances:

- ▶ Activating or deactivating an LPAR
- ▶ Changing the LPARs processing weights
- ▶ Upgrading the system through a temporary or permanent record
- ▶ Downgrading the system through deactivation of a temporary record

PR/SM schedules a global reoptimization of the resources in use. It does so by reviewing all the partitions that are active and prioritizing them based on their processing entitlement and weights, which creates a high- and low-priority rank. Then, the resources, such as logical processors and memory, can be moved from one CPC drawer to another to address the priority ranks that were created.

When partitions are activated, PR/SM tries to find a home assignment CPC drawer, home assignment node, and home assignment chip for the logical processors that are defined to them. The PR/SM goal is to allocate all the partition logical processors and memory to a single CPC drawer (the home drawer for that partition).

If all logical processors can be assigned to a home drawer and the partition-defined memory is greater than what is available in that drawer, the exceeding memory amount is allocated on another CPC drawer. If all the logical processors cannot fit in one CPC drawer, the remaining logical processors spill to another CPC drawer. When that overlap occurs, PR/SM stripes the memory (if possible) across the CPC drawers where the logical processors are assigned.

The process of reallocating memory is based on the memory copy/reassign function, which is used to allow enhanced drawer availability (EDA) and concurrent drawer replacement (CDR)¹⁷. This process was enhanced starting with z13 and IBM z13s® to provide more efficiency and speed to the process without affecting system performance.

¹⁷ In previous IBM zSystems generations (before z13), these service operations were known as enhanced book availability (EBA) and concurrent book repair (CBR).

IBM z16 A02 and IBM z16 AGZ implement a faster dynamic memory reallocation mechanism, which is especially useful during service operations (EDA and CDR)¹⁸. PR/SM controls the reassignment of the content of a specific physical memory array in one CPC drawer to a physical memory array in another CPC drawer. To accomplish this task, PR/SM uses all the available physical memory in the system. This memory includes the memory that is not in use by the system that is available but not purchased by the client if installed.

Because of the memory allocation algorithm, systems that undergo many miscellaneous equipment specification (MES) upgrades for memory can have different memory mixes and quantities in all processor drawers of the system. If the memory fails, it is technically feasible to run a POR of the system with the remaining working memory resources. After the POR completes, the memory distribution across the processor drawers is different, as is the total amount of available memory.

Large page support

By default, page frames are allocated with a 4 KB size. IBM z16 A02 and IBM z16 AGZ also support large page sizes of 1 MB or 2 GB. The first z/OS release that supports 1 MB pages is z/OS V1R9. Linux on IBM zSystems 1 MB pages support is available in SUSE Linux Enterprise Server 10 SP2 and Red Hat Enterprise Linux (RHEL) 5.2 and later.

The TLB reduces the amount of time that is required to translate a virtual address to a real address. This translation is done by dynamic address translation (DAT) when it must find the correct page for the correct address space.

Each TLB entry represents one page. As with other buffers or caches, lines are discarded from the TLB on a least recently used (LRU) basis.

The worst-case translation time occurs when a TLB miss occurs and the segment table (which is needed to find the page table) and the page table (which is needed to find the entry for the particular page in question) are not in cache. This case involves two complete real memory access delays plus the address translation delay. The duration of a processor cycle is much shorter than the duration of a memory cycle, so a TLB miss is relatively costly.

It is preferable to have addresses in the TLB. With 4 K pages, holding all of the addresses for 1 MB of storage takes 256 TLB lines. When 1 MB pages are used, it takes only one TLB line. Therefore, large page size users have a much smaller TLB footprint.

Large pages allow the TLB to better represent a large working set and suffer fewer TLB 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 see performance improvements in elapsed time and processor usage. These improvements are because DAT and memory operations are part of processor busy time, even though the processor waits for memory operations to complete without processing anything else in the meantime.

To overcome the processor usage that is associated with creating a 1 MB page, a process must run for some time. It also must maintain frequent memory access to keep the pertinent addresses in the TLB.

Short-running work does not overcome the processor usage. Short processes with small working sets are expected to receive little or no improvement. Long-running work with high memory-access frequency is the best candidate to benefit from large pages.

Long-running work with low memory-access frequency is less likely to maintain its entries in the TLB. However, when it does run, few address translations are required to resolve all of the

¹⁸ EDA and CDR are only possible for Max68 feature.

memory it needs. Therefore, a long-running process can benefit even without frequent memory access.

Weigh the benefits of whether something in this category must use large pages as a result of the system-level costs of tying up real storage. A balance exists between the performance of a process that uses large pages and the performance of the remaining work on the system.

On IBM z16 A02 and IBM z16 AGZ, 1 MB large pages become pageable if Virtual Flash Memory¹⁹ is available and enabled. They are available only for 64-bit virtual private storage, such as virtual memory that is above 2 GB.

It is easy to assume that increasing the TLB size is a feasible option to deal with TLB-miss situations. However, this process is not as simple as it seems. As the size of the TLB increases, so does the processor usage that is involved in managing the TLB's contents. Correct sizing of the TLB is subject to complex statistical modeling to find the optimal trade-off between size and performance.

Memory Encryption

Together with the new memory RAIM design the usage of memory encryption is implemented in the IBM z16 A02 and IBM z16 AGZ. The new memory interface uses transparent memory encryption technology to protect all data leaving the processor chips before it's stored in the memory DIMMs. The encryption occurs post-RAIM encoding, the decryption occurs pre-RAIM decoding. That means the data is encrypted before being distributed according to the eight channel RAIM algorithm and written to the eight DIMMs forming a RAIM group, and the data is decrypted after being read from the DIMMs.

3.6.2 Main storage

Main storage consist of memory space addressable by programs and storage that is not directly addressable by programs. Non-addressable storage includes the hardware system area (HSA).

Main storage provides the following functions:

- ▶ Data storage and retrieval for PUs and I/O
- ▶ Communication with PUs and I/O
- ▶ Communication with and control of optional expanded storage
- ▶ Error checking and correction

Main storage can be accessed by all processors, but cannot be shared between LPARs. Any system image (LPAR) must include a defined main storage size. This defined main storage is allocated exclusively to the LPAR during partition activation.

3.6.3 Hardware system area

The HSA is a non-addressable storage area that contains system LIC and configuration-dependent control blocks. On IBM z16 A02 and IBM z16 AGZ configurations, the HSA has a fixed size of 160 GB and is not part of the purchased memory that you order and install.

¹⁹ Virtual Flash Memory replaced IBM zFlash Express. No carry forward of zFlash Express exists.

The fixed size of the HSA eliminates planning for future expansion of the HSA because the hardware configuration definition (HCD)/input/output configuration program (IOCP) always reserves space for the following items:

- ▶ Three channel subsystems (CSSs)
- ▶ A total of 15 LPARs in CSSs 1 and 2, and 10 LPARs for the third CSS for a total of 40 LPARs
- ▶ Subchannel set 0 with 63.75-K devices in each CSS
- ▶ Subchannel set 1 with 64-K devices in each CSS
- ▶ Subchannel set 2 with 64-K devices in each CSS

The HSA features sufficient reserved space to allow for dynamic I/O reconfiguration changes to the maximum capability of the processor.

3.6.4 Virtual Flash Memory (FC 0644)

IBM Virtual Flash Memory (VFM, FC 0644) is the replacement for the Flash Express features that were available on the IBM zBC12 and IBM z13s. No application changes are required to change from IBM Flash Express to VFM.

For IBM z16 A02 and IBM z16 AGZ, up to 2.0 TB of virtual flash memory can be ordered, in 512 GB increments. The minimum is 0, while the maximum is 4 features. The number of VFM features ordered reduces the maximum orderable memory for the IBM z16 A02 and IBM z16 AGZ.

3.7 Logical partitioning

The logical partitioning features are described in this section.

3.7.1 Overview

Logical partitioning is a function that is implemented by the PR/SM on IBM z16 A02 and IBM z16 AGZ. IBM z16 A02 and IBM z16 AGZ can run in LPAR mode, or in Dynamic Partition Manager (DPM) mode. DPM provides a GUI for PR/SM to manage system resources (including I/O) dynamically.

PR/SM is aware of the processor drawer structure on IBM z16 A02 and IBM z16 AGZ configurations. However, LPARs do not feature this awareness. LPARs feature resources that are allocated to them from various physical resources. From a systems standpoint, LPARs have no control over these physical resources, but the PR/SM functions do have this control.

PR/SM manages and optimizes allocation and the dispatching of work on the physical topology. Most physical topology that was handled by the operating systems is the responsibility of PR/SM.

As described in 3.5.9, “Processor unit assignment” on page 111, the initial PU assignment is done during POR by using rules to optimize cache usage. This step is the “physical” step, where CPs, zIIPs, IFLs, ICFs, and SAPs are allocated on the processor drawers.

When an LPAR is activated, PR/SM builds logical processors and allocates memory for the LPAR.

PR/SM assigns all logical processors to one CPC drawer that are packed into chips of that drawer and cooperates with operating system use of HiperDispatch.

All processor types of an IBM z16 A02 and IBM z16 AGZ can be dynamically reassigned, except IFPs.

Memory allocation changed from the previous IBM zSystems servers. Partition memory is now allocated based on processor drawer affinity. For more information, see “Memory allocation” on page 114.

Logical processors are dispatched by PR/SM on physical processors. The assignment topology that is used by PR/SM to dispatch logical processors on physical PUs is also based on cache usage optimization.

Processor drawers assignment is more important because they optimize virtual L4 cache usage. Therefore, logical processors from a specific LPAR are packed into a processor drawer as much as possible.

PR/SM optimizes chip assignments within the assigned processor drawer (or drawers) to maximize virtual L3 cache efficiency. Logical processors from an LPAR are dispatched on physical processors on the same PU chip as much as possible.

PR/SM also tries to redispatch a logical processor on the same physical processor to optimize private cache (L1 and L2) usage.

HiperDispatch

PR/SM and z/OS work in tandem to use processor resources more efficiently. HiperDispatch is a function that combines the dispatcher actions and the knowledge that PR/SM has about the topology of the system.

Performance can be optimized by redispatching units of work to the same processor group, which keeps processes running near their cached instructions and data, and minimizes transfers of data ownership among processors and processor drawers.

The nested topology is returned to z/OS by the Store System Information (STSI) instruction. HiperDispatch uses the information to concentrate logical processors around shared caches (virtual L3 and virtual L4 caches at drawer level), and dynamically optimizes the assignment of logical processors and units of work.

z/OS dispatcher manages multiple queues, called *affinity queues*, with a target number of eight processors per queue, which fits well onto a single PU chip. These queues are used to assign work to as few logical processors as are needed for an LPAR workload. Therefore, even if the LPAR is defined with many logical processors, HiperDispatch optimizes this number of processors to be near the required capacity. The optimal number of processors to be used is kept within a processor drawer boundary, when possible.

Tip: z/VM V7R1^a and later also support HiperDispatch.

- a. z/VM 7.1 is NOT supported - z/VM 7.2 and later are supported on IBM z16 A02 and IBM z16 AGZ.

Logical partitions

PR/SM enables IBM z16 A02 and IBM z16 AGZ to be initialized for a logically partitioned operation, supporting up to 40 LPARs. Each LPAR can run its own operating system image in any image mode, independently from the other LPARs.

An LPAR can be added, removed, activated, or deactivated at any time. Changing the number of LPARs is not disruptive and does not require a POR. Certain facilities might not be available to all operating systems because the facilities might have software corequisites.

Each LPAR has the following resources that are the same as a real CPC:

► Processors

Called *logical processors*, they can be defined as CPs, IFLs, ICFs, or zIIPs. They can be dedicated to an LPAR or shared among LPARs. When shared, a processor weight can be defined to provide the required level of processor resources to an LPAR. Also, the capping option can be turned on, which prevents an LPAR from acquiring more than its defined weight and limits its processor consumption.

LPARs for z/OS can have CP and zIIP logical processors. The logical processor types can be defined as all dedicated or all shared. The zIIP support is available in z/OS.

The weight and number of online logical processors of an LPAR can be dynamically managed by the LPAR CPU Management function of the Intelligent Resource Director (IRD). These functions can be used to achieve the defined goals of this specific partition and of the overall system. The provisioning architecture of IBM z16 A02 and IBM z16 AGZ systems adds a dimension to the dynamic management of LPARs, as described in Chapter 8, “System upgrades” on page 317.

PR/SM supports an option to limit the amount of physical processor capacity that is used by an individual LPAR when a PU is defined as a general-purpose processor (CP) or an IFL that is shared across a set of LPARs.

This capability is designed to provide a physical capacity limit that is enforced as an absolute (versus relative) limit. It is not affected by changes to the logical or physical configuration of the system. This physical capacity limit can be specified in units of CPs or IFLs. The Change LPAR Controls and Customize Activation Profiles tasks on the HMC were enhanced to support this new function.

For the z/OS Workload License Charges (WLC) pricing metric and metrics that are based on it, such as Advanced Workload License Charges (AWLC), an LPAR *defined capacity* can be set. This defined capacity enables the soft capping function. Workload charging introduces the capability to pay software license fees that are based on the processor utilization of the LPAR on which the product is running, rather than on the total capacity of the system.

Consider the following points:

- In support of WLC, the user can specify a defined capacity in millions of service units (MSUs) per hour. The defined capacity sets the capacity of an individual LPAR when soft capping is selected.

The defined capacity value is specified on the Options tab in the Customize Image Profiles window.

- WLM keeps a four-hour rolling average of the processor usage of the LPAR. When the four-hour average processor consumption exceeds the defined capacity limit, WLM dynamically activates LPAR capping (soft capping). When the rolling four-hour average returns below the defined capacity, the soft cap is removed.

For more information about WLM, see *System Programmer's Guide to: Workload Manager*, SG24-6472.

For more information about software licensing, see 7.8, “Software licensing” on page 314.

Weight settings: When defined capacity is used to define an uncapped LPAR's capacity, carefully consider the weight settings of that LPAR. If the weight is much smaller than the defined capacity, PR/SM uses a discontinuous cap pattern to achieve the defined capacity setting. This configuration means PR/SM alternates between capping the LPAR at the MSU value that corresponds to the relative weight settings, and no capping at all. It is best to avoid this scenario and instead attempt to establish a defined capacity that is equal or close to the relative weight.

- ▶ Memory

Memory (main storage) must be dedicated to an LPAR. The defined storage must be available during the LPAR activation; otherwise, the LPAR activation fails.

Reserved storage can be defined to an LPAR, which enables nondisruptive memory addition to and removal from an LPAR by using the LPAR dynamic storage reconfiguration (z/OS and z/VM). For more information, see 3.7.4, “Logical partition storage granularity” on page 127.

- ▶ Channels

Channels can be shared between LPARs by including the partition name in the partition list of a channel-path identifier (CHPID). I/O configurations are defined by the IOCP or the HCD with the CHPID mapping tool (CMT). The CMT is an optional tool that is used to map CHPIDs onto physical channel IDs (PCHIDs). PCHIDs represent the physical location of a port on a card in an I/O cage, I/O drawer, or PCIe I/O drawer.

IOCP is available on the z/OS, z/VM, and z/VSE operating systems, and as a stand-alone program on the hardware console. For more information, see *IBM zSystems Input/Output Configuration Program User’s Guide for ICP IOCP*, SB10-7172. HCD is available on the z/OS and z/VM operating systems. Consult the appropriate 3932DEVICE Preventive Service Planning (PSP) buckets before implementation.

Fibre Channel connection (FICON) channels can be managed by the Dynamic CHPID Management (DCM) function of the Intelligent Resource Director. DCM enables the system to respond to ever-changing channel requirements by moving channels from lesser-used control units to more heavily used control units, as needed.

Modes of operation

The modes of operation are listed in Table 3-7. All available mode combinations, including their operating modes and processor types, operating systems, and addressing modes, also are listed. Only the currently supported versions of operating systems are considered.

Table 3-7 z16 modes of operation

Image mode	PU type	Operating system	Addressing mode
General ^a	CP and zIIP	<ul style="list-style-type: none"> ▶ z/OS ▶ z/VM 	64-bit
	CP	<ul style="list-style-type: none"> ▶ z/VSE and VSEn ▶ Linux on IBM Z ▶ z/TPF 	64-bit
Coupling facility	ICF or CP	CFCC	64-bit

Image mode	PU type	Operating system	Addressing mode
Linux only	IFL or CP	► Linux on IBM Z (64-bit)	64-bit
		► z/VM	
z/VM	CP, IFL, zIIP ^b , or ICF ^b	z/VM	64-bit
SSC ^c	IFL or CP	Linux-based appliance ^d	64 bit

a. General mode uses 64-bit z/Architecture

b. zIIP and ICF for guest use only

c. Secure Service Container

d. IBM Db2 Analytics Accelerator (IDAA), Hyper Protect Virtual Servers (HPVS), and others

The 64-bit z/Architecture mode has no special operating mode because the architecture mode is not an attribute of the definable images operating mode. The 64-bit operating systems are in 31-bit mode at IPL and change to 64-bit mode during their initialization. The operating system is responsible for taking advantage of the addressing capabilities that are provided by the architectural mode.

For more information about operating system support, see Chapter 7, “Operating system support” on page 241.

Logically partitioned mode

If the IBM z16 A02 and IBM z16 AGZ runs in LPAR mode, each of the 40 LPARs can be defined to operate in one of the following image modes:

- General mode to run the following systems:
 - A z/Architecture operating system, on dedicated or shared CPs
 - A Linux on Z operating system, on dedicated or shared CPs
 - z/OS, on any of the following processor units:
 - Dedicated or shared CPs
 - Dedicated CPs *and* dedicated zIIPs
 - Shared CPs *and* shared zIIPs

zIIP usage: zIIPs can be defined to General mode or z/VM mode image, as listed in Table 3-7 on page 120. However, zIIPs are used only by z/OS. Other operating systems cannot use zIIPs, even if they are defined to the LPAR. z/VM V7R1 and later support real and virtual zIIPs to guest z/OS systems.

- General mode is also used to run the z/TPF operating system on dedicated or shared CPs
- CF mode, by loading the CFCC code into the LPAR that is defined as one of the following types:
 - Dedicated or shared CPs
 - Dedicated or shared ICFs
- Linux only mode to run the following systems:

- A Linux on IBM Z operating system, on either of the following types:
 - Dedicated or shared IFLs
 - Dedicated or shared CPs
- A z/VM operating system, on either of the following types:
 - Dedicated or shared IFLs
 - Dedicated or shared CPs
- ▶ z/VM mode to run z/VM on dedicated or shared CPs or IFLs, plus zIIPs for z/OS guests and ICFs for CF guests.
- ▶ Secure Service Container (SSC) mode LPAR can run on dedicated or shared:
 - CPs
 - IFLs

All LPAR modes, required characterized PUs, operating systems, and the PU characterizations that can be configured to an LPAR image are listed in Table 3-8. The available combinations of dedicated (DED) and shared (SHR) processors are also included. For all combinations, an LPAR also can include reserved processors that are defined, which allows for nondisruptive LPAR upgrades.

Table 3-8 LPAR mode and PU usage

LPAR mode	PU type	Operating systems	PUs usage
General	CPs	<ul style="list-style-type: none"> ▶ z/Architecture operating systems ▶ Linux on Z 	CPs DED or CPs SHR
	CPs <i>and</i> zIIPs	<ul style="list-style-type: none"> ▶ z/OS ▶ z/VM (guest exploitation) 	CPs DED or zIIPs DED or CPs SHR or zIIPs SHR
General	CPs	z/TPF	CPs DED or CPs SHR
Coupling facility	ICFs <i>or</i> CPs	CFCC	ICFs DED or ICFs SHR or CPs DED or CPs SHR
Linux only	IFLs <i>or</i> CPs	<ul style="list-style-type: none"> ▶ Linux on Z ▶ z/VM 	IFLs DED or IFLs SHR or CPs DED or CPs SHR
z/VM	CPs, IFLs, zIIPs ^a , or ICFs ^a	z/VM (V7R1 and later)	All PUs must be SHR or DED
SSC ^b	IFLs, <i>or</i> CPs	Linux-based appliance	IFLs DED or IFLs SHR or CPs DED or CPs SHR

a. For guest use only

b. Secure Service Container

Dynamically adding or deleting a logical partition name

Dynamically adding or deleting an LPAR name is the ability to add or delete LPARs and their associated I/O resources to or from the configuration without a POR.

The extra channel subsystem and multiple image facility (MIF) image ID pairs (CSSID/MIFID) can be later assigned to an LPAR for use (or later removed). This process can be done through dynamic I/O commands by using the HCD. At the same time, required channels must be defined for the new LPAR.

Partition profile: Cryptographic coprocessors are not tied to partition numbers or MIF IDs. They are set up with Adjunct Processor (AP) numbers and domain indexes. These numbers are assigned to a partition profile of a given name. The client assigns these AP numbers and domains to the partitions and continues to have the responsibility to clear them out when their profiles change.

Adding logical processors to a logical partition

Logical processors can be concurrently added to an LPAR by defining them as reserved in the image profile and later configuring them online to the operating system by using the appropriate console commands. Logical processors also can be concurrently added to a logical partition dynamically by using the Support Element (SE) "Logical Processor Add" function under the **CPC Operational Customization** task. This SE function allows the initial and reserved processor values to be dynamically changed. The operating system must support the dynamic addition²⁰ of these resources.

Adding a crypto feature to a logical partition

You can plan the addition of supported Crypto Express features to an LPAR on the crypto page in the image profile by defining the Cryptographic Candidate List, and the Usage and Control Domain indexes, in the partition profile. By using the Change LPAR Cryptographic Controls task, you can add crypto adapters dynamically to an LPAR without an outage of the LPAR. Also, dynamic deletion or moving of these features does not require pre-planning. Support is provided in z/OS, z/VM, z/VSE, VSEn, Secure Service Container (based on appliance requirements), and Linux on Z.

LPAR dynamic PU reassignment

The system configuration is enhanced to optimize the PU-to-CPC drawer assignment of physical processors dynamically. The initial assignment of client-visible physical processors to physical processor drawers can change dynamically to better suit the LPAR configurations that are in use.

For more information, see 3.5.9, "Processor unit assignment" on page 111.

Swapping of specialty engines and general processors with each other, with spare PUs, or with both, can occur as the system attempts to compact LPAR configurations into physical configurations that span the least number of processor drawers.

LPAR dynamic PU reassignment can swap client processors of different types between processor drawers²¹. For example, reassignment can swap an IFL on processor drawer 0 with a CP on processor drawer 2. Swaps can also occur between PU chips within a processor drawer or a DCM and can include spare PUs. The goals are to pack the LPAR on fewer

²⁰ In z/OS, this support is available since Version 1 Release 10 (z/OS V1.10), while z/VM supports this addition since z/VM V5.4, and z/VSE since V4.3. However, z16 supports z/OS V2R2 and later, z/VSE V6R2 and z/VM V7R1 and later.

²¹ Applicable to z16 A02 and IBM z16 AGZ Max68 configurations.

processor drawers and also on fewer PU chips, based on the processor drawers' topology. The effect of this process is evident in dedicated and shared LPARs that use HiperDispatch.

LPAR dynamic PU reassignment is transparent to operating systems.

LPAR group capacity limit (LPAR group absolute capping)

The group capacity limit feature allows the definition of a group of LPARs on a z16 system, and limits the combined capacity usage by those LPARs. This process allows the system to manage the group so that the group capacity limits in MSUs per hour are not exceeded. To take advantage of this feature, you must be running z/OS V2R2 or later in all LPARs in the group.

PR/SM and WLM work together to enforce the capacity that is defined for the group and the capacity that is optionally defined for each individual LPAR.

LPAR absolute capping

Absolute capping is a logical partition control that was made available with zEC12 and is supported on IBM z16 A02 and IBM z16 AGZ. With this support, PR/SM and the HMC are enhanced to support a new option to limit the amount of physical processor capacity that is used by an individual LPAR when a PU is defined as a general-purpose processor (CP), zIIP, or an IFL processor that is shared across a set of LPARs.

Unlike traditional LPAR capping, absolute capping is designed to provide a physical capacity limit that is enforced as an absolute (versus relative) value that is not affected by changes to the virtual or physical configuration of the system.

Absolute capping provides an optional maximum capacity setting for logical partitions that is specified in the absolute processors capacity (for example, 5.00 CPs or 2.75 IFLs). This setting is specified independently by processor type (namely CPs, zIIPs, and IFLs) and provides an enforceable upper limit on the amount of the specified processor type that can be used in a partition.

Absolute capping is ideal for processor types and operating systems that the z/OS WLM cannot control. Absolute capping is not intended as a replacement for defined capacity or group capacity for z/OS, which are managed by WLM.

Absolute capping can be used with any z/OS, z/VM, or Linux on Z LPAR (that is running on an IBM zSystems server). If specified for a z/OS LPAR, absolute capping can be used concurrently with defined capacity or group capacity management for z/OS. When used concurrently, the absolute capacity limit becomes effective before other capping controls.

Dynamic Partition Manager mode

DPM is an IBM zSystems operation mode that provides a simplified approach to create and manage virtualized environments, which reduces the barriers of its adoption for new and existing customers.

The implementation provides built-in integrated capabilities that allow advanced virtualization management on IBM zSystems servers. With DPM, you can use your Linux and virtualization skills while taking advantage of the full value of IBM zSystems hardware, robustness, and security in a workload optimized environment.

DPM provides facilities to define and run virtualized computing systems by using a firmware-managed environment that coordinate the physical system resources that are shared by the partitions. The partitions' resources include processors, memory, network, storage, crypto, and accelerators.

DPM provides a new mode of operation for IBM zSystems servers that provide the following services:

- ▶ Facilitates defining, configuring, and operating PR/SM LPARs in a similar way to how someone performs these tasks on another platform.
- ▶ Lays the foundation for a general IBM zSystems new user experience.

DPM is not another hypervisor for IBM zSystems servers. DPM uses the PR/SM hypervisor infrastructure and provides an intelligent interface on top of it that allows customers to define, use, and operate the platform virtualization without IBM zSystems experience or skills.

3.7.2 Storage operations

In IBM z16 A02 and IBM z16 AGZ, memory can be assigned as main storage, supporting up to 40 LPARs. Before you activate an LPAR, main storage must be defined to the LPAR. All installed storage can be configured as main storage.

For more information about operating system main storage support, see the *PR/SM Planning Guide*, SB10-7178.

Memory *cannot* be shared between system images (LPARs). It is possible to dynamically reallocate storage resources for z/Architecture LPARs that run operating systems that support dynamic storage reconfiguration (DSR). This process is supported by z/OS, and z/VM. z/VM, in turn, virtualizes this support to its guests.

For more information, see 3.7.5, “LPAR dynamic storage reconfiguration” on page 127.

Operating systems that run as guests of z/VM can use the z/VM capability of implementing virtual memory to guest virtual machines. The z/VM dedicated real storage can be shared between guest operating systems.

LPAR main storage allocation and usage

The IBM z16 A02 and IBM z16 AGZ storage allocation and usage possibilities depend on the image mode and the operating system that is deployed in the LPAR.

Important: The memory allocation and usage depends on the operating system architecture and tested (documented for each operating system) limits. While the maximum supported memory per LPAR for IBM z16 A02 and IBM z16 AGZ is 16TB, each operating system has its own support specifications.

For more information about the amount of main memory supported by the different operation systems, see the PR/SM Planning Guide, SB10-7178, which is available at the [IBM Resource Link website](#) (log in required).

The following modes are provided:

- ▶ z/Architecture mode

In z/Architecture (General) mode, storage addressing is 64-bit, which allows for virtual addresses up to 16 exabytes (16 EB). However, the current main storage limit for LPARs on IBM z16 A02 and IBM z16 AGZ is 16 TB of main storage.

- ▶ CF mode

In CF mode, storage addressing is 64 bit for a CF image that runs CFCC. The current IBM z16 A02 and IBM z16 AGZ definition limit for CF LPARs is 16 TB of storage.

The following CFCC levels are supported in a Sysplex with IBM z16 A02 and IBM z16 AGZ:

- CFCC Level 25, available on z16 (Driver level 51)
- CFCC Level 24, available on z15 (Driver level 41)
- CFCC Level 23, available on z14 (Driver level 36)

For more information, see 7.4.3, “Coupling and clustering features and functions” on page 274.

Only IBM CFCC can run in CF mode.

► **Linux only mode**

In Linux only mode, storage addressing can be 31 bit or 64 bit, depending on the operating system architecture and the operating system configuration.

Only Linux and z/VM operating systems can run in Linux only mode. Linux on IBM Z 64-bit distributions (SUSE Linux Enterprise Server 12 SP5, SLES 15 SP4, Red Hat RHEL 7.9, RHEL 8.4, RHEL 9, Ubuntu 20.04.1 LTS and Ubuntu 22.04 LTS and later) use 64-bit addressing and operate in z/Architecture mode. z/VM also uses 64-bit addressing and operates in z/Architecture mode.

Note: For information about the (kernel) supported amount of memory, check the Linux Distribution specific documentation.

For the current supported Linux on IBM Z see the following [website](#).

► **z/VM mode**

In z/VM mode, certain types of processor units can be defined within one LPAR. This feature increases flexibility and simplifies systems management by allowing z/VM to run the following tasks in the same z/VM LPAR:

- Manage guests to operate Linux on Z on IFLs
- Operate z/VSE and z/OS on CPs
- Offload z/OS system software processor usage, such as Db2 workloads on zIIPs
- Provide an economical Java execution environment under z/OS on zIIPs

► **IBM SSC**

In IBM SSC mode, storage addressing is 64-bit for an embedded product. The amount of usable main storage by the appliance code that is deployed in the SSC LPAR is documented by the appliance code supplier.

3.7.3 Reserved storage

Reserved storage can be optionally defined to an LPAR, which allows a nondisruptive image memory upgrade for this partition. Reserved storage can be defined to central and expanded storage, and to any image mode except CF mode.

An LPAR must define an amount of main storage:

- The initial value is the storage size that is allocated to the partition when it is activated.
- The reserved value is another storage capacity that is beyond its initial storage size that an LPAR can acquire dynamically. The reserved storage sizes that are defined to an LPAR do not have to be available when the partition is activated. Instead, they are predefined storage sizes to allow a storage increase, from an LPAR perspective.

Without the reserved storage definition, an LPAR storage upgrade is a disruptive process that requires the following steps:

1. Partition deactivation.
2. An initial storage size definition change.
3. Partition activation.

The extra storage capacity for an LPAR upgrade can come from the following sources:

- ▶ Any unused available storage
- ▶ Another partition that features released storage
- ▶ A memory upgrade

A concurrent LPAR storage upgrade uses DSR. z/OS uses the reconfigurable storage unit (RSU) definition to add or remove storage units in a nondisruptive way.

z/VM V7R2 and later also support Dynamic Memory Downgrade (DMD), which allows the removal of up to 50% of the real storage from a running z/VM system. Removing memory from a z/VM guest is not disruptive to the z/VM LPAR.

3.7.4 Logical partition storage granularity

Granularity of main storage for an LPAR depends on the largest main storage amount that is defined for initial or reserved main storage, as listed in Table 3-9²².

Table 3-9 Logical partition main storage granularity (IBM z16 A02 and IBM z16 AGZ)

Logical partition: Largest main storage amount	Logical partition: Main storage granularity
Main storage amount </= 512 GB	1 GB
512 GB < main storage amount </= 1 TB	2 GB
1 TB < main storage amount </= 2 TB	4 GB
2 TB < main storage amount </= 4 TB	8 GB
4 TB < main storage amount </= 8 TB	16 GB
8 TB < main storage amount </= 16 TB	32 GB
16 TB < main storage amount </= 32TB	64 GB

LPAR storage granularity information is required for LPAR image setup and for z/OS RSU definition. On IBM z16 A02 and IBM z16 AGZ, LPARs support maximum size of 16 TB of main storage. However, the maximum amount of memory that is supported by z/OS V2R2, V2R3, and V2R4 is 4 TB. z/OS V2R5 supports up to 16 TB. z/VM V7R2 and V7R3 limit is 4 TB and it supports LPAR dynamic storage reconfiguration

3.7.5 LPAR dynamic storage reconfiguration

Dynamic storage reconfiguration on IBM z16 A02 and IBM z16 AGZ allows an operating system that is running on an LPAR to add (nondisruptively) its reserved storage amount to its configuration. This process can occur only if unused storage exists. This unused storage can be obtained when another LPAR releases storage, or when a concurrent memory upgrade occurs.

²² When defining an LPAR on the HMC, the 2G boundary should still be followed in PR/SM.

With dynamic storage reconfiguration, the unused storage does not have to be continuous.

When an operating system running on an LPAR assigns a storage increment to its configuration, PR/SM determines whether any free storage increments are available. PR/SM then dynamically brings the storage online.

PR/SM dynamically takes offline a storage increment and makes it available to other partitions when an operating system running on an LPAR releases a storage increment.

3.8 Intelligent Resource Director

Intelligent Resource Director (IRD) is an IBM zSystems capability that is used only by z/OS. IRD is a function that optimizes processor and channel resource utilization across LPARs within a single IBM zSystems server.

This feature extends the concept of goal-oriented resource management. It does so by grouping system images that are on the same z16 or Z servers that are running in LPAR mode (and in the same Parallel Sysplex) into an *LPAR cluster*. This configuration allows WLM to manage resources (processor and I/O) across the entire cluster of system images and not only in one single image.

An LPAR cluster is shown in Figure 3-21. It contains three z/OS images and one Linux image that is managed by the cluster. Included as part of the entire Parallel Sysplex is another z/OS image and a CF image. In this example, the scope over which IRD has control is the defined LPAR cluster.

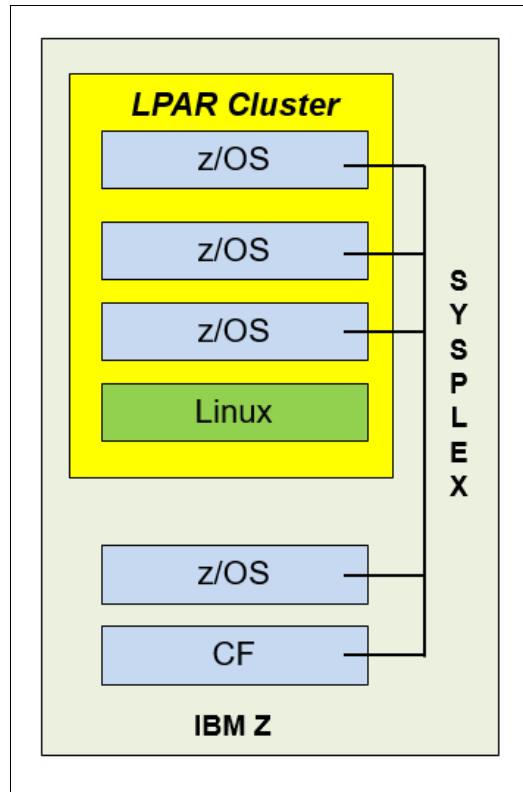


Figure 3-21 IRD LPAR cluster example

IRD features the following characteristics:

- ▶ IRD processor management

WLM dynamically adjusts the number of logical processors within an LPAR and the processor weight based on the WLM policy. The ability to move the processor weights across an LPAR cluster provides processing power where it is most needed, based on WLM goal mode policy.

The processor management function is automatically deactivated when HiperDispatch is active. However, the LPAR weight management function remains active with IRD with HiperDispatch. For more information about HiperDispatch, see 3.7, “Logical partitioning” on page 117.

HiperDispatch manages the number of logical CPs in use. It adjusts the number of logical processors within an LPAR to achieve the optimal balance between CP resources and the requirements of the workload.

HiperDispatch also adjusts the number of logical processors. The goal is to map the logical processor to as few physical processors as possible. This configuration uses the processor resources more efficiently by trying to stay within the local cache structure. Doing so makes efficient use of the advantages of the high-frequency microprocessors, and improves throughput and response times.

- ▶ Dynamic channel path management (DCM)

DCM moves FICON channel bandwidth between disk control units to address current processing needs. IBMz16 z16 A02 and IBM z16 AGZ support DCM within a channel subsystem.

- ▶ Channel subsystem priority queuing

This function allows the priority queuing of I/O requests in the channel subsystem and the specification of relative priority among LPARs. When running in goal mode, WLM sets the priority for an LPAR and coordinates this activity among clustered LPARs.

For more information about implementing LPAR processor management under IRD, see *z/OS Intelligent Resource Director*, SG24-5952.

3.9 Clustering technology

Parallel Sysplex is the clustering technology that is used with IBM zSystems servers. The components of a Parallel Sysplex as implemented within the z/Architecture are shown in Figure 3-22 on page 130. The example in Figure 3-22 on page 130 shows one of many possible Parallel Sysplex configurations.

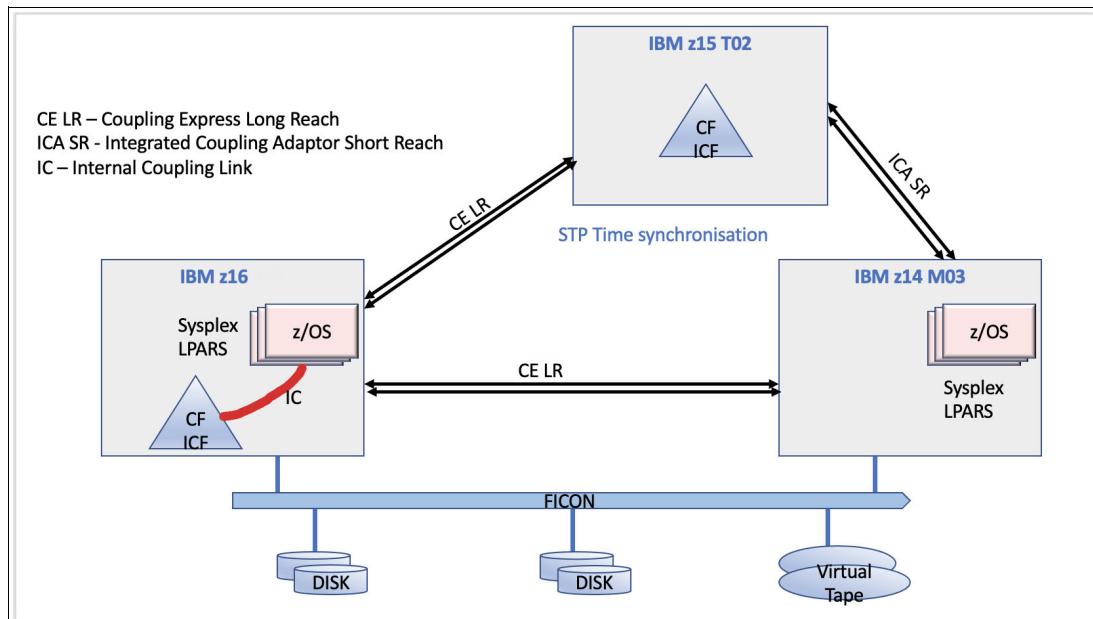


Figure 3-22 Sysplex hardware overview

Figure 3-22 shows an IBM z16 A02 or an IBM z16 AGZ system (represented by IBM z16 in the figure) that contains multiple z/OS sysplex partitions. It contains an internal CF, an IBM z15 T02 system that contains a stand-alone CF, and an IBM z14 M03 that contains multiple z/OS sysplex partitions.

STP over coupling links provides time synchronization to all systems. Selecting the suitable CF link technology Coupling Express2 Long Reach (CE2 LR) or Integrate Coupling Adapter Short Reach (ICA SR and SR1.1) depends on the system configuration and how distant they are physically.

For more information about link technologies, see “Coupling links” on page 175.

Parallel Sysplex is an enabling technology that allows highly reliable, redundant, and robust IBM zSystems technology to achieve near-continuous availability. A Parallel Sysplex consists of one or more (z/OS) operating system images that are coupled through one or more Coupling Facility LPARs.

A correctly configured Parallel Sysplex cluster maximizes availability in the following ways:

- ▶ **Continuous availability:** Changes can be introduced, such as software upgrades, one image at a time, while the remaining images continue to process work. For more information, see *Parallel Sysplex Application Considerations*, SG24-6523.
- ▶ **High capacity:** 1- 32 z/OS images in a Parallel Sysplex operating as a single system.
- ▶ **Dynamic workload balancing:** Because it is viewed as a single logical resource, work can be directed to any operating system image in a Parallel Sysplex cluster that has available capacity.
- ▶ **Systems management:** The architecture defines the infrastructure to satisfy client requirements for continuous availability. It also provides techniques for achieving simplified systems management consistent with this requirement.
- ▶ **Resource sharing:** Several base z/OS components use CF shared storage. This configuration enables sharing of physical resources with significant improvements in cost, performance, and simplified systems management.

- ▶ Single logical system: The collection of system images in the Parallel Sysplex is displayed as a single entity to the operator, user, and database administrator. A single system view means reduced complexity from operational and definition perspectives.
- ▶ N-2 support: Multiple hardware generations (normally three) are supported in the same Parallel Sysplex. This configuration provides for a gradual evolution of the systems in the Parallel Sysplex without changing all of them simultaneously. Software support for multiple releases or versions is also supported.

Note: Parallel sysplex coupling and timing links connectivity for IBM z16 A02 and IBM z16 AGZ is supported to N-2 generation CPCs (z16, z15, and z14).

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

3.9.1 CF Control Code

The LPAR that is running the CFCC can be on z16, z15, or z14 systems. For more information about CFCC requirements for supported systems and functions and feature of the different CFCC levels, see 7.4.3, “Coupling and clustering features and functions” on page 274 or the current exception letter that is published on IBM Resource Link.

Consideration: IBM z16, z15, and z14 cannot coexist in the same sysplex with IBM z13, IBM z13s or earlier generation systems.

3.9.2 Coupling Thin Interrupts

CFCC Level 19 introduced Coupling Thin Interrupts to improve performance in environments that share CF engines. Although dedicated engines are preferable to obtain the best CF performance, Coupling Thin Interrupts helps facilitate the use of a shared pool of engines, which helps to lower hardware acquisition costs.

The interrupt causes a shared logical processor CF partition to be dispatched by PR/SM (if it is not already dispatched), which allows the request or signal to be processed in a more timely manner. The CF relinquishes control when work is exhausted or when PR/SM takes the physical processor away from the logical processor.

On IBM z16 A02 and IBM z16 AGZ, the use of Coupling Thin Interrupts (DYNDISP=THIN) is now the only option that is available for shared engines in a CF LPAR. Specification of OFF or ON in CF commands and the CF configuration file will be preserved, for compatibility, but a warning message will be issued to indicate that these options are no longer supported, and that DYNDISP=THIN behavior will be used.

3.9.3 Dynamic CF dispatching

With the introduction of the Coupling Thin Interrupt support (only available option on IBM z16 A02 and IBM z16 AGZ), which is used only when the CF partition uses shared engines, the CFCC code is changed to handle these interrupts correctly. CFCC was also changed to relinquish voluntarily control of the processor whenever it runs out of work to do. It relies on

Coupling Thin Interrupts to dispatch the image again in a timely fashion when new work (or new signals) arrives at the CF to be processed.

With IBM z16 A02 and IBM z16 AGZ, DYNDISP=THIN is the only mode of operation for CF images that use shared processors.

This capability allows ICF engines to be shared by several CF images. In this environment, it provides faster and far more consistent CF service times. It can also provide performance that is reasonably close to dedicated-engine CF performance.

The use of Thin Interrupts allows a CF to run by using a shared processor while maintaining good performance. The shared engine is allowed to be undispatched when no more work exists, as in the past. The Thin Interrupt gets the shared processor that is dispatched when a command or duplexing signal is presented to the shared engine.

This function saves processor cycles and is an excellent option to be used by a production backup CF or a testing environment CF. This function is activated by default when a CF processor is shared.

The CPs can run z/OS operating system images and CF images. For software charging reasons, generally use only ICF processors to run CF images.

For more information about CF configurations, see the following resources:

- ▶ *Coupling Facility Configuration Options*, GF22-5042
- ▶ [This IBM Support web page](#)

3.10 Virtual Flash Memory

Flash Express is not supported on IBM z16 A02 and IBM z16 AGZ. This feature was replaced by Virtual Flash Memory (VFM), with IBM z14. The Virtual Flash Memory feature code is 0644 on IBM z16 A02 and IBM z16 AGZ.

3.10.1 Overview

VFM replaced the PCIe Flash Express feature with support that is based on main memory.

The “storage class memory” that is provided by Flash Express adapters is replaced with memory that is allocated from main memory (VFM).

VFM helps improve availability and handling of paging workload spikes when running z/OS. With this support, z/OS is designed to help improve system availability and responsiveness by using VFM across transitional workload events, such as market openings and diagnostic data collection.

z/OS also helps improve processor performance by supporting middleware use of pageable large (1 MB) pages, and eliminates delays that can occur when collecting diagnostic data during failures.

VFM also can be used in CF images to provide extended capacity and availability for workloads that use IBM WebSphere MQ Shared Queues structures.

3.10.2 VFM feature

A VFM feature (FC 0644) is 512 GB of memory on IBM z16 A02 and IBM z16 AGZ. The maximum number of VFM features is 4 per IBM z16 A02 and IBM z16 AGZ system.

Ordered VFM memory reduces the maximum orderable memory.

Simplification in its management is of great value because no hardware adapter is needed to manage. It also has no hardware repair and verify. It has a better performance because no I/O to attached adapter occurs. Finally, because this feature is part of memory, it is protected by RAIM and ECC.

3.10.3 VFM administration

The allocation and definition information of VFM for all partitions is viewed through the **Storage Information** panel that is under the **Operational Customization** panel.

The information is relocated during CDR in a manner that is identical to the process that was used for expanded storage. VFM is much simpler to manage (HMC task) and no hardware repair and verify (no cables and no adapters) are needed. Also, because this feature is part of internal memory, VFM is protected by RAIM and ECC and can provide better performance because no I/O to an attached adapter occurs.

Note: Use cases for Flash did not change (for example, z/OS paging and CF shared queue overflow). Instead, they transparently benefit from the changes in the hardware implementation.

No option is available for VFM plan ahead. The only option is to always include zVFM plan ahead when Flexible Memory option is selected.

3.11 Secure Service Container

Client applications are subject to several security risks in a production environment. These risks might include external risks (cyber hacker attacks) or internal risks (malicious software, system administrators that use their privileged rights for unauthorized access and many others).

The IBM Secure Service Container (SSC) is a container technology through which you can more quickly and securely deploy software appliances on IBM z16 A02 and IBM z16 AGZ.

An IBM SSC partition is a specialized container for installing and running specific appliances. An appliance is an integration of operating system, middleware, and software components that work autonomously and provide core services and infrastructures that focus on usability and security.

IBM SSC hosts most sensitive client workloads and applications. It acts as a highly protected and secured digital vault, enforcing security by encrypting the entire stack: memory, network, and data (both in-flight and at-rest). Applications that are running inside IBM SSC are isolated and protected from outsider and insider threats.

IBM SSC combines hardware, software, and middleware and is unique to IBM zSystems platform. Though it is called a container, it should not be confused with purely software open source containers (such as Kubernetes or Docker).

IBM SSC is a part of the Pervasive Encryption concept that was introduced with IBM z14, which is aimed at delivering best IBM Security hardware and software enhancements, services, and practices for 360-degree infrastructure protection.

LPAR is defined as IBM SSC by using the HMC.

The IBM SSC solution includes the following key advantages:

- ▶ Applications require zero changes to use IBM SSC; software developers do not need to write any IBM SSC-specific programming code.
- ▶ End-to-end encryption (in-flight and at-rest data):
 - Automatic Network Encryption (TLS, IPsec): Data-in-flight. Automatic File System Encryption (LUKS): Data-at-rest.
 - Linux Unified Key Setup (LUKS) is the standard way in Linux to provide disk encryption. SSC encrypts all data with a key that is stored within the appliance.
 - Protected memory: Up to 16 TB can be defined per IBM SSC LPAR.
- ▶ Encrypted Diagnostic Data
 - All diagnostic information (debug memory dump data, logs, and so on) are encrypted and do not contain any user or application data.
- ▶ No operating system access
 - After the IBM SSC appliance is built, Secure Shell (SSH) and the command line-interface (CLI) are disabled, which ensures that even system administrators cannot access the contents of the IBM SSC and do not know which application is running there.
- ▶ Applications that run inside IBM SSC are being accessed externally by REST APIs only, in a transparent way to user.
- ▶ Tamper-proof SSC Secure Boot:
 - IBM SSC-eligible applications are booted into IBM SSC by using verified booting sequence, where only trusted and digitally signed and verified by IBM software code is uploaded into the IBM SSC.
 - Vertical workload isolation, certified by EAL5+ Common Criteria Standard, which is the highest level that ensures workload separation and isolation.
 - Horizontal workload isolation: Separation from the rest of the host environment.

IBM z16 A02 and IBM z16 AGZ technology provides built-in data encryption with excellent vertical scalability and performance that protects against data breach threats and data manipulation by privileged users. IBM SSC is a powerful IBM technology for providing the extra protection of the most sensitive workloads.

The following IBM solutions and offerings, and more to come, can be deployed in an IBM SSC environment:

- ▶ IBM Hyper Protect Virtual Servers (HPVS) solution is available for running Linux-based virtual servers with sensitive data and applications delivering a confidential computing environment to address your top security concerns.
For more information, see this IBM Cloud® [web page](#).
- ▶ IBM Db2 Analytics Accelerator (IDAA) is a high-performance component that is tightly integrated with Db2 for z/OS. It delivers high-speed processing for complex Db2 queries to support business-critical reporting and analytic workloads. The accelerator transforms the mainframe into a hybrid transaction and analytic processing (HTAP) environment.
For more information, see this IBM [web page](#).

- ▶ IBM Cloud Hyper Protect Data Base as a Service (DBaaS) for PostgreSQL or MongoDB offers enterprise cloud database environments with high availability for sensitive data workloads.
For more information, see this IBM Cloud [web page](#).
- ▶ IBM Cloud Hyper Protect Crypto Services is a key management service and cloud hardware security module (HSM) that supports industry standards such as PKCS #11.
For more information, see this IBM Cloud [web page](#).
- ▶ IBM Security® Guardium® Data Encryption (GDE) consists of a unified suite of products that are built on a common infrastructure. These highly scalable solutions provide data encryption, tokenization, data masking, and key management capabilities to help protect and control access to data across the hybrid multicloud environment.
For more information, see this [web page](#).
- ▶ IBM Blockchain™ platform can be deployed on an IBM z16 A02 and IBM z16 AGZ by using IBM SSC to host the IBM Blockchain network.
For more information, see this [web page](#).



I/O structure

This chapter describes the I/O system structure and connectivity options that are available on the IBM z16 A02 and IBM z16 AGZ.

This chapter includes the following topics:

- ▶ 4.1, “Introduction to I/O infrastructure” on page 138
- ▶ 4.2, “I/O system overview” on page 140
- ▶ 4.3, “PCIe+ I/O drawer” on page 142
- ▶ 4.4, “CPC drawer fanouts” on page 145
- ▶ 4.5, “I/O features” on page 148
- ▶ 4.6, “Connectivity” on page 152
- ▶ 4.7, “Cryptographic functions” on page 180
- ▶ 4.8, “Integrated Firmware Processor” on page 183

4.1 Introduction to I/O infrastructure

This section describes the I/O features available on the IBM z16 A02 and IBM z16 AGZ. Both the BM z16 A02 and IBM z16 AGZ configurations support PCIe+ I/O drawers only.

I/O cage, I/O drawer, and PCIe I/O drawer are not supported.

Note: Throughout this chapter, the terms *adapter* and *card* refer to a PCIe I/O feature that is installed in a PCIe+ I/O drawer.

4.1.1 I/O infrastructure

IBM zSystems I/O is based on industry standard Peripheral Component Interconnect Express Generation 3 (PCIe Gen3) I/O infrastructure. The PCIe I/O infrastructure that is provided by the central processor complex (CPC) enhances I/O capability and flexibility, while allowing for the future integration of PCIe adapters and features.

The PCIe I/O infrastructure in IBM z16 A02 and IBM z16 AGZ consists of the following components:

- ▶ PCIe+ Gen3 dual port fanouts that support 16 GBps I/O bus for CPC drawer connectivity to the PCIe+ I/O drawers. It connects to the PCIe Interconnect Gen3 in the PCIe+ I/O drawers.
- ▶ Integrated Coupling Adapter Short Reach (ICA SR and ICA SR1.1), which are PCIe Gen3 features that support short distance coupling links. The ICA SR and ICA SR1.1 features have two ports, each port supporting 8 GBps.
- ▶ The 8U, 16-slot, and 2-domain PCIe+ I/O drawer for PCIe I/O features.

Features installed in the PCIe+ I/O drawer

The I/O infrastructure of IBM z16 A02 and IBM z16 AGZ provides the following benefits:

- ▶ The bus connecting the CPC drawer to the I/O domain in the PCIe+ I/O drawer bandwidth is 16 GBps.
- ▶ Up to 32 channels (16 PCIe I/O cards) are supported in the PCIe+ I/O drawer.
- ▶ Storage connectivity:
 - Storage Area Network (SAN) connectivity:
 - The FICON Express32S
 - FICON Express16S+ (carry forward)

These cards provide two channels per feature for Fibre Channel connection (FICON), High-Performance FICON on Z (zHPF), and Fibre Channel Protocol (FCP) storage area networks.

- IBM zHyperLink Express 1.1 - two ports per feature (new build and carry forward)
Ultra high speed, direct connection to Select DS8000; works in tandem with FICON Express channels
- IBM zHyperLink Express - two ports per feature (carry forward)
- ▶ Local area network (LAN) connectivity:
 - The Open Systems Adapter (OSA)-Express7S 1.2 GbE, OSA-Express7S 1.2 1000BASE-T, OSA-Express6S GbE, and the OSA-Express6S 1000BASE-T features include two ports each.

- The OSA-Express7S 1.2 25 GbE, and OSA-Express7S 1.2 10 GbE features have one port each.
- ▶ Native PCIe features (plugged into the PCIe+ I/O drawer):
 - 25GbE and 10GbE Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE) Express3 (two ports per feature)
 - 25GbE and 10GbE Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE) Express2.1 (two ports per feature, carry forward)
 - 25GbE and 10GbE RoCE Express2 (two ports per feature, carry forward)
 - Coupling Express2 Long Reach (CE2 LR) - two ports per feature
 - Crypto Express8S (single/dual HSM)
 - Crypto Express7S (single/dual ports/HSM, carry forward)
 - Crypto Express6S (single HSM, carry forward)

4.1.2 PCIe Generation 3

The PCIe Generation 3 uses 128b/130b encoding for data transmission. This configuration reduces the encoding overhead to about 1.54% versus the PCIe Generation 2 overhead of 20% that uses 8b/10b encoding.

The PCIe standard uses a low-voltage differential serial bus. Two wires are used for signal transmission, and a total of four wires (two for transmit and two for receive) form a lane of a PCIe link, which is full-duplex. Multiple lanes can be aggregated into a larger link width. PCIe supports link widths of 1, 2, 4, 8, 12, 16, and 32 lanes (x1, x2, x4, x8, x12, x16, and x32).

The data transmission rate of a PCIe link is determined by the link width (numbers of lanes), the signaling rate of each lane, and the signal encoding rule. The signaling rate of one PCIe Generation 3 lane is eight gigatransfers per second (GTps), which means that nearly 8 gigabits are transmitted per second (Gbps).

Note: I/O infrastructure for IBM z16 A02 and IBM z16 AGZ, as well as for IBM z16 A01 and IBM z15, is implemented as PCIe Generation 3. The PU chip PCIe interface is PCIe Generation 4 (x16 @32 GBps), but the CPC I/O Fanout infrastructure provides external connectivity as PCIe Generation 3 @16GBps

A PCIe Gen3 x16 link features the following data transmission rates:

- ▶ The maximum theoretical data transmission rate per lane:

$$8 \text{ Gbps} * 128/130 \text{ bit (encoding)} = 7.87 \text{ Gbps} = 984.6 \text{ MBps}$$
- ▶ The maximum theoretical data transmission rate per link:

$$984.6 \text{ MBps} * 16 \text{ (lanes)} = 15.75 \text{ GBps}$$

Considering that the PCIe link is full-duplex mode, the data throughput rate of a PCIe Gen3 x16 link is 31.5 GBps (15.75 GBps in both directions).

Link performance: The link speeds do not represent the performance of the link. The performance depends on many factors, including latency through the adapters, cable lengths, and the type of workload.

PCIe Gen3 x16 links are used in IBM z16 A02 and IBM z16 AGZ systems for driving the PCIe+ I/O drawers, and for coupling links for CPC to CPC communications.

Note: Unless specified otherwise, PCIe refers to PCIe Generation 3 in remaining sections of this chapter.

4.2 I/O system overview

The IBM z16 A02 and IBM z16 AGZ I/O characteristics and supported features are described in this section.

4.2.1 Characteristics

The IBM z16 A02 and IBM z16 AGZ I/O subsystem is designed to provide great flexibility, high availability, and the following excellent performance characteristics:

- ▶ High bandwidth

IBM z16 A02 and IBM z16 AGZ use PCIe Gen3 protocol to drive PCIe+ I/O drawers and CPC to CPC (coupling) connections. The I/O bus infrastructure data rate of up to 128 GBps¹ per system (12 PCIe+ Gen3 fanout slots). For more information about coupling link connectivity, see 4.6.4, “Parallel Sysplex connectivity” on page 175.

- ▶ Connectivity options:

- IBM z16 A02 and IBM z16 AGZ configurations can be connected to an extensive range of interfaces, such as FICON/FCP for SAN connectivity, OSA features for LAN connectivity and zHyperLink Express for storage connectivity (low latency compared to FICON).
- For CPC to CPC connections, IBM z16 A02 and IBM z16 AGZ configurations use Integrated Coupling Adapter (ICA SR and ICA SR 1.1) and the Coupling Express2 Long Reach (CE2 LR). The Parallel Sysplex InfiniBand **is not supported**.
- The 25GbE and 10GbE RoCE Express3, 25GbE and 10 GbE RoCE Express2.1, 25GbE and 10GbE RoCE Express2 provide high-speed memory-to-memory data exchange to a remote CPC by using the Shared Memory Communications over RDMA (SMC-R) protocol for TCP (socket-based) communications.

The RoCE Express3 features can also provide local area network (LAN) connectivity for Linux on IBM Z, and comply with IEEE standards. In addition, RoCE Express features assume several functions of the TCP/IP stack that normally are performed by the PU, which allows significant performance benefits by offloading processing from the operating system.

- ▶ Concurrent I/O upgrade

You can concurrently add I/O features to IBM z16 A02 and IBM z16 AGZ configurations if unused I/O slot positions are available.

- ▶ Concurrent PCIe+ I/O drawer upgrade

Additional PCIe+ I/O drawers can be installed concurrently if free frame slots for the PCIe+ I/O drawers and PCIe fanouts in the CPC drawer are available.

- ▶ Dynamic I/O configuration

Dynamic I/O configuration supports the dynamic addition, removal, or modification of the channel path, control units, and I/O devices without a planned outage.

- ▶ Pluggable optics:

¹ The link speeds do not represent the performance of the link. The performance depends on many factors, including latency through the adapters, cable lengths, and the type of workload.

- The FICON Express32S, FICON Express16S+, OSA Express7S 1.2, OSA Express6S, RoCE Express3, RoCE Express2.1, RoCE Express2, and RoCE Express features include Small Form-Factor Pluggable (SFP) optics². These optics allow each channel to be individually serviced in a fiber optic module failure. The traffic on the other channels on the same feature can continue to flow if a channel requires servicing.
 - The zHyperLink Express feature uses fiber optics cable with MTP³ connector and the cable uses a CXP connection to the adapter. The CXP⁴ optics are provided with the adapter.
- Concurrent I/O card maintenance
- Every I/O card that is plugged in a PCIe+ I/O drawer supports concurrent card replacement during a repair action.

4.2.2 Supported I/O features

The following I/O features are supported on an IBM z16 A02 and IBM z16 AGZ (max. for each individual adapter type):

- Up to 96 FICON Express32S channels
- Up to 96 FICON Express16S+ channels
- Up to 48 OSA-Express7S 1.2 25GbE ports
- Up to 48 OSA-Express7S 1.2 10GbE ports
- Up to 96 OSA-Express7S 1.2 GbE ports
- Up to 96 OSA-Express7S 1.2 1000BASE-T ports
- Up to 48 OSA-Express6S 10GbE ports
- Up to 96 OSA-Express6S GbE ports
- Up to 96 OSA-Express6S 1000BASE-T ports
- Up to 16 25GbE RoCE Express3 features
- Up to 8 25GbE RoCE Express2.1 features
- Up to 8 25GbE RoCE Express2 features
- Up to 8 10GbE RoCE Express3 features
- Up to 8 10GbE RoCE Express2.1 features
- Up to 8 10GbE RoCE Express2 features
- Up to 16 zHyperLink Express features
- Up to 16 zHyperLink Express1.1 features
- Up to 24 ICA SR1.1 and ICA SR features (combined) with up to 96 ports
- Up to 32 CE2 LR features with up to 64 ports

² OSA-Express 1000BASE-T features do not have optics (copper only, RJ45 connectors).

³ Multifiber Termination Push-On.

⁴ For more information, see this web page: <https://cw.infinibandta.org/document/dl/7157>

Notes: Consider the following points:

- ▶ IBM z16 A02 and IBM z16 AGZ support a maximum of 3 PCIe+ I/O drawers
- ▶ The maximum number of coupling CHPIDs on an IBM z16 A02 and IBM z16 AGZ is 384 in a combination of the following (not all combinations are possible; subject to I/O configuration options):
 - Up to 48 ICA SR1.1 and ICA SR ports (24 ICA SR features)
 - Up to 64 CE2 LR ports (32 CE2 LR features)
- ▶ zEDC PCIe features are not supported. These have been replaced by the IBM Integrated Accelerator for zEDC (on PU chip).
- ▶ The maximum combined number of RoCE features that can be installed is eight (16 ports); that is, any combination of 25GbE RoCE Express3, 25GbE RoCE Express2.1, 25GbE RoCE Express2, 10GbE RoCE Express3, 10GbE RoCE Express2.1, 10GbE RoCE Express2 features.
- ▶ 25GbE RoCE Express features should not be configured in the same SMC-R link group with 10GbE RoCE Express features.

4.3 PCIe+ I/O drawer

The PCIe+ I/O drawers (see Figure 4-1) are attached to the CPC drawer through a PCIe cable and use PCIe Gen3 as the infrastructure bus within the drawer. The PCIe Gen3 I/O bus infrastructure data rate is up to 16 GBps.

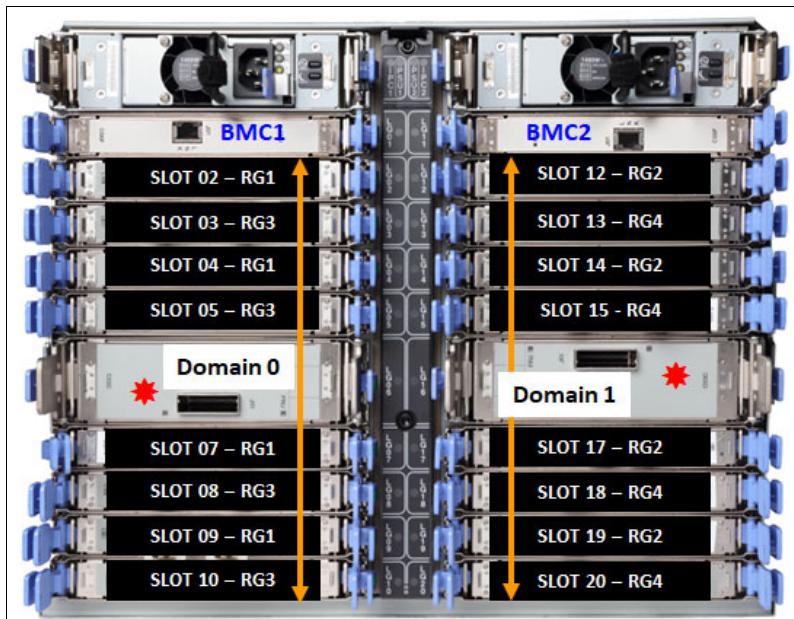


Figure 4-1 Rear view of PCIe+ I/O drawer

PCIe switch application-specific integrated circuits (ASICs) are used to fan out the host bus from the CPC drawer through the PCIe+ I/O drawer to the individual I/O features. Maximum 16 PCIe I/O features (up to 32 channels) per PCIe+ I/O drawer are supported.

The PCIe+ I/O drawer is a one-sided drawer (all I/O cards on one side, in the rear of the drawer) that is 8U high. The PCIe+ I/O drawer contains the 16 I/O slots for PCIe features, two switch cards, and two power supply units (PSUs) to provide redundant power, as shown in Figure 4-1 on page 142.

The PCIe+ I/O drawer slots numbers are shown in Figure 4-2.

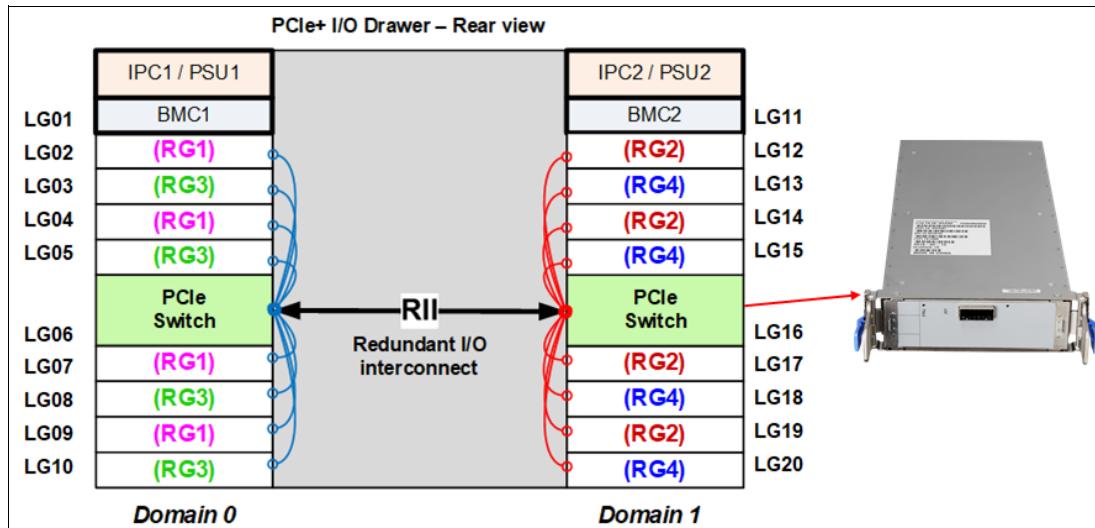


Figure 4-2 PCIe+ I/O drawer slots numbers

The I/O structure in an IBM z16 A02 and IBM z16 AGZ CPCs is shown in Figure 4-3 on page 144. The PCIe switch card provides the fanout from the high-speed x16 PCIe host bus to eight individual card slots. The PCIe switch card is connected to the CPC drawer through a single x16 PCIe Gen3 bus from a PCIe fanout card (PCIe+ fanout cards).

In the PCIe+ I/O drawer, the eight I/O feature cards that directly attach to the switch card constitute an I/O domain. The PCIe+ I/O drawer supports concurrent add and replace I/O features with which you can increase I/O capability as needed, depending on the CPC drawer.

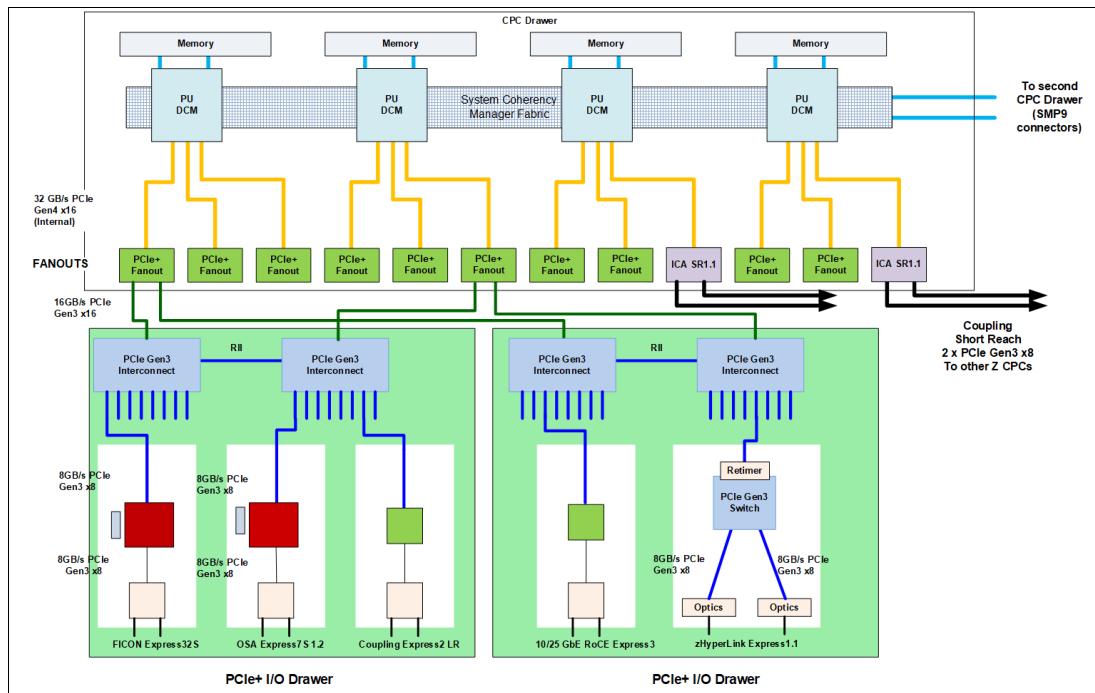


Figure 4-3 IBM z16 A02 and IBM z16 AGZ I/O connectivity - Max32 feature with two PCIe+ I/O drawers

The PCIe slots in a PCIe+ I/O drawer are organized into two I/O domains. Each I/O domain supports up to eight features and is driven through a PCIe switch card. Two PCIe switch cards always provide a backup path for each other through the passive connection in the PCIe+ I/O drawer backplane. During a PCIe fanout card or cable failure, 16 I/O cards in two domains can be driven through a single PCIe switch card. It is not possible to drive 16 I/O cards after one of the PCIe switch cards is removed.

The two switch cards are interconnected through the PCIe+ I/O drawer board (Redundant I/O Interconnect, or RII). In addition, switch cards in same PCIe+ I/O drawer are connected to PCIe fanouts across clusters in CPC drawer for higher availability.

The RII design provides a failover capability during a PCIe fanout card failure. Both domains in one of these PCIe+ I/O drawers are activated with two fanouts. The Base Management Cards (BMCS) are used for system control.

The domains and their related I/O slots are shown in Figure 4-2 on page 143.

Each I/O domain supports up to eight features (FICON, OSA, Crypto, and so on.) All I/O cards connect to the PCIe switch card through the backplane board. The I/O domains and slots are listed in Table 4-1.

Table 4-1 I/O domains of PCIe+ I/O drawer

Domain	I/O slot in the domain
0	LG02, LG03, LG04, LG05, LG07, LG08, LG09, and LG10
1	LG12, LG13, LG14, LG15, LG17, LG18, LG19, and LG20

4.3.1 PCIe+ I/O drawer offering

Up to three PCIe+ I/O drawers can be installed for supporting up to 48 PCIe I/O features.

For an upgrade to IBM z16 A02 and IBM z16 AGZ, only the following PCIe features can be carried forward:

- ▶ FICON Express16S+
- ▶ zHyperLink Express (all features)
- ▶ OSA-Express6S (all features)
- ▶ 25GbE RoCE Express2.1
- ▶ 25GbE RoCE Express2
- ▶ 10GbE RoCE Express2.1
- ▶ 10GbE RoCE Express2
- ▶ Crypto Express7S (one or two ports/HSMs)
- ▶ Crypto Express6S

Note: On an IBM z16 A02 and IBM z16 AGZ system, only PCIe+ I/O drawers are supported. Older generation I/O drawers cannot be carried forward.

IBM z16 A02 and IBM z16 AGZ support the following PCIe I/O new features that are hosted in the PCIe+ I/O drawers:

- ▶ FICON Express32S
- ▶ OSA-Express7S 1.2 25GbE
- ▶ OSA-Express7S 1.2 10GbE
- ▶ OSA-Express7S 1.2 GbE
- ▶ OSA-Express7S 1.2 1000BASE-T
- ▶ 25GbE RoCE Express3
- ▶ 10GbE RoCE Express3
- ▶ Crypto Express8S (one or two HSMs)
- ▶ Coupling Express2 Long Reach (CE2 LR)
- ▶ zHyperLink Express1.1

4.4 CPC drawer fanouts

The IBM z16 A02 and IBM z16 AGZ use PCIe+ Gen3 fanout cards to connect the I/O subsystem in the CPC drawer to the PCIe+ I/O drawers. The fanout cards also include the ICA SR (ICA SR and ICA SR1.1) coupling links for Parallel Sysplex. All fanout cards support concurrent add, remove, and move.

The IBM z16 A02 and IBM z16 AGZ CPC drawer I/O infrastructure consist of the following features:

- ▶ The PCIe+ Generation 3 fanout cards: Two ports per card (feature) that connect to PCIe+ I/O drawers.
- ▶ ICA SR (ICA SR and ICA SR1.1) fanout cards: Two ports per card (feature) that connect to other (external) CPCs.

Note: IBM z16 A02 and IBM z16 AGZ do not support Parallel Sysplex InfiniBand (PSIFB) links.

Also, if and IBM z16 A02 or IBM z16 AGZ is part of a Parallel Sysplex or Coordinated Timing Network, InfiniBand links cannot be used on older IBM zSystems even if installed.

Unless otherwise noted, ICA SR is used for ICA SR and ICA SR1.1 for the rest of the chapter.

The PCIe fanout cards are installed in the rear of the CPC drawers. Each CPC drawer features 12 PCIe+ Gen3 fanout slots.

The PCIe fanouts and ICA SR fanouts are installed in locations LG01 - LG12 at the rear in the CPC drawers (see Figure 2-7 on page 31).

On the CPC drawer there are two BMC/OSC cards, each being a combination of BMC card and OSC card. BMC stands for Base Management and OSC for Oscillator Card. Each BMC/OSC card has one PPS port and one ETS port (RJ45 Ethernet, for both PTP and NTP).

An I/O connection diagram is shown in Figure 4-3 on page 144.

4.4.1 PCIe+ Generation 3 fanout (FC 0175)

The PCIe+ Gen3 fanout card provides connectivity to a PCIe+ I/O drawer by using a copper cable. This PCIe fanout card supports a link rate of 16 GBps (with two links per card).

A 16x PCIe copper cable of 1.5 meters (4.92 feet) to 4.0 meters (13.1 feet) is used for connection to the PCIe switch card in the PCIe+ I/O drawer. PCIe fanout cards are always plugged in pairs and provide redundancy for I/O domains within the PCIe+ I/O drawer.

Note: The PCIe fanout is used exclusively for I/O and cannot be shared for any other purpose.

4.4.2 Integrated Coupling Adapter (FC 0172 and FC 0176)

The IBM ICA SR (FC 0172) is a two-port fanout feature that is used for short distance coupling connectivity and uses channel type CS5. For IBM z16 A02 and IBM z16 AGZ, the new build feature is ICA SR1.1 (FC 0176).

The ICA SR (FC 0172) and ICA SR1.1 (FC 0176) use PCIe Gen3 technology, with x16 lanes that are bifurcated into x8 lanes for coupling.

Both cards are designed to drive distances up to 150 meters (492 feet) with a link data rate of 8 GBps. ICA SR supports up to four channel-path identifiers (CHPIIDs) per port and eight subchannels (devices) per CHPID.

The coupling links can be defined as shared between images (z/OS) within a CSS. They also can be spanned across multiple CSSs in a CPC. For ICA SR features, a maximum four CHPIIDs per port can be defined.

When STP⁵ (FC 1021) is available, ICA SR coupling links can be defined as timing-only links to other IBM z16 A02 and IBM z16 AGZ, IBM z15, IBM z14 ZR1, or IBM z14 M0x systems. The ICA SR cannot be connected to InfiniBand coupling fanouts.

These two fanouts features are housed in the PCIe+ Gen3 I/O fanout slot on the IBM z16 A02 and IBM z16 AGZ CPC drawers. Up to 24 ICA SR and ICA SR1.1 features (up to 48 ports) are supported.

OM3 fiber optic can be used for distances up to 100 meters (328 feet). OM4 fiber optic cables can be used for distances up to 150 meters (492 feet). For more information, see the following manuals:

- ▶ *Planning for Fiber Optic Links*, GA23-1409
- ▶ *3931 Installation Manual for Physical Planning*, GC28-7015

4.4.3 Fanout considerations

Fanout slots in the CPC drawer can be used to plug different fanouts. On IBM z16 A02 and IBM z16 AGZ, the CPC drawers can hold up to 24 PCIe fanout cards for two-CPC drawers configuration, i.e. Max68, and 12 PCIe fanout cards for one CPC drawer, i.e. Max5, Max16 and Max32.

Adapter ID number assignment

PCIe fanouts and ports are identified by an Adapter ID (AID) that is initially dependent on their physical locations, which is unlike channels that are installed in a PCIe+ I/O drawer. Those channels are identified by a physical channel ID (PCHID) number that is related to their physical location. This AID must be used to assign a CHPID to the fanout in the IOCDs definition. The CHPID assignment is done by associating the CHPID to an AID port (see Table 4-2).

Table 4-2 Fanout locations and their AIDs for the CPC drawer (IBM z16 A02 and IBM z16 AGZ)

Fanout locations	CPC0 Location A10 AID (Hex)	CPC1 Location A15 AID (Hex)
LG01	00	0C
LG02	01	0D
LG03	02	0E
LG04	03	0F
LG05	04	10
LG06	05	11
LG07	06	12
LG08	07	13
LG09	08	14
LG10	09	15
LG11	0A	16
LG12	0B	17

⁵ Server Time Protocol

Fanout slots

The fanout slots are numbered LG01 - LG12, from left to right, as listed in Table 4-2 on page 147. All fanout locations and their AIDs for the CPC drawer are shown for reference only.

Important: The AID numbers that are listed in Table 4-2 on page 147 are valid only for a new build system. If a fanout is moved, the AID follows the fanout to its new physical location.

The AID assignment is listed in the PCHID REPORT that is provided for each new server or for an MES upgrade on existing servers. Part of a PCHID REPORT for a z15 T02 is shown in Example 4-1. In this example, four fanout cards are installed at in CPC drawer at location A10B, in slots LG03, LG06, LG07, and LG10 with AIDs 02, 05, 06, and 09.

Example 4-1 AID assignments PCHID REPORT sample

CHPIDSTART						Nov 10,2022
31463036 PCHID REPORT						
Machine: 3932-A02 SN1						
<hr/>						
Source	Drwr	Slot	F/C	PCHID/Ports or AID	Comment	
A10/LG03	A10B	LG03	0176	AID=02		
A10/LG06	A10B	LG06	0176	AID=05		
A10/LG07	A10B	LG07	0176	AID=06		
A10/LG10	A10B	LG10	0176	AID=09		

Fanout features that are supported by the IBM z16 A02 and IBM z16 AGZ are listed in Table 4-3, which includes the feature type, feature code, and information about the link supported by the fanout feature.

Table 4-3 Fanout summary

Fanout feature	Feature code	Use	Cable type	Connector type	Maximum distance	Link data rate ^a
PCIe+ Gen3 fanout	0175	PCIe I/O drawer conn.	Copper	N/A	4 m (13.1 ft.)	16 GBps
ICA SR	0172	Coupling link	OM4	MTP	150 m (492 ft.)	8 Gbps
			OM3	MTP	100 m (328 ft.)	8 Gbps
ICA SR1.1	0176	Coupling link	OM4	MTP	150 m (492 ft.)	8 Gbps
			OM3	MTP	100 m (328 ft.)	8 Gbps

a. The link data rates do not represent the actual performance of the link. The actual performance depends on many factors, including latency through the adapters, cable lengths, and the type of workload.

4.5 I/O features

I/O features (adapters) include ports⁶ to connect the IBM z16 A02 and IBM z16 AGZ to external devices, networks, or other zSystems. I/O features are plugged into the PCIe+ I/O drawer, based on the machine's configuration rules. Different types of I/O cards are available, one for each channel or link type. I/O cards can be installed or replaced concurrently.

⁶ Certain I/O features do not have external ports, such as Crypto Express.

4.5.1 I/O feature card ordering information

The I/O features that are supported by IBM z16 A02 and IBM z16 AGZ configurations and the ordering information for them are listed in Table 4-4.

Table 4-4 I/O features and ordering information

Channel feature	Feature code	New build	Carry-forward
FICON Express32S LX	0461	Y	N/A
FICON Express32S SX	0462	Y	N/A
FICON Express16S+ LX	0427	N	Y
FICON Express16S+ SX	0428	N	Y
OSA-Express7S 1.2 25GbE LR	0460	Y	N/A
OSA-Express7S 1.2 25GbE SR	0459	Y	N/A
OSA-Express7S 1.2 10GbE LR	0456	Y	N/A
OSA-Express7S 1.2 10GbE SR	0457	Y	N/A
OSA-Express7S 1.2 GbE LX	0454	Y	N/A
OSA-Express7S 1.2 GbE SX	0455	Y	N/A
OSA-Express7S 1.2 1000BASE-T	0458	Y	N/A
OSA-Express6S 10GbE LR	0424	N	Y
OSA-Express6S 10GbE SR	0425	N	Y
OSA-Express6S GbE LX	0422	N	Y
OSA-Express6S GbE SX	0423	N	Y
OSA-Express6S 1000BASE-T Ethernet	0426	N	Y
PCIe+ Gen3 fanout ^a	0175	Y	Y
Integrated Coupling Adapter (ICA SR1.1) ^b	0176	Y	Y
Integrated Coupling Adapter (ICA SR) ^b	0172	N	Y
Coupling Express2 LR	0434	Y	N/A
Crypto Express8S (dual HSM)	0908	Y	N/A
Crypto Express8S (single HSM)	0909	Y	N/A
Crypto Express7S (2 ports)	0898	N	Y
Crypto Express7S (1 port)	0899	N	Y
Crypto Express6S	0893	N	Y
25GbE RoCE Express3 SR	0452	Y	N/A
25GbE RoCE Express3 LR	0453	Y	N/A
10GbE RoCE Express3 SR	0440	Y	N/A
10GbE RoCE Express3 LR	0441	Y	N/A
25GbE RoCE Express2.1	0450	N	Y

Channel feature	Feature code	New build	Carry-forward
25GbE RoCE Express2	0430	N	Y
10GbE RoCE Express2.1	0432	N	Y
10GbE RoCE Express2	0412	N	Y
zHyperLink Express1.1	0451	Y	Y
zHyperLink Express	0431	N	Y

- a. Installed in the CPC Drawer; provides connectivity for the PCIe+ I/O Drawer
- b. Installed in the CPC Drawer; provides coupling connectivity (short distance - up to 150m).

Coupling links connectivity support:

- ▶ z13 and z13s and older systems are not supported in same Parallel Sysplex or STP CTN with IBM z16 A02 and IBM z16 AGZ.
- ▶ InfiniBand coupling Links (if available on IBM z14 M0x) are NOT supported in a Parallel Sysplex or CTN for connections to an IBM z16 A02 and IBM z16 AGZ member.

4.5.2 Physical channel ID report

A physical channel ID (PCHID) reflects the physical location of a channel-type interface. A PCHID number is based on the following factors:

- ▶ PCIe+ I/O drawer location
- ▶ Channel feature slot number
- ▶ Port number of the channel feature

A CHPID does not directly correspond to a hardware channel port. Instead, it is assigned to a PCHID in the hardware configuration definition (HCD) or IOCP.

A PCHID REPORT is created for each new build configuration and for upgrades. The report lists all I/O features that are installed, the physical slot location, and the assigned PCHID. A portion of a sample PCHID REPORT is shown in Example 4-2. For more information about the AID numbering rules for coupling links, see Example 4-2

Example 4-2 PCHID REPORT

CHPIDSTART					PCHID REPORT	Nov 10,2021
Machine: 3932-A02 SN1						
Source	Drwr	Slot	F/C	PCHID/Ports or AID	Comment	
A10/LG03	A10B	LG03	0176	AID=02		
A10/LG06	A10B	LG06	0176	AID=05		
A10/LG07	A10B	LG07	0176	AID=06		
A10/LG10	A10B	LG10	0176	AID=09		
A10/LG01/J02	A01B	02	0439	100/D1 101/D2		
A10/LG01/J02	A01B	03	0439	104/D1 105/D2		
A10/LG01/J02	A01B	04	0439	108/D1 109/D2		
A10/LG01/J02	A01B	05	0908	10C/P00 10D/P01		
A10/LG01/J02	A01B	07	0425	110/		
.....<< snippet >>.....						

The PCHID REPORT that is shown in Example 4-2 on page 166 includes the following components (among others):

- ▶ Feature code 0176 (Integrated Coupling Adapters (ICA SR1.1) is installed in the CPC drawer (location A10B, slots LG03, LG06, LG07, and LG10), and have AIDs 02, 05, 06, and 09 assigned.
- ▶ Feature codes 0439 (FICON Express32S+ SX) are installed in PCIe+ I/O drawer 1:
 - Location A01B, slot 02 with PCHIDs 100/D1 and 101/D2 assigned
 - Location A01B, slot 03 with PCHIDs 104/D1 and 105/D2 assigned
 - Location A01B, slot 04 with PCHIDs 108/D1 and 109/D2 assigned
- ▶ Feature code 0908 (Crypto Express8S 2 port) installed in PCIe+ I/O drawer 1 in location A01B, slot 05.
- ▶ Feature code 0457 (OSA-Express7S 10 GbE SR 1.2) installed in PCIe+ I/O drawer 1 in location A01B, slot 07 with PCHID110/D1

A resource group (RG) parameter is also shown in the PCHID REPORT for native PCIe features. A balanced plugging of native PCIe features exists between four resource groups (RG1, RG2, RG3, and RG4).

The preassigned PCHID number of each I/O port relates directly to its physical location (jack location in a specific slot).

4.6 Connectivity

I/O channels are part of the CSS. They provide connectivity for data exchange between systems, between systems and external control units (CUs) and devices, or between networks.

For more information about connectivity to external I/O subsystems (for example, disks), see 4.6.2, “Storage connectivity” on page 155.

For more information about communication to LANs, see 4.6.3, “Network connectivity” on page 162.

Communication between systems is implemented by using CE LR, ICA SR, or channel-to-channel (FICON CTC) connections. For more information, see 4.6.4, “Parallel Sysplex connectivity” on page 175.

4.6.1 I/O feature support and configuration rules

The supported I/O features are listed in Table 4-5. Also listed in Table 4-5 are the number of ports per card, port increments, the maximum number of feature cards, and the maximum number of channels for each feature type. The CHPID definitions that are used in the IOCDs also are listed

Table 4-5 IBM z16 A02 and IBM z16 AGZ A01 supported I/O features

I/O feature	Ports per card	Port increments	Max. ports ^a	Max. I/O slots ^a	PCHID	CHPID definition
Storage access						
FICON Express32S LX/SX	2	2	96	48	Yes	FC, FCP ^b
FICON Express16S+ LX/SX	2	2	96	48	Yes	FC, FCP ^b
zHyperLink Express 1.1	2	2	32	16	Yes	N/A ^c
zHyperLink Express	2	2	32	16	Yes	N/A ^c
OSA-Express features ^d						
OSA-Express7S 1.2 25GbE LR/SR	1	1	48	48	Yes	OSD
OSA-Express7S 1.2 10GbE LR/SR	1	1	48	48	Yes	OSD
OSA-Express7S 1.2 GbE LR/SR	2	2	96	48	Yes	OSC, OSD
OSA-Express7S 1.2 ^e 1000BASE-T	2	2	96	48	Yes	OSC, OSD, OSE
OSA-Express6S 10 GbE LR/SR	1	1	48	48	Yes	OSD
OSA-Express6S GbE LX/SX	2	2	96	48	Yes	OSD
OSA-Express6S 1000BASE-T	2	2	96	48	Yes	OSC, OSD, OSE

I/O feature	Ports per card	Port increments	Max. ports ^a	Max. I/O slots ^a	PCHID	CHPID definition
RoCE Express features						
25GbE RoCE Express3	2	2	16	8	Yes	N/A
10GbE RoCE Express3	2	2	16	8	Yes	N/A
25GbE RoCE Express2.1	2	2	16	8	Yes	N/A
10GbE RoCE Express2.1	2	2	16	8	Yes	N/A
25GbE RoCE Express2	2	2	16	8	Yes	N/A ^c
10GbE RoCE Express2	2	2	16	8	Yes	N/A ^c
Coupling features						
Coupling Express2 LR	2	2	64	32	Yes	CL5
Integrated Coupling Adapter (ICA SR1.1) ^f	2	2	48	24	N/A ^g	CS5
Integrated Coupling Adapter (ICA SR) ^f	2	2	48	24	N/A ^g	CS5

- a. Max. number depends on the feature: Max5, Max16 and Max32 have one CPC drawer and up to three I/O drawers. Max68 has two CPC drawers and up to three I/O drawers.
- b. Both ports must be defined with the same CHPID type.
- c. These features are defined by using Virtual Functions ID (FIDs).
- d. On IBM z16, the OSX and OSM type CHPIDs cannot be defined. IBM z16 cannot be part of an ensemble managed by zManager.
- e. IBM z16 is planned to be the last IBM zSystems generation to support OSA-Express 1000BASE-T adapters
- f. Installed in the CPC Drawer
- g. ICA SR 1.1 and ICA SR features are characterized by Adapter ID (AID).

At least one I/O feature (FICON) or one coupling link feature (ICA SR or CE LR) must be present in the minimum configuration.

The following features can be shared and spanned:

- ▶ FICON channels that are defined as FC or FCP
- ▶ OSA-Express features that are defined as OSC, OSD, or OSE
- ▶ Coupling links that are defined as CS5 or CL5
- ▶ HiperSockets that are defined as IQD

The following features are plugged into a PCIe+ I/O drawer and do not require the definition of a CHPID and CHPID type:

- ▶ Each Crypto Express (8S/7S/6S) feature occupies one I/O slot, but does not include a PCHID type. However, LPARs in all CSSs can access the features. Each Crypto Express adapter can support up to 40 domains.
- ▶ Each 25GbE RoCE Express(3 / 2.1 / 2) feature occupies one I/O slot but does not include a CHPID type. However, LPARs in all CSSs can access the feature. The 25GbE RoCE Express3 can be defined to up to 126 virtual functions (VFs) per feature (port is defined in z/OS Communications Server). The 25GbE RoCE Express3 feature support up to 63 VFs per port (up to 126 VFs per feature).
- ▶ Each 10GbE RoCE Express(3 / 2.1 / 2) feature occupies one I/O slot but does not include a CHPID type. However, LPARs in all CSSs can access the feature. The 10GbE RoCE

Express3 can be defined to up to 126 virtual functions (VFs) per feature (port is defined in z/OS Communications Server). The 10GbE RoCE Express3 feature support up to 63 VFs per port (up to 126 VFs per feature).

- ▶ Each zHyperLink Express/zHyperlink Express1.1 feature occupies one I/O slot but does not include a CHPID type. However, LPARs in all CSSs can access the feature. The zHyperLink Express adapter works as native PCIe adapter and can be shared by multiple LPARs. Each port supports up to 127 Virtual Functions (VFs), with one or more VFs/PFIDs being assigned to each LPAR. This support gives a maximum of 254 VFs per adapter.

I/O feature cables and connectors

The IBM Facilities Cabling Services fiber transport system offers a total cable solution service to help with cable ordering requirements. These services can include the requirements for all of the protocols and media types that are supported (for example, FICON, Coupling Links, and OSA). The services can help whether the focus is the data center, SAN, LAN, or the end-to-end enterprise.

Cables: All fiber optic cables, cable planning, labeling, and installation are client responsibilities for new IBM z16 A02 and IBM z16 AGZ installations and upgrades. Fiber optic conversion kits and mode conditioning patch cables are not orderable as features on IBM z16 A02 and IBM z16 AGZ. All other cables must be sourced separately.

The required connector and cable type for each I/O feature on IBM z16 A02 and IBM z16 AGZ servers are listed in Table 4-6.

Table 4-6 Feature connector and cable types

Feature code	Feature name	Connector type	Cable type
0461	FICON Express32S LX	LC Duplex	9 µm SM
0462	FICON Express32S SX	LC Duplex	50 µm MM ^a
0427	FICON Express16S+ LX 10 Km	LC Duplex	9 µm SM
0428	FICON Express16S+ SX	LC Duplex	50 ^a , 62.5 µm MM
0459	OSA-Express7S 1.2 25 GbE SR	LC Duplex	50 µm MM
0460	OSA-Express7S 1.2 25 GbE LR	LC Duplex	9 µm SM
0456	OSA-Express7S 1.2 10 GbE LR	LC Duplex	9 µm SM
0457	OSA-Express7S 1.2 10 GbE SR	LC Duplex	50 µm MM ^a
0454	OSA-Express7S 1.2 GbE LX	LC Duplex	9 µm SM
0455	OSA-Express7S 1.2 GbE SX	LC Duplex	50, 62.5 µm MM
0458	OSA-Express7S 1.2 1000BASE-T	RJ-45	Category 5 UTP ^b
0424	OSA-Express6S 10GbE LR	LC Duplex	9 µm SM
0425	OSA-Express6S 10 GbE SR	LC Duplex	50, 62.5 µm MM
0422	OSA-Express6S GbE LX	LC Duplex	9 µm SM
0423	OSA-Express6S GbE SX	LC Duplex	50, 62.5 µm MM
0426	OSA-Express6S 1000BASE-T	RJ-45	Category 5 UTP ^b

Feature code	Feature name	Connector type	Cable type
0452	25GbE RoCE Express3 SR	LC Duplex	50, 62.5 µm MM
0453	25GbE RoCE Express3 LR	LC Duplex	9 µm SM
0440	10GbE RoCE Express3 SR	LC Duplex	50, 62.5 µm MM
0441	10GbE RoCE Express3 LR	LC Duplex	9 µm SM
0450	25GbE RoCE Express 2.1	LC Duplex	50 µm MM ^a
0430	25GbE RoCE Express2	LC Duplex	50 µm MM ^a
0412	10GbE RoCE Express2	LC Duplex	50, 62.5 µm MM
0434	CE2 LR	LC Duplex	9 µm SM
0176	Integrated Coupling Adapter SR1.1 (ICA SR1.1)	MTP	50 µm MM OM3/OM4
0172	Integrated Coupling Adapter (ICA SR)	MTP	50 µm MM OM3/OM4
0451	zHyperLink Express1.1	MPO	50 µm MM OM3/OM4
0431	zHyperLink Express	MPO	50 µm MM OM3/OM4

- a. 50 µm core Multi Mode (MM) fiber - OM2, OM3, or OM4 (OM4 is highly recommended)
- b. UTP is unshielded twisted pair. Consider the use of category 6 UTP for 1000 Mbps connections.

4.6.2 Storage connectivity

Connectivity to external I/O subsystems (for example, disks) is provided by FICON channels and zHyperLink⁷.

FICON channels

IBM z16 A02 and IBM z16 AGZ support the following FICON features:

- ▶ FICON Express32S LX and SX (FC 0461/0462)
- ▶ FICON Express16S+ LX and SX (FC 0427/0428)

The FICON Express32S and FICON Express16S+ features conform to the following architectures:

- ▶ Fibre Connection (FICON)
- ▶ High Performance FICON on Z (zHPF)
- ▶ Fibre Channel Protocol (FCP)

The FICON features provide connectivity between any combination of servers, directors, switches, and devices (control units, disks, tapes, and printers) in a SAN.

Each FICON Express feature occupies one I/O slot in the PCIe+ I/O drawer. Each feature includes two ports, each supporting an LC Duplex connector, with one PCHID and one CHPID that is associated with each port.

⁷ zHyperLink feature operates with a FICON channel.

Each FICON Express feature uses SFP (SFP+ for FICON Express32S) optics that allow for concurrent repairing or replacement for each SFP. The data flow on the unaffected channels on the same feature can continue. A problem with one FICON Express port does not require replacement of a complete feature.

Each FICON Express feature also supports cascading, which is the connection of two FICON Directors in succession. This configuration minimizes the number of cross-site connections and helps reduce implementation costs for disaster recovery applications, IBM Geographically Dispersed Parallel Sysplex™ (GDPS), and remote copy.

IBM z16 A02 and IBM z16 AGZ configurations support 32K devices per FICON channel for all FICON features.

Each FICON Express channel can be defined independently for connectivity to servers, switches, directors, disks, tapes, and printers, by using the following CHPID types:

- ▶ CHPID type FC: The FICON, zHPF, and FCTC protocols are supported simultaneously.
- ▶ CHPID type FCP: Fibre Channel Protocol that supports attachment to SCSI devices directly or through Fibre Channel switches or directors.

FICON channels (CHPID type FC or FCP) can be shared among LPARs and defined as spanned. All ports on a FICON feature must be of the same type (LX or SX). The features are connected to a FICON capable control unit (point-to-point or switched point-to-point) through a Fibre Channel switch.

FICON Express32S

The FICON Express32S feature is installed in the PCIe+ I/O drawer. Each of the two independent ports is capable of 8 Gbps, 16Gbps or 32 Gbps. The link speed depends on the capability of the attached switch or device. The link speed is auto-negotiated, point-to-point, and is transparent to users and applications.

The following types of FICON Express32S optical transceivers are supported (no mix on same card):

- ▶ FICON Express32S LX feature, FC 0461, with two ports per feature, LC Duplex connectors
- ▶ FICON Express32S SX feature, FC 0462, with two ports per feature, LC Duplex connectors

For supported distances, see Table 4-6 on page 154.

Consideration: FICON Express32S features do not support auto-negotiation to a data link rate of 2 or 4 Gbps (only 8, 16, or 32 Gbps) for point to point connections. To connect to lower speed devices a compatible switch must be used.

FICON Express16S+

The FICON Express16S+ feature is installed in the PCIe+ I/O drawer. Each of the two independent ports is capable of 4 Gbps, 8 Gbps, or 16 Gbps. The link speed depends on the capability of the attached switch or device. The link speed is auto-negotiated, point-to-point, and is transparent to users and applications.

The following types of FICON Express16S+ optical transceivers are supported (no mix on same card):

- ▶ FICON Express16S+ LX feature, FC 0427, with two ports per feature, supporting LC Duplex connectors
- ▶ FICON Express16S+ SX feature, FC 0428, with two ports per feature, supporting LC Duplex connectors

For more information, see the FICON Express chapter *IBM Z Connectivity Handbook*, SG24-5444.

Consideration: FICON Express16S+ features do not support auto-negotiation to a data link rate of 2 Gbps (only 4, 8, or 16 Gbps). To connect to lower speed devices a compatible switch must be used.

FICON features and built in functions

Together with the FICON Express32S and FICON Express16S+, IBM z16 A02 and IBM z16 AGZ provide enhancements for FICON in functional and performance aspects with IBM Endpoint Security solution.

IBM Fibre Channel Endpoint Security

IBM z16 A02 and IBM z16 AGZ support IBM Fibre Channel Endpoint Security feature (FC 1146). FC 1146 provides FC/FCP link encryption and endpoint authentication. It is an end-to-end solution that helps ensure all data flowing on the Fiber Channel links within and across datacenters flows between trusted entities. This is an optional priced feature which requires the following:

- ▶ FICON Express32S for both link encryption and endpoint authentication
- ▶ FICON Express16S for endpoint authentication

Note: FICON Express16S+ supports endpoint authentication only (no data-in-flight encryption).

- ▶ Select DS8000 storage models and firmware
- ▶ Supporting infrastructure - IBM Security Guardium Key Lifecycle Manager
- ▶ CPACF enablement (FC 3863)

Forward Error Correction

Forward Error Correction (FEC) is a technique that is used for reducing data errors when transmitting over unreliable or noisy communication channels (improving signal to noise ratio). By adding redundancy error-correction code (ECC) to the transmitted information, the receiver can detect and correct several errors without requiring retransmission. This process improves signal reliability and bandwidth use by reducing retransmissions because of bit errors, especially for connections across long distance, such as an inter-switch link (ISL) in a GDPS Metro Mirror environment.

The FICON Express32S and FICON Express16S+ are designed to support FEC coding on top of its 64b/66b data encoding for 16 and 32 Gbps connections. This design can correct up to 11 bit errors per 2112 bits transmitted. Therefore, while connected to devices that support

FEC at 16 Gbps connections, the FEC design allows FICON Express32 and FICON Express16S+ channels to operate at higher speeds, over longer distances, with reduced power and higher throughput while retaining the same reliability and robustness for which FICON channels are traditionally known.

With the IBM DS8870 or newer, IBM z16 A02 and IBM z16 AGZ can extend the use of FEC to the fabric N_Ports for a completed end-to-end coverage of 32 Gbps FC links.

FICON dynamic routing

With the IBM z14 and newer IBM zSystems, FICON channels are no longer restricted to the use of static SAN routing policies for ISLs for cascaded FICON directors. The IBM zSystems now support dynamic routing in the SAN with the FICON Dynamic Routing (FIDR) feature. FDR is designed to support the dynamic routing policies that are provided by the FICON director manufacturers; for example, Brocade's exchange-based routing (EBR) and Cisco's originator exchange ID (OxID)⁸ routing.

A static SAN routing policy normally assigns the ISL routes according to the incoming port and its destination domain (port-based routing), or the source and destination ports pairing (device-based routing).

The port-based routing (PBR) assigns the ISL routes statically that is based on "first come, first served" when a port starts a fabric login (FLOGI) to a destination domain. The ISL is round-robin that is selected for assignment. Therefore, I/O flow from same incoming port to same destination domain always is assigned the same ISL route, regardless of the destination port of each I/O. This setup can result in some ISLs overloaded while some are under-used. The ISL routing table is changed whenever IBM zSystem undergoes a power-on-reset (POR), so the ISL assignment is unpredictable.

Device-based routing (DBR) assigns the ISL routes statically that is based on a hash of the source and destination port. That I/O flow from same incoming port to same destination is assigned to same ISL route. Compared to PBR, the DBR is more capable of spreading the load across ISLs for I/O flow from the same incoming port to different destination ports within a destination domain.

When a static SAN routing policy is used, the FICON director features limited capability to assign ISL routes based on workload. This limitation can result in unbalanced use of ISLs (some might be overloaded, while others are under-used).

The dynamic routing ISL routes are dynamically changed based on the Fibre Channel exchange ID, which is unique for each I/O operation. ISL is assigned at I/O request time, so different I/Os from same incoming port to same destination port are assigned different ISLs.

With FIDR, IBM z16 A02 and IBM z16 AGZ, feature the following advantages for performance and management in configurations with ISL and cascaded FICON directors:

- ▶ Support sharing of ISLs between FICON and FCP (PPRC or distributed)
- ▶ I/O traffic is better balanced between all available ISLs
- ▶ Improved use of FICON director and ISL
- ▶ Easier to manage with a predictable and repeatable I/O performance

FICON dynamic routing can be enabled by defining dynamic routing-capable switches and control units in HCD. Also, z/OS implemented a health check function for FICON dynamic routing.

⁸ Check with the switch provider for their support statement.

Improved zHPF I/O execution at distance

By introducing the concept of pre-deposit writes, zHPF reduces the number of round trips of standard FCP I/Os to a single round trip. Originally, this benefit is limited to writes that are less than 64 KB. zHPF on IBM z14 and newer IBM zSystems were enhanced to allow all large write operations (> 64 KB) at distances up to 100 kilometers to be run in a single round trip to the control unit. This improvement avoids elongating the I/O service time for these write operations at extended distances.

Read Diagnostic Parameter Extended Link Service support

To improve the accuracy of identifying a failed component without unnecessarily replacing components in a SAN fabric, a new Extended Link Service (ELS) command called Read Diagnostic Parameters (RDP) was added to the Fibre Channel T11 standard to allow IBM zSystems to obtain extra diagnostic data from the SFP optics that are throughout the SAN fabric.

IBM z14 and newer IBM zSystems can read this extra diagnostic data for all the ports that are accessed in the I/O configuration and make the data available to an LPAR. For z/OS LPARs that use FICON channels, z/OS displays the data with a [new message and display](#) command. For Linux on IBM Z, z/VM, z/VSE, and LPARs that use FCP channels, this diagnostic data is available in a new window in the SAN Explorer tool.

N_Port ID Virtualization

N_Port ID Virtualization (NPIV) allows multiple system images (in LPARs or z/VM guests) to use a single FCP channel as though each were the sole user of the channel. First introduced with IBM z9® EC, this feature can be used with earlier FICON features that were carried forward from earlier servers.

By using the FICON Express16S (or newer) as an FCP channel with NPIV enabled, the maximum numbers of the following aspects for one FCP physical channel are doubled:

- ▶ Maximum number of NPIV hosts defined: Increased from 32 to 64
- ▶ Maximum number of remote N_Ports communicated: Increased from 512 to 1024
- ▶ Maximum number of addressable LUNs: Increased from 4096 to 8192
- ▶ Concurrent I/O operations: Increased from 764 to 1528

For more information about operating systems that support NPIV, see “[N_Port ID Virtualization](#)” on page 289.

Export and import physical port WWPNs for FCP Channels

IBM zSystems automatically assign worldwide port names (WWPNs) to the physical ports of an FCP channel that is based on the PCHID. This WWPN assignment changes when an FCP channel is moved to a different physical slot position.

IBM z14 and newer IBM zSystems allow for the modification of these default assignments, which also allows FCP channels to keep previously assigned WWPNs, even after being moved to a different slot position. This capability can eliminate the need for reconfiguration of the SAN in many situations, and is especially helpful during a system upgrade (FC 0099 - WWPN Persistence).

FICON support for multiple-hop cascaded SAN configurations

Before the introduction of z13 and z13s, IBM zSystems FICON SAN configurations supported a single ISL (a single hop) in a cascaded FICON SAN environment only.

IBM z14 and newer IBM zSystems support up to three hops in a cascaded FICON SAN environment. This support allows clients to more easily configure a three- or four-site disaster recovery solution.

For more information about the FICON multi-hop, see the [FICON Multihop: Requirements and Configurations white paper](#) at the IBM Techdocs Library website

FICON feature summary

The FICON feature codes, cable type, maximum unrepeated distance, and the link data rate on an IBM z16 A02 and IBM z16 AGZ are listed in Table 4-7. All FICON features use LC Duplex connectors.

Table 4-7 FICON features

Channel feature	Feature codes	Bit rate	Cable type	Maximum unrepeated distance ^a (MHz -km)
FICON Express32S LX ^{bc}	0461	8, 16 or 32 ^d Gbps	SM 9 µm	10 km
FICON Express32S SX ^{cd}	0462	32 Gbps	MM 50 µm	20m (500) 70m (2000) 100m (4700)
		16 Gbps	MM 62.5 µm MM 50 µm	15m (200) 35 m (500) 100 m (2000) 125 m (4700)
		8 Gbps	MM 62.5 µm MM 50 µm	21 m (200) 50 m (500) 150 m (2000) 190 m (4700)
FICON Express16S+ 10km LX	0427	4, 8, or 16 Gbps	SM 9 µm	10 km
FICON Express16S+ SX	0428	16 Gbps	MM 50 µm	35 m (500) 100 m (2000) 125 m (4700)
		8 Gbps	MM 62.5 µm MM 50 µm	21 m (200) 50 m (500) 150 m (2000) 190 m (4700)
		4 Gbps	MM 62.5 µm MM 50 µm	70 m (200) 150 m (500) 380 m (2000) 400 m (4700)

a. Minimum fiber bandwidths in MHz/km for multimode fiber optic links are included in parentheses, where applicable.

b. 2 and 4 Gbps connectivity is not supported for point to point connections

c. 2 and 4 Gbps connectivity is supported through a SAN switch

d. For Single Mode fiber, at 1310nm, link running at 32Gbps are limited for point to point connectivity to 5 km.

zHyperLink Express1.1 (FC 0451)

zHyperLink is a new technology that provides up to 5x reduction in I/O latency times for Db2 read requests with the qualities of service IBM zSystem clients expect from I/O infrastructure for Db2 v11 plus fixes (for read support) and v12 plus fixes (for write support) with z/OS.

The z/OS supported versions for zHyperLink are:

- ▶ z/OS V2.5
- ▶ z/OS V2.4 with PTFs.
- ▶ z/OS V2.3 with PTFs.

The zHyperLink Express1.1 feature (FC 0451) provides a low latency direct connection between IBM z16 A02 and IBM z16 AGZ and DS8000 storage system.

The zHyperLink Express1.1 is the result of new business requirements that demand fast and consistent application response times. It dramatically reduces latency by interconnecting the IBM z16 A02 and IBM z16 AGZ directly to I/O Bay of the DS8K by using PCIe Gen3 x 8 physical link (up to 150-meter [492-foot] distance). A new transport protocol is defined for reading and writing IBM CKD data records⁹, as documented in the zHyperLink interface specification.

On IBM z16 A02 and IBM z16 AGZ, zHyperLink Express1.1 card is a PCIe Gen3 adapter, which installed in the PCIe+ I/O drawer. HCD definition support was added for new PCIe function type with PORT attributes.

Requirements of zHyperLink Express1.1

The zHyperLink Express1.1 feature is available on IBM z16 A02 and IBM z16 AGZ, and includes the following requirements:

- ▶ z/OS 2.3 or later
- ▶ 150 m maximum distance in a point to point configuration
- ▶ Supported DS8000 (see *Getting Started with IBM zHyperLink for z/OS*, REDP-5493)
- ▶ zHyperLink Express1.1 adapter (FC 0451) installed
- ▶ FICON channel as a driver
- ▶ Only ECKD supported
- ▶ z/VM is not supported

Up to 16 zHyperLink Express1.1 adapters can be installed in an IBM z16 A02 or IBM z16 AGZ, up to 32 links).

The zHyperLink Express1.1 is virtualized as a native PCIe adapter and can be shared by multiple LPARs. Each port can support up to 127 Virtual Functions (VFs), with one or more VFs/PFIDs being assigned to each LPAR. This configuration gives a maximum of 254 VFs per adapter. The zHyperLink Express requires the following components:

- ▶ zHyperLink connector on DS8K I/O Bay

For DS8880 firmware R8.3 or newer, the I/O Bay planar is updated to support the zHyperLink interface. This update includes the update of the PEX 8732 switch to PEX8733 that includes a DMA engine for the zHyperLink transfers, and the upgrade from a copper to optical interface by a CXP connector (provided).

- ▶ Cable

The zHyperLink Express1.1 uses optical cable with MTP connector. Maximum supported cable length is 150 meters (492 feet).

zHyperLink Express (FC 0431)

zHyperLink is a new technology that provides up to 5x reduction in I/O latency times for Db2 read requests with the qualities of service IBM zSystem clients expect from I/O infrastructure for Db2 v12 with z/OS. The z/OS supported versions are the same ad for zHyperLink Express 1.1.

⁹ CKD data records are handled by using IBM Enhanced Count Key Data (ECKD™) command set.

The zHyperLink Express feature (FC 0431) provides a low latency direct connection between IBM z16 A02 and IBM z16 AGZ and DS8000 I/O Port.

On IBM z16 A02 and IBM z16 AGZ, zHyperLink Express card is a carry forward PCIe adapter, which installed in the PCIe+ I/O drawer. HCD definition support was added for new PCIe function type with PORT attributes.

Requirements of zHyperLink

The zHyperLink Express feature is available on IBM z16 A02 and IBM z16 AGZ, and includes the following requirements:

- ▶ z/OS 2.2 or later
- ▶ Supported DS8000 (see *Getting Started with IBM zHyperLink for z/OS*, REDP-5493)
- ▶ zHyperLink Express adapter (FC 0431 or FC0451) installed
- ▶ FICON channel as a driver
- ▶ Only ECKD supported
- ▶ z/VM is not supported

Up to 16 zHyperLink Express adapters can be installed in an IBM z16 A02 or IBM z16 AGZ (up to 32 links).

The zHyperLink Express is virtualized as a native PCIe adapter and can be shared by multiple LPARs. Each port can support up to 127 Virtual Functions (VFs), with one or more VFs/PFIDs being assigned to each LPAR. This configuration gives a maximum of 254 VFs per adapter. The zHyperLink Express requires the following components:

- ▶ zHyperLink connector on supported DS8000 I/O Bay.
- ▶ Optic Fiber Cable

The zHyperLink Express and zHyperlink Express1.1 use optical cable with MTP connector. Maximum supported cable length is 150 meters (492 feet).

4.6.3 Network connectivity

Communication for LANs is provided by the OSA-Express7S 1.2, OSA-Express6S, 25GbE and 10GbE RoCE Express3, 25GbE and 10GbE RoCE Express2.1, and 25GbE and 10 GbE RoCE Express2 features.

OSA-Express7S 1.2 25GbE SR (FC 0459)

OSA-Express7S 1.2 25Gigabit Ethernet SR (FC 0459) is installed in the PCIe+ I/O Drawer.

OSA-Express7S 1.2 25Gigabit Ethernet Short Reach (SR) feature includes one PCIe Gen3 adapter and one port per feature. The port supports CHPID types OSD.

The OSA-Express7S 1.2 25GbE SR feature is designed to support attachment to a multimode fiber 25 Gbps Ethernet LAN or Ethernet switch that is capable of 25 Gbps. The port can be defined as a spanned channel and shared among LPARs within and across logical channel subsystems.

The OSA-Express7S 1.2 25GbE SR feature supports the use of an industry standard small form factor (SFP+) LC Duplex connector. Ensure that the attaching or downstream device has an SR transceiver. The sending and receiving transceivers must be the same (SR to SR).

The OSA-Express7S 1.2 25GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 50 μm multimode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

OSA-Express7S 1.2 25GbE LR (FC 0460)

The OSA-Express7S 1.2 25Gigabit Ethernet (GbE) Long Reach (LR) feature includes one PCIe Gen3 adapter and one port per feature. The port supports CHPID types OSD. The 25GbE feature is designed to support attachment to a single-mode fiber 10 Gbps Ethernet LAN or Ethernet switch that is capable of 25Gbps. The port can be defined as a spanned channel and can be shared among LPARs within and across logical channel subsystems.

The OSA-Express7S 1.2 25GbE LR feature supports the use of an industry standard small form factor (SFP+) LC Duplex connector. Ensure that the attaching or downstream device includes an LR transceiver. The transceivers at both ends must be the same (LR to LR).

The OSA-Express7S 1.2 25GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 9 μm single-mode fiber optic cable that ends with an LC Duplex connector is required for connecting this feature to the selected device.

OSA-Express7S 1.2 10GbE LR (FC 0456)

The OSA-Express7S 1.2 10Gigabit Ethernet (GbE) Long Reach (LR) feature includes one PCIe Gen3 adapter and one port per feature. The port supports CHPID types OSD. The 10GbE feature is designed to support attachment to a single-mode fiber 10Gbps Ethernet LAN or Ethernet switch that is capable of 10Gbps. The port can be defined as a spanned channel and can be shared among LPARs within and across logical channel subsystems.

The OSA-Express7S 1.2 10GbE LR feature supports the use of an industry standard small form factor (SFP+) LC Duplex connector. Ensure that the attaching or downstream device includes an LR transceiver. The transceivers at both ends must be the same (LR to LR).

The OSA-Express7S 1.2 10GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 9 μm single-mode fiber optic cable that ends with an LC Duplex connector is required for connecting this feature to the selected device.

OSA-Express7S 1.2 10GbE SR (FC 0457)

The OSA-Express7S 1.2 10GbE Short Reach (SR) feature includes one PCIe Gen3 adapter and one port per feature. The port supports CHPID types OSD. The 10 GbE feature is designed to support attachment to a multimode fiber 10 Gbps Ethernet LAN or Ethernet switch that is capable of 10 Gbps. The port can be defined as a spanned channel and shared among LPARs within and across logical channel subsystems.

The OSA-Express7S 1.2 10GbE SR feature supports the use of an industry standard small form factor (SFP+) LC Duplex connector. Ensure that the attaching or downstream device has an SR transceiver. The sending and receiving transceivers must be the same (SR to SR).

The OSA-Express7S 1.2 10GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 50 or a 62.5 μm multimode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

OSA-Express7S 1.2 GbE LX (FC 0454)

The OSA-Express7S 1.2 GbE LX feature includes one PCIe adapter and two ports. The two ports share a channel path identifier (CHPID type OSD). The ports support attachment to a 1 Gbps Ethernet LAN. Each port can be defined as a spanned channel and can be shared among LPARs and across logical channel subsystems.

The OSA-Express7S 1.2 GbE LX feature supports the use of an LC Duplex connector. Ensure that the attaching or downstream device has an LX transceiver. The sending and receiving transceivers must be the same (LX to LX).

A 9 µm single-mode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device. If multimode fiber optic cables are being reused, a pair of Mode Conditioning Patch cables is required, with one cable for each end of the link.

OSA-Express7S 1.2 GbE SX (FC 0455)

The OSA-Express7S 1.2 GbE SX feature includes one PCIe adapter and two ports. The two ports share a channel path identifier (CHPID type OSD). The ports support attachment to a 1 Gbps Ethernet LAN. Each port can be defined as a spanned channel and shared among LPARs and across logical channel subsystems.

The OSA-Express7S 1.2 GbE SX feature supports the use of an LC Duplex connector. Ensure that the attaching or downstream device has an SX transceiver. The sending and receiving transceivers must be the same (SX to SX).

A multi-mode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

OSA-Express7S 1.2 1000BASE-T (FC 0458)

Feature code 0458 occupies one slot in the PCIe+ I/O drawer. It features two ports that connect to a 1000 Mbps (1 Gbps) Ethernet LAN. Each port has an SFP+ with an RJ-45 receptacle for cabling to an Ethernet switch. The RJ-45 receptacle is required to be attached by using an EIA/TIA Category 5 or Category 6 UTP cable with a maximum length of 100 meters (328 feet). The SFP allows a concurrent repair or replace action.

The OSA-Express7S 1.2 1000BASE-T Ethernet feature does not support auto-negotiation. It supports links at 1000 Mbps in full duplex mode only.

The OSA-Express7S 1.2 1000BASE-T Ethernet feature can be configured as CHPID type OSC, OSD, OSE. Non-QDIO operation mode requires CHPID type OSE.

Note: CHPID types OSM, OSN and OSX are not supported on IBM z16 A02 and IBM z16 AGZ.

OSA-Express6S

The OSA-Express6S feature is installed in the PCIe+ I/O drawer. The following OSA-Express6S features can be installed on IBM z16 A02 and IBM z16 AGZ (when carried forward):

- ▶ OSA-Express6S 10 Gigabit Ethernet LR, FC 0424
- ▶ OSA-Express6S 10 Gigabit Ethernet SR, FC 0425
- ▶ OSA-Express6S Gigabit Ethernet LX, FC 0422
- ▶ OSA-Express6S Gigabit Ethernet SX, FC 0423

- ▶ OSA-Express6S 1000BASE-T Ethernet, FC 0426

The supported OSA-Express6S features are listed in Table 4-8 on page 166.

OSA-Express6S 10 Gigabit Ethernet LR (FC 0424)

The OSA-Express6S 10 Gigabit Ethernet (GbE) Long Reach (LR) feature includes one PCIe adapter and one port per feature. The port supports CHPID types OSD and OSX. The 10 GbE feature is designed to support attachment to a single-mode fiber 10 Gbps Ethernet LAN or Ethernet switch that is capable of 10 Gbps. The port can be defined as a spanned channel and can be shared among LPARs within and across logical channel subsystems.

The OSA-Express6S 10 GbE LR feature supports the use of an industry standard small form factor LC Duplex connector. Ensure that the attaching or downstream device includes an LR transceiver. The transceivers at both ends must be the same (LR to LR).

The OSA-Express6S 10 GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 9 µm single-mode fiber optic cable that ends with an LC Duplex connector is required for connecting this feature to the selected device.

For supported distances, see Table 4-8 on page 166.

OSA-Express6S Gigabit Ethernet SX (FC 0423)

The OSA-Express6S GbE SX feature includes one PCIe adapter and two ports. The two ports share a channel path identifier (CHPID type OSD). The ports support attachment to a 1 Gbps Ethernet LAN. Each port can be defined as a spanned channel and shared among LPARs and across logical channel subsystems.

The OSA-Express6S GbE SX feature supports the use of an LC Duplex connector. Ensure that the attaching or downstream device has an SX transceiver. The sending and receiving transceivers must be the same (SX to SX).

A multi-mode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

For supported distances, see Table 4-8 on page 166.

OSA-Express6S 1000BASE-T Ethernet feature (FC 0426)

Feature code 0426 occupies one slot in the PCIe+ I/O drawer. It features two ports that connect to a 1000 Mbps (1 Gbps) or 100 Mbps Ethernet LAN. Each port has an SFP with an RJ-45 receptacle for cabling to an Ethernet switch. The RJ-45 receptacle is required to be attached by using an EIA/TIA Category 5 or Category 6 UTP cable with a maximum length of 100 meters (328 feet). The SFP allows a concurrent repair or replace action.

The OSA-Express6S 1000BASE-T Ethernet feature supports auto-negotiation when attached to an Ethernet router or switch. If you allow the LAN speed and duplex mode to default to auto-negotiation, the OSA-Express port and the attached router or switch auto-negotiate the LAN speed and duplex mode settings between them. They then connect at the highest common performance speed and duplex mode of interoperation. If the attached Ethernet router or switch does not support auto-negotiation, the OSA-Express port examines the signal that it is receiving and connects at the speed and duplex mode of the device at the other end of the cable.

The OSA-Express6S 1000BASE-T Ethernet feature can be configured as CHPID type OSC, OSD, OSE, or OSM. Non-QDIO operation mode requires CHPID type OSE.

Note: CHPID types OSM, OSN and OSX are not supported on IBM z16 A02 and IBM z16 AGZ.

The following settings are supported on the OSA-Express6S 1000BASE-T Ethernet feature port:

- ▶ Auto-negotiate
- ▶ 100 Mbps half-duplex or full-duplex
- ▶ 1000 Mbps full-duplex

If auto-negotiate is not used, the OSA-Express port attempts to join the LAN at the specified speed and duplex mode. If this specified speed and duplex mode do not match the speed and duplex mode of the signal on the cable, the OSA-Express port does not connect.

For supported distances, see Table 4-8.

OSA-Express features summary

The OSA-Express feature codes, cable type, maximum unrepeated distance, and the link rate on an IBM z16 A02 and IBM z16 AGZ are listed in Table 4-8.

Table 4-8 OSA features

Channel feature	Feature code	Bit rate in Gbps	Cable type	Maximum unrepeated distance ^a (MHz - km)
OSA-Express7S 1.2 25GbE SR	0459	25	MM 50 µm	70 m (2000) 100 m (4700)
OSA-Express7S 1.2 25GbE LR	0460	25	SM 9 µm	10 km (6.8 miles)
OSA-Express7S 1.2 10GbE LR	0456	10	SM 9 µm	10 km (6.8 miles)
OSA-Express7S 1.2 10GbE SR	0457	10	MM 62.5 µm MM 50 µm	33 m (200) 82 m (500) 300 m (2000)
OSA-Express7S 1.2 GbE LX	0454	1.25	SM 9 µm	5 km (3.1 miles)
OSA-Express7S 1.2 GbE SX	0455	1.25	MM 62.5 µm MM 50 µm	275 m (200) 550 m (500)
OSA-Express7S 1.2 1000BASE-T	0458	1000Mbps	Cat 5, Cat 6 unshielded twisted pair (UTP)	100 m
OSA-Express6S 10GbE LR	0424	10	SM 9 µm	10 km (6.8 miles)
OSA-Express6S 10GbE SR	0425	10	MM 62.5 µm MM 50 µm	33 m (200) 82 m (500) 300 m (2000)
OSA-Express6S GbE LX	0422	1.25	SM 9 µm	5 km (3.1 miles)

Channel feature	Feature code	Bit rate in Gbps	Cable type	Maximum unrepeated distance ^a (MHz - km)
OSA-Express6S GbE SX	0423	1.25	MM 62.5 µm MM 50 µm	275 m (200) 550 m (500)
OSA-Express6S 1000BASE-T	0426	100 or 1000 Mbps	Cat 5, Cat 6 unshielded twisted pair (UTP)	100 m

a. Minimum fiber bandwidths in MHz/km for multimode fiber optic links are included in parentheses, where applicable

25GbE RoCE Express3 SR (FC 0452)

25GbE RoCE Express3 SR (FC 0452) is installed in the PCIe+ I/O drawer and is supported only on IBM z16 A02 and IBM z16 AGZ. The 25GbE RoCE Express3 SR is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD). The maximum supported unrepeated distance, point-to-point, is 100 meters (328 feet). A client-supplied cable is required. Two types of cables can be used for connecting the port to the selected 25GbE switch or to another 25GbE RoCE Express3 SR feature:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector, which supports 70 meters (229 feet)
- ▶ OM4 50-micron multimode fiber optic cable that is rated at 4700 MHz-km that ends with an LC Duplex connector, which supports 100 meters (328 feet)

The virtualization capabilities for IBM z16 A02 and IBM z16 AGZ are 63 Virtual Functions per port (126 VFs per feature/PCHID). One RoCE Express feature can be shared by up to 126 partitions (LPARs) (one adapter is one PCHID). The 25GbE RoCE Express3 feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector.

25GbE RoCE requirements:

- ▶ The 25GbE RoCE Express3 SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.
- ▶ 25GbE/10GbE RoCE features should not be mixed in a z/OS SMC-R Link Group.

Both point-to-point connections and switched connections with an enterprise-class 25GbE switch are supported.

25GbE RoCE Express3 LR (FC 0453)

25GbE RoCE Express3 LR (FC 0453) is installed in the PCIe+ I/O drawer and is supported on IBM z16 A02 and IBM z16 AGZ. The 25GbE RoCE Express3 LR is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

The maximum supported unrepeated distance, point-to-point, is 100 meters (328 feet). A client-supplied cable is required. Two types of cables can be used for connecting the port to the selected 25GbE switch or to another 25GbE RoCE Express3 LR feature:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector, which supports 70 meters (229 feet)

- ▶ OM4 50-micron multimode fiber optic cable that is rated at 4700 MHz-km that ends with an LC Duplex connector, which supports 100 meters (328 feet)

The virtualization capabilities for IBM z16 A02 and IBM z16 AGZ are 63 Virtual Functions per port (126 VFs per feature/PCHID). One RoCE Express feature can be shared by up to 126 partitions (LPARs) (one adapter is one PCHID). The 25GbE RoCE Express3 feature uses LR optics and supports the use of a single mode fiber optic cable that ends with an LC Duplex connector.

25GbE RoCE requirements:

- ▶ The 25GbE RoCE Express3 LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.
- ▶ 25GbE/10GbE RoCE features should not be mixed in a z/OS SMC-R Link Group.

Both point-to-point connections and switched connections with an enterprise-class 25GbE switch are supported.

10GbE RoCE Express3 SR (FC 0440)

10GbE RoCE Express3 SR(FC 0440) is installed in the PCIe+ I/O drawer and is supported only on IBM z16 A02 and IBM z16 AGZ. The 10GbE RoCE Express3 SR is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

The maximum supported unrepeated distance, point-to-point, is 100 meters (328 feet). A client-supplied cable is required. Two types of cables can be used for connecting the port to the selected 10GbE switch or to another 10GbE RoCE Express3 SR feature:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector, which supports 70 meters (229 feet)
- ▶ OM4 50-micron multimode fiber optic cable that is rated at 4700 MHz-km that ends with an LC Duplex connector, which supports 100 meters (328 feet)

The virtualization capabilities for IBM z16 A02 and IBM z16 AGZ are 63 Virtual Functions per port (126 VFs per feature/PCHID). One RoCE Express feature can be shared by up to 126 partitions (LPARs) (one adapter is one PCHID). The 10GbE RoCE Express3 feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector.

10 GbE RoCE Express3 requirements:

- ▶ The 10GbE RoCE Express3 SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.
- ▶ 25GbE/10GbE RoCE features should not be mixed in a z/OS SMC-R Link Group.

Both point-to-point connections and switched connections with an enterprise-class 10GbE switch are supported.

10GbE RoCE Express3 LR (FC 0441)

10GbE RoCE Express3 LR(FC 0441) is installed in the PCIe+ I/O drawer and is supported only on IBM z16 A02 and IBM z16 AGZ. The 10GbE RoCE Express3 LR is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

The maximum supported unrepeated distance, point-to-point, is 100 meters (328 feet). A client-supplied cable is required. Two types of cables can be used for connecting the port to the selected 10GbE switch or to another 10GbE RoCE Express3 LR feature:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector, which supports 70 meters (229 feet)
- ▶ OM4 50-micron multimode fiber optic cable that is rated at 4700 MHz-km that ends with an LC Duplex connector, which supports 100 meters (328 feet)

The virtualization capabilities for IBM z16 A02 and IBM z16 AGZ are 63 Virtual Functions per port (126 VFs per feature/PCHID). One RoCE Express feature can be shared by up to 126 partitions (LPARs) (one adapter is one PCHID). The 10GbE RoCE Express3 feature uses LR optics and supports the use of a single mode fiber optic cable that ends with an LC Duplex connector.

10 GbE RoCE Express3 requirements:

- ▶ The 10GbE RoCE Express3 LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.
- ▶ 25GbE/10GbE RoCE features should not be mixed in a z/OS SMC-R Link Group.

Both point-to-point connections and switched connections with an enterprise-class 10GbE switch are supported.

25GbE RoCE Express2.1 (FC 0450)

25GbE RoCE Express2.1 (FC 0450) is installed in the PCIe+ I/O drawer and is supported on IBM z16 A02 and IBM z16 AGZ configurations when carried forward. The 25GbE RoCE Express2.1 is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

Switch configuration for 25 GbE RoCE Express2.1: The switches must meet the following requirements:

- ▶ Global Pause function enabled
- ▶ Priority flow control (PFC) disabled

The maximum supported unrepeated distance, point-to-point, is 100 meters (328 feet). A client-supplied cable is required. Two types of cables can be used for connecting the port to the selected 25GbE switch or to the 25GbE RoCE Express2.1 feature on the attached server:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector, which supports 70 meters (229 feet).
- ▶ OM4 50-micron multimode fiber optic cable that is rated at 4700 MHz-km that ends with an LC Duplex connector, which supports 100 meters (328 feet).

The virtualization capabilities for IBM z16 A02 and IBM z16 AGZ are 63 Virtual Functions per port (126 VFs per feature/PCHID). One RoCE Express feature can be shared by up to 126 partitions (LPARs) (one adapter is one PCHID). The 25GbE RoCE Express2.1 feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector.

25 GbE RoCE Express2.1 requirements:

- ▶ The 25GbE RoCE Express2.1 feature does not support auto-negotiation to any other speed and runs in full duplex mode only.
- ▶ 25GbE and 10GbE RoCE features should not be mixed in a z/OS SMC-R Link Group.

Both point-to-point connections and switched connections with an enterprise-class switch are supported (ports running at matching speeds).

10GbE RoCE Express2.1 (FC 0432)

10GbE RoCE Express2.1 (FC 0432) is installed in the PCIe+ I/O drawer and is supported on IBM z16 A02 and IBM z16 AGZ configurations (carry forward). The 10GbE RoCE Express2.1 is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP FUNCTION statement or in the hardware configuration definition (HCD).

Switch configuration for 10GbE RoCE Express2.1: The 10GbE switches must meet the following requirements:

- ▶ Global Pause function enabled
- ▶ Priority flow control (PFC) disabled

The maximum supported unrepeatable distance, point-to-point, is 100 meters (328 feet). A client-supplied cable is required. Two types of cables can be used for connecting the port to the selected 10GbE switch or to another 10GbE RoCE Express2 feature:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector, which supports 70 meters (229 feet)
- ▶ OM4 50-micron multimode fiber optic cable that is rated at 4700 MHz-km that ends with an LC Duplex connector, which supports 100 meters (328 feet)

The virtualization capabilities for IBM z16 A02 and IBM z16 AGZ are 63 Virtual Functions per port (126 VFs per feature/PCHID). One RoCE Express feature can be shared by up to 126 partitions (LPARs) (one adapter is one PCHID). The 10GbE RoCE Express2.1 feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector.

10 GbE RoCE Express2.1 requirements:

- ▶ The 10GbE RoCE Express2.1 feature does not support auto-negotiation to any other speed and runs in full duplex mode only.
- ▶ 25GbE and 10GbE RoCE features should not be mixed in a z/OS SMC-R Link Group.

Both point-to-point connections and switched connections with an enterprise-class 10GbE switch are supported.

25GbE RoCE Express2 (FC 0430)

25GbE RoCE Express2 (FC 0430) is installed in the PCIe I/O drawer and is supported on IBM z16 A02 and IBM z16 AGZ configurations (carry forward). The 25GbE RoCE Express2 is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP FUNCTION statement or in the hardware configuration definition (HCD).

Switch configuration for RoCE Express2: If the IBM 25GbE RoCE Express2 features are connected to 25GbE switches, the switches must meet the following requirements:

- ▶ Global Pause function enabled
- ▶ Priority flow control (PFC) disabled.

The maximum supported unrepeated distance, point-to-point, is 300 meters (984 feet). A client-supplied cable is required. The following types of cables can be used for connecting the port to the selected 10 GbE switch or to the 10GbE RoCE Express2 feature on the attached server:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector; supports 300 meters (984 feet)
- ▶ OM2 50-micron multimode fiber optic cable that is rated at 500 MHz-km that ends with an LC Duplex connector; supports 82 meters (269 feet)
- ▶ OM1 62.5-micron multimode fiber optic cable that is rated at 200 MHz-km that ends with an LC Duplex connector; supports 33 meters (108 feet)

The virtualization capabilities for IBM z16 A02 and IBM z16 AGZ are 63 Virtual Functions per port (126 VFs per feature/PCHID). One RoCE Express feature can be shared by up to 126 partitions (LPARs) (one adapter is one PCHID). The 25GbE RoCE Express2.1 feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector.

25 GbE RoCE Express2 requirements:

- ▶ The 25GbE RoCE Express2 feature does not support auto-negotiation to any other speed and runs in full duplex mode only.
- ▶ 25GbE and 10GbE RoCE features should not be mixed in a z/OS SMC-R Link Group.

Both point-to-point connections and switched connections with an enterprise-class switch are supported (ports running at matching speeds).

10GbE RoCE Express2 (FC 412)

10GbE RoCE Express2 (FC 0412) is installed in the PCIe I/O drawer and is supported on IBM z16 A02 and IBM z16 AGZ configurations as carry forward. The 10GbE RoCE Express2 is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP FUNCTION statement or in the hardware configuration definition (HCD).

Switch configuration for RoCE Express2: The switches must meet the following requirements:

- ▶ Global Pause function enabled
- ▶ Priority flow control (PFC) disabled.

The maximum supported unrepeated distance, point-to-point, is 300 meters (984 feet). A client-supplied cable is required. The following types of cables can be used for connecting the port to the selected 10 GbE switch or to another 10GbE RoCE Express2 feature:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector; supports 300 meters (984 feet)
- ▶ OM2 50-micron multimode fiber optic cable that is rated at 500 MHz-km that ends with an LC Duplex connector; supports 82 meters (269 feet)

- ▶ OM1 62.5-micron multimode fiber optic cable that is rated at 200 MHz-km that ends with an LC Duplex connector; supports 33 meters (108 feet)

The virtualization capabilities for IBM z16 A02 and IBM z16 AGZ are 63 Virtual Functions per port (126 VFs per feature/PCHID). One RoCE Express feature can be shared by up to 126 partitions (LPARs) (one adapter is one PCHID). The 10GbE RoCE Express2 feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector.

10GbE RoCE Express2 requirements:

- ▶ The 10GbE RoCE Express2 feature does not support auto-negotiation to any other speed and runs in full duplex mode only.
- ▶ 25GbE and 10GbE RoCE features should not be mixed in a z/OS SMC-R Link Group.

Both point-to-point connections and switched connections with an enterprise-class switch are supported (ports running at matching speeds).

Shared Memory Communications functions

The Shared Memory Communication (SMC) capabilities of the IBM z16 A02 and IBM z16 AGZ help optimize the communications between applications for server-to-server (SMC-R) or LPAR-to-LPAR (SMC-D) connectivity.

Shared Memory Communications Version 1

SMC-R

SMC-R provides application transparent use of the RoCE-Express feature. This feature reduces the network overhead and latency of data transfers, which effectively offers the benefits of optimized network performance across processors.

SMC-D

SMC-D was used with the introduction of the Internal Shared Memory (ISM) virtual PCI function. ISM is a virtual PCI network adapter that enables direct access to shared virtual memory, which provides a highly optimized network interconnect for IBM zSystem intra-CPC communications.

SMC-D maintains the socket-API transparency aspect of SMC-R so that applications that use TCP/IP communications can benefit immediately without requiring any application software or IP topology changes. SMC-D completes the overall SMC solution, which provides synergy with SMC-R.

SMC-R and SMC-D use shared memory architectural concepts, which eliminates the TCP/IP processing in the data path, yet preserves TCP/IP Qualities of Service for connection management purposes.

Internal Shared Memory (ISM)

ISM is a function that is supported by IBM z16 A02 and IBM z16 AGZ, IBM z15, and IBM z14 systems. It is the firmware that provides connectivity by using shared memory access between multiple operating system images within the same CPC. ISM creates virtual adapters with shared memory that is allocated for each OS image.

ISM is defined by the FUNCTION statement with a virtual CHPID (VCHID) in hardware configuration definition (HCD)/IOCDS. Identified by the PNETID parameter, each ISM VCHID defines an isolated, internal virtual network for SMC-D communication, without any hardware component required. Virtual adapters are defined by virtual function (VF) statements. Multiple

LPARs can access the same virtual network for SMC-D data exchange by associating their VF with same VCHID.

Applications that use HiperSockets can realize network latency and CPU reduction benefits and performance improvement by using the SMC-D over ISM.

IBM z16 A02 and IBM z16 AGZ support up to 32 ISM VCHIDs per CPC. Each VCHID supports up to 255 VFs, with a total maximum of 8,000 VFs.

Shared Memory Communications Version 2

Shared Memory Communications v2 is available in z/OS V2R4 (with PTFs) and z/OS V2R5.

The initial version of SMC was limited to TCP/IP connections over the same layer 2 network and therefore was not routable across multiple IP subnets. The associated TCP/IP connection was limited to hosts within a single IP subnet requiring the hosts to have direct access to the same physical layer 2 network (i.e. same Ethernet LAN over a single VLAN ID). The scope of eligible TCP/IP connections for SMC was limited to and defined by the single IP subnet.

SMC Version 2 (SMCv2) provides support for SMC over multiple IP subnets for both SMC-D and SMC-R and is referred to as SMC-Dv2 and SMC-Rv2. SMCv2 requires updates to the underlying network technology. SMC-Dv2 requires ISMv2 and SMC-Rv2 requires RoCEv2.

The SMCv2 protocol is downward compatible allowing SMCv2 hosts to continue to communicate with SMCv1 down-level hosts.

While SMCv2 changes the SMC connection protocol enabling multiple IP subnet support, SMCv2 does not change how actual user TCP socket data is transferred, which preserves the benefits of SMC to TCP workloads.

TCP/IP connections that require IPSec are not eligible for any form of SMC.

HiperSockets

The HiperSockets function of IBM z16 A02 and IBM z16 AGZ provides up to 32 high-speed virtual LAN attachments.

HiperSockets can be customized to accommodate varying traffic sizes. Because HiperSockets does not use an external network, it can free up system and network resources. This advantage can help eliminate attachment costs and improve availability and performance.

HiperSockets eliminates the need to use I/O subsystem operations and traverse an external network connection to communicate between LPARs in the same IBM z16 A02 or IBM z16 AGZ CPC. HiperSockets offers significant value in server consolidation when connecting many virtual servers. It can be used instead of certain coupling link configurations in a Parallel Sysplex.

HiperSockets internal networks support the following transport modes:

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

Traffic can be IPv4 or IPv6, or non-IP, such as AppleTalk, DECnet, IPX, NetBIOS, or SNA.

HiperSockets devices are protocol-independent and Layer 3-independent. Each HiperSockets device (Layer 2 and Layer 3 mode) features its own Media Access Control (MAC) address. This address allows 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 helps facilitate server consolidation, and can reduce complexity and simplify network configuration. It also allows LAN administrators to maintain the mainframe network environment similarly to non-mainframe environments.

Packet forwarding decisions are based on Layer 2 information instead of Layer 3. The HiperSockets device can run automatic MAC address generation to create uniqueness within and across LPARs and servers. The use of Group MAC addresses for multicast is supported, and broadcasts to all other Layer 2 devices on the same HiperSockets networks.

Datagrams are delivered only between HiperSockets devices that use the same transport mode. A Layer 2 device cannot communicate directly to a Layer 3 device in another LPAR network. A HiperSockets device can filter inbound datagrams by VLAN identification, the destination MAC address, or both.

Analogous to the Layer 3 functions, HiperSockets Layer 2 devices can be configured as primary or secondary connectors, or multicast routers. This configuration enables the creation of high-performance and high-availability link layer switches between the internal HiperSockets network and an external Ethernet network. It also can be used to connect to the HiperSockets Layer 2 networks of different servers.

HiperSockets Layer 2 is supported by Linux on IBM zSystems, and by z/VM for Linux guest use.

IBM z16 A02 and IBM z16 AGZ support the HiperSockets Completion Queue function that is designed to allow HiperSockets to transfer data synchronously (if possible) and asynchronously, if necessary. This feature combines ultra-low latency with more tolerance for traffic peaks.

With the asynchronous support, data can be temporarily held until the receiver has buffers that are available in its inbound queue during high volume situations. The HiperSockets Completion Queue function requires the following **minimum** Operating Systems support¹⁰:

- ▶ z/OS V2.2 with PTFs
- ▶ Linux on IBM Z distributions:
 - Red Hat Enterprise Linux (RHEL) 6.2
 - SUSE Linux Enterprise Server (SLES) 11 SP2
 - Ubuntu server 16.04 LTS
- ▶ z/VSE V6.2
- ▶ z/VM V6.4¹¹ with maintenance

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

- ▶ Other guest virtual machines on the virtual switch

External network hosts through the virtual switch OSA UPLINK port

¹⁰ Minimum OS support for IBM z16 A02 and IBM z16 AGZ can differ. For more information, see Chapter 7, “Operating system support” on page 241.

¹¹ z/VM V6 is not supported on IBM z16 A02 and IBM z16 AGZ. z/VM V7.2 or newer is needed.

RoCE Express features summary

The RoCE Express feature codes, cable type, maximum unrepeated distance, and the link rate on an IBM z16 A02 and IBM z16 AGZ are listed in Table 4-9.

Table 4-9 RoCE Express features summary

Channel feature	Feature code	Bit rate in Gbps	Cable type	Maximum unrepeated distance ^a (MHz - km)
25GbE RoCE Express3 SR	0452	25	MM 50 µm	70 m (2000) 100 m (4700)
25GbE RoCE Express3 LR	0453	25	SM 9 µm	10 km
10GbE RoCE Express3 SR	0440	10	MM 50 µm	70 m (2000) 100 m (4700)
10GbE RoCE Express3 LR	0441	10	SM 9 µm	10 km
25GbE RoCE Express2.1	0450	25	MM 50 µm	70 m (2000) 100 m (4700)
10GbE RoCE Express2.1	0432	10	MM 62.5 µm MM 50 µm	33 m (200) 82 m (500) 300 m (2000)
25GbE RoCE Express2	0430	25	MM 50 µm	70 m (2000) 100 m (4700)
10GbE RoCE Express2	0412	10	MM 62.5 µm MM 50 µm	33 m (200) 82 m (500) 300 m (2000)

a. Minimum fiber bandwidths in MHz/km for multimode fiber optic links are included in parentheses, where applicable

4.6.4 Parallel Sysplex connectivity

Coupling links are required in a Parallel Sysplex configuration to provide connectivity from the z/OS images to the coupling facility (CF). A properly configured Parallel Sysplex provides a highly reliable, redundant, and robust IBM zSystems technology solution to achieve near-continuous availability. A Parallel Sysplex is composed of one or more z/OS operating system images that are coupled through one or more CFs.

This section describes coupling link features supported in a Parallel Sysplex in which an IBM z16 A02 or IBM z16 AGZ may participate.

Coupling links

The type of coupling link that is used to connect a CF to an operating system LPAR is important. The link performance significantly affects response times and coupling processor usage. For configurations that extend over large distances, the time that is spent on the link can be the largest part of the response time.

IBM z16 A02, IBM z16 AGZ, IBM z15 and IBM z14¹² support three coupling link types:

¹² IBM z14 M0x (M/T 3906) also supports Infiniband coupling links, however these are not supported on IBM z16 A02 and IBM z16 AGZ, IBM z15, and IBM z14 ZR1. Careful connectivity planning is needed if InfiniBand links are present in the supported systems.

- ▶ Integrated Coupling Adapter Short Reach (ICA SR1.1 and ICA SR) links connect directly to the CPC drawer and are intended for short distances between CPCs of up to 150 meters.
- ▶ Coupling Express2 Long Reach (CE2 LR) adapters for IBM z16 A02 and IBM z16 AGZ and Coupling Express Long Reach (CE LR) are located in the PCIe+ drawer and support unrepeated distances of up to 10 km or up to 100 km over qualified WDM services.
- ▶ Internal Coupling (IC) links are for internal links within a CPC.

Note: Parallel Sysplex supports connectivity between systems that differ by up to two generations (n-2). For example, an IBM z16 A02 and IBM z16 AGZ can participate in an IBM Parallel Sysplex cluster with IBM z15, and IBM z14 systems.

Only Integrated Coupling Adapter Short Reach (ICA SR) and Coupling Express2 Long Reach (CE2 LR) features are supported on IBM z16 A02 and IBM z16 AGZ.

Figure 4-4 shows the supported Coupling Link connections for the IBM z16 A02¹³. Only ICA SR and CE LR links are supported on IBM z16 A02 and IBM z16 AGZ, IBM z15, and IBM z14 ZR1 systems.

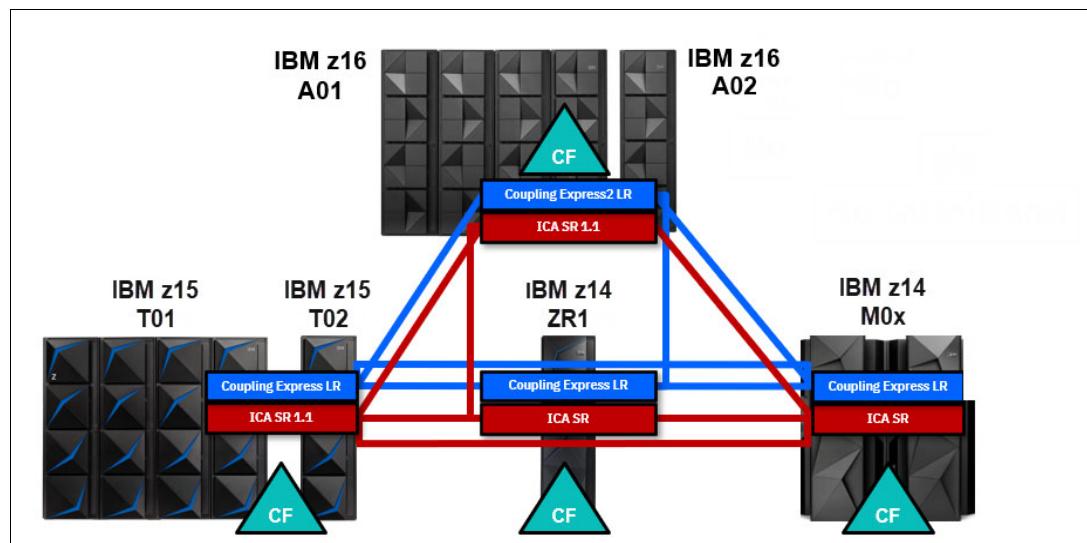


Figure 4-4 Parallel Sysplex connectivity options

The coupling link options that are listed in Table 4-10. Also listed are the coupling link support for each IBM zSystems platform. Restrictions on the maximum numbers can apply, depending on the configuration. Always check with your IBM support team for more information.

¹³ Connections shown for IBM z16 A02 are also valid for the IBM z16 AGZ.

Table 4-10 Coupling link options that are supported on IBM z16 A02 and IBM z16 AGZ

Type	Description	Feature Code	Link rate	Max unrepeated distance	Maximum number of supported links					
					IBM z16 A02 and IBM z16 AGZ A01	IBM z16 A02 and IBM z16 AGZ A02 ^a	IBM z15 T01	IBM z15 T02	IBM z14 ZR1	IBM z14 M0x
CE2 LR	Coupling Express2 LR	0434	10 Gbps	10 km (6.2 miles)	64	64	N/A	N/A	N/A	N/A
CE LR	Coupling Express LR	0433	10 Gbps	10 km (6.2 miles)	N/A	N/A	64	64	32	64
ICA SR 1.1	Integrated Coupling Adapter	0176	8 GBps	150 meters (492 feet)	48	48	96	48	N/A	N/A
ICA SR	Integrated Coupling Adapter	0172	8 GBps	150 meters (492 feet)	48	48	96	48	16	80
IC	Internal Coupling	N/A	Internal speeds	N/A	64	64	64	64	32	32

a. IBM z16 AGZ supports the same features as IBM z16 A02.

The maximum number of combined external coupling links (active CE LR, ICA SR links) is 106 per IBM z16 A02 or IBM z16 AGZ system. The IBM z16 A02 and IBM z16 AGZ coupling link support summary is shown in Table 4-10. Consider the following points:

- ▶ The maximum supported links depends on the IBM zSystem model or capacity feature code.
- ▶ IBM z16 A02 and IBM z16 AGZ ICA SR maximum depends on the number of CPC drawers. A total of 12 PCIe+ fanouts are used per CPU drawer, which gives a maximum of 24 ICA SR ports.

For more information about distance support for coupling links, see *System z End-to-End Extended Distance Guide*, SG24-8047.

Internal Coupling link

IC links are Licensed Internal Code-defined links to connect a CF to a z/OS logical partition in the same CPC. These links are available on all IBM zSystem platforms. The IC link is an IBM zSystems coupling connectivity option that enables high-speed, efficient communication between a CF partition and one or more z/OS logical partitions that are running on the same CPC. The IC is a linkless connection (implemented in LIC) and does not require any hardware or cabling.

An IC link is a fast coupling link that uses memory-to-memory data transfers. IC links do not have PCHID numbers, but do require CHPIDs.

IC links have the following attributes:

- ▶ They provide the fastest connectivity that is significantly faster than external link alternatives.
- ▶ They result in better coupling efficiency than with external links, effectively reducing the CPU cost that is associated with Parallel Sysplex.
- ▶ They can be used in test or production configurations, reduce the cost of moving into Parallel Sysplex technology, and enhance performance and reliability.
- ▶ They can be defined as spanned channels across multiple channel subsystems.
- ▶ They are available at no extra hardware cost (no feature code). Employing ICFs with IC links results in considerable cost savings when configuring a cluster.

IC links are enabled by defining CHPID type ICP. A maximum of 64 IC links can be defined on an IBM z16 A02 or IBM z16 AGZ.

Integrated Coupling Adapter Short Reach

The ICA SR (FC 0172) was introduced with the IBM z13. ICA SR1.1 (FC 0176) was introduced with IBM z15. ICA SR and ICA SR1.1 are two-port, short-distance coupling features that allow the supported IBM zSystems systems to connect to each other. ICA SR and ICA SR1.1 use coupling channel type: CS5. The ICA SR uses PCIe Gen3 technology, with x16 lanes that are bifurcated into x8 lanes for coupling. ICA SR1.1 utilizes PCIe Gen4 technology, with x16 lanes that are bifurcated into x8 lanes for coupling.

The ICA SR & SR1.1 are designed to drive distances up to 150 m and supports a link data rate of 8 GBps. It is designed to support up to four CHPIDs per port and eight subchannels (devices) per CHPID.

For more information, see *IBM Z Planning for Fiber Optic Links (FICON/FCP, Coupling Links, and Open System Adapters)*, GA23-1409. This publication is available in [the Library section of Resource Link](#).

Coupling Express2 Long Reach

The Coupling Express2 LR (FC 0434) occupies one slot in a PCIe I/O drawer or PCIe+ I/O drawer¹⁴. It allows the supported IBM zSystems to connect to each other over extended distance. The Coupling Express2 LR (FC 0434) is a two-port card that uses coupling channel type CL5.

The Coupling Express2 LR utilizes 10GbE RoCE technology and is designed to drive distances up to 10km unrepeated and support a link data rate of 10 Gigabits per second (Gbps). For distance requirements greater than 10km, clients must utilize a Wavelength Division Multiplexer (WDM). The WDM vendor must be qualified by IBM zSystems.

Coupling Express2 LR is designed to support up to four CHPIDs per port, 32 buffers (that is, 32 subchannels) per CHPID. The Coupling Express2 LR feature resides in the PCIe+ I/O drawer on IBM z16 A02 and IBM z16 AGZ.

For more information, see *IBM Z Planning for Fiber Optic Links (FICON/FCP, Coupling Links, Open Systems Adapters, and zHyperLink Express)*, GA23-1409. This publication is available in [the Library section of Resource Link](#).

¹⁴ PCIe+ I/O drawer (FC 4023 on IBM z16 A02 and IBM z16 AGZ, FC 4021 on IBM z15, and FC 4001 on IBM z14 ZR1) is installed in a 19" rack. PCIe+ I/O Drawers contains and can host up to 16 PCIe I/O features (adapters). They are not supported on IBM z14 M0x or older zSystems. Also, the PCIe I/O drawer cannot be carried forward during and MES upgrade to IBM z14 ZR1 or newer. IBM z16 A02 and IBM z16 AGZ, IBM z15 and IBM z14 ZR1 support ONLY PCIe+ I/O drawers.

Extended distance support

For more information about extended distance support, see *System z End-to-End Extended Distance Guide*, SG24-8047.

Migration considerations

Upgrading from previous generations of IBM zSystems in a Parallel Sysplex to IBM z16 A02 or IBM z16 AGZ in that same Parallel Sysplex requires proper planning for coupling connectivity. Planning is important because of the change in the supported type of coupling link adapters and the number of available fanout slots of the IBM z16 A02 and IBM z16 AGZ CPC drawers.

The ICA SR fanout features provide short-distance connectivity to another IBM z16 A01, IBM z16 A02, IBM z16 AGZ, IBM z15, or IBM z14.

The CE LR adapter provides long-distance connectivity to IBM z16 A01, IBM z16 A02, IBM z16 AGZ, IBM z15, or IBM z14.

The IBM z16 A02 and IBM z16 AGZ fanout slots in the CPC drawer provide coupling link connectivity through the ICA SR fanout cards. In addition to coupling links for Parallel Sysplex, the fanout cards provide connectivity for the PCIe+ I/O drawer (PCIe+ Gen3 fanout).

Up to 12 PCIe fanout cards can be installed in an IBM z16 A02 and IBM z16 AGZ CPC drawer.

To migrate from an older generation machine to an IBM z16 A02 or IBM z16 AGZ without disruption in a Parallel Sysplex environment requires that the older machines are no more than n-2 generation (namely, at least IBM z14) and that they carry enough coupling links to connect to the existing systems while also connecting to the new machine. This is necessary to allow individual components (z/OS LPARs, CFs) to be shut down and moved to the target machine and continue connect to the remaining systems.

It is beyond the scope of this book to describe all possible migration scenarios. Always consult with subject matter experts to help you to develop your migration strategy.

Coupling links and Server Time Protocol

All external coupling links can be used to pass time synchronization signals by using Server Time Protocol (STP). STP is a message-based protocol in which timing messages are passed over data links between servers. The same coupling links can be used to exchange time and CF messages in a Parallel Sysplex.

The use of the coupling links to exchange STP messages has the following advantages:

- ▶ By using the same links to exchange STP messages and CF messages in a Parallel Sysplex, STP can scale with distance. Servers that are exchanging messages over short distances (ICA SR links), can meet more stringent synchronization requirements than servers that exchange messages over long distance (CE2 LR links), with distances up to 100 kilometers (62 miles)¹⁵. This advantage is an enhancement over the IBM Sysplex Timer implementation, which does not scale with distance.
- ▶ Coupling links also provide the connectivity that is necessary in a Parallel Sysplex. Therefore, a potential benefit can be realized of minimizing the number of cross-site links that is required in a multi-site Parallel Sysplex.

Between any two servers that are intended to exchange STP messages, configure each server so that at least two coupling links exist for communication between the servers. This

¹⁵ 10 km (6.2 miles) without DWDM extender, 100 km (62 miles) with certified DWDM equipment.

configuration prevents the loss of one link from causing the loss of STP communication between the servers. If a server does not have a CF LPAR, timing-only links can be used to provide STP connectivity.

IBM z16 A02 and IBM z16 AGZ PTP¹⁶ support

Precision Time Protocol (PTP) is introduced as an alternative to NTP.

- ▶ PTP provides more accurate timestamps to connected devices
- ▶ Initially used for Power Distribution Systems, Telecommunications, and Laboratories
- ▶ Requires Customer Network Infrastructure to be PTP capable
- ▶ IBM z16 A02 and IBM z16 AGZ provide PTP connectivity direct to the CPC

4.7 Cryptographic functions

Cryptographic functions are provided by the CP Assist for Cryptographic Function (CPACF) and the PCI Express cryptographic adapters. IBM z16 A02 and IBM z16 AGZ support the Crypto Express8S, and as carry forward, Crypto Express7S and crypto Express6S features.

4.7.1 CPACF functions (FC 3863)

FC 3863¹⁷ is required to enable Cryptographic functions.

4.7.2 Crypto Express8S feature (FC 0908 and FC 0909)

The Crypto Express8S represents the newest generation of the Peripheral Component Interconnect Express (PCIe) cryptographic coprocessors, which are an optional feature that is available on the IBM z16 A02 and IBM z16 AGZ. These coprocessors are Hardware Security Modules (HSMs) that provide high-security cryptographic processing as required by banking and other industries. This feature provides a secure programming and hardware environment wherein crypto processes are performed. Each cryptographic coprocessor includes general-purpose processors, non-volatile storage, and specialized cryptographic electronics, which are all contained within a tamper-sensing and tamper-responsive enclosure that eliminates all keys and sensitive data on any attempt to tamper with the device. The Crypto Express8S hardware is designed to meet the requirements of FIPS 140 Level 4 for Cryptographic modules and includes Quantum-save encryption technologies.

The Crypto Express8S (2 HSM) includes two IBM PCIe Cryptographic Co-processors (PCIeCC), while the Crypto Express8S (1 HSM) includes one PCIeCC per feature. For availability reasons, a minimum of two features is required. Up to 20 Crypto Express8S (2 HSM) features are supported. The maximum number of the 1 HSM features is 16. The Crypto Express8S feature occupies one I/O slot in a 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.

Crypto Express8S provides domain support for up to 40 logical partitions.

¹⁶ Precision Time Protocol - IEEE 1588 v2

¹⁷ Subject to export regulations.

The accelerator function is designed for maximum-speed Secure Sockets Layer and Transport Layer Security (SSL/TLS) acceleration, rather than for specialized financial applications for secure, long-term storage of keys or secrets. The Crypto Express8S can also be configured as one of the following configurations:

- ▶ The Secure IBM CCA coprocessor includes secure key functions with emphasis on the specialized functions that are required for banking and payment card systems. It is optionally programmable to add custom functions and algorithms by using User Defined Extensions (UDX).

Payment Card Industry (PCI) PIN Transaction Security (PTS) Hardware Security Module (HSM) (PCI-HSM), is available for Crypto Express6S and newer in CCA mode. PCI-HSM mode simplifies compliance with PCI requirements for hardware security modules.

- ▶ The Secure IBM Enterprise PKCS #11 (EP11) coprocessor implements 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 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 Express6S when configured as an Enterprise PKCS #11 coprocessor or managing the CCA mode PCI-HSM.

When the Crypto Express8S PCI Express adapter is configured as a secure IBM CCA co-processor, it still provides accelerator functions. However, up to 3x better performance for those functions can be achieved if the Crypto Express8S PCI Express adapter is configured as an accelerator.

CCA enhancements include the ability to use triple-length (192-bit) Triple-DES (TDES) keys for operations, such as data encryption, PIN processing, and key wrapping to strengthen security. CCA also extended the support for the cryptographic requirements of the German Banking Industry Committee Deutsche Kreditwirtschaft (DK).

Several features that support the use of the AES algorithm in banking applications also were added to CCA. These features include the addition of AES-related key management features and the AES ISO Format 4 (ISO-4) PIN blocks as defined in the ISO 9564-1 standard. PIN block translation is supported as well as usage of AES PIN blocks in other CCA callable services. IBM continues to add enhancements as AES finance industry standards are released

4.7.3 Crypto Express7S feature (FC 0898 and FC 0899) as carry forward only

The Crypto Express7S are supported on IBM z16 A02 and IBM z16 AGZ. These coprocessors are Hardware Security Modules (HSMs) that provide high-security cryptographic processing as required by banking and other industries. This feature provides a secure programming and hardware environment wherein crypto processes are performed. Each cryptographic coprocessor includes general-purpose processors, non-volatile storage, and specialized cryptographic electronics, which are all contained within a tamper-sensing and tamper-responsive enclosure that eliminates all keys and sensitive data on any attempt to tamper with the device. The security features of the HSM are designed to meet the requirements of FIPS 140, Level 4, which is the highest security level defined.

The Crypto Express7S (2 port), FC 0898 includes two IBM PCIe Cryptographic Coprocessors (PCIeCC) per feature. The IBM PCIeCC is a hardware security module (HSM). The Crypto Express7S (1 port), FC 0899 includes one IBM PCIe Cryptographic Coprocessors (PCIeCC)

per feature. For availability reasons, a minimum of two features is required for the one port feature. Up to 20 Crypto Express7S (2 port) features are supported on IBM z16 A02 and IBM z16 AGZ. The maximum number of the one-port features is 16. The total number of HSMs supported on IBM z16 A02 and IBM z16 AGZ is 60 in a combination of Crypto Express8S (2 HSM), Crypto Express8S (1 HSM), Crypto Express7S (2 port), Crypto Express7S (1 port) or Crypto Express6S.

The Crypto Express7S feature occupies one I/O slot in a PCIe+ I/O drawer.

Each adapter can be configured as a Secure IBM CCA coprocessor, Secure IBM Enterprise PKCS #11 (EP11) coprocessor, or accelerator.

Crypto Express7S provides domain support for up to 85 logical partitions.

The accelerator function is designed for maximum-speed Secure Sockets Layer and Transport Layer Security (SSL/TLS) acceleration, rather than for specialized financial applications for secure, long-term storage of keys or secrets. The Crypto Express7S can also be configured as one of the following configurations:

- ▶ The Secure IBM CCA coprocessor includes secure key functions with emphasis on the specialized functions that are required for banking and payment card systems. It is optionally programmable to add custom functions and algorithms by using User Defined Extensions (UDX).

A new mode, called Payment Card Industry (PCI) PIN Transaction Security (PTS) Hardware Security Module (HSM) (PCI-HSM), is available exclusively for Crypto Express6S in CCA mode. PCI-HSM mode simplifies compliance with PCI requirements for hardware security modules.

- ▶ The Secure IBM Enterprise PKCS #11 (EP11) coprocessor implements 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 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 Express8S when configured as an Enterprise PKCS #11 coprocessor or managing the CCA mode PCI-HSM.

When the Crypto Express7S PCI Express adapter is configured as a secure IBM CCA co-processor, it still provides accelerator functions. However, up to 3x better performance for those functions can be achieved if the Crypto Express7S PCI Express adapter is configured as an accelerator.

CCA enhancements include the ability to use triple-length (192-bit) Triple-DES (TDES) keys for operations, such as data encryption, PIN processing, and key wrapping to strengthen security. CCA also extended the support for the cryptographic requirements of the German Banking Industry Committee Deutsche Kreditwirtschaft (DK).

Several features that support the use of the AES algorithm in banking applications also were added to CCA. These features include the addition of AES-related key management features and the AES ISO Format 4 (ISO-4) PIN blocks as defined in the ISO 9564-1 standard. PIN block translation is supported as well as usage of AES PIN blocks in other CCA callable services. IBM continues to add enhancements as AES finance industry standards are released

4.7.4 Crypto Express6S feature (FC 0893) as carry forward only

Crypto Express6S was introduced with IBM z14. On the initial configuration, a minimum of two features are installed (for availability). The number of features then increases one at a time up to a maximum of 16 features.

Each Crypto Express6S feature holds one PCI Express cryptographic adapter. Each adapter can be configured by the installation as a Secure IBM Common Cryptographic Architecture (CCA) coprocessor, as a Secure IBM Enterprise Public Key Cryptography Standards (PKCS) #11 (EP11) coprocessor, or as an accelerator.

The tamper-resistant hardware security module, which is contained on the Crypto Express6S feature, conforms to the Federal Information Processing Standard (FIPS) 140-2 Level 4 Certification. It supports User Defined Extension (UDX) services to implement cryptographic functions and algorithms (when defined as an IBM CCA coprocessor).

The following CCA compliance levels are available:

- ▶ Non-compliant (default)
- ▶ PCI-HSM 2016
- ▶ PCI-HSM 2016 (migration, key tokens while migrating to compliant)

The following EP11 compliance levels are available (Crypto Express6S and Crypto Express5S):

- ▶ FIPS 2009 (default)
- ▶ FIPS 2011
- ▶ BSI 2009
- ▶ BSI 2011

Each Crypto Express6S feature occupies one I/O slot in the PCIe I/O drawer, and features no CHPID assigned. However, it includes one PCHID.

4.8 Integrated Firmware Processor

The Integrated Firmware Processor (IFP) was introduced with the zEC12 and zBC12. The IFP is used to support firmware partitions and for managing PCIe native features. The following features installed in the PCIe+ I/O drawer are managed by the resource group firmware (running on IFP):

- ▶ 25GbE RoCE Express3 SR
- ▶ 25GbE RoCE Express3 LR
- ▶ 10GbE RoCE Express3 SR
- ▶ 10GbE RoCE Express3 LR
- ▶ 25GbE RoCE Express2.1
- ▶ 10GbE RoCE Express2.1
- ▶ 25GbE RoCE Express2
- ▶ 10GbE RoCE Express2
- ▶ Coupling Express2 Long Reach (CE LR)

All native PCIe features should be ordered in pairs for redundancy. The features are assigned to one of the four resource groups (RGs) that are running on the IFP according to their physical location in the PCIe+ I/O drawer, which provides management functions and virtualization functions.

If two features of the same type are installed, one always is managed by resource group 1 (RG 1) or resource group 3 (RG3) while the other feature is managed by resource group 2 (RG 2) or resource group 4 (RG 4). This configuration provides redundancy if one of the features or resource groups needs maintenance or fails.

The IFP and RGs support the following infrastructure management functions:

- ▶ Firmware update of adapters and resource groups
- ▶ Error recovery and failure data collection
- ▶ Diagnostic and maintenance tasks



Logical I/O - Channel Subsystem

This chapter describes the concepts of the IBM z16 A02 and IBM z16 AGZ channel subsystem, including multiple channel subsystems and multiple subchannel sets. It also describes the technology, terminology, and implementation aspects of the channel subsystem.

This chapter includes the following topics:

- ▶ 5.1, “Channel subsystem” on page 186
- ▶ 5.2, “I/O configuration management” on page 194
- ▶ 5.3, “Channel subsystem summary” on page 195

5.1 Channel subsystem

Channel subsystem (CSS) is a collective name of facilities that IBM zSystems use to control I/O operations.

The channel subsystem directs the flow of information between I/O devices and main storage. It allows data processing to proceed concurrently with I/O processing, which relieves data processors (central processor (CP) and Integrated Facility for Linux [IFL]) of the task of communicating directly with I/O devices.

The channel subsystem includes subchannels, I/O devices that are attached through control units, and channel paths between the subsystem and control units. For more information about the channel subsystem, see 5.1.1, “Multiple logical channel subsystems”.

The design of IBM zSystems offers considerable processing power, memory size, and I/O connectivity. In support of the larger I/O capability, the CSS structure is scaled up by introducing the multiple logical channel subsystem (LCSS) since IBM z990, and multiple subchannel sets (MSS) since IBM z9.

An overview of the channel subsystem for IBM z16 A02 and IBM z16 AGZ is shown in Figure 5-1. IBM z16 A02 and IBM z16 AGZ configurations are designed to support up to three logical channel subsystems, each with three subchannel sets and up to 256 channels.

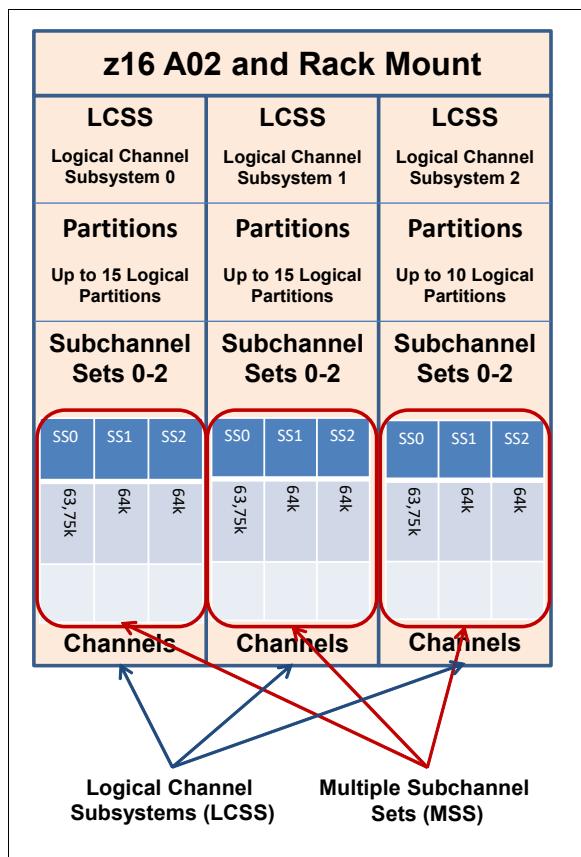


Figure 5-1 Multiple channel subsystem and multiple subchannel sets

All channel subsystems are defined within a single configuration, which is called I/O configuration data set (IOCDs). The IOCDs is loaded into the hardware system area (HSA).

during a central processor complex (CPC) power-on reset (POR) to start all of the channel subsystems.

On IBM z16 A02 and IBM z16 AGZ, the HSA is pre-allocated in memory with a fixed size of 160 GB, which is in addition to the customer purchased memory. This fixed size memory for HSA eliminates the requirement for more planning of the initial I/O configuration and pre-planning for future I/O expansions.

CPC drawer repair: The HSA can be moved from one CPC drawer to the additional CPC drawer, if available, in an enhanced availability configuration as part of a concurrent CPC drawer repair (CDR) action (available only on Max68 feature).

The following objects are always reserved in the IBM z16 A02 and IBM z16 AGZ HSA during POR, whether they are defined in the IOCDs for use:

- ▶ Three LCSSs
- ▶ A total of 15 LPARs in each LCSS0 and LCSS1
- ▶ A total of 10 LPARs in LCSS2
- ▶ Subchannel set 0 with 63.75 K devices in each LCSS
- ▶ Subchannel set 1 with 64 K minus one device in each LCSS
- ▶ Subchannel set 2 with 64 K minus one device in each LCSS

5.1.1 Multiple logical channel subsystems

In the z/Architecture, a *single channel subsystem* can have up to 256 channel paths that are defined, which limited the total numbers of I/O connectivity on older IBM zSystems to 256.

The introduction of *multiple LCSSs* enabled an IBM zSystem to have more than one channel subsystems logically, while each logical channel subsystem maintains the same manner of I/O processing. Also, a logical partition (LPAR) is now attached to a specific logical channel subsystem, which makes the extension of multiple logical channel subsystems not apparent to the operating systems and applications. The multiple image facility (MIF) in the structure enables resource sharing across LPARs within a single LCSS or across the LCSSs.

The multiple LCSS structure extended the IBM zSystems' total number of I/O connectivity to support a balanced configuration for the growth of processor and I/O capabilities.

A one-digit number ID starting from 0 (CSSID) is assigned to an LCSS, and a one-digit hexadecimal ID (MIF ID) starting from 0 is assigned to an LPAR within the LCSS.

Note: The phrase channel subsystem has same meaning as logical channel subsystem in this section, unless otherwise stated.

Subchannels

A *subchannel* provides the logical appearance of a device to the program and contains the information that is required for sustaining a single I/O operation. Each device is accessible by using one subchannel in a channel subsystem to which it is assigned according to the active IOCDs of the IBM zSystem.

A subchannel set (SS) is a collection of subchannels within a channel subsystem. The maximum number of subchannels of a subchannel set determines how many devices are accessible to a channel subsystem.

In z/Architecture, the first subchannel set of an LCSS can have 63.75 K subchannels (with 0.25 K reserved), with a subchannel set ID (SSID) of 0. By enabling the multiple subchannel sets, extra subchannel sets are available to increase the device addressability of a channel subsystem. For more information about multiple subchannel sets, see 5.1.2, “Multiple subchannel sets” on page 206.

Channel paths

A *channel path* provides a connection between the channel subsystem and control units that allows the channel subsystem to communicate with I/O devices. Depending on the type of connections, a channel path might be a physical connection to a control unit with I/O devices, such as FICON, or an internal logical control unit, such as HiperSockets.

Each channel path in a channel subsystem features a unique 2-digit hexadecimal identifier that is known as a *channel-path identifier* (CHPID), which ranges 00 - FF. Therefore, a total of 256 CHPIDs are supported by a CSS, and a maximum of 768 CHPIDs are available on an IBM z16 A02 and IBM z16 AGZ with three logical channel subsystems.

By assigning a CHPID to a physical port of an I/O feature adapter, such as FICON Express32S, or a fanout adapter (ICA SR) port, the channel subsystem connects to the I/O devices through these physical ports.

A port on an I/O feature card features a unique physical channel identifier (PCHID) according to the physical location of this I/O feature adapter, and the sequence of this port on the adapter.

In addition, a port on a fanout adapter has a unique adapter identifier (AID), according to the physical location of this fanout adapter, and the sequence of this port on the adapter.

A CHPID is assigned to a physical port by defining the corresponding PCHID or AID in the I/O configuration definitions.

Control units

A *control unit* provides the logical capabilities that are necessary to operate and control an I/O device. It adapts the characteristics of each device so that it can respond to the standard form of control that is provided by the CSS.

A control unit can be housed separately or can be physically and logically integrated with the I/O device, channel subsystem, or within the IBM zSystem.

I/O devices

An *I/O device* provides external storage, a means of communication between data-processing systems, or a means of communication between a system and its environment. In the simplest case, an I/O device is attached to one control unit and is accessible through one or more channel paths that are connected to the control unit.

5.1.2 Multiple subchannel sets

A subchannel set is a collection of subchannels within a channel subsystem. The maximum number of subchannels of a subchannel set determines how many I/O devices a channel

subsystem can access. The maximum number of subchannels also determines the number of addressable devices to the program (for example, an operating system) running in the LPAR.

Each subchannel has a unique four-digit hexadecimal number 0x0000 - 0xFFFF. Therefore, a single subchannel set can address and access up to 64 K I/O devices.

The IBM z16 A02 and IBM z16 AGZ configurations support three subchannel sets for each logical channel subsystem. The IBM z16 A02 and IBM z16 AGZ can access a maximum of 191.74 K devices for a logical channel subsystem and a logical partition and the programs that are running on it.

Note: Do not confuse the multiple subchannel sets function with multiple channel subsystems.

Subchannel number

The subchannel number is a four-digit hexadecimal number 0x0000 - 0xFFFF, which is assigned to a subchannel within a subchannel set of a channel subsystem. Subchannels in each subchannel set are always assigned subchannel numbers within a single range of contiguous numbers.

The lowest-numbered subchannel is subchannel 0, and the highest-numbered subchannel includes a subchannel number equal to one less than the maximum numbers of subchannels that are supported by the subchannel set. Therefore, a subchannel number is always unique within a subchannel set of a channel subsystem and depends on the sequence of assigning.

With the subchannel numbers, a program that is running on an LPAR (for example, an operating system) can specify all I/O functions relative to a specific I/O device by designating a subchannel that is assigned to the I/O devices.

Normally, subchannel numbers are used only in communication between the programs and the channel subsystem.

Subchannel set identifier

While introducing the MSS, the channel subsystem is extended to assign a value 0 - 2 for each subchannel set, which is the SSID. A subchannel can be identified by its SSID and subchannel number.

Device number

A device number is an arbitrary number 0x0000 - 0xFFFF, which is defined by a system programmer in an I/O configuration for naming an I/O device. The device number must be unique within a subchannel set of a channel subsystem. It is assigned to the corresponding subchannel by channel subsystem when an I/O configuration is activated. Therefore, a subchannel in a subchannel set of a channel subsystem includes a device number together with a subchannel number for designating an I/O operation.

The device number provides a means to identify a device, independent of any limitations that are imposed by the system model, configuration, or channel-path protocols.

A device number also can be used to designate an I/O function to a specific I/O device. Because it is an arbitrary number, it can easily be fit into any configuration management and operating management scenarios. For example, a system administrator can set all of the z/OS systems in an environment to device number 1000 for their system RES volumes.

With multiple subchannel sets, a subchannel is assigned to a specific I/O device by the channel subsystem with an automatically assigned subchannel number and a device number that is defined by user. An I/O device can always be identified by an SSID with a subchannel number or a device number. For example, a device with device number AB00 of subchannel set 1 can be designated as 1AB00.

Normally, the subchannel number is used by the programs to communicate with the channel subsystem and I/O device, whereas the device number is used by a system programmer, operator, and administrator.

Device in subchannel set 0 and extra subchannel sets

An LCSS always includes the first subchannel set (SSID 0), which can have up to 63.75 K subchannels with 256 subchannels that are reserved by the channel subsystem. Users can always define their I/O devices in this subchannel set for general use.

For the extra subchannel sets enabled by the MSS facility, each has 65535 subchannels (64 K minus one) for specific types of devices. These extra subchannel sets are referred as *alternative subchannel sets* in z/OS. Also, a device that is defined in an alternative subchannel set is considered a special device, which normally features a special device type in the I/O configuration.

Currently, an IBM z16 A02 or IBM z16 AGZ that is running z/OS defines the following types of devices in another subchannel set, with proper APAR or PTF installed:

- ▶ Alias devices of the parallel access volumes (PAV).
- ▶ Secondary devices of GDPS Metro Mirror Copy Service (formerly Peer-to-Peer Remote Copy [PPRC]).
- ▶ FlashCopy SOURCE and TARGET devices with program temporary fix (PTF) OA46900.
- ▶ Db2 data backup volumes with PTF OA24142.

The use of another subchannel set for these special devices helps reduce the number of devices in the subchannel set 0, which increases the growth capability for accessing more devices.

Initial program load from an alternative subchannel set

IBM z16 A02 and IBM z16 AGZ support initial program load (IPL) from alternative subchannel sets in addition to subchannel set 0. Devices that are used early during IPL processing now can be accessed by using subchannel set 1 or subchannel set 2 on an IBM z16 A02 and IBM z16 AGZ.

This configuration allows the users of Metro Mirror (formerly PPRC) secondary devices that are defined by using the same device number and a new device type in an alternative subchannel set to be used for IPL, an I/O definition file (IODF), and stand-alone memory dump volumes, when needed.

The display ios,config command

The z/OS `display ios,config(a11)` command that is shown in Figure 5-2 on page 191 includes information about the MSSs.

```
D IOS,CONFIG(ALL)
RESPONSE=SC76
IOS506I 10.35.55 I/O CONFIG DATA 560
ACTIVE IODF DATA SET = SYS9.IODF04
CONFIGURATION ID = ITSO          EDT ID = 01
TOKEN: PROCESSOR DATE      TIME      DESCRIPTION
      SOURCE: VELA    22-12-19 11:12:34 SYS9      IODF04
ACTIVE CSS: 2   SUBCHANNEL SETS CONFIGURED: 0, 1, 2
CHANNEL MEASUREMENT BLOCK FACILITY IS ACTIVE
SUBCHANNEL SET FOR PPRC PRIMARY: INITIAL = 0   ACTIVE = 0
HYPERSWAP FAILOVER HAS OCCURRED: NO
LOCAL SYSTEM NAME (LSYSTEM): VELA
HARDWARE SYSTEM AREA AVAILABLE FOR CONFIGURATION CHANGES
PHYSICAL CONTROL UNITS           8097
CSS  0 - LOGICAL CONTROL UNITS   3995
  SS  0  SUBCHANNELS             53313
  SS  1  SUBCHANNELS             65535
  SS  2  SUBCHANNELS             65535
CSS  1 - LOGICAL CONTROL UNITS   4009
  SS  0  SUBCHANNELS             53761
  SS  1  SUBCHANNELS             65535
  SS  2  SUBCHANNELS             65535
CSS  2 - LOGICAL CONTROL UNITS   4010
  SS  0  SUBCHANNELS             53678
  SS  1  SUBCHANNELS             65535
  SS  2  SUBCHANNELS             65535
ELIGIBLE DEVICE TABLE LATCH COUNTS
  0 OUTSTANDING BINDS ON PRIMARY EDT
```

Figure 5-2 Output for display ios,config(all) command with MSS

5.1.3 Channel path spanning

With the implementation of multiple LCSSs, a channel path can be available to LPARs as dedicated, shared, and spanned.

While a shared channel path can be shared by LPARs within a same LCSS, a spanned channel path can be shared by LPARs within and across LCSSs.

By assigning the same CHPID from different LCSSs to the same channel path (for example, a PCHID), the channel path can be accessed by any LPARs from these LCSSs at the same time. The CHPID is spanned across those LCSSs. The use of spanned channels paths decreases the number of channels that are needed in an installation of IBM zSystems.

A sample of channel paths that are defined as dedicated, shared, and spanned is shown in Figure 5-3 on page 192.

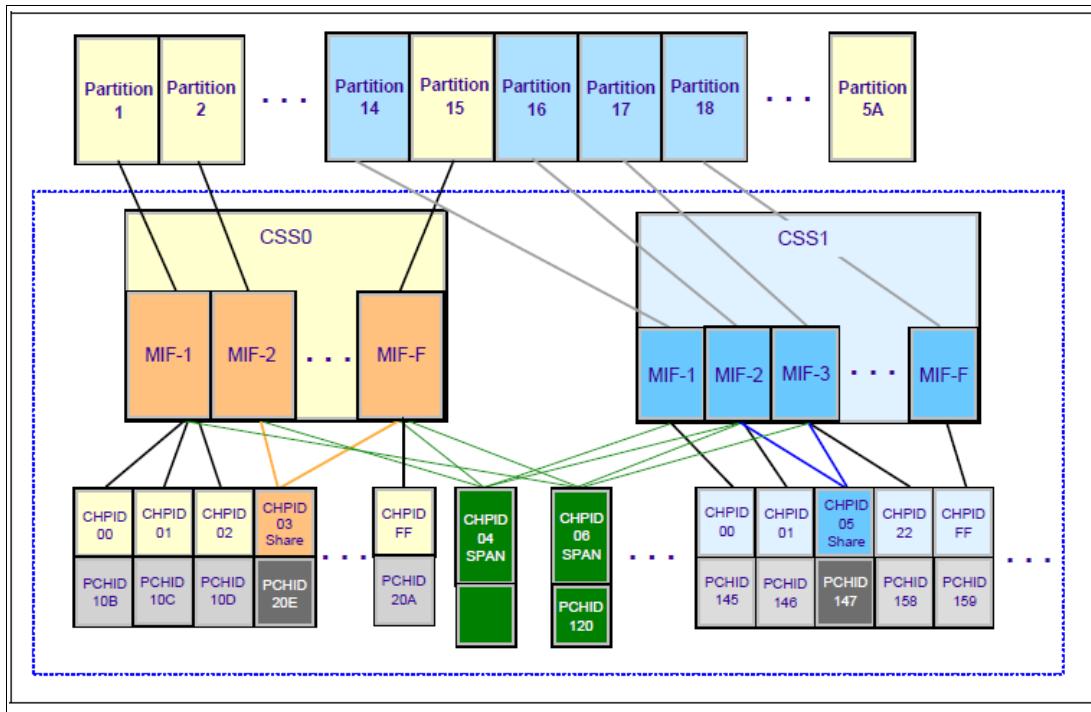


Figure 5-3 IBM zSystems CSS: Channel subsystems with spanning channels

In the sample, the following definitions of a channel path are shown:

- ▶ CHPID FF, assigned to PCHID 20A, is dedicated access for partition 15 of LCSS0. The same applies to CHPID 00,01,02 of LCSS0, and CHPID 00,01,FF of LCSS1.
- ▶ CHPID 03, assigned to PCHID 20E, is shared access for partition 2, and 15 of LCSS0. The same applies to CHPID 05 of LCSS1.
- ▶ CHPID 06, assigned to PCHID 120 is spanned access for partition 1, 15 of LCSS0, and partition 16, 17 of LCSS1. The same applies to CHPID 04.

Channel spanning is supported for internal links (HiperSockets and IC links) and for certain types of external links. External links that are supported on IBM z16 A02 and IBM z16 AGZ configurations include FICON Express32S, FICON Express16S+, OSA-Express7S 1.2, OSA-Express6S, and Coupling Links.

The definition of LPAR name, MIF image ID, and LPAR ID are used to identify an LPAR by the channel subsystem to identify I/O functions from different LPARs of multiple LCSSs, which support the implementation of these dedicated, shared, and spanned paths.

An example of definition of these LPAR-related identifications is shown in Figure 5-4 on page 193.

CSS0			CSS1			CSS2		
Logical Partition Name TST1	Logical Partition Name PROD1	Logical Partition Name PROD2	Logical Partition Name TST2	Logical Partition Name PROD3	Logical Partition Name PROD4	LPAR Name TST3	LPAR Name TST4	Specified in HCD / IOCP
Logical Partition ID 02	Logical Partition ID 04	Logical Partition ID 0A	Logical Partition ID 14	Logical Partition ID 16	Logical Partition ID 1D	LPAR ID 22	LPAR ID 26	Specified in HCD / IOCP
MIF ID 2	MIF ID 4	MIF ID A	MIF ID 4	MIF ID 6	MIF ID D	MIF ID 2	MIF ID 6	Specified in Image Profile

Figure 5-4 CSS, LPAR, and identifier example

LPAR name

The LPAR name is defined as partition name parameter in the RESOURCE statement of an I/O configuration. The LPAR name must be unique across the server.

MIF image ID

The MIF image ID is defined as a parameter for each LPAR in the RESOURCE statement of an I/O configuration. It ranges 1 - F, and must be unique within an LCSS. However, duplicates are allowed in different LCSSs.

If a MIF image ID is not defined, an arbitrary ID is assigned when the I/O configuration activated. The IBM z16 A02 and IBM z16 AGZ support a maximum of three LCSSs, with a total of 40 LPARs that can be defined.

Each LCSS of an IBM z16 A02 or IBM z16 AGZ can support the following numbers of LPARs:

- ▶ LCSS0 and LCSS1 support 15 LPARs each, and the MIF image ID is 1 - F.
- ▶ LCSS2 supports 10 LPARs, and the MIF image IDs are 1 - A.

LPAR ID

The LPAR ID is defined by a user in an image activation profile for each LPAR. It is a 2-digit hexadecimal number 00 - 7F. The LPAR ID must be unique across the server. Although it is arbitrarily defined by the user, an LPAR ID often is the CSS ID concatenated to its MIF image ID, which makes the value more meaningful for the system administrator. For example, an LPAR with LPAR ID 1A defined in that manner means that the LPAR is defined in LCSS1, with the MIF image ID A.

5.2 I/O configuration management

The following tools are available to help maintain and optimize the I/O configuration:

- ▶ IBM Configurator for e-business (eConfig)

The eConfig tool is used by your IBM representative. It is used to create configurations or upgrades of a configuration, and maintains tracking to the installed features of those configurations. eConfig produces reports that help you understand the changes that are being made for a new system, or a system upgrade, and what the target configuration looks like.

- ▶ Hardware configuration definition (HCD)

HCD supplies an interactive dialog to generate the IODF, and later the IOCDS. Generally, use HCD or Hardware Configuration Manager (HCM) to generate the I/O configuration rather than writing I/O configuration program (IOCP) statements. The validation checking that HCD runs against a IODF source file helps minimize the risk of errors before an I/O configuration is activated.

HCD support for multiple channel subsystems is available with z/VM and z/OS. HCD provides the capability to make dynamic hardware and software I/O configuration changes.

Note: Certain functions might require specific levels of an operating system, PTFs, or both.

Consult the appropriate fix categories:

- IBM z16 A02 and IBM z16 AGZ A01: IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ -3931
- IBM z16 A02 and IBM z16 AGZ: IBM.Device.Server.IBMz16A02-3932.*
- IBM z15 T01: IBM.Device.Server.IBM z15-8561
- IBM z15 T02: IBM.Device.Server.IBM z15-8562
- IBM z14 M0x: IBM.Device.Server.IBM z14-3906
- IBM z14 ZR1: IBM.Device.Server.IBM z14ZR1-3907
- ▶ HCM

HCM is a priced optional feature that supplies a graphical interface of HCD. It is installed on a PC and allows you to manage the physical and logical aspects of a mainframe's hardware configuration.

- ▶ CHPID Mapping Tool (CMT)

The CMT helps to map CHPIDs onto PCHIDs that are based on an IODF source file and the eConfig configuration file of a mainframe. It provides a CHPID to PCHID mapping with high availability for the targeted I/O configuration. It also features built-in mechanisms to generate a mapping according to customized I/O performance groups. More enhancements are implemented in CMT to support IBM z16 A02 and IBM z16 AGZ configurations.

The CMT is available for download from the IBM Resource Link website.

The configuration file for a new machine or upgrade is also available from [IBM Resource Link](#). Ask your IBM technical sales representative for the name of the file to download.

5.3 Channel subsystem summary

IBM z16 A02 and IBM z16 AGZ support the channel subsystem features of multiple LCSS, MSS, and the channel spanning that is described in this chapter. The channel subsystem capabilities of IBM z16 A02 and IBM z16 AGZ are listed in Table 5-1.

Table 5-1 IBM z16 A02 and IBM z16 AGZ CSS overview

Maximum number of CSSs	3
Maximum number of LPARs per CSS	CSS0 - CSS1: 15 CSS2: 10
Maximum number of LPARs per system	40
Maximum number of subchannel sets per CSS	3
Maximum number of subchannels per CSS	191.74 K SS0: 65280 SS1 - SS2: 65535
Maximum number of CHPIIDs per CSS	256



Cryptographic features

This chapter describes the hardware cryptographic functions that are available on IBM z16 A02 and IBM z16 AGZ. The CP Assist for Cryptographic Function (CPACF), together with the Peripheral Component Interconnect Express (PCIe) cryptographic coprocessors (PCIeCC), offer a balanced use of processing resources and unmatched scalability for fulfilling pervasive encryption demands.

IBM recognizes that with any new technology, new threats exist, and as such, suitable counter measures must be taken. Quantum technology can be used for incredible good, but in the hands of an adversary, it has the potential to weaken or break core cryptographic primitives that were used to secure systems and communications. Quantum-safe cryptography aims to provide protection against attacks that can be started by quantum computers.

The IBM z16 A02 and IBM z16 AGZ use quantum-safe technologies to help protect your business-critical infrastructure and data from potential attacks.

The IBM z16 A02 and IBM z16 AGZ are designed for delivering a transparent and consumable approach that enables extensive (pervasive) encryption of data in flight and at rest, with the goal of substantially simplifying data security and reducing the costs that are associated with protecting data while achieving compliance mandates.

This chapter also introduces the principles of cryptography and describes the implementation of cryptography in the hardware and software architecture of IBM Z. It also describes the features that IBM z16 A02 and IBM z16 AGZ offer. Finally, the chapter summarizes the cryptographic features and required software.

This chapter includes the following topics:

- ▶ 6.1, “Cryptography enhancements on IBM z16 A02 and IBM z16 AGZ” on page 199
- ▶ 6.2, “Cryptography overview” on page 200
- ▶ 6.3, “Cryptography on IBM z16 A02 and IBM z16 AGZ” on page 204
- ▶ 6.4, “CP Assist for cryptographic functions” on page 209
- ▶ 6.5, “Crypto Express8S” on page 213
- ▶ 6.6, “Trusted Key Entry workstation” on page 228
- ▶ 6.7, “Cryptographic functions comparison” on page 231
- ▶ 6.8, “Cryptographic operating system support for IBM z16 A02 and IBM z16 AGZ” on page 233

- ▶ 6.9, “Further use of cryptography on IBM z16 A02 and IBM z16 AGZ” on page 235

6.1 Cryptography enhancements on IBM z16 A02 and IBM z16 AGZ

Attention: Many older cryptographic algorithms like DES or RSA, and hashing algorithms such as SHA1 are considered weak algorithms and do not provide sufficient protection against today's cyberattacks.

This risk can be mitigated by switching to stronger algorithms, such as AES-256, SHA-256, SHA-3, and CRYSTALS-Dilithium.

IBM provides several tools that can aid in the discovery process:

- ▶ IBM z/OS Integrated Cryptographic Service Facility (ICSF)
- ▶ IBM Application Discovery and Delivery Intelligence (ADDI)
- ▶ IBM Crypto Analytics Tool (CAT)
- ▶ IBM z/OS Encryption Readiness Technology (zERT)

These tools can help you identify certificates, encryption protocols, algorithms, and key lengths that are at risk in your IBM zSystems environment.

IBM z16 A02 and IBM z16 AGZ introduced the new PCIe Crypto Express8S feature, together with a further improved CPACF Coprocessor, that can be managed by a new Trusted Key Entry (TKE) workstation. In addition, the IBM Common Cryptographic Architecture (CCA) and the IBM Enterprise PKCS #11 (EP11) Licensed Internal Code (LIC) were enhanced.

The functions support new standards and are designed to meet the following compliance requirements:

- ▶ Payment Card Industry (PCI) Hardware Security Module (HSM) certification to strengthen the cryptographic standards for attack resistance in the payment card systems area.
PCI HSM certification is exclusive for Crypto Express7S and Crypto Express6S.
- ▶ National Institute of Standards and Technology (NIST) through the Federal Information Processing Standard (FIPS) standard to implement guidance requirements.
- ▶ Common Criteria EP11 EAL4.
- ▶ German Banking Industry Commission (GBIC).
- ▶ Visa Format Preserving Encryption (VFPE) for credit card numbers.
- ▶ Enhanced public key Elliptic Curve Cryptography (ECC) for users such as Chrome, Firefox, and Apple's iMessage.
- ▶ Accredited Standards Committee X9 Inc Technical Report-34 (ASC X9 TR-34)

These enhancements are described in this chapter.

IBM z16 A02 and IBM z16 AGZ include standard cryptographic hardware and optional cryptographic features for flexibility and growth capability. IBM has a long history of providing hardware cryptographic solutions. This history stretches from the development of the Data Encryption Standard (DES) in the 1970s to the Crypto Express tamper-sensing and tamper-responding programmable features.

The Crypto Express8S hardware is designed to meet the Federal Information Processing Standards (FIPS) 140 Level 4 for cryptographic modules and includes Quantum-safe encryption technologies. It also meets several other security ratings, such as the Common Criteria for Information Technology Security Evaluation, the PCI HSM criteria, and the criteria for German Banking Industry Commission (formerly known as Deutsche Kreditwirtschaft evaluation).

The cryptographic functions include the full range of cryptographic operations that are necessary for local and global business and financial institution applications. User Defined Extensions (UDX) allow you to add custom cryptographic functions to the functions that IBM z16 A02 and IBM z16 AGZ systems offer.

6.2 Cryptography overview

Throughout history, a need existed for secret communication between people that cannot be understood by outside parties.

Also, it is necessary to ensure that a message cannot be corrupted (message integrity), while ensuring that the sender and the receiver really are the persons who they claim to be. Over time, several methods were used to achieve these objectives, with more or less success. Many procedures and algorithms for encrypting and decrypting data were developed that are increasingly complicated and time-consuming.

6.2.1 Modern cryptography

With the development of computing technology, the encryption and decryption algorithms can be performed by computers, which enables the use of complicated mathematical algorithms. Most of these algorithms are based on the prime factorization of large numbers.

Cryptography is used to meet the following requirements:

- ▶ Data protection

The protection of data usually is the main concept that is associated with cryptography. Only authorized persons should be able to read the message or get information about it. Data is encrypted by using a known algorithm and secret keys, such that the intended party can de-scramble the data, but an interloper cannot. This concept is also referred to as *confidentiality*.

- ▶ Authentication (identity validation)

This process decides whether the communication partners are who they claim to be, which can be done by using certificates and signatures. It must be possible to clearly identify the owner of the data or the sender and the receiver of the message.

- ▶ Message (data) Integrity

The verification of data ensures that what was received is identical to what was sent. It must be proven that the data is complete and was not altered during the moment it was transmitted (by the sender) and the moment it was received (by the receiver).

- ▶ Non-repudiation

It must be impossible for the owner of the data or the sender of the message to deny authorship. Non-repudiation ensures that both sides of a communication know that the other side agreed to what was exchanged, and not someone else. This specification implies a legal liability and contractual obligation, which is the same as a signature on a contract.

These goals should all be possible without unacceptable overhead to the communication. The goal is to keep the system secure, manageable, and productive.

The basic data protection method is achieved by encrypting and decrypting the data, while hash algorithms, message authentication codes (MACs), digital signatures, and certificates are used for authentication, data integrity, and non-repudiation.

When encrypting a message, the sender transforms the clear text into a secret text. Doing so requires the following main elements:

- ▶ The *algorithm* is the mathematical or logical formula that is applied to the key and the clear text to deliver a ciphered result, or to take a ciphered text and deliver the original clear text.
- ▶ The *key* ensures that the result of the encrypting data transformation by the algorithm is only the same when the same key is used. That decryption of a ciphered message results only in the original clear message when the correct key is used. Therefore, the receiver of a ciphered message must know which algorithm and key must be used to decrypt the message.

6.2.2 Kerckhoffs' principle

In modern cryptography, the algorithm is published and known to everyone, whereas the keys are kept secret. This configuration corresponds to Kerckhoffs' principle, which is named after Auguste Kerckhoffs, a Dutch cryptographer, who formulated it in 1883:

“A system should not depend on secrecy, and it should be able to fall into the enemy’s hands without disadvantage.”

In other words, the security of a cryptographic system should depend on the security of the key, so the key must be kept secret. Therefore, the secure management of keys is the primal task of modern cryptographic systems.

Adhering to Kerckhoffs' Principle is done for the following reasons:

- ▶ It is much more difficult to keep an algorithm secret than a key.
- ▶ It is harder to exchange a compromised algorithm than to exchange a compromised key.
- ▶ Secret algorithms can be reconstructed by reverse engineering software or hardware implementations.
- ▶ Errors in public algorithms can generally be found more easily, when many experts examine it.
- ▶ In history, most secret encryption methods proved to be weak and inadequate.
- ▶ When a secret encryption method is used, it is possible that a back door was built in.
- ▶ If an algorithm is public, many experts can form an opinion about it. Also, the method can be more thoroughly investigated for potential weaknesses and vulnerabilities.

6.2.3 Keys

The keys that are used for the cryptographic algorithms often are sequences of numbers and characters, but can also be any other sequence of bits. The length of a key influences the security (strength) of the cryptographic method. The longer the used key, the more difficult it is to compromise a cryptographic algorithm.

For example, the DES (symmetric key) algorithm uses keys with a length of 56 bits, Triple-DES (TDES) uses keys with a length of 112 bits, and Advanced Encryption Standard (AES) uses keys of 128, 192, 256, or 512 bits. The asymmetric key RSA algorithm (named after its inventors Rivest, Shamir, and Adleman) uses keys with a length of 1024 - 4096 bits.

In modern cryptography, keys must be kept secret. Depending on the effort that is made to protect the key, keys are classified into the following levels:

- ▶ A *clear key* is a key that is transferred from the application in clear text to the cryptographic function. The key value is stored in the clear (at least briefly) somewhere in unprotected memory areas. Therefore, the key can be made available to someone under certain circumstances who is accessing this memory area.

This risk must be considered when clear keys are used. However, many applications exist where this risk can be accepted. For example, the transaction security for the (widely used) encryption methods Secure Sockets Layer (SSL) and Transport Layer Security (TLS) is based on clear keys.

- ▶ The value of a *protected key* is stored only in clear in memory areas that cannot be read by applications or users. The key value does not exist outside of the physical hardware, although the hardware might not be tamper-resistant. The principle of protected keys is unique to IBM Z. For more information, see 6.4.2, “CPACF protected key” on page 211.
- ▶ For a *secure key*, the key value does not exist in clear format outside of a special hardware device (HSM), which must be secured and tamper-resistant. A secure key is protected from disclosure and misuse, and can be used for the trusted execution of cryptographic algorithms on highly sensitive data. If used and stored outside of the HSM, a secure key must be encrypted with a *master key*, which is created within the HSM and never leaves the HSM.

Because a secure key must be handled in a special hardware device, the use of secure keys usually is far slower than the use of clear keys, as shown in Figure 6-1.

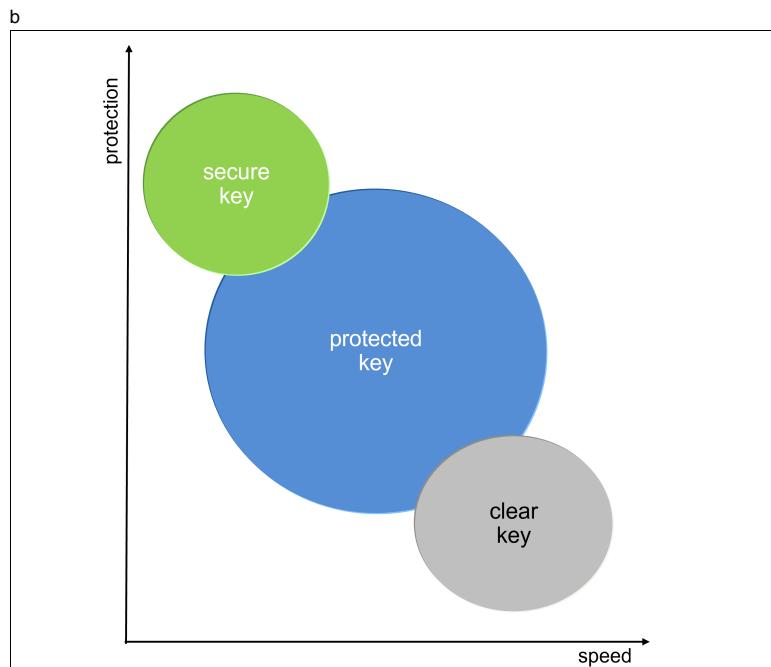


Figure 6-1 Three levels of protection with three levels of speed

6.2.4 Algorithms

The following algorithms of modern cryptography are differentiated based on whether they use the same key for the encryption of the message as for the decryption:

- ▶ *Symmetric algorithms* use the same key to encrypt and to decrypt data. The function that is used to decrypt the data is the opposite of the function that is used to encrypt the data. Because the same key is used on both sides of an operation, it must be negotiated between both parties and kept secret. Therefore, symmetric algorithms are also known as *secret key algorithms*.

The main advantage of symmetric algorithms is that they are fast and therefore can be used for large amounts of data, even if they are not run on specialized hardware. The disadvantage is that the key must be known by both sender and receiver of the messages, which implies that the key must be exchanged between them. This key exchange is a weak point that can be attacked.

Prominent examples for symmetric algorithms are DES, TDES, and AES.

- ▶ *Asymmetric algorithms* use two distinct but related key: the *public key* and the *private key*. As the names imply, the private key must be kept secret, whereas the public key is shown to everyone. However, with asymmetric cryptography, it is not important who sees or knows the public key. Whatever is done with one key can be undone by the other key only.

For example, data that is encrypted by the public key can be decrypted by the associated private key only, and vice versa. Unlike symmetric algorithms, which use distinct functions for encryption and decryption, only one function is used in asymmetric algorithms.

Depending on the values that are passed to this function, it encrypts or decrypts the data. Asymmetric algorithms are also known as *public key algorithms*.

Asymmetric algorithms use complex calculations and are relatively slow (about 100 - 1000 times slower than symmetric algorithms). Therefore, such algorithms are not used for the encryption of bulk data.

Because the private key is never exchanged between the parties in communication, they are less vulnerable than symmetric algorithms. Asymmetric algorithms mainly are used for authentication, digital signatures, and for the encryption and exchange of secret keys, which in turn are used to encrypt bulk data with a symmetric algorithm.

Examples for asymmetric algorithms are RSA and the elliptic curve algorithms.

- ▶ *One-way algorithms* are not cryptographic functions. They do not use keys, and they can scramble data only, not de-scramble it. These algorithms are used extensively within cryptographic procedures for digital signing and tend to be developed and governed by using the same principles as cryptographic algorithms. One-way algorithms are also known as *hash algorithms*.

The most prominent one-way algorithms are the Secure Hash Algorithms (SHA).

6.3 Cryptography on IBM z16 A02 and IBM z16 AGZ

In principle, cryptographic algorithms can run on processor hardware. However, these workloads are compute-intensive, and the handling of secure keys also requires special hardware protection. Therefore, IBM zSystems offer several cryptographic hardware features, which are specialized to meet the requirements for cryptographic workload.

The cryptographic hardware that is supported on IBM z16 A02 and IBM z16 AGZ is shown in Figure 6-2. These features are described in this chapter.

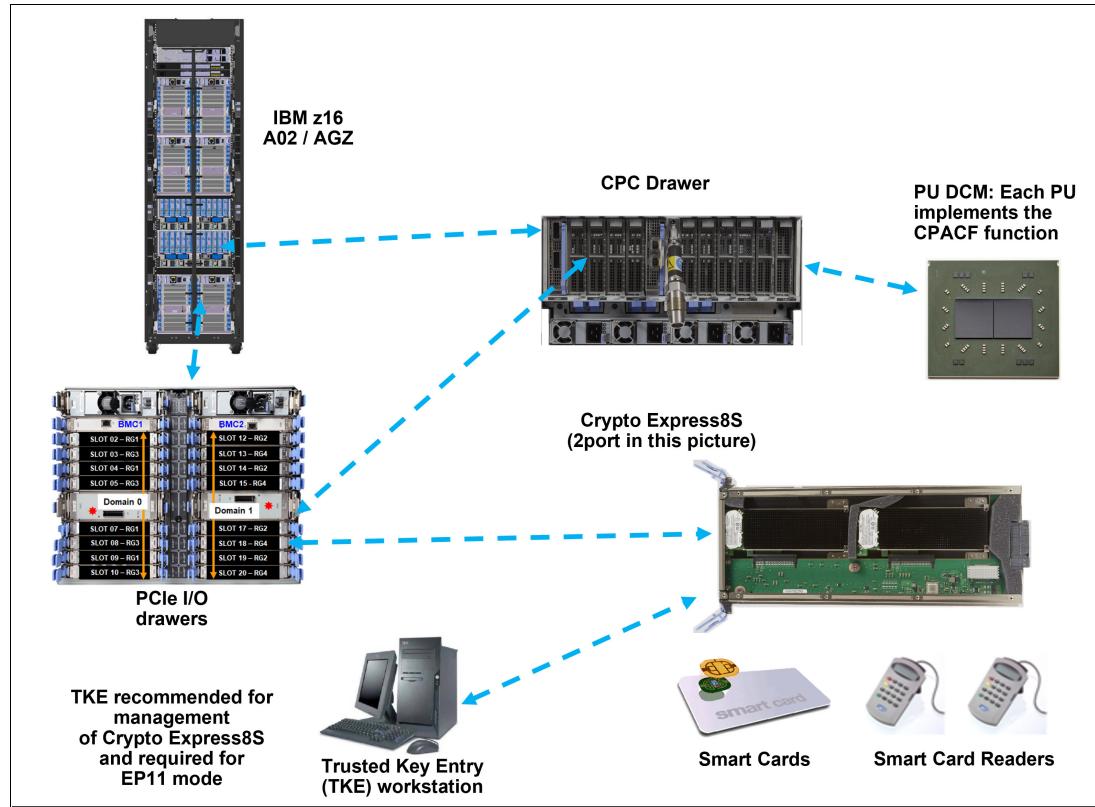


Figure 6-2 Cryptographic hardware that is supported in IBM z16 A02 and IBM z16 AGZ

Implemented in every processor unit (PU) or core in a central processor complex (CPC) is a cryptographic coprocessor that can be used¹ for cryptographic algorithms that use clear keys or protected keys. For more information, see 6.4, “CP Assist for cryptographic functions” on page 209.

Crypto Express coprocessor adapters contain one or two hardware security modules (HSMs) and are placed in the PCIe+ I/O drawer of IBM z16 A02 or IBM z16 AGZ. These features also support cryptographic algorithms by using secret keys. There are three generations of cryptographic coprocessors that are supported for IBM z16 A02 and IBM z16 AGZ systems:

- ▶ Crypto Express6S, Feature Code 0893, carry forward only (miscellaneous equipment specification (MES) from IBM z14 ZR1 or IBM z15 T02 systems)
- ▶ Crypto Express7S, Feature Codes 0899 (one adapter on card) and 0898 (two adapters on card), carry forward only (MES from z15 systems)

¹ CPACF enablement feature must be ordered (FC 3863).

- ▶ Crypto Express8S, Feature Codes 0909 (one adapter on card) and 0908 (two adapters on card)

For more information about the Crypto Express8S feature, see 6.5, “Crypto Express8S” on page 213.

Finally, a TKE workstation is required for entering keys in a secure way into the Crypto Express8S HSMs, which often also is equipped with smart card readers. For more information, see 6.6, “Trusted Key Entry workstation” on page 228.

The feature codes and purpose of the cryptographic hardware features that are available for IBM z16 A02 and IBM z16 AGZ are listed in Table 6-1.

Table 6-1 Cryptographic features for IBM z16 A02 and IBM z16 AGZ

Feature code	Description
3863	CP Assist for Cryptographic Function (CPACF) enablement This feature is a prerequisite to use CPACF (except for SHA-1, SHA-224, SHA-256, SHA-384, and SHA-512) and the PCIe Crypto Express features.
0908	Crypto Express8S feature (dual HSM) ^a These features are optional. The 2-port feature contains two IBM 4770 PCIe Cryptographic Coprocessors (HSMs), which can be independently defined as Coprocessor or Accelerator. New feature. Not supported on previous generations IBM Z systems. A TKE, Smart Card Reader and latest available level Smart cards are required to operate the Crypto adapter card in EP11 mode.
0909	Crypto Express8S feature (single HSM) ^a These features are optional. The 2-port feature contains two IBM 4770 PCIe Cryptographic Coprocessors (HSMs), which can be independently defined as Coprocessor or Accelerator. New feature. Not supported on previous generations IBM Z systems. A TKE, Smart Card Reader and latest available level Smart cards are required to operate the Crypto adapter card in EP11 mode.
0898	Crypto Express7S feature (2-port) Carry forward from IBM z15 T02. This feature contains two IBM 4769 PCIe Cryptographic Coprocessors (HSMs), which can be independently defined as Coprocessor or Accelerator. Supported on IBM z15, IBM z16 A01, IBM z16 A02 and IBM z16 AGZ.
0899	Crypto Express7S feature (1-port) ^a Carry forward from IBM z15 T02. This feature contains one IBM 4769 PCIe Cryptographic Coprocessor (HSM), which can be defined as Coprocessor or Accelerator. Supported on IBM z15, IBM z16 A01, IBM z16 A02 and IBM z16 AGZ.
0893	Crypto Express6S adapter ^a This feature is available as a carry forward MES from IBM z14 ZR1. This feature is optional. Each feature one IBM 4768 PCIe Cryptographic Coprocessor (HSM). This feature is supported in IBM z16 A01, IBM z16 A02, IBM z16 AGZ, IBM z15, and IBM z14.

Feature code	Description
0058	<p>TKE tower workstation</p> <p>A TKE provides basic key management (key identification, exchange, separation, update, and backup) and security administration. It is optional for running a Crypto Express feature in CCA mode in non PCI-compliant environment. It is required for running in EP11 mode and CCA mode with full PCI compliance. The TKE workstation has one 1000BASE-T Ethernet port, and supports connectivity to an Ethernet local area network (LAN). Up to 10 features combined (0057/0058) can be ordered per IBM z16 A02 and IBM z16 AGZ.</p>
0157	<p>TKE Table Top Keyboard/Monitor/Mouse</p> <p>A table top monitor with a US English language keyboard. There is a touchpad for pointing, and a country-specific power cord.</p>
0057	<p>TKE rack-mounted workstation</p> <p>The rack-mounted version of the TKE, which needs a customer-provided standard 19-inch rack. It features a 1u TKE unit and an (optional) 1u console tray (screen, keyboard, and pointing device). When smart card readers are used, another customer-provided tray is needed. Up to 10 features combined (0057/0058) can be ordered per IBM z16 A02 and IBM z16 AGZ.</p>
0156	<p>TKE Rack Keyboard/Monitor/Mouse</p> <p>A 1U rack-mounted display and keyboard with a built-in pointing device. The keyboard comes in the English language.</p>
0851	<p>4770 TKE Crypto Adapter (IBM PCIeCC)</p> <p>The stand-alone crypto adapter is required for TKE upgrade from FC 0085 and FC 0086 TKE tower, or FC 0087 and FC 0088 TKE rack mount when carry forward these features to IBM z16 A02 and IBM z16 AGZ.</p>
0144	<p>TKE Tower carry forward to IBM z16 A02 and IBM z16 AGZ</p> <p>TKE Tower FC 0088 can be carried forward to IBM z16 A02 and IBM z16 AGZ. It requires IBM 4770 PCIeCC (FC 0851) for compatibility with TKE LIC 10.0 (FC 0882) and for managing Crypto Express8S. (FC 0144 = FC 0088 + FC 0851 + FC 0882).</p>
0145	<p>TKE rack mount carry forward to IBM z16 A02 and IBM z16 AGZ</p> <p>TKE rack mount FC 0087 can be carried forward to IBM z16 A02 and IBM z16 AGZ. It requires IBM 4770 PCIeCC (FC 0851) for compatibility with TKE LIC 10.0 (FC 0882) and for managing Crypto Express8S. (FC 0145 = FC 0087 + FC 0851 + FC 0882).</p>
0233	<p>TKE rack mount carry forward to IBM z16 A02 and IBM z16 AGZ</p> <p>TKE rack mount FC 0085 can be carried forward to IBM z16 A02 and IBM z16 AGZ. It requires IBM 4770 PCIeCC (FC 0851) for compatibility with TKE LIC 10.0 (FC 0882) and for managing Crypto Express8S. (FC 0233 = FC 0085 + FC 0851 + FC 0882).</p>
0234	<p>TKE Tower Carry forward to IBM z16 A02 and IBM z16 AGZ</p> <p>TKE Tower FC 0086 can be carried forward to IBM z16 A02 and IBM z16 AGZ. It requires IBM 4770 PCIeCC (FC 0851) for compatibility with TKE LIC 10.0 (FC 0882) and for managing Crypto Express8S. (FC 0234 = FC 0086 + FC 0851 + FC 0882).</p>

Feature code	Description
0882	<p>TKE 10.0 Licensed Internal Code (LIC)</p> <p>Included with the TKE tower workstation FC 0058 and the TKE rack-mounted workstation FC 0057 for IBM z16 A02 and IBM z16 AGZ. Earlier versions of TKE features (feature codes: 0088, 0087, 0086, and 0085) can also be upgraded to TKE 10.0 LIC, adding FC 0851 (IBM 4770 PCIeCC) if the TKE is assigned to an manages Crypto Express8S</p>
0891	<p>TKE Smart Card Reader</p> <p>Access to information in the smart card is protected by a PIN. One feature code includes two smart card readers, two cables to connect to the TKE workstation, and 20 smart cards.</p>
0900	<p>TKE additional smart cards</p> <p>This card allows the TKE to support zones with EC 521 key strength (EC 521 strength for Logon Keys, Authority Signature Keys, and EP11 signature keys). When one feature code is ordered, 10 smart cards are included. The order increment is 1 - 99 (990 blank smart cards).</p>

- a. The maximum number of combined features of all types cannot exceed 40 HSMs on a IBM z16 A02 and IBM z16 AGZ. Therefore, the maximum number for feature code 0898 is 20; all other (single HSM) types is 16 when installed exclusively.

A TKE includes support for the AES encryption algorithm with 256-bit master keys and key management functions to load or generate master keys to the cryptographic coprocessor.

If the TKE workstation is chosen to operate the Crypto Express8S adapter in a IBM z16 A02 and IBM z16 AGZ, TKE workstation with the TKE 10.0 LIC is required. For more information, see 6.6, “Trusted Key Entry workstation” on page 228.

Important: Products that include any of the cryptographic feature codes contain cryptographic functions that are subject to special export licensing requirements by the United States Department of Commerce. It is your responsibility to understand and adhere to these regulations when you are moving, selling, or transferring these products.

To access and use the cryptographic hardware devices that are provided by IBM z16 A02 and IBM z16 AGZ, the application must use an application programming interface (API) that is provided by the operating system. In z/OS, the Integrated Cryptographic Service Facility (ICSF) provides the APIs and is managing the access to the cryptographic devices, as shown in Figure 6-3 on page 208.

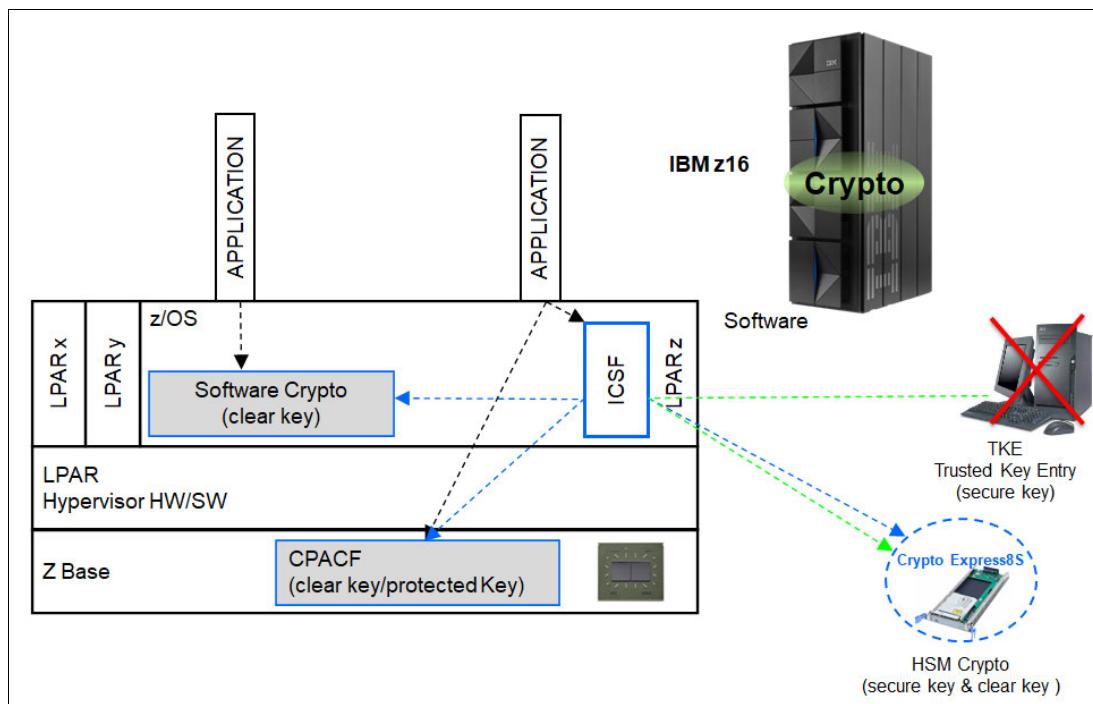


Figure 6-3 z16 Cryptographic Support in z/OS

ICSF is a software component of z/OS. ICSF works with the hardware cryptographic features and the Security Server (IBM Resource Access Control Facility [IBM RACF®] element) to provide secure, high-speed cryptographic services in the z/OS environment. ICSF provides the APIs by which applications request the cryptographic services, and from the CPACF and the Crypto Express features.

ICSF transparently routes application requests for cryptographic services to one of the integrated cryptographic engines (CPACF or a Crypto Express feature), depending on performance or requested cryptographic function. ICSF is also the means by which the secure Crypto Express features are loaded with master key values, which allows the hardware features to be used by applications.

The cryptographic hardware that is installed in IBM z16 A02 and IBM z16 AGZ determines the cryptographic features and services that are available to the applications.

The users of the cryptographic services call the ICSF API. Some functions are performed by the ICSF software without starting the cryptographic hardware features. Other functions result in ICSF going into routines that contain proprietary IBM zSystems crypto instructions. These instructions are run by a CPU engine and result in a work request that is generated for a cryptographic hardware feature.

6.4 CP Assist for cryptographic functions

Attached to every PU (core) of a IBM z16 A02 and IBM z16 AGZ are two independent engines, one for compression and one for cryptographic functions, as shown in Figure 6-4.

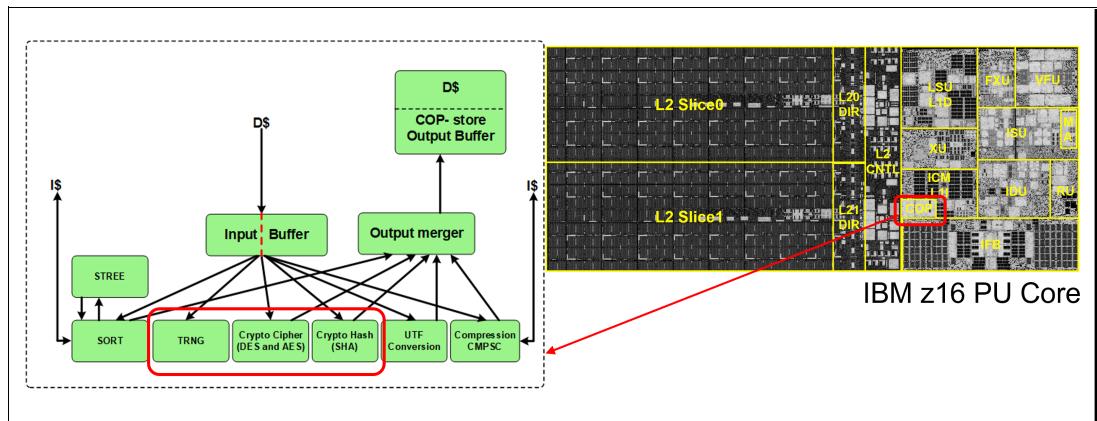


Figure 6-4 The cryptographic coprocessor CPACF

This cryptographic coprocessor, which is known as the CPACF, is not qualified as an HSM; therefore, it is not suitable for handling algorithms that use secret keys. However, the coprocessor can be used for cryptographic algorithms that use clear keys or protected keys. The CPACF works synchronously with the PU, which means that the owning processor is busy when its coprocessor is busy. This setup provides a fast device for cryptographic services.

CPACF supports pervasive encryption. Simple policy controls allow businesses to enable encryption to protect data in mission-critical databases without the need to stop the database or re-create database objects. Pervasive encryption includes z/OS Dataset Encryption, z/OS Coupling Facility Encryption, z/VM encrypted hypervisor paging, and z/TPF transparent database encryption, which use performance enhancements in the hardware.

The CPACF offers a set of symmetric cryptographic functions that enhances the encryption and decryption performance of clear key operations. These functions are for SSL, virtual private network (VPN), and data-storing applications that do not require FIPS 140 Level 4 security.

CPACF is designed to facilitate the privacy of cryptographic key material when used for data encryption through key wrapping implementation. It ensures that key material is not visible to applications or operating systems during encryption operations. For more information, see 6.4.2, “CPACF protected key” on page 211.

The CPACF feature provides hardware acceleration for the following cryptographic services:

- ▶ Symmetric ciphers
 - DES
 - Triple-DES
 - AES-128
 - AES-192
 - AES-256 (all for clear and protected keys)
- ▶ Elliptic curves cryptography (ECC)
 - ECDSA, ECDH, support for the NIST P256, NIST P386, NIST P521
 - EdDSA for Ed25519, Ed448 Curves
 - ECDH for X25519, X448 Curves

- Key generation for NIST, Ed and X curves
- ▶ Hashes/MACs
 - SHA-1
 - SHA-224 (SHA-2 or SHA-3 standard)
 - SHA-256 (SHA-2 or SHA-3 standard)
 - SHA-384 (SHA-2 or SHA-3 standard)
 - SHA-512 (SHA-2 or SHA-3 standard)
 - SHAKE-128
 - SHAKE-256
 - GHASH
- ▶ Random number generator
 - PRNG (3DES based)
 - DRNG (NIST SP-800-90A SHA-512 based)
 - TRNG (true random number generator)

It provides high-performance hardware encryption, decryption, hashing, and random number generation support. The following instructions support the cryptographic assist function:

- ▶ KMAC: Compute Message Authentic Code
- ▶ KM: Cipher Message
- ▶ KMC: Cipher Message with Chaining
- ▶ KMF: Cipher Message with CFB
- ▶ KMCTR: Cipher Message with Counter
- ▶ KMO: Cipher Message with OFB
- ▶ KIMD: Compute Intermediate Message Digest
- ▶ KLMD: Compute Last Message Digest
- ▶ PCKMO: Provide Cryptographic Key Management Operation

These functions are provided as problem-state z/Architecture instructions that are directly available to application programs. These instructions are known as Message-Security Assist (MSA). When enabled, the CPACF runs at processor speed for every CP, IFL, and zIIP. For more information about MSA instructions, see *z/Architecture Principles of Operation*, SA22-7832.

For activating these functions, the *CPACF must be enabled by using feature code (FC) 3863*, which is available for no extra charge. Support for hashing algorithms SHA-1, SHA-256, SHA-384, and SHA-512 is always enabled.

6.4.1 Cryptographic synchronous functions

Because the CPACF works synchronously with the PU, it provides cryptographic synchronous functions. For IBM and client-written programs, CPACF functions can be started by using the MSA instructions. z/OS ICSF callable services on z/OS, in-kernel crypto APIs, and a *libica* cryptographic functions library that is running on Linux on IBM Z can also start CPACF synchronous functions.

The following tools might benefit from the throughput improvements for IBM z16 A02 and IBM z16 AGZ CPACF:

- ▶ Db2/IMS encryption tool
- ▶ Db2 built-in encryption
- ▶ z/OS Communication Server: IPsec/IKE/AT-TLS
- ▶ z/OS System SSL
- ▶ z/OS Network Authentication Service (Kerberos)
- ▶ DFDSS Volume encryption
- ▶ z/OS Java SDK

- ▶ z/OS Encryption Facility
- ▶ Linux on IBM Z: Kernel, openSSL, openCryptoki, and GSKIT

The IBM z16 A02 and IBM z16 AGZ hardware include the implementation of algorithms as hardware synchronous operations. This configuration holds the PU processing of the instruction flow until the operation completes.

IBM z16 A02 and IBM z16 AGZ offers the following synchronous functions:

- ▶ Data encryption and decryption algorithms for data privacy and confidentiality:
 - Data Encryption Standard (DES):
 - Single-length key DES
 - Double-length key DES
 - Triple-length key DES (also known as Triple-DES)
 - Advanced Encryption Standard (AES) for 128-bit, 192-bit, and 256-bit keys
 - Elliptic Curve supported operations:
 - ECDH[E]: P256, P384, P52, X25519, X448
 - ECDSA: Keygen, sign, verify, P256, P384, P521
 - EdDSA: Keygen, sign, verify, Ed25519, Ed448
- ▶ Hashing algorithms for data integrity, such as SHA-1 and SHA-2. New for z14 ZR1 is SHA-3 support for SHA-224, SHA-256, SHA-384, and SHA-512 and the two extendable output functions as described by the standard SHAKE-128 and SHAKE-256.
- ▶ Message authentication code (MAC):
 - Single-length key MAC
 - Double-length key MAC
- ▶ Pseudo-Random Number Generator (PRNG), Deterministic Random Number Generation (DRNG), and True Random Number Generation (TRNG) for cryptographic key generation.
- ▶ Galois Counter Mode (GCM) encryption, which is enabled by a single hardware instruction.

For the SHA hashing algorithms and the random number generation algorithms, only clear keys are used. For the symmetric encryption and decryption DES and AES algorithms and clear keys, protected keys can also be used. On IBM z16 A02 and IBM z16 AGZ, protected keys require a Crypto Express adapter that is running in CCA mode. For more information, see 6.5.2, “Crypto Express8S as a CCA coprocessor” on page 217.

The hashing algorithms SHA-1, SHA-2, and SHA-3 support for SHA-224, SHA-256, SHA-384, and SHA-512, are enabled on all systems and do not require the CPACF enablement feature. For all other algorithms, the no-charge CPACF enablement feature (FC 3863) is required.

The CPACF functions are implemented as processor instructions and require operating system support for use. Operating systems that use the CPACF instructions include z/OS, z/VM, z/VSE, z/TPF, and Linux on IBM Z.

6.4.2 CPACF protected key

IBM z16 A02 and IBM z16 AGZ support the protected key implementation. Secure keys are processed on the PCIeCC adapters (HSMs)². This process requires an asynchronous

² PCIeCC - IBM PCIe Cryptographic Coprocessor - this is the Hardware Security Module (HSM)

operation to move the data and keys from the general-purpose central processor (CP) to the crypto adapters.

Clear keys process faster than secure keys because the process is done synchronously on the CPACF. Protected keys blend the security of Crypto Express7S, Crypto Express6S, or Crypto Express5S coprocessors and the performance characteristics of the CPACF. This process allows it to run closer to the speed of clear keys.

CPACF facilitates the continued privacy of cryptographic key material when used for data encryption. In Crypto Express8S, Crypto Express7S, or Express6S coprocessors, a secure key is encrypted under a master key. However, a protected key is encrypted under a wrapping key that is unique to each LPAR.

Because the wrapping key is unique to each LPAR, a protected key cannot be shared with another LPAR. By using key wrapping, CPACF ensures that key material is not visible to applications or operating systems during encryption operations.

CPACF code generates the wrapping key and stores it in the protected area of the hardware system area (HSA). The wrapping key is accessible only by firmware. It cannot be accessed by operating systems or applications. DES/T-DES and AES algorithms are implemented in CPACF code with the support of hardware assist functions. Two variations of wrapping keys are generated: one for DES/T-DES keys and another for AES keys.

Wrapping keys are generated during the clear reset each time an LPAR is activated or reset. No customizable option is available at Support Element (SE) or Hardware Management Console (HMC) that permits or avoids the wrapping key generation. This function flow for the Crypto Express8S, Crypto Express7S, and Crypto Express6S adapters is shown in Figure 6-5.

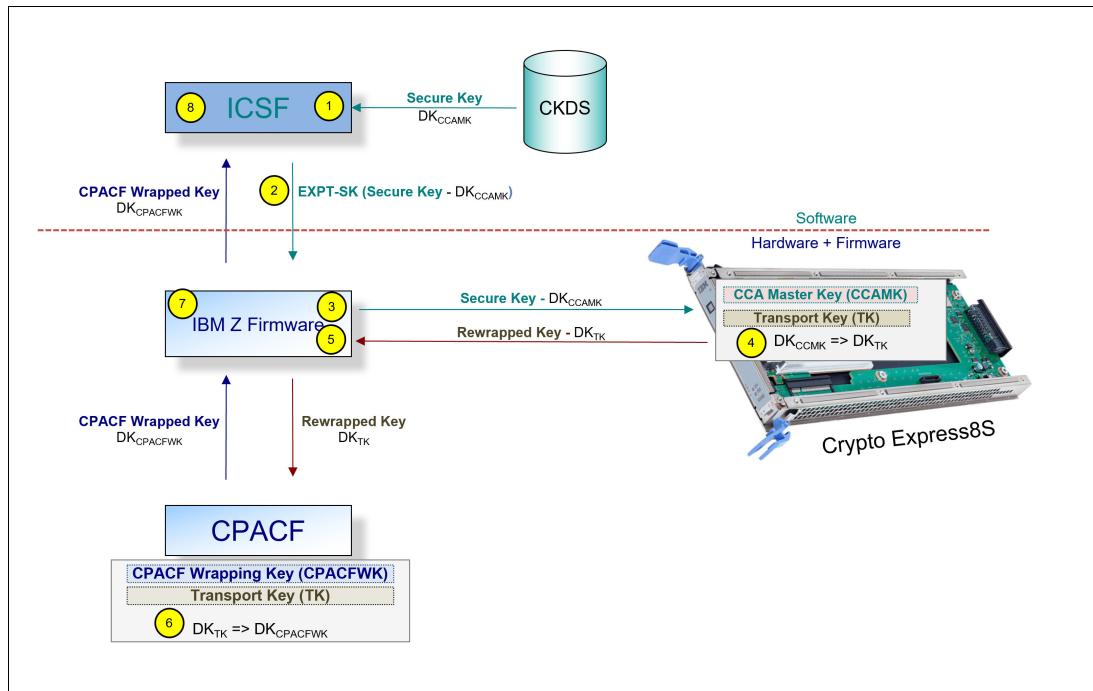


Figure 6-5 CPACF key wrapping for Express8S, Crypto Express7S, and Crypto Express6S

The CPACF Wrapping Key and the Transport Key for use with Crypto Express8S, Crypto Express7S or Crypto Express6S are in a protected area of the HSA that is not visible to operating systems or applications.

If a Crypto Express coprocessor (CEX7C, CEX6C, or CEX5C) is available, a protected key can begin its life as a secure key. Otherwise, an application is responsible for creating or loading a clear key value, and then uses the PCKMO instruction to wrap the key. ICSF is not called by the application if the CEX is not available.

A new segment in the profiles of the CSFKEYS class in IBM RACF restricts which secure keys can be used as protected keys. By default, all secure keys are considered not eligible to be used as protected keys. The process that is shown in Figure 6-5 on page 212 considers a secure key as the source of a protected key.

The source key in this case is stored in the ICSF Cryptographic Key Data Set (CKDS) as a secure key, which was encrypted under the master key. This secure key is sent to CEX8C, CEX7C, or CEX6C to be deciphered and then, sent to the CPACF in clear text. At the CPACF, the key is wrapped under the LPAR wrapping key, and is then returned to ICSF. After the key is wrapped, ICSF can keep the protected value in memory. It then passes it to the CPACF, where the key is unwrapped for each encryption or decryption operation.

The protected key is designed to provide substantial throughput improvements for a large volume of data encryption and low latency for encryption of small blocks of data. A high-performance secure key solution, also known as a *protected key solution*, requires the ICSF HCR7770 as a minimum release.

6.5 Crypto Express8S

The Crypto Express8S feature (FC 0908 or FC 0909) is an optional feature that is exclusive to IBM z16 A02 and IBM z16 AGZ. Each feature FC 0909 has one IBM 4770 PCIe cryptographic adapter (hardware security module - HSM), whereas FC 0908 has two IBM 4770 PCIe cryptographic adapters (two HSMs). The Crypto Express8S (CEX8S) feature occupies one I/O slot in PCIe+ I/O drawer. This feature provides one or two HSMs and for a secure programming and hardware environment on which crypto processes are run.

Each cryptographic coprocessor includes a general-purpose processor, non-volatile storage, and specialized cryptographic electronics. The Crypto Express8S feature provides tamper-sensing and tamper-responding, high-performance cryptographic operations.

Each Crypto Express8S PCI Express adapter is available in one of the following configurations:

- ▶ Secure IBM CCA coprocessor (CEX8C) - This configuration includes secure key functions. It is optionally programmable to deploy more functions and algorithms by using UDX. For more information, see 6.5.2, "Crypto Express8S as a CCA coprocessor" on page 217.

A TKE workstation is required to support the administration of the Crypto Express8S when it is configured in CCA mode when in full PCI³-compliant mode for the necessary certificate management in this mode. The TKE is optional in all other use cases for CCA.

- ▶ Secure IBM Enterprise PKCS #11 (EP11) coprocessor (CEX8P) implements 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 public sector requirements. This new cryptographic coprocessor mode introduced the PKCS #11 secure key function. For more information, see 6.5.3, "Crypto Express8S as an EP11 coprocessor" on page 223.

³ Payment Card Industry

A TKE workstation is always required to support the administration of the Crypto Express7S when it is configured in EP11 mode.

- ▶ Accelerator (CEX8A) for acceleration of public key and private key cryptographic operations that are used with SSL/TLS processing. For more information, see 6.5.4, “Crypto Express8S as an accelerator” on page 224.

These modes can be configured by using the SE. The PCIe adapter must be configured offline to change the mode.

Attention: Switching between configuration modes erases all adapter secrets. The exception is when you are switching from Secure CCA to accelerator, and vice versa.

The Crypto Express8S feature is released for enhanced cryptographic performance. Clients who migrated to variable-length AES key tokens cannot take advantage of faster encryption speeds by using CPACF. Support is being added to translate a secure variable-length AES CIPHER token to a protected key token (protected by the system wrapping key). This support allows for faster AES encryption speeds when variable-length tokens are used while maintaining strong levels of security.

The Crypto Express8S feature does not include external ports and does not use optical fiber or other cables. It does not use channel path identifiers (CHIDs), but requires one slot in the PCIe I/O drawer and one physical channel ID (PCHID) for each PCIe cryptographic adapter. Removal of the feature or adapter *zeroizes* its content. Access to the PCIe cryptographic adapter is controlled through the setup in the image profiles on the SE.

Adapter: Although PCIe cryptographic adapters include no CHID type and are not identified as external channels, all logical partitions (LPARs) in all channel subsystems can access the adapter. In IBM z16 A02 or IBM z16 AGZ, up to 40 LPARs are supported per adapter. Accessing the adapter requires a setup in the image profile for each partition. The adapter must be in the candidate list.

Each IBM z16 A02 and IBM z16 AGZ supports up to 40 Hardware Security Modules in total (a combination of Crypto Express8S (1 or 2 HSM), Crypto Express7S (1 or 2 port), and Crypto Express6S). Crypto Express7S (1 or 2 port) and Crypto Express6S features (single HSM) *are not orderable* for a new build IBM z16 A02 or IBM z16 AGZ but can be carried forward from an IBM z14 ZR1 or IBM z15 T02 by using an MES. Configuration information for Crypto Express7S is listed in Table 6-2.

Table 6-2 *Crypto features supported on IBM A02 and IBM z16 AGZ*

Feature	Quantity
Minimum number of orderable features 0908 for IBM z16 A02 and IBM z16 AGZ	2
Minimum number of orderable features 0909 for IBM z16 A02 and IBM z16 AGZ ^a	2
Order increment (above two features for features 0908 and 0909)	1
Maximum number of HSMs for IBM z16 A02 and IBM z16 AGZ (combining all CEX8S, CEX7S, and CEX6S)	40 ^b
Number of PCIe cryptographic adapters for each feature 0908 (coprocessor or accelerator)	2
Number of PCIe cryptographic adapters for each feature 0909 (coprocessor or accelerator)	1
Number of cryptographic domains at IBM z16 A02 and IBM z16 AGZ for each PCIe adapter ^c	40

- a. The minimum initial order of Crypto Express8S feature 0909 is two. After the initial order, more Crypto Express7S features can be ordered one feature individually, for a total of 40 HSMs (combined).
- b. Crypto Express8S (dual HSM) has two hardware security modules (HSMs) per feature. The HSM is one IBM 4770 PCIe Cryptographic Coprocessor (PCIeCC). The max. number of HSMs per IBM z16 A02 or IBM z16 AGZ, combining all cryptographic features is 40, while the max. number of single HSM (port) cryptographic features is 16 (CEX8S (single HSM), CEX7S (1 port), and CEX6S)
- c. More than one partition, which is defined to the same channel subsystem (CSS) or to different CSSs, can use the same domain number when assigned to different PCIe cryptographic adapters.

The concept of *dedicated processor* does not apply to the PCIe cryptographic adapter. Whether configured as a coprocessor or an accelerator, the PCIe cryptographic adapter is made available to an LPAR. It is made available as directed by the domain assignment and the candidate list in the LPAR image profile. This availability is not changed by the shared or dedicated status that is given to the PUs in the partition.

When installed non-concurrently, Crypto Express8S features are assigned PCIe cryptographic adapter numbers sequentially during the power-on reset (POR) that follows the installation. When a Crypto Express8S feature is installed concurrently, the installation can select an out-of-sequence number from the unused range. When a Crypto Express8S (or Crypto Express7S or Crypto Express6S) feature is removed concurrently, the PCIe adapter numbers are automatically freed.

The definition of domain indexes and PCIe cryptographic adapter numbers in the candidate list for each LPAR must be planned to allow for nondisruptive changes. Consider the following points:

- ▶ Operational changes can be made by using the Change LPAR Cryptographic Controls task from the SE, which reflects the cryptographic definitions in the image profile for the partition. With this function, adding and removing the cryptographic feature without stopping a running operating system can be done dynamically.
- ▶ The same usage domain index can be defined more than once across multiple LPARs. However, the PCIe cryptographic adapter number that is coupled with the usage domain index that is specified must be unique across all active LPARs.

The same PCIe cryptographic adapter number and usage domain index combination can be defined for more than one LPAR (up to 40 for IBM z16 A02 and IBM z16 AGZ). For example, you might define a configuration for backup situations. However, only one of the LPARs can be active at a time.

For more information, see 6.5.5, “Managing Crypto Express8S” on page 224.

6.5.1 Cryptographic asynchronous functions

The optional PCIe cryptographic coprocessors Crypto Express8S provides asynchronous cryptographic functions to IBM z16 A02 and IBM z16 AGZ. Over 300 Cryptographic algorithms and modes are supported, including the following algorithms and modes:

- ▶ DES/TDES w DES/TDES MAC/CMAC: The Data Encryption Standard is a widespread symmetrical encryption algorithm. DES, along with its double-length and triple length variations, TDES today are considered to be not sufficient secure for many applications. They were replaced by the AES as the official US standard, but it is still used in the industry with the MAC and the Cipher-based Message Authentication Code (CMAC) for verifying the integrity of messages. The Enhanced Wrapping Method for TDES key tokens

WRAPENH3 is supported, which is very important for independently reviewed key block protection for keys used in PCI PIN audited workloads.

- ▶ AES, AESKW, AES GMAC, AES GCM, AES XTS, AES CIPHER mode, and CMAC: AES replaced DES as the official US standard in October 2000. The enhanced standards for AES Key Warp (AESKW), the AES Galois Message Authentication Code (AES GMAC) and Galois/Counter Mode (AES GCM), the XEX-based tweaked-codebook mode with ciphertext stealing (AES XTS), CMAC, AES-DUKPT (unique key per transaction for AES-based PIN and transaction protection) are supported.
- ▶ MD5, SHA-1, SHA-2, or SHA-3⁴ (224, 256, 384, and 512), and HMAC: The Secure Hash Algorithm (SHA-1 and the enhanced SHA-2 or SHA-3 for different block sizes), the older message-digest (MD5) algorithm, and the advanced keyed-hash method authentication code (HMAC) are used for verifying the data integrity and the authentication of a message.
- ▶ VFPE: A method of encryption in which the resulting cipher text features the same form as the input clear text, which is developed for use with credit cards. Three algorithms are standardized by NIST and in X9.124: FF1, FF2 and FF2.1
- ▶ RSA (512, 1024, 2048, and 4096): RSA was published in 1977. It is widely used asymmetric public-key algorithm, which means that the encryption key is public whereas the decryption key is kept secret. It is based on the difficulty of factoring the product of two large prime numbers. The number describes the length of the keys.
- ▶ ECDSA (192, 224, 256, 384, and 521 Prime/NIST): ECC is a family of asymmetric cryptographic algorithms that are based on the algebraic structure of elliptic curves. ECC can be used for encryption, pseudo-random number generation, and digital certificates. The Elliptic Curve Digital Signature Algorithm (ECDSA) Prime/NIST method is used for ECC digital signatures, which are recommended for government use by NIST.
- ▶ ECDSA (160, 192, 224, 256, 320, 384, and 512 BrainPool): ECC BrainPool is a workgroup of companies and institutions that collaborate on developing ECC algorithms. The ECDSA algorithms that are recommended by this group are supported.
- ▶ ECDH (192, 224, 256, 384, and 521 Prime/NIST): Elliptic Curve Diffie-Hellman (ECDH) is an asymmetric protocol that is used for key agreement between two parties by using ECC-based private keys. The recommendations by NIST are supported.
- ▶ ECDH (160, 192, 224, 256, 320, 384, and 512 BrainPool): ECDH according to the BrainPool recommendations.
- ▶ Montgomery Modular Math Engine: The Montgomery Modular Math Engine is a method for fast modular multiplication. Many crypto systems, such as RSA and Diffie-Hellman key Exchange, can use this method.
- ▶ Random Number Generator (RNG): The generation of random numbers for cryptographic key generation is supported.
- ▶ Prime Number Generator (PNG): The generation of prime numbers is also supported.
- ▶ Clear Key Fast Path (Symmetric and Asymmetric): This mode of operation gives a direct hardware path to the cryptographic engine and provides high performance for public-key cryptographic functions.

Several of these algorithms require a secure key and must run on an HSM. Some of these algorithms can also run with a clear key on the CPACF. Many standards are supported only when Crypto Express8S is running in CCA mode. Others are supported only when the adapter is running in EP11 mode.

⁴ SHA-3 was standardized by NIST in 2015. SHA-2 is still acceptable and no indication exists that SHA-2 is vulnerable or that SHA-3 is more or less vulnerable than SHA-2.

The three modes for Crypto Express features are described next. For more information, see 6.7, “Cryptographic functions comparison” on page 231.

6.5.2 Crypto Express8S as a CCA coprocessor

A Crypto Express8S adapter that is running in CCA mode supports IBM CCA. CCA is an architecture and a set of APIs. It provides cryptographic algorithms, secure key management, and many special functions that are required for banking. Over 129 APIs with more than 600 options are provided, with new functions and algorithms always being added.

The IBM CCA provides functions for the following tasks:

- ▶ Encryption of data (DES/TDES/AES)
- ▶ Key management:
 - Using TDES or AES keys
 - Using RSA or Elliptic Curve keys
- ▶ Message authentication for MAC/HMAC/AES-CMAC
- ▶ Key generation
- ▶ Digital signatures
- ▶ Random number generation
- ▶ Hashing (SHA, MD5, and others)
- ▶ ATM PIN generation and processing
- ▶ Credit card transaction processing
- ▶ Visa Data Secure Platform (DSP) Point to Point Encryption (P2PE)
- ▶ Europay, MasterCard, and Visa (EMV) card transaction processing
- ▶ Card personalization
- ▶ Other financial transaction processing
- ▶ Integrated role-based access control system
- ▶ Compliance support for:
 - All DES services
 - AES services
 - RSA services, including full use of X.509 certificates
- ▶ TR-34 Remote Key Load

User-defined extensions support

User-defined extension (UDX) allows a developer to add customized operations to IBM's CCA Support Program. UDXs to the CCA support customized operations that run within the Crypto Express features when defined as a coprocessor.

UDX is supported under a special contract through an IBM or approved third-party service offering. The Crypto Cards website directs your request to an IBM Global Services location for your geographic location. A special contract is negotiated between IBM Global Services and you for the development of the UDX code by IBM Global Services according to your specifications and an agreed-upon level of the UDX.

A UDX toolkit for IBM zSystems is tied to specific versions of the CCA code and the related host code. UDX is available for the Crypto Express7S and Crypto Express6S (Secure IBM CCA coprocessor mode only) features. An UDX migration is no more disruptive than a normal Microcode Change Level (MCL) or ICSF release migration.

In IBM z16 A02 and IBM z16 AGZ, up to four UDX files can be imported. These files can be imported from a USB media stick or an FTP server. The UDX configuration window is updated to include a Reset to IBM Default button.

Consideration: CCA features a new code level starting with z13/z13s systems, and the UDX clients require a new UDX.

On IBM z16 A02 and IBM z16 AGZ, Crypto Express8S is delivered with CCA Level 8.1 firmware. A new set of cryptographic functions and callable services is provided by the IBM CCA LIC to enhance the functions that secure financial transactions and keys. The Crypto Express8S includes the following features:

- ▶ Greater than 16 domains support up to 40 LPARs on IBM z16 A02 and IBM z16 AGZ.
- ▶ Payment Card Industry (PCI) PIN Transaction Security (PTS) HSM Certification that is available to IBM z16 A02 and IBM z16 AGZ in combination with CEX8S, CEX7S or CEX6S features, to IBM z15 in combination with CEX7S or CEX6S features, and to IBM z14 with CEX6S features.
- ▶ VFPE support, which was introduced with z13/z13s systems.
- ▶ AES PIN support for the German banking industry.
- ▶ PKA Translate UDX function into CCA.
- ▶ Verb Algorithm Currency.

CCA improvements

- ▶ CCA Quantum Safe Algorithm enhancements
 - updated support for Dilithium signatures
 - Round 2: Level 2 (6 5) and 3 (8 7)
 - Round 3: Level 3 (6 5) and 5 (8 7)
 - add support for Kyber key encapsulation
 - Round 2: Level 5 (1024)
- ▶ Quantum Safe protected key support for CCA
 - Host Firmware and CCA now employ a hybrid scheme combining ECDH and Kyber to accomplish a quantum safe transport key exchange for protected key import.

CCA 8.1 improvements

- ▶ Local/native support for ANSI X9.143-2022 / ASC X9 TR-31-2018 key blocks in operational use with the CCA interface. AES, DES/TDES, HMAC keys are supported. TR-31 key blocks can be created and used with the CCA API in over 70 interfaces that also use CCA proprietary key blocks. Key storage is also updated for TR-31 key block support.

CCA 8.0 improvements

- ▶ ASC X9 TR-31 key exchange support update in CSNBT31X and CSNBT31I: Prepares customers for payment network mandate to use enhanced TR-31 wrapping method 'B' when exchanging keys with major payment card brand networks such as Visa and MasterCard.
- ▶ Performance enhancement for mixed workloads: better performance when one partition focuses on RSA/ECC and another partition focuses on AES/DES/TDES or financial operations
- ▶ Hardware accelerated key unwrap for AES wrapped keys
 - Trusted Key Entry workstation (TKE) controlled selection of WRAPENH3 as the default TDES key token wrapping method for easier management.

CCA 7.4 and CCA 6.7 improvements

- ▶ German banking API updates for program currency

- ▶ X9.23 random padding for AES encryption for important cryptographic operation protection
- ▶ Enhanced triple length TDES PIN encryption key support for PIN change workloads
- ▶ New service to compare encrypted PINs, required for ISO 4 PIN block verification inside the HSM
- ▶ EC SDSA signature support useful for new EMV certificate formats
- ▶ PKCS#11 update to CCA API export of AES and RSA keys using RSA public key and AES ephemeral keys, for key exchange with Cloud service key management APIs
- ▶ Australian Payment Network Acquirer function for key derivation and MAC chaining added for interoperability with Australian audited payment networks

CCA 7.3 and CCA 6.6 improvements

- ▶ WRAPENH3 Enhanced Wrapping Method for TDES key tokens

CCA 7.2 and CCA 6.5 improvements

- ▶ AES DUKPT unique key per transaction for AES based PIN and transaction protection
- ▶ ISO 4 enhanced support adding AES protected PIN block support to all remaining services that process PIN blocks
- ▶ Format Preserving Encryption (FPE) support 3 algorithms standardized by NIST and in X9.124: FF1, FF2, and FF2.1.
- ▶ Elliptic Curve support for the Koblitz curve secp256k1, to all ECC services including native support for X.509 certificates

CCA Version 7.1 improvements

- ▶ Supported curves:
 - NIST Prime Curves: P192, P224, P256, P384, P521
 - BrainPool Curves: 160, 192, 224, 256, 320, 384, 512
- ▶ Support in the CCA coprocessor for these Edwards curves:
 - ED25519 (128-bit security strength) and ED448 (224-bit security strength).
 - ED25519 is faster but ED448 is more secure. Practically though, 128-bit security strength is very secure.
- ▶ Edwards curves are used for digitally signing documents and verifying those signatures. They also are less susceptible to side channel attacks when compared to Prime and BrainPool curves.
- ▶ ECC Protected Keys

Crypto Express7S provides support in CCA coprocessors to take advantage of fast DES, AES data encryption speeds in CPACF while maintaining high levels of security for the secure key material. The key remains encrypted and the key encrypting key never appears in host storage.

When using CCA ECC services, ICSF can now take advantage of ECC support in CPACF (protected key support) for these curves:

- Prime: P256, P384, P521
- Edwards: ED25519, ED448

CPACF can achieve much faster crypto speeds compared to the coprocessor

The translation to protected key happens automatically once the attribute is set in the key token. No application change is required.

- ▶ New signatures

Support for the Cryptographic Suite for Algebraic Lattices signatures algorithm with the largest key sizes (MODE=3)

- Public Key size: 1760 bytes
- Private Key Size: 3856 bytes
- Signature Size: 3366 bytes

Lattice-based cryptographic keys will be protected by the 256-bit AES MK. The lattice-based key has a security strength of 128 bits.

► TR-31 for Hash-based Message Authentication Code (HMAC)

HMAC keys are used to verify the integrity and authenticity of a message. This support provides a standard method of exchanging HMAC keys with a partner using symmetric key techniques. The key is exchanged in the standard TR-31 key block format which can be consumed by any crypto system supporting the standard

CCA Version 6.3 improvements⁵

- Compliance support for:
 - All DES services
 - AES services
 - RSA services, including full use of X.509 certificates
- TR-34 Remote Key Load

Greater than 16 domains support

IBM z16 A02 and IBM z16 AGZ supports up to 40 LPARs. The early IBM zSystems crypto architecture was designed to support 16 domains, which matched the LPAR maximum at the time. Before IBM z13 systems, crypto workload separation can be complex in customer environments where the number of LPARs was larger than 16. These customers mapped a large set of LPARs to a small set of crypto domains.

Starting with IBM z14, the IBM zSystems crypto architecture can support up to 256 domains in an adjunct processor (AP) with the AP extended addressing (APXA) facility that is installed. As such, the Crypto Express adapters are enhanced to handle 256 domains. The IBM system z firmware provides up to 40 domains for IBM z16 A02 and IBM z16 AGZ to customers (to match the current LPAR maximum). Customers can map individual LPARs to unique crypto domains or continue to share crypto domains across LPARs.

The following requirements must be met to support 40 domains:

- Hardware: IBM z16 A02 and IBM z16 AGZ and Express8S, Crypto Express7S, or Crypto Express6S.
- Operating systems:
 - z/OS
 - New ICSF support is required to administer a CEX8 coprocessor using a TKE workstation, due to exploitation of quantum algorithms. Otherwise, existing workloads will run on IBM z16 A02 and IBM z16 AGZ without requiring ICSF support.
 - Exploitation of new function is supplied in ICSF PTFs on z/OS V2.2 V2.4 (Web deliverable HCR77D1) or V2.5 (base, which is HCR77D2)
 - When exploiting new Quantum Safe Algorithms and sharing a KDS in a sysplex, ensure all ICSF PTFs are installed on all systems.

⁵ A TKE is required to manage a PCI-compliant coprocessor and for certificate management

Tip: All supported levels of ICSF automatically detect what HW cryptographic capabilities are available where it is running, then enables functions accordingly. No toleration of new HW is necessary. If you want to exploit new capabilities, then ICSF support is necessary.

- z/VM Version 6.4 and 7.1 with PTFs or newer for guest use.

Payment Card Industry-HSM certification

Payment Card Industry (PCI) standards are developed to help ensure security in the PCI. PCI defines their standards as a set of security standards that is designed to ensure that all companies that accept, process, store, or transmit credit card information that is maintained a secure environment.

Compliance with the PCI-HSM standard is valuable for customers, particularly those customers who are in the banking and finance industry. This certification is important to clients for the following fundamental reasons:

- ▶ Compliance is increasingly becoming mandatory.
- ▶ The requirements in PCI-HSM make the system more secure.

Industry requirements for PCI-HSM compliance

The PCI organization cannot require compliance with its standards. Compliance with PCI standards is enforced by the payment card brands, such as Visa, Master Card, American Express, JCB International, and Discover.

If you are a bank, acquirer, processor, or other participant in the payment card systems, the card brands can impose requirements on you if you want to process their cards. One set of requirements they are increasingly enforcing is the PCI standards.

The card brands work with PCI in developing these standards, and they focused first on the standards they considered most important, particularly the PCI Data Security Standard (PCI-DSS). Some of the other standards were written or required later, and PCI-HSM is one of the last standards to be developed. In addition, the standards themselves were increasing the strength of their requirements over time. Some requirements that were optional in earlier versions of the standards are now mandatory.

In general, the trend is for the card brands to enforce more of the PCI standards and to enforce them more rigorously. The trend in the standards is to impose more and stricter requirements in each successive version. The net result is that companies subject to these requirements can expect that they eventually must comply with all of the requirements.

Improved security through use of PCI-HSM

PCI-HSM was developed primarily to improve security in payment card systems. It imposes requirements in key management, HSM API functions, and device physical security. It also controls during manufacturing and delivery, device administration, and several other areas. It prohibits many things that were in common use for many years, but are no longer considered secure.

The result of these requirements is that applications and procedures often must be updated because they used some of the things that are now prohibited. Although this issue is inconvenient and imposes some costs, it does increase the resistance of the systems to attacks of various kinds. Updating a system to use PCI-HSM compliant HSMs is expected to reduce the risk of loss for the institution and its clients.

The following requirements must be met to use PCI-HSM:

- ▶ Hardware: IBM z16 A02 or IBM z16 AGZ⁶ and Crypto Express8S, Crypto Express7S, or Crypto Express6S
- ▶ Operating systems:
 - z/OS - ICSF Web deliverable 19 (HCR77D1), unless otherwise noted. WD19 supports z/OS V2R2, V2R3, and V2R4.
 - WD 20 supports z/OS V2R5 (base, which is HCR77D2)
 - z/VM Version 7.1 or newer for guest use

Visa Format Preserving Encryption

VFPE refers to a method of encryption in which the resulting cipher text features the same form as the input clear text. The form of the text can vary according to use and application. One of the classic examples is a 16-digit credit card number. After VFPE is used to encrypt a credit card number, the resulting cipher text is another 16-digit number. This process helps older databases contain encrypted data of sensitive fields without having to restructure the database or applications.

VFPE allows customers to add encryption to their applications in such a way that the encrypted data can flow through their systems without requiring a massive redesign of their application. In our example, if the credit card number is VFPE-encrypted at the point of entry, the cipher text still behaves as a credit card number. It can flow through business logic until it meets a back-end transaction server that can VFPE-decrypt it to get the original credit card number to process the transaction.

Note: VFPE technology forms part of Visa, Inc.'s, Data Secure Platform (DSP). The use of this function requires a service agreement with Visa. You must maintain a valid service agreement with Visa when you use DSP/FPE.

AES PIN support for the German banking industry

The German banking industry organization, DK, defined a new set of PIN processing functions to be used on the internal systems of banks and their servers. CCA is designed to support the functions that are essential to those parts of the German banking industry that are governed by DK requirements. The functions include key management support for new AES key types, AES key derivation support, and several DK-specific PIN and administrative functions.

This support includes PIN method APIs, PIN administration APIs, new key management verbs, and new access control points support that is needed for DK-defined functions.

Support for the updated German Banking standard (DK)

Update support requires ICSF WD19 (HCR77D1) for z/OS V2R2, V2R3, and V2R4.

PKA Translate UDX function into CCA

UDX is custom code that allows the client to add unique operations or extensions to the CCA firmware. Certain UDX functions are integrated into the base CCA code over time to accomplish the following tasks:

- ▶ Remove headaches and challenges that are associated with UDX management and currency.
- ▶ Make available popular UDX functions to a wider audience to encourage adoption.

⁶ Always check the latest information about security certification status for your specific model.

UDX is integrated into the base CCA code to support translating an external RSA CRT key into new formats. These formats use tags to identify key components. Depending on which new rule array keyword is used with the PKA Key Translate callable service, the service TDES encrypts those components in CBC or ECB mode. In addition, AES CMAC support is delivered.

Verb Algorithm Currency

Verb Algorithm Currency is a collection of CCA verb enhancements that are related to customer requirements, with the intent of maintaining currency with cryptographic algorithms and standards. It is also intended for customers who want to maintain the following latest cryptographic capabilities:

- ▶ Secure key support AES GCM encryption
- ▶ Key Check Value (KCV) algorithm for service CSNBKYT2 Key Test 2
- ▶ Key derivation options for CSNDEDH EC Diffie-Hellman service

6.5.3 Crypto Express8S as an EP11 coprocessor

A Crypto Express8S adapter that is configured in Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode provides PKCS #11 secure key support for public sector requirements. Before EP11, the ICSF PKCS #11 implementation supported only clear keys. In EP11, keys can now be generated and securely wrapped under the EP11 Master Key. The secure keys never leave the secure coprocessor boundary decrypted.

The secure IBM Enterprise PKCS #11 (EP11) coprocessor runs the following tasks:

- ▶ Encrypt and decrypt (AES, DES, TDES, and RSA)
- ▶ Sign and verify (DSA, RSA, EdDSA, ECDSA and ECSDSA)
- ▶ Generate keys and key pairs (DES, AES, DSA, ECC, and RSA)
- ▶ HMAC (SHA1, SHA2 or SHA3 [SHA224, SHA256, SHA384, and SHA512])
- ▶ Digest (SHA1, SHA2 or SHA3 [SHA224, SHA256, SHA384, and SHA512])
- ▶ Wrap and unwrap keys
- ▶ Random number generation
- ▶ Get mechanism list and information
- ▶ Attribute values
- ▶ Key Agreement (Diffie-Hellman)

The function extension capability through UDX is not available to the EP11.

When defined in EP11 mode, the TKE workstation is required to manage the Crypto Express features.

Enterprise PKCS #11 (EP11) with IBM z16 A02 and IBM z16 AGZ provides the following updates⁷:

- ▶ Quantum Safe Algorithm enhancements provides:
 - Updated support for Dilithium signatures
 - Round 2: Level 2 (6 5) and 3 (8 7)
 - Round 3: Level 2 (4 4), 3 (6 5) and 5 (8 7)
 - Add support for Kyber key encapsulation
 - Round 2: Level 3 (768) and 5 (1024)
- ▶ Quantum Safe protected key support for EP11
 - Host Firmware and EP11 now employ a hybrid scheme combining ECDH and Kyber to accomplish a quantum safe transport key exchange for protected key import
- ▶ Quantum Safe host firmware management support for EP11

⁷ At the time of this writing (December 2022)

- Host Firmware and EP11 now employ a hybrid scheme for authenticating management functions initiated from the SE/HMC
- ▶ EP11 for all of CEX8S (5.8.x), CEX7S (4.8.x) and CEX6S (3.8.x)
 - Support for HSM backed Hierarchical Deterministic Wallets for Bitcoin (BIP 0032 and SLIP 0010)
 - Hash collision resistant Schnorr signature scheme BSI TR 03111, two variants:
 - plain BSI TR 03111
 - with compressed keys and signing party's public key as additional input
 - Support for Edwards and Montgomery elliptic curves: EdDSA (Ed25519, Ed448) and ECDH (X25519, X448) (8s and 7s)
 - RSA OAEP with SHA 2 and SHA 3 (8s and 7s only)
 - Extensive IBM Cloud Crypto support:
 - Domains fully manageable by clients without cloud admin's assistance
 - Do not Disturb: actively prohibit cloud admin's from domain management
 - HSM internal re encrypt support for block based cipher modes
- ▶ EP11 for CEX8S (5.8.x) only
 - Enhanced concurrent update support now includes kernel modules
 - Enhanced maximum performance for digest and random number generation
 - Allow for regular extractable keys to be tagged as protected key exportable
- ▶ EP11 Support Program 4.0 for Linux
 - enables PQC Dilithium and Kyber exploitation by Linux/opencryptoki
 - adds TLS support for TKE to ep11tked connections

6.5.4 Crypto Express8S as an accelerator

A Crypto Express8S adapter that is running in accelerator mode supports only RSA clear key and SSL Acceleration. A request is processed fully in hardware. The Crypto Express accelerator is a coprocessor that is reconfigured by the installation process so that it uses only a subset of the coprocessor functions at a higher speed. Reconfiguration is disruptive to coprocessor and accelerator operations. The coprocessor or accelerator must be deactivated before you begin the reconfiguration.

FIPS 140 certification is not relevant to the accelerator because it operates with clear keys only. The function extension capability through UDX is not available to the accelerator.

The functions that remain available when the Crypto Express6S feature is configured as an accelerator are used for the acceleration of modular arithmetic operations. That is, the RSA cryptographic operations are used with the SSL/TLS protocol. The following operations are accelerated:

- ▶ PKA Decrypt (CSNDPKD) with PKCS-1.2 formatting
- ▶ PKA Encrypt (CSNDPKE) with zero-pad formatting
- ▶ Digital Signature Verify

The RSA encryption and decryption functions support key lengths of 512 - 4,096 bits in the Modulus-Exponent (ME) and Chinese Remainder Theorem (CRT) formats.

6.5.5 Managing Crypto Express8S

Each cryptographic coprocessor has 40 physical sets of registers or queue registers, which corresponds to the maximum number of LPARs that are running on IBM z16 A02 and IBM z16 AGZ, which is also 40. Each of these sets belongs to the following domains:

- ▶ A cryptographic domain index, in the range of 0 - 39 for IBM z16 A02 and IBM z16 AGZ, is allocated to a logical partition in its image profile. The same domain must also be allocated to the ICSF instance that is running in the logical partition that uses the Options data set.

- ▶ Each ICSF instance accesses only the Master Keys or queue registers that correspond to the domain number that is specified in the logical partition image profile at the SE and in its Options data set. Each ICSF instance sees a logical cryptographic coprocessor that consists of the physical cryptographic engine and the unique set of registers (the domain) that is allocated to this logical partition.

The installation of CP Assist for Cryptographic Functions (CPACF) DES/TDES enablement (FC 3863) is required to use the Crypto Express8S feature.

Each Crypto Express8S FC 0908 includes two IBM 4770 PCIe Cryptographic Coprocessors (PCIeCC - which is a hardware security module - HSM); FC 0909 includes one IBM 4770 PCIeCC. The adapters are available in the following configurations:

- ▶ IBM Enterprise Common Cryptographic Architecture (CCA) Coprocessor (CEX8C)
- ▶ IBM Enterprise Public Key Cryptography Standards #11 (PKCS) Coprocessor (CEX8P)
- ▶ IBM Crypto Express7S Accelerator (CEX8A)

During the feature installation, the PCIeCC is configured by default as the CCA coprocessor.

The configuration of the Crypto Express8S adapter as EP11 coprocessor requires a TKE workstation (FC 0057/0058) with TKE 10.0 (FC 0882) LIC. The same requirement applies to CCA mode for a full PCI-compliant environment

The Crypto Express8S feature does not use CHPIIDs from the channel subsystem pool. However, the Crypto Express8S feature requires one slot in a PCIe+ I/O drawer, and one PCHID for each PCIe cryptographic adapter.

For enabling an LPAR to use a Crypto Express7S adapter, the following cryptographic resources in the image profile must be defined for each partition:

- ▶ Usage domain index
- ▶ Control domain index
- ▶ PCI Cryptographic Coprocessor Candidate List
- ▶ PCI Cryptographic Coprocessor Online List

This task is accomplished by using the Customize/Delete Activation Profile task, which is in the Operational Customization Group, from the HMC or from the SE. Modify the cryptographic initial definition from the Crypto option in the image profile, as shown in Figure 6-6 on page 226.

Important: After this definition is modified, any change to the image profile requires a DEACTIVATE and ACTIVATE of the logical partition for the change to take effect. Therefore, this cryptographic definition is disruptive to a running system.

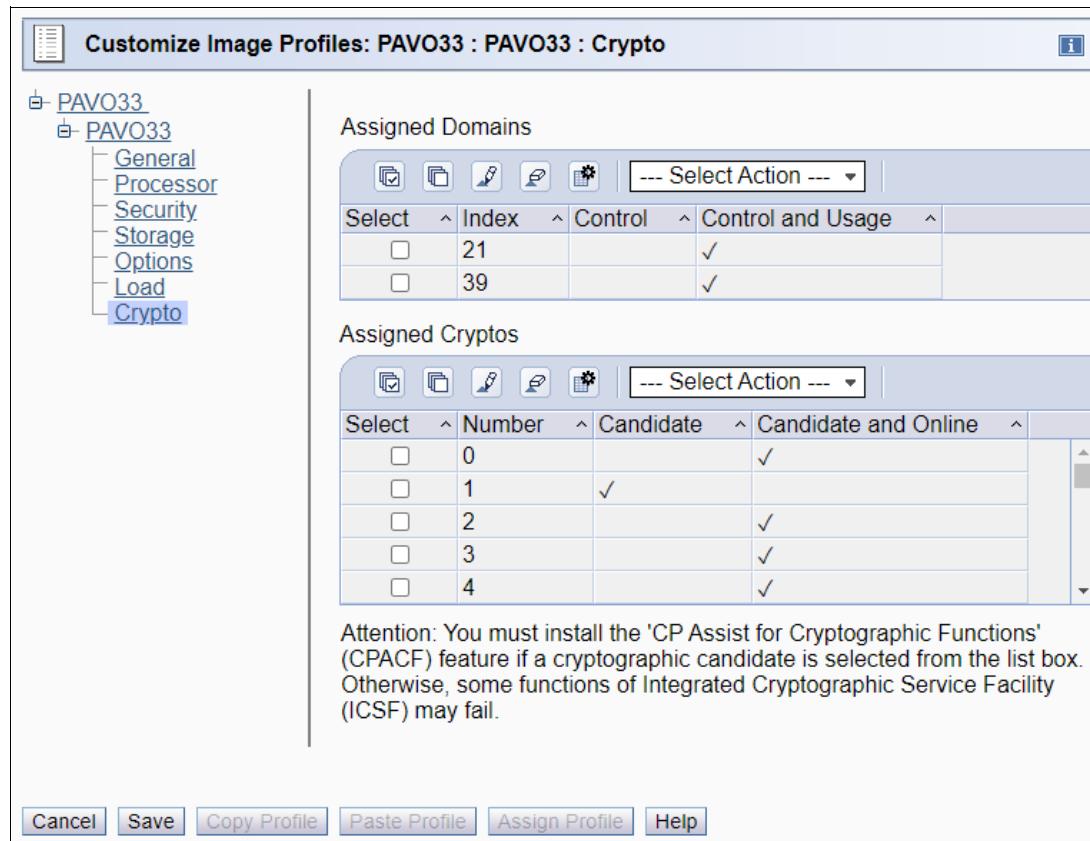


Figure 6-6 Customize Image Profiles: Crypto

The following cryptographic resource definitions are used:

- ▶ Control Domain

Identifies the cryptographic coprocessor domains that can be administered from this logical partition if it is set up as the TCP/IP host for the TKE.

If you are setting up the host TCP/IP in this logical partition to communicate with the TKE, the partition is used as a path to other domains' Master Keys. Indicate all the control domains that you want to access (including this partition's own control domain) from this partition.

- ▶ Control and Usage Domain

Identifies the cryptographic coprocessor domains that are assigned to the partition for all cryptographic coprocessors that are configured on the partition. The usage domains cannot be removed if they are online. The numbers that are selected must match the domain numbers that are entered in the Options data set when you start this partition instance of ICSF.

The same usage domain index can be used by multiple partitions, regardless to which CSS they are defined. However, the combination of PCIe adapter number and usage domain index number must be unique across all active partitions.

- ▶ Cryptographic Candidate list

Identifies the cryptographic coprocessor numbers that can be accessed by this logical partition. From the list, select the coprocessor numbers (in the range 0 - 15) that identify the PCIe adapters to be accessed by this partition.

- ▶ Cryptographic Online list

Identifies the cryptographic coprocessor numbers that are automatically brought online during logical partition activation. The numbers that are selected in the online list must also be part of the candidate list.

After they are activated, the active partition cryptographic definitions can be viewed from the HMC. Select the CPC, and click **View LPAR Cryptographic Controls** in the CPC Operational Customization window. The resulting window displays the definition of Usage and Control domain indexes, and PCI Cryptographic candidate and online lists, as shown in Figure 6-7. Information is provided for active logical partitions only.

The screenshot shows the 'View LPAR Cryptographic Controls - PAVO' window. At the top, it displays 'Installed Crypto Express8S: 00 01 02 03 04 05 06 07'. Below this are two tables:

Partition	Active	Crypto Numbers	Conflicts
PAVO31	Yes		
PAVO32	Yes		
PAVO33	Yes	0-4	
PAVO34	Yes		
PAVO35	Yes		
PAVO36	Yes		
PAVO37	No		
PAVO38	Yes		
PAVO39	Yes		
PAVO41	No		
PAVO42	No		
PAVO43	No		

Partition	Active	Indexes	Conflicts
PAVO31	Yes		
PAVO32	Yes		
PAVO33	Yes	21, 39	
PAVO34	Yes		
PAVO35	Yes		
PAVO36	Yes		
PAVO37	No		
PAVO38	Yes		
PAVO39	Yes		
PAVO41	No		
PAVO42	No		
PAVO43	No		

At the bottom left are buttons for 'Close', 'Refresh', and 'Help'. To the right of the tables is a vertical column labeled 'Summary' containing the names of the partitions listed in the tables: PAVO01, PAVO02, PAVO3A, PAVO31, PAVO32, PAVO33, PAVO34, PAVO35, PAVO36, PAVO38, and PAVO39.

Figure 6-7 View LPAR Cryptographic Controls

Operational changes can be made by using the Change LPAR Cryptographic Controls task, which reflects the cryptographic definitions in the image profile for the partition. With this

function, the cryptographic feature can be added and removed dynamically, without stopping a running operating system.

For more information about the management of Crypto Express8S, see the *IBM z16 A02 and IBM z16 AGZ Configuration Setup*, SG24-8960.

6.6 Trusted Key Entry workstation

The TKE workstation is an optional feature that offers key management functions. It can be a TKE tower workstation (FC 0058) or TKE rack-mounted workstation (FC 0077) for IBM z16 A02 and IBM z16 AGZ to manage Crypto Express8S, Crypto Express7S, or Crypto Express6S.

The TKE contains a combination of hardware and software. A mouse, keyboard, flat panel display, PCIe adapter, and a writable USB media to install the TKE Licensed Internal Code (LIC) are included with the system unit. The TKE workstation requires an IBM 4770 crypto adapter.

A TKE workstation is part of a customized solution for the use of the Integrated Cryptographic Service Facility for z/OS (ICSF for z/OS) or Linux for IBM Z. This program provides a basic key management system for the cryptographic keys of a IBM z16 A02 or IBM z16 AGZ that has Crypto Express features installed.

TKE provides compliant-level hardware-based HSM management mechanisms that clients want or must use when managing IBM zSystems and LinuxONE Crypto Express Hardware Security Modules (HSMs). TKE also provides mechanisms that simplify HSM management in complex IBM z Systems and LinuxONE cryptographic environments as much as regulations and policy will allow. The TKE does not have to be in the same location as the IBM zSystems or LinunxONE system that has the HSMs. However, the users need to have physical access to the TKE when managing IBM zSystems or LinuxONE HSMs.

The TKE product is a Workstation with an HSM running a specific level of TKE Licensed internal code. In addition, smart card readers and smart cards are required to use most compliant-level HSM management mechanisms. The TKE workstation communicates with the IBM zSystems through a TCP/IP connection. The TKE workstation is available with Ethernet LAN connectivity only. Up to 10 TKE workstations can be ordered.

TKE FCs 0057 and 0058 can be used to control any supported Crypto Express feature on IBM z16 A02 and IBM z16 AGZ, IBM z15, IBM z14 systems, and the Crypto adapters on older, still supported systems.

The TKE 10.0 LIC (FC 0882) feature requires a 4770 HSM. The following features are supported:

- ▶ Managing the Crypto Express8S HSMs (CCA normal mode, CCA PCI mode, and EP11)
- ▶ QSC (Quantum Safe Cryptography) used when:
 - TKE authenticates Crypto Express Next HSMs
 - Deriving a Transport Key between TKE's HSM and target Crypto Express8S HSM
 - On demand HSM dual validation check.
- ▶ CCA domain group limitations. All CCA HSMs in a group must:
 - Support QSC (Can only include Crypto Express8S HSMs)
 - Not support QSC (Can't include Crypto Express8S HSMs)
- ▶ Configuration migration tasks support:
 - Can collect and apply data to a Crypto Express8S HSM
 - Can apply data from a pre Crypto Express8S HSM.

- Can use EC-521 strength migration zones when working with EP11 HSMs
- ▶ Launch Point For CHIM - For clients that also have IBM 4769 HSMs on linux x86, AIX x86, or Power AIX servers those HSMs can be managed using the Cryptographic Hardware Initialization and Maintenance (CHIM) feature of TKE.
- ▶ New default wrapping method for the Crypto Express8S HSM.
- ▶ New AES DUKPT key attribute on AES DKYGENKY parts.
- ▶ New EP11 attributes and control points can be managed from the TKE.
- ▶ Smart Card Reader Buzzer – For Identiv smart card readers, the reader's buzzer can be controlled by the TKE. When it is on, a sound is heard when it is time to provide input, and each time a button is pressed.

Tip: There are a number of TKE videos in the IBM Media Center. The introduction to TKE can be found at https://mediacenter.ibm.com/media/1_csb6z99p

For a full list of TKE videos, go to the IBM media center at <https://mediacenter.ibm.com/> and search using the key word **TKE**.

6.6.1 Logical partition, TKE host, and TKE target

If one or more LPARs are configured to use Crypto Express coprocessors, the TKE workstation can be used to manage DES, AES, ECC, and PKA master keys. This management can be done for all cryptographic domains of each Crypto Express coprocessor feature that is assigned to the LPARs that are defined to the TKE workstation.

Each LPAR in the same system that uses a domain that is managed through a TKE workstation connection is a TKE host or TKE target. An LPAR with a TCP/IP connection to the TKE is referred to as the *TKE host*; all other partitions are *TKE targets*.

The cryptographic controls that are set for an LPAR through the SE determine whether the workstation is a TKE host or a TKE target.

6.6.2 Optional smart card reader

An optional smart card reader (FC 0891) can be added to the TKE workstation. One FC 0891 includes two smart card readers, two cables to connect them to the TKE workstation, and 20 smart cards. The reader supports the use of smart card parts shipped by the smart card reader or extra smart cards feature codes. The part numbers from newest to oldest are 00RY790, 00JA710, 74Y0551, 45D3398. Smart cards contain a microprocessor that provides cryptographic functions and storage for signing keys and key parts. The 00JA710 and 00RY790 smart cards part is FIPS certified.

Smart card readers from FC 0885 or FC 0891 can be carried forward. Smart cards can be used on TKE 10.0 with these readers. Access to and use of confidential data on the smart card are protected by a user-defined PIN. Up to 990 other smart cards can be ordered for backup. (The extra smart card feature code is FC 0900). When one feature code is ordered, 10 smart cards are included. The order increment is 1 - 99 (10 - 990 blank smart cards).

If you are using Gemalto CT700 smart card readers:

- ▶ MCA smart cards must be at the minimum applet version 0.4. This applet version was first available in TKE 8.1.
- ▶ IA smart cards must be at the minimum applet version of 0.4. This applet version was first available in TKE 8.1.
- ▶ KPH smart cards must be at the minimum applet version of 0.4. This applet version was first available in TKE 8.1.

The current smart card shipped with the features for smart cards or extra smart cards is 00RY790. These smart cards have the strongest Elliptic Curve Cryptography (ECC) levels. Trusted Key Entry (TKE) allows stronger Elliptic Curve Cryptography (ECC) levels. More TKE Smart Cards (FC 0900, packs of 10, FIPS certified blanks) require TKE 9.1 LIC or up.

6.6.3 TKE hardware support and migration information

The new TKE 10.0 LIC (FC 0882) is originally shipped with a new IBM z16 A02 and IBM z16 AGZ server. The following TKE workstations can be ordered with a new IBM z16 A02 and IBM z16 AGZ:

- ▶ TKE 10.0 tower workstation (FC 0058)
- ▶ TKE 10.0 rack-mounted workstation (FC 0057)

Note: Several options for ordering the TKE with or without ordering Keyboard, Mouse, and Display are available. Ask your IBM Representative for more information about which option is the best option for you.

The TKE 10.0 LIC requires the 4770 crypto adapter. The TKE 9.x and TKE 8.x⁸ workstations can be upgraded to the TKE 10.0 tower workstation by purchasing a 4770 crypto adapter.

The Omnikey Cardman 3821 smart card readers can be carried forward to any TKE 10.0 workstation. Smart cards 45D3398, 74Y0551, 00JA710 and 00RY790 can be used on TKE 10.0.

When performing a MES upgrade from TKE 8.x, or TKE 9.x to a TKE 10.0 installation, the following steps must be completed:

1. Save Upgrade Data on an old TKE to USB memory to save client data.
2. Replace the 4768 crypto adapter with the 4770 crypto adapter.
3. Upgrade the firmware to TKE 10.0.
4. Install the Frame Roll to apply Save Upgrade Data (client data) to the TKE 10.0 system.
5. Run the TKE Workstation Setup wizard.

TKE upgrade considerations

If you are migrating your configuration with Crypto Express6S Crypto Express7S and TKE Release 9.x to a IBM z16 A02 and IBM z16 AGZ, you do not need to upgrade the TKE LIC.

Note: If your IBM z16 A02 and IBM z16 AGZ includes Crypto Express7S or Crypto Express6S, you can use TKE V9.2, which requires the 4768 cryptographic adapter.

For more information about TKE hardware support, see Table 6-3 on page 231. For some functionality, requirements must be considered; for example, the characterization of a Crypto Express adapter in EP 11 mode always requires the use of a TKE.

⁸ TKE 8.x can be upgraded to TKE 10.0 in a limited number of cases. Please verify possibilities in e-Config.

Table 6-3 TKE Compatibility Matrix

TKE workstation	TKE Release LIC	8.0 ^a	8.1 ^a	9.0 ^a	9.1 ^a	9.2 ^a	10.0
Manage Host Crypto Module	HW Feature Code	0847	0847 or 0097	0085 or 0086	0085 or 0086	0087 or 0088	0057 or 0058
	LICC	0877	0878	0879	0880	0881	0882
	Smart Card Read- er	0891	0891	0895	0895	0895	0891
	Smart Card	0892	0892	0892	0892	0892	0900
Manage Host Crypto Module	CEX8C (CCA)	No	No	No	No	No	Yes
	CEX8P (EP11)	No	No	No	No	No	Yes
	CEX7C (CCA)	No	No	No	No	Yes	Yes
	CEX7P (EP11)	No	No	No	No	Yes	Yes
	CEX6C (CCA)	No	No	Yes	Yes	Yes	Yes
	CEX6P (EP11)	No	No	Yes	Yes	Yes	Yes

a. The TKE workstation can be upgraded to TKE LIC V10.0 by adding a 4770 cryptographic adapter.

Attention: The TKE is unaware of the CPC type where the host crypto module is installed. That is, the TKE does not consider whether a Crypto Express is running on IBM z16 A01, IBM z16 A02, IBM z16 AGZ, IBM z15, or IBM z14 system. Therefore, the LIC can support any CPC where the coprocessor is supported, but the TKE LIC must support the specific crypto module.

6.7 Cryptographic functions comparison

The functions or attributes on IBM z16 A02 and IBM z16 AGZ for the two cryptographic hardware features are listed in Table 6-4, where “X” indicates that the function or attribute is supported.

Table 6-4 Cryptographic functions on IBM z16 A02 and IBM z16 AGZ

Functions or attributes	CPACF	CEX8C	CEX8P	CEX8A
Supports z/OS applications that use CSF	X	X	X	X
Supports Linux on IBM Z CCA applications	X	X	-	X
Encryption and decryption by using secret-key algorithm	-	X	X	-
Provides the highest SSL/TLS handshake performance	-	-	-	X
Supports SSL/TLS functions	X	X	X	X
Provides the highest symmetric (clear key) encryption performance	X	-	-	-
Provides the highest asymmetric (clear key) encryption performance	-	-	-	X

Functions or attributes	CPACF	CEX8C	CEX8P	CEX8A
Provides the highest asymmetric (encrypted key) encryption performance	-	X	X	-
Nondisruptive process to enable ^a	-	X	X	X
Requires IOCDs definition	-	-	-	-
Uses CHPID numbers	-	-	-	-
Uses PCHIDs (one PCHID)	-	X	X	X
Requires CPACF enablement (FC 3863) ^b	X	X	X	X
Requires ICSF to be active	-	X	X	X
Offers UDX	-	X	-	-
Usable for data privacy: Encryption and decryption processing	X	X	X	-
Usable for data integrity: Hashing and message authentication	X	X	X	-
Usable for financial processes and key management operations	-	X	X	-
Crypto performance IBM RMF monitoring	-	X	X	X
Requires system master keys to be loaded	-	X	X	-
System (master) key storage	-	X	X	-
Retained key storage	-	X	-	-
Tamper-resistant hardware packaging	-	X	X	X ^c
Hardware Designed for FIPS 140 Level 4 certification	-	X	X	X
Supports Linux applications that perform SSL handshakes	-	-	-	X
RSA functions	-	X	X	X
High-performance SHA-1, SHA-2, and SHA-3	X	X	X	-
Clear key DES or triple DES	X	-	-	-
Advanced Encryption Standard (AES) for 128-bit, 192-bit, and 256-bit keys	X	X	X	-
True random number generator (TRNG)	X	X	X	-
Deterministic random number generator (DRNG)	X	X	X	-
Pseudo random number generator (PRNG)	X	X	X	-
Clear key RSA	-	-	-	X
Payment Card Industry (PCI) PIN Transaction (PTS) Hardware Security Module (HSM) PCI-HSM		X	X	
Europay, MasterCard, and Visa (EMV) support	-	X	-	-
Public Key Decrypt (PKD) support for Zero-Pad option for clear RSA private keys	-	X	-	-

Functions or attributes	CPACF	CEX8C	CEX8P	CEX8A
Public Key Encrypt (PKE) support for Mod_Raised_to Power (MRP) function	-	X	X	-
Remote loading of initial keys in ATM	-	X	-	-
Improved key exchange with non-CCA systems	-	X	-	-
ISO 16609 CBC mode triple DES message authentication code (MAC) support	-	X	-	-
AES GMAC, AES GCM, AES XTS mode, CMAC	-	X	-	-
SHA-2, SHA-3 (384,512), HMAC	-	X	-	-
Visa Format Preserving Encryption	-	X	-	-
AES PIN support for the German banking industry	-	X	-	-
ECDSA (192, 224, 256, 384, 521 Prime/NIST)	-	X	-	-
ECDSA (160, 192, 224, 256, 320, 384, 512 BrainPool)	-	X	-	-
ECDH (192, 224, 256, 384, 521 Prime/NIST)	-	X	-	-
ECDH (160, 192, 224, 256, 320, 384, 512 BrainPool)	-	X	-	-
PNG (Prime Number Generator)	-	X	-	-

- a. To make adding the Crypto Express features nondisruptive, the logical partition must be predefined with the appropriate PCI Express cryptographic adapter number. This number must be selected from its candidate list in the partition image profile.
- b. This feature is not required for Linux if only RSA clear key operations are used. DES or triple DES encryption requires CPACF to be enabled.
- c. This feature is physically present, but is not used when configured as an accelerator (clear key only).

6.8 Cryptographic operating system support for IBM z16 A02 and IBM z16 AGZ

The following section gives an overview of the operating systems requirements in relation to cryptographic elements.

6.8.1 Crypto Express8S Exploitation

For full exploitation of the new functions of Crypto Express8S (0908/0909) is supplied in ICFS PTFs on:

- ▶ z/OS V2.5 (base, which is HCR77D2)
- ▶ z/OS V2.2 to V2.4 (Web Deliverable HCR77D1)

These ICSF support is required to administer a Crypto Express8S coprocessor using a TKE workstation, due to exploitation of quantum algorithms. All supported levels of ICSF automatically detect what HW cryptographic capabilities are available where it is running, then enables functions accordingly. No toleration support of new hardware is necessary.

For other operating systems than z/OS, the following prerequisites must be met.

- ▶ z/VM V7.1 and V7.2 for guest use

- ▶ z/VSE V6.2 with PTFs
- ▶ z/TPF V1.1 with PTFs
- ▶ Linux on IBM Z: IBM is working with its Linux distribution partners to provide support by way of maintenance or future releases for the following distributions:
 - SUSE Linux Enterprise Server 12 and SLES 11
 - Red Hat Enterprise Linux (RHEL) 8 and Red Hat Enterprise Linux 7
 - Ubuntu 16.04 LTS (or higher)

The KVM hypervisor, which is offered supported Linux distributions. For more information about minimal and recommended distribution levels, see the [Tested platforms for Linux web page](#) of the IBM IT infrastructure website.

6.8.2 Crypto Express8S support of VFPE

The following minimum prerequisites must be met to use this element:

- ▶ z/OS V2.5 (base, which is HCR77D2)
- ▶ z/OS V2.2 to V2.4 (Web Deliverable HCR77D1)
- ▶ z/VM V7.3 for guest use
- ▶ z/VM V7.2 for guest use
- ▶ Linux on IBM Z:
 - SUSE SLES 15 SP1 with service, SUSE SLES 12 SP4 with service, and SUSE SLES 11 SP4 with service.
 - Red Hat RHEL 9.0, RHEL 8.0 with service, Red Hat RHEL 7.9 with service.
 - Ubuntu 18.04.1 LTS with service and newer.
 - The support statements for IBM z16 A02 and IBM z16 AGZ also cover the KVM hypervisor on distribution levels that have KVM support.

For more information about the minimum required and recommended distribution levels, see the [IBM zSystems website](#).

6.8.3 Crypto Express8S support of greater than 16 domains

The following prerequisites must be met to support more than 16 domains:

- ▶ z/OS 2.5
- ▶ z/OS V2.4 with PTFs
- ▶ z/OS V2.3 with PTFs
- ▶ z/OS V2.2 with the Enhanced Cryptographic Support for z/OS V1.13-V2.2 Web deliverable installed
- ▶ z/VM V7.3 and V7.2 for guest use
- ▶ z/VSE V6.2 with PTFs
- ▶ Linux on IBM Z:
 - SUSE SLES 15 SP4 with service, SUSE SLES 12 SP4 with service.
 - Red Hat RHEL 9.0, RHEL 8.0 with service, Red Hat RHEL 7.9 with service.
 - Ubuntu 18.04.1 LTS with service and newer.

- The support statements for IBM z16 A02 and IBM z16 AGZ also cover the KVM hypervisor on distribution levels that have KVM support.

For more information about the minimum required and recommended distribution levels, see the [IBM zSystems website](#).

For more information about the software support levels for cryptographic functions, see Chapter 7, “Operating system support” on page 241.

6.9 Further use of cryptography on IBM z16 A02 and IBM z16 AGZ

Besides the common cryptography using CPACF and Crypto Express8S cards, which is ready to be exploited by the operating systems running in a logical partition, the IBM z16 A02 and IBM z16 AGZ offer several other features that use cryptography to securely transfer and store data.

The secure communication between SEs and HMAs, as well as the secure communication from HMA to the outside is described in Chapter 10, “Hardware Management Console and Support Element” on page 391.

The communication to consoles attached to an OSA-Express 1000BaseT Ethernet feature or to an OSA-Express GbE feature running in OSA-ICC mode can be secured by using Transport Layer Security/Secure Sockets Layer (TLS/SSL) with Certificate Authentication. This is described in chapter 7.4.5, “Networking features and functions” on page 290.

Encrypting the data traffic between a storage control unit and a FICON port in the IBM z16 A02 and IBM z16 AGZ by using IBM Fibre channel Endpoint Security is shortly described in 4.6.2, “Storage connectivity” on page 155.

Together with the new memory RAIM design the usage of memory encryption is implemented in the IBM z16 A02 and IBM z16 AGZ. This is described in 2.5.3, “Memory Encryption” on page 45.

6.9.1 Validated boot

IBM z16 A02 and IBM z16 AGZ Validated Boot technology is a security feature that verifies the authenticity of the boot process, including the operating system and all other components that are loaded during boot. It uses digital signatures to provide a check at the time of an Initial Program Load (IPL) that all IPL data like the executables residing on an IPL volume is intact, unaltered, and originates from a trusted source. Thus the detection of unauthorized changes to software executables is enabled, and if any component is found to be modified, the system will halt the boot process, preventing any malicious code from executing.

With Validated Boot customers are able to meet regulatory compliance required for certain secure software deployment scenarios. Any accidental IPL data changes are detected early, which can reduce the impact of outages. Also any malicious IPL data changes can be detected and certain types of attacks can be stopped.

Figure 6-8 on page 236 is showing the principles of Validated Boot. The client builds all needed load module executables, and signs them with the client's private key. The IPL Data is stored with this signature. At IPL time the IBM zSystems platform firmware (the Z Bootloader) validates the signed IPL text, which contains the validation function code the operating

system will use in subsequent load module verification steps. The operating system loads individual authorized load modules during the IPL, and uses the validation function code to validate their signatures using the client's public keys. The firmware (SE/HMC and LPAR) provides the trusted validation Certificate Store for use in validation processing done by both the Z Bootloader and the operating system. If the verification is successful, the IPL is performed with the validated IPL data.

Doing this a chain of trust is build, through signature validation, at every step of this cascading process, anchored in the firmware validation of the IPL Text and the secure firmware repository for the validation keys.

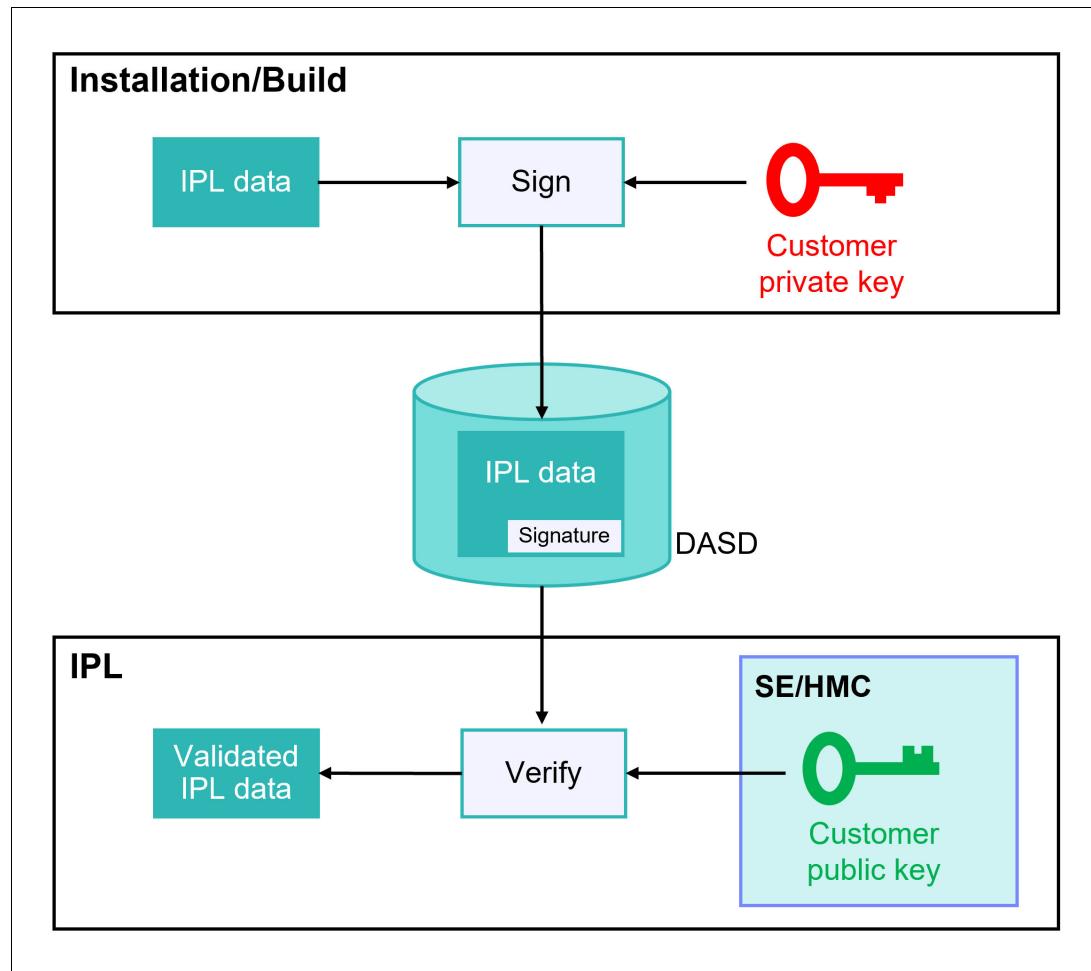


Figure 6-8 Validated Boot

At the time of writing this book, there is a Statement of Direction (SoD)⁹ for Validated Boot and for Package Signing for z/OS 2.5, see IBM United States Software Announcement 222-214, dated June 21, 2022. See Figure 6-9 on page 237

⁹ All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Validated Boot for z/OS

IBM plans to deliver a solution providing Validated Boot, also known as Secure Boot or Boot Integrity Validation, capability for z/OS IPLs. This solution is intended to validate digital signatures for loaded z/OS executables that have been built and signed as part of the solution. This solution is designed to meet the requirements for achieving the National Information Assurance Partnership (NIAP) OS Protection Profile 4.2.1 Certification.

Package signing

IBM plans to provide the capability to digitally sign electronically and physically delivered software packages. This new capability is designed to allow a user to ensure the package hasn't been tampered with and that the package was signed by the expected provider of the package by verifying the signature of the package.

Software packages from IBM that are intended to be signed include: ServerPac, CBPDO, Shopz PTF orders, SMP/E RECEIVE ORDER PTFs, and HOLDDATA. This support for signing and verification is planned to be available in both SMP/E and z/OSMF Software Management on all supported z/OS releases.

Figure 6-9 Statement of Direction in IBM United States Software Announcement 222-214

The support for Validated Boot for z/OS 2.5 will be delivered via z/OS Apart, also available for ServerPac installation.

The process of building a secure Code Package for Validated IPL will be open to include all forms of z/OS code provided by ISVs (Independent Software Vendors). This process is illustrated in Figure 6-10.

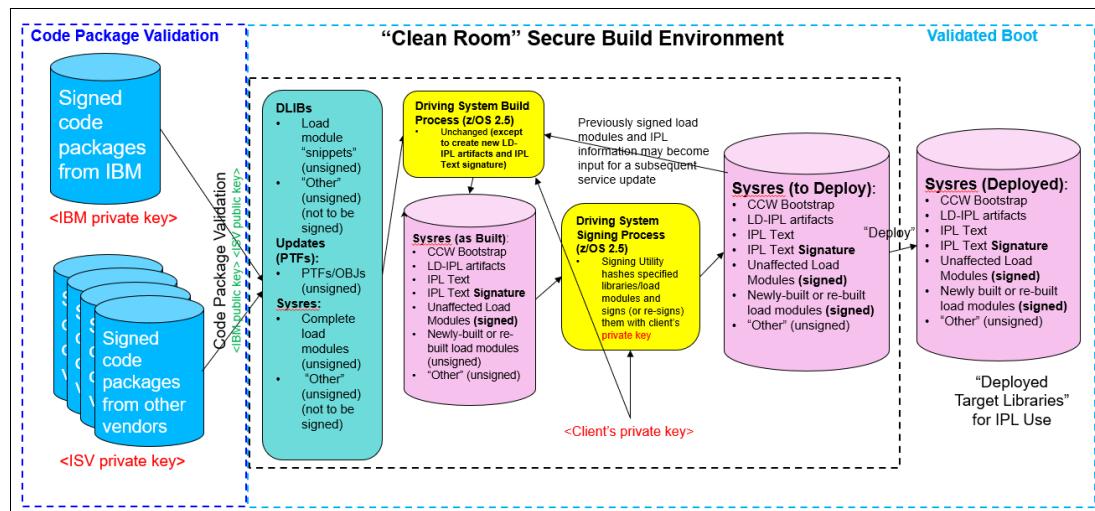


Figure 6-10 Client Secure Build Process

The IBM z16 A02 and IBM z16 AGZ hardware prerequisites include

- CPACF digital signature support with Elliptic Curve ECDSA-P521 support and SHA-512 hashing support and
- Virtual Flash Memory (VFM), also known as Storage Class Memory (SCM), for use in z/OS paging, because LPA pages will be paged to SCM/VFM Flash memory only.

The IBM z16 A02 and IBM z16 AGZ firmware for Validated Boot provides SE and HMC panels to perform a List-Directed IPL (LD-IPL) from ECKD DASD and also from SCSI DASD (for Linux on IBM Z), as well as to store certificates for Validated Boot on a per-LPAR basis.

With z/OS 2.5, the IPL must be done with CLPA (cold start) on all Validated Boot IPLs. This avoids any tampering with executable code on the PLPA page dataset while residing on disk. Building the LPA via load module loads at cold start time can be validated, but re-building LPA via page-in from PLPA page dataset on disk cannot be validated and would be inherently insecure.

The process of performing a Validated Boot via List-Directed IPL is shown in Figure 6-11.

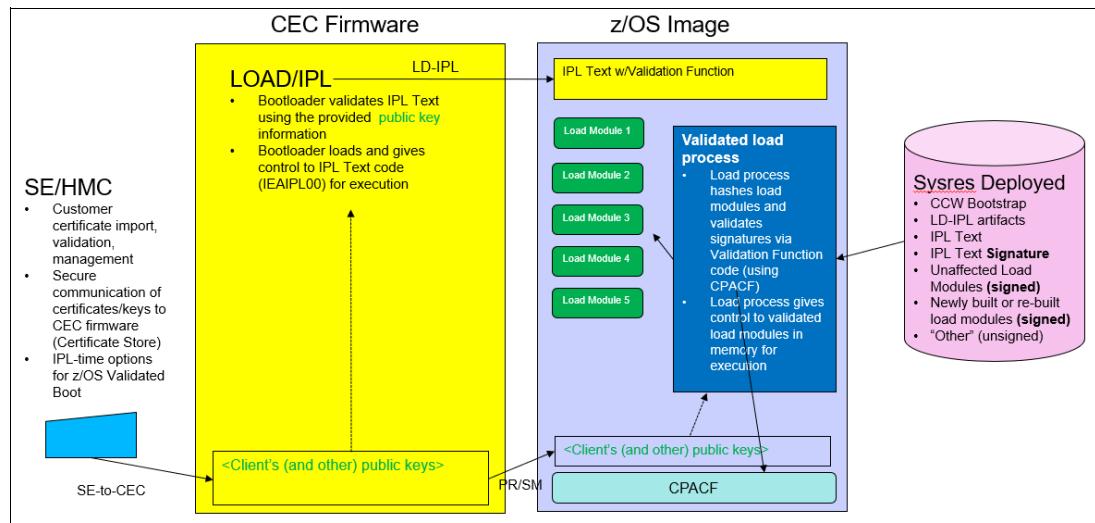


Figure 6-11 z/OS Validated Boot Process (via LD-IPL)

6.9.2 Secure Boot for ECKD devices

Linux on IBM Z has supported secure boot from FCP-attached devices since IBM z15. With support for a Validated Boot for z/OS, support in IBM z16 extends existing Linux secure boot capabilities to ECKD devices and allows client-provided validation certificates provided through the SE/HMC to be used for validation purposes during Linux secure boot. This new function is embedded in the baseline offering functionality. In addition, z/VM(R) 7.3 has been enhanced with support to securely boot a Linux guest.

Before you can use an ECKD type DASD as a disk for Linux on IBM Z, you must format it with a suitable disk layout. You must then create a file system or define a swap space. Please review [Preparing an ECKD type DASD for use - IBM document](#).

With secure boot enabled, an IPL fails if a component containing code is not signed or cannot be verified. IBM Secure Boot for Linux allows validation of signed Linux kernels to prevent non-approved Linux kernels from booting, based on a previously validated root of trust established for the firmware Bootloader. For details about how to prepare a device for secure boot, see [zipl modes and syntax overview](#) in Linux on IBM zSystems documentation.

Kernel interfaces are restricted in a kernel that is prepared for secure boot. In particular, in a kernel prepared for secure boot, all kernel modules must be signed.

Note: KVM: You can IPL a KVM guest from a device with the secure boot format, but signatures are not verified.

Instructions about how to boot Linux from DASD (ECKD) are detailed in [Booting from DASD](#) Linux on IBM zSystems documentation website.

6.9.3 z/VM 7.3 Guest Secure-IPL

z/VM V7.3 extends support for hardware secure boot by providing guest secure boot for both Load and Dump operations from ECKD and SCSI devices. This support provides the ability for a Linux guest to exploit hardware to validate the code being booted, ensuring it is signed by the client or its supplier. Secure Boot uses digital signatures to provide an IPL-time check that helps ensure IPL data is intact, unaltered, and originates from a trusted build-time source, enabling detection of unauthorized changes to those software executables.

The machine loader verifies the digitally signed hashes of the code, using verification certificates that the client has loaded into the HMC certificate store. z/VM makes these same certificates available to guests so that the OS can verify the authenticity of additional code loaded after IPL.

z/OS as a z/VM guest can only be securely IPLed in AUDIT mode because full exploitation requires Virtual Flash Memory support, which is not available to a guest.

z/VM V7.3 support is provided with the PTFs for APARs VM66434, VM66424, and VM66650.

Support for Linux Guest Secure IPL

Guest Exploitation support for Secure IPL uses digital signatures to provide an IPL-time check that IPL data is intact, unaltered, and originates from a trusted build-time source. Enables detection of unauthorized changes to those software executables.

Note: z/VM host secure boot is currently not supported; the CP nucleus is not hashed or signed.

This support provides ability for a guest to use (or exploit) hardware to validate the code being booted. Booted code must be signed by the customer or its supplier. The zBootLoader verifies the digitally signed hashes of the code, using verification certificates that the customer has loaded into the HMC certificate store. The machine loader will use the certificates in the HMC certificate store to validate that the signature of the code being loaded matches the signature on the program

z/VM makes these same certificates available to the guest OS (Linux or z/OS), so that the OS can verify additional code loaded after IPL. z/OS as a z/VM guest supports AUDIT mode only.

Enablement

The program to be loaded must be signed and at a level that supports secure IPL. The HMC certificate store must be loaded with any necessary matching security certificates. Then the **SET LOADDEV** or **SET DUMPDEV** command must be issued to set the appropriate LOADDEV or DUMPDEV parameters that includes the **SECURE** option. The IPL command is then issued to request a secure list-directed IPL.

Environment variable name: **CP.FUNCTION.SECURITY.IPL** - available with APAR VM66434 - indicates what level of Secure IPL is installed on the system.

ISV impact

Directory maintenance products like *DirMaint* and *SMAPI* may be affected although the changes made are upwardly compatible. Check with your vendor for updates.

Linux or hardware interaction

New levels of both hardware and Linux are required if the guest is to do a secure IPL. Guest Secure IPL requires an IBM z16 family server with driver D51C bundle 19 applied.



Operating system support

This chapter describes the minimum operating system requirements and support considerations for the IBM z16™ Machine Type 3932. It addresses IBM z/OS, z/VM, z/VSE, z/TPF, Linux on IBM Z and the KVM hypervisor.

Because this information is subject to change, see the following hardware fix categories for most current information:

- ▶ IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ -3931.* and
- ▶ IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ A02-3932.* specific for IBM z16 A02 and IBM z16 AGZ.

Support of IBM z16 A02 and IBM z16 AGZ functions depends on the operating system and version and release.

This chapter includes the following topics:

- ▶ 7.1, “Operating systems summary” on page 242
- ▶ 7.2, “Support by operating system” on page 242
- ▶ 7.3, “IBM z16 A02 and IBM z16 AGZ features and functions support overview” on page 246
- ▶ 7.4, “Support by features and functions” on page 258
- ▶ 7.5, “z/OS migration considerations” on page 307
- ▶ 7.6, “IBM z/VM migration considerations” on page 313
- ▶ 7.7, “z/VSE migration considerations” on page 313
- ▶ 7.8, “Software licensing” on page 314
- ▶ 7.9, “References” on page 316

7.1 Operating systems summary

The minimum operating system levels that are required on IBM z16 A02 and IBM z16 AGZ are listed in Table 7-1.

End of service operating systems: Operating system levels that are no longer in service are not covered in this publication.

Table 7-1 IBM z16 A02 and IBM z16 AGZ *minimum operating systems requirements*

Operating systems	Supported Version and release on IBM z16 ^a
z/OS	V2R2 ^b
z/VM	V7R1
z/VSE	V6.2
21 st Century Software z/VSE ^{n c}	V6.3
z/TPF	V1R1
Linux on IBM Z ^d	See Table 7-2 on page 245

a. Service is required.

b. z/OS V2R2 - Toleration mode only. The IBM Software Support Services for z/OS V2R2 offered as October 1st, 2020, provides the ability for customers to purchase extended defect support service for z/OS V2.R2.

c. z/VSEⁿ is supported by 21st Century Software.

d. KVM hypervisor is supported by Linux distribution partners.

The use of certain features depends on the operating system. In all cases, program temporary fixes (PTFs) might be required with the operating system level that is indicated. Check the z/OS fix categories, or the subsets of the 3932DEVICE (IBM z16 A02 and IBM z16 AGZ) PSP buckets for z/VM and z/VSE. The fix categories and the PSP buckets are continuously updated, and contain the latest information about maintenance:

- ▶ Hardware and software buckets contain installation information, hardware and software service levels, service guidelines, and cross-product dependencies.
- ▶ For more information about Linux on IBM Z distributions and KVM hypervisor, see the distributor's support information.

7.2 Support by operating system

IBM z16 A02 and IBM z16 AGZ introduce several new functions. This section describes the support of those functions by the current operating systems. Also included are some of the functions that were introduced in previous IBM zSystems servers and carried forward or enhanced in IBM z16 A02 and IBM z16 AGZ servers. Features and functions that are available on previous servers, but no longer supported by IBM z16 A02 and IBM z16 AGZ are not documented here. Previous versions of this document are available on the IBM Redbooks site.

For more information about supported functions that are based on operating systems, see 7.3, "IBM z16 A02 and IBM z16 AGZ features and functions support overview" on page 246. Tables are built by function and feature classification to help you determine, by a quick scan, what is supported and the minimum operating system level that is required.

7.2.1 z/OS

z/OS Version 2 Release 3 is the earliest in-service release that supports IBM z16 A02 and IBM z16 AGZ. Consider the following points:

- ▶ Service support for z/OS Version 2 Release 2 ended in September of 2020; however, a fee-based extension for defect support (for up to three years) can be obtained by ordering IBM Software Support Services - Service Extension for z/OS V2.R2.
- ▶ IBM z16 A02 and IBM z16 AGZ capabilities differ depending on the z/OS release. Toleration support is provided on z/OS V2R2. Exploitation support is provided on z/OS V2R3 and later only¹.

How to get the latest fix information for z/OS systems

For the latest information on z/OS PTFs that apply to the IBM z16 A02 and IBM z16 AGZ M/T 3932 consult the Fix Categories (FIXCATs).

Important: z16 M/T 3932 users **must** use the FIXCATs for both the 3932 and the 3931. Only unique capabilities for the 3932 will be identified with the 3932 FIXCATs.

Fixes are grouped into three categories:

- ▶ Base support is provided by PTFs identified by:
[IBM.Device.Server.z16-3931.RequiredService](#)
[IBM.Device.Server.z16A02-3932.RequiredService](#)
 - Fixes that are required to run z/OS on the IBM z16 A02 and IBM z16 AGZ and must be installed before migration
- ▶ Exploitation of many functions is provided by PTFs identified by:
[IBM.Device.Server.z16-3931.Exploitation](#)
[IBM.Device.Server.z16A02-3932.Exploitation](#)
 - Fixes that are required to exploit the capabilities of the IBM z16 A02 and IBM z16 AGZ. Only necessary to install if you are exploiting the function
- ▶ Recommended service is identified by:
[IBM.Device.Server.z16-3931.RecommendedService](#)
[IBM.Device.Server.z16A02-3932.RecommendedService](#)
 - Fixes that are recommended to run z/OS on the IBM z16 A02 and IBM z16 AGZ. These fixes are also listed in the Recommended Service section of the hardware PSP bucket. They represent fixes that have been recommended by IBM Service. It is recommended that you review and install these PTFs.

For more information about supported functions and their minimum required support levels, see 7.3, “IBM z16 A02 and IBM z16 AGZ features and functions support overview” on page 246.

7.2.2 z/VM

IBM z16 A02 and IBM z16 AGZ M/T 3932 support is provided with PTFs for z/VM 7.2 (compatibility only) and 7.3 with PTFs for IOCP, HCD, and HLASM.

z/VM Compatibility Support will enable Guest Exploitation for several new facilities:

¹ Use support for select features by way of PTFs. Toleration support for new hardware might also require PTFs.

- ▶ Imbedded Artificial Intelligence Acceleration:
designed to reduce the overall time required to execute CPU operations for neural networking processing functions, and help support real-time applications like fraud detection.
- ▶ Compliance-ready CPACF Counters support:
means for guests to track crypto compliance & instruction usage
- ▶ Breaking Event Address Register (BEAR) Enhancement Facility;
facilitates the debug of wild branches
- ▶ Vector Packed Decimal Enhancements 2:
new instructions intended to provide performance improvements
- ▶ Reset DAT protection Facility:
provides a more efficient way to disable DAT protection, such as during copy-on-write or page change tracking operations
- ▶ RoCE Express3 adapter
allows guests to exploit Routable RoCE, Zero Touch RoCE, and SMC-R V2 support
- ▶ Guest Enablement for the CEX8S crypto adapter and assorted crypto enhancements
Including Quantum Safe API Guest Exploitation Support available to dedicated guests
- ▶ CPU/Core topology location information within z/VM monitor data:
providing a better picture of the system for diagnostic and tuning purposes
- ▶ Consolidated Boot Loader for guest IPL from SCSI
- ▶ Guest Exploitation support for Secure IPL (Support for Linux Guest Secure IPL)
Uses digital signatures to provide an IPL-time check that IPL data is intact, unaltered, and originates from a trusted build-time source.
- ▶ Secure boot (a.k.a. validated boot) for z/OS as a z/VM guest supports AUDIT mode only
- ▶ Crypto stateless command filtering
Support new capability in Crypto Express8s when configured in CCA Co-processor mode to enforce restrictions on classes of requests
- ▶ Remove Obsolete IOCP Parameters
- ▶ z/VM Security Settings and Compliance interface (API and Command)
Provides support for a compliance status extractor for future IBM Z Security and Compliance Center (zSCC) exploitation

The following IBM z16 A02 and IBM z16 AGZ support will be transparent to z/VM:

- ▶ Dynamic Partition Mode (DPM) enhancements SMC-R, SMC-D
- ▶ OSA-Express7S 1.2 (GbE, 10GbE, 1000BASE-T, OSA-Express7S 1.1 (25GbE) Adapters
- ▶ Coupling Express2 LR Adapter
- ▶ 32Gbps 2 port FICON Adapter
- ▶ Coupling facility scalability and performance improvements

For more information about supported functions and their minimum required support levels, see 7.3, “IBM z16 A02 and IBM z16 AGZ features and functions support overview” on page 246.

7.2.3 z/VSE

IBM z16 A02 and IBM z16 AGZ support is provided by z/VSE V6R2 and later, with the following considerations:

- ▶ z/VSE runs in z/Architecture mode only.
- ▶ z/VSE supports 64-bit real and virtual addressing.

For more information about supported functions and their minimum required support levels, see 7.3, “IBM z16 A02 and IBM z16 AGZ features and functions support overview” on page 246.

7.2.4 21st Century Software z/VSEⁿ V6.3

21st Century Software VSEⁿ V6.3 was announced in March 2022 and is based on an IBM licensed copy of IBM z/VSE. For more information about this product visit the [21st Century Software website](#).

7.2.5 z/TPF

IBM z16 A02 and IBM z16 AGZ support is provided by z/TPF V1R1 with PTFs. For more information about supported functions and their minimum required support levels, see 7.3, “IBM z16 A02 and IBM z16 AGZ features and functions support overview” on page 246.

7.2.6 Linux on IBM Z

Generally, a new machine is not apparent to Linux on IBM Z. For IBM z16 A02 and IBM z16 AGZ, toleration support is required for the following functions and features:

- ▶ IPL in “z/Architecture” mode
- ▶ Crypto Express8S cards
- ▶ RoCE Express3 adapters
- ▶ 8-byte LPAR offset

The service levels of SUSE, Red Hat, and Ubuntu releases that are supported at the time of this writing are listed in Table 7-2.

Table 7-2 Linux on IBM Z distributions

Linux on IBM Z distribution ^a	Supported Version and Release on IBM z16 A02 and IBM z16 AGZ ^b
SUSE Linux Enterprise Server	15 SP5 and later
SUSE Linux Enterprise Server	12 SP5 ^c
Red Hat RHEL	9.1
Red Hat RHEL	8.4 ^c and later with service
Red Hat RHEL	7.9 ^c with service
Ubuntu	22.04 LTS
Ubuntu	20.04 LTS ^c
KVM Hypervisor ^d	Offered with the supported Linux distributions.

a. Only z/Architecture (64-bit mode) is supported. IBM testing identifies the minimum required level and the recommended levels of the tested distributions.

b. Fix installation is required for toleration.

c. Maintenance is required.

d. For more information about minimal and recommended distribution levels, see [the Linux on IBM Z website](#).

For more information about supported Linux distributions on IBM Z servers, see the [Tested platforms for Linux page](#) of the IBM IT infrastructure website.

IBM is working with Linux distribution Business Partners to provide further use of selected IBM z16 A02 and IBM z16 AGZ functions in future Linux on IBM Z distribution releases.

Consider the following guidelines:

- ▶ Use SUSE Linux Enterprise Server 15, Red Hat RHEL 9, or Ubuntu 22.10 LTS or newer in any new projects for IBM z16 A02 and IBM z16 AGZ.
- ▶ Update any Linux distribution to the latest service level before migrating to IBM z16 A02 and IBM z16 AGZ.
- ▶ Adjust the capacity of any Linux on IBM Z and z/VM guests, in terms of the number of IFLs and CPs, real or virtual, according to the PU capacity of the IBM z16 A02 and IBM z16 AGZ.

7.2.7 KVM hypervisor

KVM is offered through our Linux distribution partners to help simplify delivery and installation. Linux and KVM is provided from a single source. With KVM being included in the Linux distribution, ordering and installing KVM is easier.

For KVM support information, see [the IBM Z website](#).

7.3 IBM z16 A02 and IBM z16 AGZ features and functions support overview

The following list the IBM z16 A02 and IBM z16 AGZ features and functions and their minimum required operating system support levels:

- ▶ Table 7-3 on page 247
- ▶ Table 7-4 on page 248
- ▶ Table 7-5 on page 250
- ▶ Table 7-6 on page 250
- ▶ Table 7-7 on page 251
- ▶ Table 7-8 on page 253
- ▶ Table 7-9 on page 255
- ▶ Table 7-10 on page 257
- ▶ Table 7-11 on page 258

Information about Linux on IBM Z refers exclusively to the appropriate distributions of SUSE, Red Hat, and Ubuntu.

The tables in this section list but do not explicitly mark all the features that require fixes that are required by the corresponding operating system for toleration or exploitation.

All tables use the following conventions:

- ▶ Y: The function is supported.
- ▶ N: The function is not supported.
- ▶ -: The function is not applicable to that specific operating system.

7.3.1 Supported CPC functions

The supported Base CPC Functions or z/OS and z/VM are listed in Table 7-3.

Statements of Direction:

- ▶ In a future IBM Z hardware system family, the transactional execution and constrained transactional execution facility will no longer be supported. Users of the facility on current servers should always check the facility indications before use.

Note: z/OS V2R2 support has ended on as of September 2020. No new function is provided for exploiting the new HW features (toleration support only). Although extended (fee-based) support for z/OS V2.R2 can be obtained, support for z/OS V2.R2 is not covered extensively in this document.

Table 7-3 Supported Base CPC Functions or z/OS and z/VM

Function ^a	z/OS V2R5	z/OS V2R4	z/OS V2R3	z/VM V7R3	z/VM V7R2
IBM z16 A02 and IBM z16 AGZ servers	Y	Y	Y	Y	Y
Maximum processor unit (PUs) per system image	200	200	200	80 ^b	80 ^b
Maximum main storage size	16 TB	4 TB	4 TB	4 TB ^c	4 TB ^c
Dynamic PU add	Y	Y	Y	Y	Y
Dynamic LPAR memory add	Y	Y	Y	Y	Y
Dynamic LPAR memory removal	Y	Y	Y	Y	Y ^c
LPAR group absolute capping	Y	Y	Y	Y	Y
Capacity Provisioning Manager	Y	Y	Y	N	N
Program-directed re-IPL	-	-	-	Y	Y
Transactional Execution ^d	Y	Y	Y	Y ^e	Y ^e
Java Exploitation of Transactional Execution	Y	Y	Y	Y ^d	Y ^d
Simultaneous multithreading (SMT)	Y	Y	Y	Y	Y
Single Instruction Multiple Data (SIMD)	Y	Y	Y	Y ^f	Y ^f
2 GB large page support	Y	Y	Y	N	N
Large page (1 MB) support	Y	Y	Y	Y ^g	Y ^g
Db2 exploitation of IBM Z zAIU	Y ^h	Y ^h	N	-	-
CPUMF (CPU measurement facility) for IBM z16 A02 and IBM z16 AGZ	Y	Y	Y	Y	Y
Flexible Capacity	Y	Y	Y	Y	Y
IBM Virtual Flash Memory (VFM)	Y	Y	Y	N	N
1 MB pageable large pages	Y	Y	Y	N	N
Guarded Storage Facility (GSF)	Y	Y	Y	Y ^g	Y ^g

Function ^a	z/OS V2R5	z/OS V2R4	z/OS V2R3	z/VM V7R3	z/VM V7R2
Instruction Execution Protection (IEP)	Y	Y	Y	Y ^g	Y ^g
Co-processor Compression Enhancements (CMPSC)	Y	Y ^h	Y ^h	N	N
IBM Integrated Accelerator for zEDC (on-chip compression)	Y	Y	Y	Y ⁱ	Y ⁱ
CPU MF Extended Counters	Y	Y	Y	Y	Y
CF level 25 Enhancements	Y ^j	Y ^j	Y ^j	-	-
HiperDispatch Optimization	Y	Y	Y	Y ^g	Y ^g
System Recovery Boost Enhancements (IBM z16 A02 and IBM z16 AGZ)	Y ^j	Y ^j	N/A	N/A	N/A
IBM Integrated Accelerator for Z SORT	Y	Y	Y ^j	Y ^g	Y ^g
ICSF Enhancements	Y	Y	Y	-	-
Guest support for Breaking Event Address Register (BEAR) Enhancement Facility	Y	Y	Y	Y	Y
Guest support for Reset DAT protection Facility	Y	Y	Y	Y	Y
zDNN library enablement for IBM Z Integrated Accelerator for AI	Y	Y ^k	N	-	-
z/OS Validated boot (a.k.a. Secure boot)	Y ^h	N	N	Y ^l	Y ^l
Guest Exploitation support for Secure IPL (Support for Linux Guest Secure IPL)	N/A	N/A	N/A	Y ^h	Y ^h

- a. PTFs might be required for toleration support or exploitation of IBM z16 A02 and IBM z16 AGZ features and functions.
- b. 80-way without multithreading; 40-way with multithreading enabled
- c. With Service
- d. Statement of Direction: In a future IBM Z hardware system family, the transactional execution and constrained transactional execution facility will no longer be supported. Users of the facility on current servers should always check the facility indications before use.
- e. Guests are informed that TX facility is available for use
- f. Guests are informed that SIMD is available for use
- g. Guest Exploitation support
- h. With PTFs for exploitation
- i. Transparent for Guest support use of the gzip acceleration; guest support for z/OS Storage Compression
- j. With Exploitation PTFs
- k. With Required PTFs
- l. z/OS validated boot for z/VM guest supported in audit mode only (no enforcement)

The supported base CPC functions for z/VSE, z/TPF, and Linux on IBM Z are listed in Table 7-4.

Table 7-4 Supported base CPC functions for z/VSE, z/TPF, and Linux on IBM Z

Function ^a	z/VSE V6R2	z/TPF V1R1	Linux on IBM Z ^b
IBM z16 A02 and IBM z16 AGZ servers	Y	Y	Y
Maximum processor unit (PUs) per system image	10	86	200 ^c

Function ^a	z/VSE V6R2	z/TPF V1R1	Linux on IBM Z ^b
Maximum main storage size	32 GB	4 TB	16 TB ^d
Dynamic PU add	Y	N	Y
Dynamic LPAR memory upgrade	N	N	Y
LPAR group absolute capping	Y	N	N
Program-directed re-IPL	Y	N	Y
HiperDispatch	N	N	Y
IBM Z Integrated Information Processors (zIIPs)	N	N	N
Java Exploitation of Transactional Execution	N	N	Y
Simultaneous multithreading (SMT)	N	N	Ye
Single Instruction Multiple Data (SIMD)	Y	N	Y
Hardware decimal floating point ^f	N	N	Y
2 GB large page support	N	Y	Y
Large page (1 MB) support	Y	Y	Y
CPUMF (CPU measurement facility) for IBM z16 A02 and IBM z16 AGZ	N	Y	N ^g
AI accelerator exploitation	N	N	Y ^h
IBM Virtual Flash Memory (VFM)	N	N	Y
Guarded Storage Facility (GSF)	N	N	Y
Instruction Execution Protection (IEP)	N	N	Y
System Recovery Boost	Y ⁱ	Y ⁱ	N
Secure Boot (code integrity check)	-	-	Y ^j
Secure Execution Support for Linux	N/A	N/A	Y ^k
IBM Integrated Accelerator for zEDC (on-chip compression)	N	N	Y ^l
IBM Integrated Accelerator for Z SORT	N	N	N

- a. PTFs might be required for toleration support or exploitation of IBM z16 A02 and IBM z16 AGZ features and functions
- b. Support statement varies based on Linux on IBM Z distribution and release
- c. Linux kernel supports 256 cores without SMT and 128 cores with SMT (= 256 threads)
- d. IBM z16 A02 and IBM z16 AGZ supports defining up to 32 TB per LPAR (OS support is required)
- e. On IFL only
- f. Packed decimal conversion support
- g. IBM is working with its Linux distribution Business Partners to provide this feature
- h. Delivered with Linux distributions as a new package: libzdn
- i. Subcapacity CP speed boost (no zIIP boost)
- j. For SCSI IPL
- k. For second level guests running under KVM
- l. Requires Linux kernel exploitation support for gzip/zlib compression.

7.3.2 Coupling and clustering

The supported coupling and clustering functions for z/OS and z/VM are listed in Table 7-5.

Note: z/OS V2R2 support ended as of September 2020. No new function is provided for exploiting the new HW features (toleration support only). Although extended (fee-based) support for z/OS V2.R2 can be obtained, support for z/OS V2.R2 is not covered extensively in this document.

Table 7-5 Supported coupling and clustering functions for z/OS and z/VM

Function ^a	z/OS V2R5	z/OS V2R4	z/OS V2R3	z/VM V7R3	z/VM V7R2
CFCC Level 25 ^b	Y	Y	Y	Y	Y
CFCC Level 24 ^c	Y	Y	Y	Y	Y
CFCC Level 23 ^d	Y	Y	Y	Y	Y
CFCC Level 25 Coupling Facility Processor Scalability Enhancements	Y	Y	Y	Y	Y
RMF coupling channel reporting	Y	Y	Y	N	N
z/VM Dynamic I/O support for ICA SR CHPIDs	-	-	-	Y	Y
Asynchronous CF Duplexing for lock structures	Y	Y	Y	Y	Y
Cache residency time metrics ^e	Y	Y	Y	N	N
Dynamic I/O activation for stand-alone CF CPCs ^f	Y	Y	Y	Y	Y

a. PTFs are required for toleration support or exploitation of IBM z16 A02 and IBM z16 AGZ features and functions

b. CFCC Level 25 with Driver 51 (IBM z16 A02 and IBM z16 AGZ)

c. CFCC Level 24 with Driver 41 (IBM z15)

d. CFCC Level 23 with Driver 36 (IBM z14)

e. With APAR OA60650

f. Requires HMC 2.14.1(Driver 36) or newer and various OS fixes (HCD, HCM, IOS, IOCP)

In addition to operating system support that is listed in Table 7-5, Server Time Protocol is supported on z/TPF V1R1 and Linux on IBM Z. Also, CFCC Level 23, Level 24, and Level 25 are supported for z/TPF V1R1.

7.3.3 Storage connectivity

The supported storage connectivity functions for z/OS and z/VM are listed Table 7-6.

Table 7-6 Supported storage connectivity functions for z/OS and z/VM

Function ^a	z/OS V2R5	z/OS V2R4	z/OS V2R3	z/VM V7R3	z/VM V7R2
zHyperLink Read Support for Db2 and VSAM	Y	Y	Y ^b	N	N
zHyperLink Write Support for Db2 logs	Y	Y	Y ^b	N	N
zHyperLink Writes support for asynchronous mirroring ^c	Y	Y ^b	Y ^b	N	N

Function ^a	z/OS V2R5	z/OS V2R4	z/OS V2R3	z/VM V7R3	z/VM V7R2
zHyperLink Consistent read from metro mirror secondary ^d	Y	Y ^b	Y ^b	N	N
z/VM Dynamic I/O support for FICON FC and FCP CHPIDs	-	-	-	Y	Y
CHPID (Channel-Path Identifier) type FC					
FICON support for zHPF (IBM Z High-Performance FICON)	Y	Y	Y	Y ^e	Y ^e
IBM Fibre Channel Endpoint Security ^f	Y	Y	Y	Y	Y
FICON when using CTC (channel-to-channel)	Y	Y	Y	Y	Y
IPL from an alternative subchannel set	Y	Y	Y	N	N
32 K subchannels for FICON Express	Y	Y	Y	Y	Y
Request node identification data (RNID)	Y	Y	Y	N	N
FICON link incident reporting	Y	Y	Y	N	N
CHPID (Channel-Path Identifier) type FCP					
FICON Express support of SCSI devices	-	-	-	Y	Y
FICON Express support of hardware data router	-	-	-	Y ^e	Y ^e
FICON Express support of T10 Data Integrity Field (DIF)	-	-	-	Y ^e	Y ^e
N_Port ID Virtualization (NPIV)	-	-	-	Y	Y
Worldwide port name tool	-	-	-	Y	Y

a. PTFs might be required for toleration support or exploitation of IBM z16 A02 and IBM z16 AGZ features and functions.

b. With PTFs

c. DS8900F only, 9.1 release with z/OS PTFs, does not include XRC

d. DS8900F only, 9.2 release with z/OS PTFs

e. For guest use

f. Requires FC 1146. Minimum, DS8910 or DS8890 storage, CPACF enablement and FICON Express32S LX/SX

The supported storage connectivity functions for z/VSE, z/TPF, and Linux on IBM Z are listed in Table 7-7.

Table 7-7 Supported storage connectivity functions for z/VSE, z/TPF, and Linux on IBM Z

Function ^a	z/VSE V6R2	z/TPF V1R1	Linux on IBM Z ^b
The 63.75-K subchannels	N	N	Y
Six logical channel subsystems (LCSSs)	Y	N	Y
Four subchannel set per LCSS	Y	N	Y
CHPID (Channel-Path Identifier) type FC			
FICON Express support of zHPF (IBM Z High-Performance FICON) ^c	Y	Y	Y

Function ^a	z/VSE V6R2	z/TPF V1R1	Linux on IBM Z ^b
IBM Fibre Channel Endpoint Security ^d	Y ^e	Y ^e	Y ^e
MIDAW (Modified Indirect Data Address Word)	N	N	N
FICON Express support for CTC (channel-to-channel)	Y	Y	Y
IPL from an alternative subchannel set	N	N	N
32 K subchannels for FICON Express	N	N	Y
Request node identification data (RNID)	N	N	N
CHPID (Channel-Path Identifier) type FCP			
FICON Express support of SCSI devices	Y	-	Y
FICON Express support of hardware data router	N	N	Y
FICON Express support of T10 Data Integrity Field (DIF)	N	N	Y
N_Port ID Virtualization (NPIV)	Y	N	Y
Worldwide port name tool	-	-	Y

- a. PTFs might be required for toleration support or exploitation of IBM z16 A02 and IBM z16 AGZ features and functions
- b. Support statement varies based on Linux on IBM Z distribution and release
- c. Transparent to operating systems
- d. Requires FC 1146. Minimum, DS8910 or DS8890 storage, CPACF enablement and FICON Express32S LX/SX
- e. Feature is OS independent (transparent to OS); OS support is only needed for displaying configuration and monitoring (fixes may be required)

7.3.4 Network connectivity

The supported network connectivity functions for z/OS and z/VM are listed in Table 7-8.

Statements of Direction^a:

- ▶ IBM z16 A02 and IBM z16 AGZ generation will be the last IBM zSystems to support the OSE CHPID type.
- ▶ IBM z16 A02 and IBM z16 AGZ generation will be the last IBM zSystems to support OSA Express 1000BASE-T hardware adapters.

a. Statements by IBM regarding its plans, directions, and intent are subject to change or withdrawal without notice at the sole discretion of IBM.

Note: z/OS V2R2 support ended as of September 2020. No new function is provided for exploiting the new HW features (toleration support only). Although extended (fee-based) support for z/OS V2.R2 can be obtained, support for z/OS V2.R2 is not covered extensively in this document.

Table 7-8 Supported network connectivity functions for z/OS and z/VM

Function ^a	z/OS V2R5	z/OS V2R4	z/OS V2R3	z/VM V7R3	z/VM V7R2
Checksum offload for IPV6 packets	Y	Y	Y	Y ^b	Y ^b
Checksum offload for LPAR-to-LPAR traffic with IPv4 and IPv6	Y	Y	Y	Y ^b	Y ^b
QDIO data connection isolation for z/VM	-	-	-	Y	Y
QDIO interface isolation for z/OS	Y	Y	Y	-	-
QDIO OLM (Optimized Latency Mode)	Y	Y	Y	-	-
QDIO Diagnostic Synchronization	Y	Y	Y	N	N
IWQ (Inbound Workload Queuing) for OSA	Y	Y	Y	Y ^b	Y ^b
GARP VLAN Registration Protocol	Y	Y	Y	Y	Y
Link aggregation support for z/VM	-	-	-	Y	Y
Multi-vSwitch Link Aggregation	-	-	-	Y	Y
Large send for IPV6 packets	Y	Y	Y	Y ^b	Y ^b
z/VM Dynamic I/O Support for OSA-Express OSD CHPIDs	-	-	-	Y	Y
OSA Dynamic LAN idle	Y	Y	Y	N	N
OSA Layer 3 virtual MAC for z/OS environments	Y	Y	Y	-	-
Network Traffic Analyzer	Y	Y	Y	N	N
Hipersockets					
HiperSockets ^c	Y	Y	Y	Y	Y
HiperSockets Completion Queue	Y	Y	Y	Y	Y
HiperSockets Virtual Switch Bridge	-	-	-	Y	Y

Function ^a	z/OS V2R5	z/OS V2R4	z/OS V2R3	z/VM V7R3	z/VM V7R2
HiperSockets Multiple Write Facility	Y	Y	Y	N	N
HiperSockets support of IPV6	Y	Y	Y	Y	Y
HiperSockets Layer 2 support	Y	Y	Y	Y	Y
SMC-D and SMC-R					
SMC-D ^d over ISM (Internal Shared Memory)	Y	Y	Y	Y ^b	Y ^b
25GbE and 10GbE RoCE Express3 for SMC-R	Y	Y	Y	Y ^b	Y ^b
25GbE and 10GbE RoCE Express3 and Express2/2.1 for Ethernet communications ^e including Single Root I/O Virtualization (SR-IOV)	N	N	N	Y ^b	Y ^b
z/VM Dynamic I/O support for RoCE Express	-	-	-	Y	Y
Shared RoCE environment	Y	Y	Y	Y	Y
Open Systems Adapter (OSA)^f					
OSA-Express7S 1.2 1000BASE-T Ethernet ^g CHPID type OSC and OSD	y	Y	Y	Y	Y
OSA-Express7S 1000BASE-T Ethernet CHPID type OSC and OSD	Y	Y	Y	Y	Y
OSA-Express6S 1000BASE-T Ethernet CHPID types OSC and OSD	Y	Y	Y	Y	Y
OSA-Express7S 1.2 25GbE CHPID type OSD	Y	Y	Y	Y	Y
OSA-Express7S 25-Gigabit Ethernet Short Reach (SR and SR1.1) CHPID type OSD	Y	Y	Y ^h	Y	Y
OSA-Express7S 1.2 10GbE CHPID type OSD	Y	Y	Y	Y	Y
OSA-Express7S 10-Gigabit Ethernet Long Reach (LR) and Short Reach (SR) CHPID type OSD	Y	Y	Y	Y	Y
OSA-Express6S 10-Gigabit Ethernet Long Reach (LR) and Short Reach (SR) CHPID type OSD	Y	Y	Y	Y	Y
OSA-Express7S 1.2 GbE CHPID type OSD and OSC	Y	Y	Y	Y	Y
OSA-Express7S GbE CHPID type OSD and OSC	Y	Y	Y	Y	Y
OSA-Express6S GbE CHPID type OSD	Y	Y	Y	Y	Y
OSA-Express7S 1.2 1000BASE-T Ethernet CHPID type OSE ⁱ	Y	Y	Y	Y	Y
OSA-Express7S 1000BASE-T Ethernet CHPID type OSE ⁱ	Y	Y	Y	Y	Y
OSA-Express6S 1000BASE-T Ethernet CHPID type OSE ⁱ	Y	Y	Y	Y	Y

a. PTFs might be required for toleration support or exploitation of IBM z16 A02 and IBM z16 AGZ features and functions

b. For guest use or exploitation

- c. On IBM z16 A02 and IBM z16 AGZ, the CHPID statement of HiperSockets devices requires the keyword VCHID. If you are migrating from a zEC12 or earlier, the IBM z16 A02 and IBM z16 AGZ IOCP definitions must be migrated to support the HiperSockets definitions (from CHPID type IQD). VCHID specifies the virtual channel identification number that is associated with the channel path (valid range is 7C0 - 7FF)
- d. Shared Memory Communications - Direct Memory Access.
- e. Does not require a peer OSA.
- f. CHPID types OSM, OSN, OSX are no longer supported
- g. Statements of Direction: IBM z16 A02 and IBM z16 AGZ server generation will be the last IBM zSystems to support OSA Express 1000BASE-T hardware adapters
- h. Require PTFs for APARs OA55256 (IBM VTAM®) and PI95703 (TCP/IP).
- i. Statements of Direction: IBM z16 A02 and IBM z16 AGZ will be the last IBM Z server to support the OSE CHPID type

The supported network connectivity functions for z/VSE, z/TPF, and Linux on IBM Z are listed in Table 7-9.

Table 7-9 Supported network connectivity functions for z/VSE, z/TPF, and Linux on IBM Z

Function ^a	z/VSE V6R2	z/TPF V1R1	Linux on IBM Z ^b
Checksum offload for IPV6 packets	N	N	Y
Checksum offload for LPAR-to-LPAR traffic with IPv4 and IPv6	N	N	Y
QDIO Diagnostic Synchronization	N	N	N
IWQ (Inbound Workload Queuing) for OSA	N	N	N
GARP VLAN Registration Protocol	N	N	Y ^c
Multi-vSwitch Link Aggregation	N	N	N
Large send for IPV6 packets	N	N	Y
OSA Dynamic LAN idle	N	N	N
Hipersockets			
HiperSockets ^d	Y	N	Y
HiperSockets Completion Queue	Y	N	Y
HiperSockets Virtual Switch Bridge	-	-	Y ^e
HiperSockets support of IPV6	Y	N	Y
HiperSockets Layer 2 support	Y	N	Y
HiperSockets Network Traffic Analyzer for Linux on IBM Z	N	N	Y
SMC-D and SMC-R			
SMC-D ^f over ISM (Internal Shared Memory)	N	N	Y ^g
10GbE RoCE ^h Express	N	N	Y ^{gi}
25GbE and 10GbE RoCE Express3 for SMC-R	N	N	Y ^{gi}
25GbE and 10GbE RoCE Express3 for Ethernet communications ^j including Single Root I/O Virtualization (SR-IOV)	N	N	Y ^{gi}
Shared RoCE environment	N	N	Y
Open Systems Adapter (OSA)			

Function ^a	z/VSE V6R2	z/TPF V1R1	Linux on IBM Z ^b
OSA-Express7S 1.2 1000BASE-T Ethernet ^k CHPID type OSC and OSD	Y	Y	Y
OSA-Express7S 1000BASE-T Ethernet CHPID type OSC and OSD	Y	Y	Y
OSA-Express6S 1000BASE-T Ethernet CHPID types OSC and OSD	Y	Y	Y
OSA-Express7S 1.2 25GbE CHPID type OSD	Y	Y	Y
OSA-Express7S 25-Gigabit Ethernet Short Reach (SR and SR1.1) CHPID type OSD	Y	Y	Y
OSA-Express7S 1.2 10GbE CHPID type OSD	Y	Y	Y
OSA-Express7S 10-Gigabit Ethernet Long Reach (LR) and Short Reach (SR) CHPID type OSD	Y	Y	Y
OSA-Express6S 10-Gigabit Ethernet Long Reach (LR) and Short Reach (SR) CHPID type OSD	Y	Y	Y
OSA-Express7S 1.2 GbE CHPID type OSD and OSC	Y	Y	Y
OSA-Express7S GbE CHPID type OSD and OSC	Y	Y	Y
OSA-Express6S GbE CHPID type OSD	Y	Y	Y
OSA-Express7S 1.2 1000BASE-T Ethernet CHPID type OSE ⁱ	Y	N	N
OSA-Express7S 1000BASE-T Ethernet CHPID type OSE ⁱ	Y	N	N
OSA-Express6S 1000BASE-T Ethernet CHPID type OSE ⁱ	Y	N	N

- a. PTFs might be required for toleration support or exploitation of IBM z16 A02 and IBM z16 AGZ features and functions.
- b. Support statement varies based on Linux on IBM Z distribution and release.
- c. By using VLANs.
- d. On IBM z16 A02 and IBM z16 AGZ, the CHPID statement of HiperSockets devices requires the keyword VCHID. Therefore, the IBM z16 A02 and IBM z16 AGZ IOCP definitions must be migrated to support the HiperSockets definitions (CHPID type IQD). VCHID specifies the virtual channel identification number that is associated with the channel path (valid range is 7C0 - 7FF). VCHID is not valid on IBM zSystems before IBM z13.
- e. Applicable to guest operating systems.
- f. Shared Memory Communications - Direct Memory Access.
- g. SMC-R and SMC-D are supported on Linux kernel; see:
<https://linux-on-z.blogspot.com/p-smc-for-linux-on-ibm-z.html>
- h. Remote Direct Memory Access (RDMA) over Converged Ethernet.
- i. Linux can also use RocE Express as a standard NIC (Network Interface Card) for Ethernet.
- j. Does not require a peer OSA.
- k. Statements of Direction: IBM z16 A02 and IBM z16 AGZ server generation will be the last IBM zSystems to support OSA Express 1000BASE-T hardware adapters
- l. Statements of Direction: IBM z16 A02 and IBM z16 AGZ server generation will be the last IBM zSystems to support the OSE CHPID type

7.3.5 Cryptographic functions

The IBM z16 A02 and IBM z16 AGZ supported cryptography functions for z/OS and z/VM are listed in Table 7-10.

Note: z/OS V2R2 support has ended as of September 2020. No new function is provided for exploiting the new HW features (toleration support only). Although extended (fee-based) support for z/OS V2.R2 can be obtained, support for z/OS V2.R2 is not covered extensively in this document.

Table 7-10 Supported cryptography functions for z/OS and z/VM

Function ^a	z/OS V2R5	z/OS V2R4	z/OS V2R3	z/VM V7R3	z/VM V7R2
CP Assist for Cryptographic Function (CPACF)	Y	Y	Y	Y ^b	Y ^b
Support for 40 Domains	Y	Y	Y	Y ^b	Y ^b
CPACF support for: ▶ Encryption (DES, TDES, AES) ▶ Hashing (SHA-1, SHA-2, SHA-3, SHAKE) ▶ Random Number Generation (PRNG, DRNG, TRNG)	Y	Y	Y	Y ^b	Y ^b
Crypto Express8S	Y	Y	Y	Y ^b	Y ^b
Crypto Express7S	Y	Y	Y	Y ^b	Y ^b
Crypto Express7S Support for Visa Format Preserving Encryption	Y	Y	Y	Y ^b	Y ^b
Crypto Express7S Support for Coprocessor in PCI-HSM Compliance Mode ^c	Y	Y	Y	Y ^b	Y ^b
Crypto Express6S	Y	Y	Y	Y ^b	Y ^b
Crypto Express6S Support for Visa Format Preserving Encryption	Y	Y	Y	Y ^b	Y ^b
Crypto Express6S Support for Coprocessor in PCI-HSM Compliance Mode ^d	Y	Y	Y	Y ^b	Y ^b
Elliptic Curve Cryptography (ECC)	Y	Y	Y	Y ^b	Y ^b
Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode	Y	Y	Y	Y ^b	Y ^b
z/OS Data Set Encryption	Y	Y	Y	-	-
z/VM Encrypted paging support	-	-	-	Y	Y
RMF Support for Crypto Express8, Express7, and Express6	Y	Y	Y	-	-
z/OS SMF Enhancements for CPACF	Y ^e	Y ^e	N	-	-
z/OS encryption readiness technology (zERT)	Y	Y	Y	-	-
ICFSF New Function Support	Y	Y ^e	Y ^e	-	-
Quantum-Safe Cryptography (QSC) for signing and key negotiation	Y	Y ^f	Y ^f	Y	Y

- a. PTFs might be required for toleration support or exploitation of IBM z16 A02 and IBM z16 AGZ features and functions.
- b. For guest use or exploitation.
- c. Requires TKE 9.1 or newer.
- d. Requires TKE 9.2 or newer.
- e. Requires z/OS Exploitation support via APAR
- f. Requires Web deliverable HCR77D1

The IBM z16 A02 and IBM z16 AGZ supported cryptography functions for z/VSE, z/TPF, and Linux on IBM Z are listed in Table 7-11.

Table 7-11 Supported cryptography functions for z/VSE, z/TPF, and Linux on IBM Z

Function ^a	z/VSE V6R2	z/TPF V1R1	Linux on IBM Z ^b
CP Assist for Cryptographic Function (CPACF)	Y	Y	Y
Support for 85 Domains	Y	N	Y
CPACF support for: <ul style="list-style-type: none"> ► Encryption (DES, TDES, AES) ► Hashing (SHA-1, SHA-2, SHA-3, SHAKE) ► Random Number Generation (PRNG, DRNG, TRNG) 	Y	Y ^{cd}	Y
CPACF protected key	N	N	N
Crypto Express8S	Y	Y	Y
Crypto Express7S	Y	Y	Y
Crypto Support for Visa Format Preserving Encryption	N	N	N
Crypto Support for Coprocessor in PCI-HSM Compliance Mode	N	N	N
Crypto Express6S	Y	Y	Y
Elliptic Curve Cryptography (ECC)	Y	N	Y
Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode	N	N	Y
z/TPF transparent database encryption	-	Y	-

- a. PTFs might be required for toleration support or exploitation of IBM z16 A02 and IBM z16 AGZ features and functions.
- b. Support statement varies based on Linux on IBM Z distribution and release.
- c. z/TPF supports only AES-128 and AES-256
- d. z/TPF supports only SHA-1 and SHA-256

7.4 Support by features and functions

This section addresses operating system support by function. Only the currently in-support releases are covered.

Tables in this section use the following convention:

- N/A: Not applicable
- NA: Not available

7.4.1 LPAR Configuration and Management

A single system image can control multiple processor units (PUs), such as CPs, zIIPs, or IFLs.

Note: z/OS V2R2 support ended as of September 2020. No new function is provided for exploiting the new HW features (toleration support only). Although extended (fee-based) support for z/OS V2.R2 can be obtained, support for z/OS V2.R2 is not covered extensively in this document.

Maximum number of PUs per system image

The maximum number of PUs that is supported by each operating system image and by special-purpose LPARs are listed in Table 7-12.

Table 7-12 Maximum number of PUs per system image

Operating system	Maximum number of PUs per system image
z/OS V2R5	256 ^{a,b}
z/OS V2R4	256 ^{a,b}
z/OS V2R3	256 ^{a,b}
z/VM V7R3	80 ^c
z/VM V7R2	80 ^d
z/VSE V6.2 and later	z/VSE Turbo Dispatcher can use up to 4 CPs, and tolerates up to 10-way LPARs
z/TPF V1R1	86 CPs
CFCC Level 25	16 CPs or ICFs (CPs and ICFs cannot be mixed)
Linux on IBM Z	SUSE Linux Enterprise Server 12 and later: 256 CPs or IFLs. Red Hat RHEL 7 and later: 256 CPs or IFLs. Ubuntu 20.04.1 LTS and later: 256 CPs or IFLs.
KVM Hypervisor	The KVM hypervisor is offered with the following Linux distributions -- 256CPs or IFLs--: SLES 12 SP5 and later RHEL 7.9 and later Ubuntu 20.04.1 LTS and later
Secure Service Container	80
GDPS Virtual Appliance	80

- a. IBM z16 A02 and IBM z16 AGZ A01 LPARs support 200-way without multithreading; 128-way with multithreading (SMT), however, z16 M/T 3932 supports max. 6 CPs, max. 67 zIIPs, or 68 IFLs.
- b. Total characterizable PUs, including zIIPs and CPs. z16 M/T 3932 supports max. 6 CPs, and max. 67 zIIPs.
- c. 80-way without multithreading and 40-way with multithreading enabled
- d. 80-way without multithreading and 40-way with multithreading enabled

Maximum main storage size

The maximum amount of main storage that is supported by current operating systems is listed in Table 7-13 on page 260. A maximum of 16 TB of main storage can be defined for an LPAR on an IBM z16 A02 and IBM z16 AGZ.

Table 7-13 Maximum memory that is supported by the operating system

Operating system	Maximum supported main storage ^a
z/OS	z/OS V2R5 supports 16 TB. Prior z/OS releases support 4TB
z/VM	z/VM V7R1 supports 2TB while z/VM 7.2 ^b and z/VM 7.3 support 4TB
z/VSE	z/VSE V6R2 supports 32 GB
z/TPF	z/TPF supports 4 TB
CFCC	Levels 23, 24 and 25 support up to 3 TB
Secure Service Containers	Supports up to 16TB ^a
Linux on IBM Z (64-bit)	16TB ^{ac}

- a. On IBM z16 A02 and IBM z16 AGZ LPAR storage definition supports 32TB (IBM z16 A02 and IBM z16 AGZ supports up to 40 TB of usable memory).
- b. With fix for APAR VM66173.
- c. Support may vary by distribution. Check with your distribution provider.

IBM z16 A02 and IBM z16 AGZ - Up to 40 LPARs

The IBM z16 A02 and IBM z16 AGZ can be configured with up to 40 LPARs (same as previous air cooled systems). Because channel subsystems can be shared by up to 15 LPARs, it is necessary to configure three channel subsystems to reach the 40 LPARs limit.

Remember: A virtual appliance that is deployed in a Secure Service Container runs in a dedicated LPAR. When activated, it reduces the maximum number of available LPARs by one.

Dynamic PU add

Planning an LPAR configuration includes defining reserved PUs that can be brought online when extra capacity is needed. Operating system support is required to use this capability without an IPL; that is, nondisruptively. This support is available in z/OS for some time.

The dynamic PU add function enhances this support by allowing you to dynamically define and change the number and type of reserved PUs in an LPAR profile, which removes any planning requirements. The new resources are immediately made available to the operating system and in the case of z/VM, to its guests.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248.

Dynamic LPAR memory upgrade

An LPAR can be defined with 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 be added later, partially or totally. Although these two memory zones do not have to be contiguous in real memory, they appear as logically contiguous to the operating system that runs in the LPAR.

z/OS can take advantage of this support and non disruptively acquire and release memory from the reserved area. z/VM V7R2 and later can acquire memory non disruptively and immediately make it available to guests. z/VM virtualizes this support to its guests, which now also can increase their memory non disruptively if supported by the guest operating system. Currently, releasing memory from z/VM is supported on z/VM 7.3 and z/VM V7.2 with PTFs². Releasing memory from the z/VM guest depends on the guest's operating system support.

Linux on IBM Z also supports acquiring and releasing memory non disruptively. This feature is enabled for SUSE Linux Enterprise Server 12 and RHEL 7.9 and later releases.

LPAR group absolute capping

Group absolute capping allows you to limit the amount of physical processor capacity that is used by an individual LPAR when a PU that is defined as a CP or an IFL is shared across a set of LPARs. This facility is designed to provide a physical capacity limit that is enforced as an absolute (versus a relative) limit. It is not affected by changes to the logical or physical configuration of the system. This physical capacity limit can be specified in units of CPs or IFLs.

Capacity Provisioning Manager

The provisioning architecture enables clients to better control the configuration and activation of the On/Off CoD. For more information, see [Chapter 8, “System upgrades” on page 317](#). This process is inherently more flexible and can be automated. This capability can result in easier, faster, and more reliable management of the processing capacity.

The Capacity Provisioning Manager, which is a feature that was first available with z/OS V1R9, interfaces with z/OS Workload Manager (WLM) and implements capacity provisioning policies. Several implementation options are available, from an analysis mode that issues only guidelines, to an autonomic mode that provides fully automated operations.

Replacing manual monitoring with autonomic management or supporting manual operation with guidelines can help ensure that sufficient processing power is available with the least possible delay. The supported operating systems are listed in Table 7-3 on page 247.

Program-directed re-IPL

Program directed re-IPL allows an operating system to re-IPL without operator intervention. This function is supported for SCSI and IBM extended count key data (IBM ECKD) devices.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248.

IOCP

All IBM Z servers require a description of their I/O configuration. This description is stored in I/O configuration data set (IOCDS) files. The I/O configuration program (IOCP) allows for the creation of the IOCDS file from a source file that is known as the I/O configuration source (IOCS).

The IOCS file contains definitions of LPARs and channel subsystems. It also includes detailed information for each channel and path assignment, control unit, and device in the configuration.

IOCP for IBM z16 A02 and IBM z16 AGZ provides support for the following features:

- ▶ IBM z16 A02 and IBM z16 AGZ Base machine definition
- ▶ PCI function adapter for zHyperLink (HYL)
- ▶ PCI function adapter for RoCE Express3 (CX6)
- ▶ New hardware (announced with Driver 51)
- ▶ IOCP support for Dynamic I/O for stand-alone CF (Driver 36 and later)

² z/VM Dynamic Memory Downgrade (releasing memory from z/VM LPAR) made available with PTFs for APAR VM66271. For more information, see: <http://www.vm.ibm.com/newfunction/#dmd>

IOCP required level for IBM z16 A02 and IBM z16 AGZ:

- ▶ z/OS V2.R4 and older releases use the same IOCP FMID HIO1104
- ▶ z/OS V2.R5 uses IOCP FMID HIO1105

For more information, see the following publications:

- ▶ *Stand-Alone Input/Output Configuration Program User's Guide*, SB10-7180
- ▶ *Input/Output Configuration Program User's Guide for ICP IOCP*, SB10-7177.

Dynamic Partition Manager V5.1: Dynamic Partition Manager V5.1 is available for managing IBM z16 A02 and IBM z16 AGZ that are running Linux. DPM 5.0 is available with HMC Driver Level 51 (HMC Version 2.16.0). IOCP does not need to configure a server that is running in DPM mode. For more information, see *IBM Dynamic Partition Manager (DPM) Guide*, SB10-7182.

7.4.2 Base CPC features and functions

In this section, we describe the features and functions of Base CPC.

HiperDispatch

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

In z/OS, the IEAOPTxx keyword **HIPERDISPATCH** defaults to YES when it is running on an IBM z16 A01, IBM z16 A02, IBM z16 AGZ, IBM z15, or IBM z14.

The use of SMT on IBM z16 A02 and IBM z16 AGZ requires that HiperDispatch is enabled on the operating system. For more information, see “Simultaneous multithreading” on page 266.

Additionally, any LPAR that is running with more than 64 logical processors is required to operate in HiperDispatch Management Mode.

The following rules control this environment:

- ▶ If an LPAR is defined at IPL with more than 64 logical processors, the LPAR automatically operates in HiperDispatch Management Mode, regardless of the **HIPERDISPATCH=** specification.
- ▶ If logical processors are added to an LPAR that has 64 or fewer logical processors and the extra logical processors raise the number of logical processors to more than 64, the LPAR automatically operates in HiperDispatch Management Mode, regardless of the **HIPERDISPATCH=YES/NO** specification. That is, even if the LPAR has the **HIPERDISPATCH=NO** specification, that LPAR is converted to operate in HiperDispatch Management Mode.
- ▶ An LPAR with more than 64 logical processors that are running in HiperDispatch Management Mode cannot be reverted to run in non-HiperDispatch Management Mode.

HiperDispatch on IBM z16 A02 and IBM z16 AGZ systems uses chip and CPC drawer configuration to improve the access cache performance. It optimizes the system PU allocation with Chip/cluster/drawer cache structure on IBM Z servers. The base support for IBM z16 A02 and IBM z16 AGZ is provided by PTFs that are identified by:

- ▶ IBM.device.server.IBM z16 A02 and IBM z16 AGZ -3931.requiredservice
- ▶ IBM.device.server.IBM z16 A02 and IBM z16 AGZ A02-3932.requiredservice

Important: z16 M/T 3932 users must use the FIXCATs for both the 3932 and the 3931. The Required FIXCAT is absolutely critical, and cannot be understated. Only unique capabilities for the 3932 will be identified with the 3932 FIXCATs.

PR/SM on IBM z16 A02 and IBM z16 AGZ preferentially assigns memory for a system in one CPC drawer that is striped across the DCMs of that drawer to take advantage of the lower latency memory access in a drawer. Also, PR/SM tries to consolidate storage onto drawers with the most processor entitlement.

With HiperDispatch enabled, PR/SM seeks to assign logical processors of a partition to the smallest number of PU chips within a drawer in cooperation with operating system to optimize shared cache usage.

PR/SM automatically keeps a partition's memory and logical processors on the same CPC drawer where possible. This arrangement looks simple for a partition, but it is a complex optimization for multiple logical partitions because some must be split among processors drawers.

All IBM z16 A02 and IBM z16 AGZ processor types can be dynamically reassigned except IFPs.

To use HiperDispatch effectively, WLM goal adjustment might be required. Review the WLM policies and goals and update them as necessary. WLM policies can be changed without turning off HiperDispatch. A health check is provided to verify whether HiperDispatch is enabled on a system image.

z/VM V7R3 and V7R2

z/VM also uses the HiperDispatch facility for improved processor efficiency by better use of the processor cache to take advantage of the cache-rich PU chip, node, and drawer design of the IBM z16 A02 and IBM z16 AGZ.

CPU polarization support in Linux on IBM Z

You can optimize the operation of a vertical SMP environment by adjusting the SMP factor based on the workload demands. For more information about CPU polarization support in Linux on IBM Z, see the [CPU polarization page](#) of IBM Knowledge Center.

z/TPF

z/TPF on IBM z16 A02 and IBM z16 AGZ can use more processors immediately without reactivating the LPAR or IPLing the *z/TPF* system.

When *z/TPF* is running in a shared processor configuration, the achieved MIPS is higher when *z/TPF* is using a minimum set of processors.

In low-utilization periods, *z/TPF* minimizes the processor footprint by compressing TPF workload onto a minimal set of I-streams (engines), which reduces the effect on other LPARs and allows the entire CPC to operate more efficiently.

As a consequence, *z/OS* and *z/VM* experience less contention from the *z/TPF* system when the *z/TPF* system is operating at periods of low demand.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248.

zIIP support

zIIPs do not change the model capacity identifier of IBM z16 A02 and IBM z16 AGZ configurations. IBM software product license charges that are based on the model capacity identifier are not affected by the addition of zIIPs.

2:1 zIIP:CP ratio restriction removal: Starting with IBM z16 A02 and IBM z16 AGZ M/T 3932 announcement as of April 4, 2023, the 2:1 zIIP:CP ratio restriction has been removed. The IBM z16 A01 now supports up to 199 zIIPs, while IBM z16 A02 and IBM z16 AGZ supports up to 67 zIIPs.

No changes to applications are required to use zIIPs. They can be used by the following applications:

- ▶ Db2 V8 and later for z/OS data serving for applications that use data Distributed Relational Database Architecture (DRDA) over TCP/IP, such as data serving, data warehousing, and selected utilities.
- ▶ z/OS XML services.
- ▶ z/OS CIM Server.
- ▶ z/OS Communications Server for network encryption (Internet Protocol Security [IPSec]) and for large messages that are sent by HiperSockets.
- ▶ IBM GBS Scalable Architecture for Financial Reporting.
- ▶ IBM z/OS Global Mirror (formerly XRC) and System Data Mover.
- ▶ IBM z/OS Container Extensions.
- ▶ IBM OMEGAMON® XE on z/OS, OMEGAMON XE on Db2 Performance Expert, and Db2 Performance Monitor.
- ▶ Any Java application that uses the current IBM SDK.
- ▶ Java IBM Semeru Runtime offloading enablement for DLC models exploiting Integrated Accelerator for AI
- ▶ WebSphere Application Server V5R1 and later, and products that are based on it, such as WebSphere Portal, WebSphere Enterprise Service Bus (WebSphere ESB), and WebSphere Business Integration (WBI) for z/OS.
- ▶ CICS/TS V2R3 and later.
- ▶ Db2 UDB for z/OS Version 8 and later.
- ▶ IMS Version 8 and later.
- ▶ zIIP Assisted HiperSockets for large messages.
- ▶ z/OSMF (z/OS Management Facility).
- ▶ IBM z/OS Platform for Apache Spark.
- ▶ IBM Watson® Machine Learning for z/OS.
- ▶ z/OS System Recovery Boost.
- ▶ IBM Python for z/OS Version 3.11
- ▶ Approved 3rd party vendor products.

The use of a zIIP is transparent to application programs. The supported operating systems are listed in Table 7-3 on page 247.

On IBM z16 A02 and IBM z16 AGZ servers, the zIIP processor is designed to run in SMT mode, with up to two threads per processor. This function is designed to help improve

throughput for zIIP workloads and provide appropriate performance measurement, capacity planning, and SMF accounting data. zIIP support is available on all currently supported z/OS versions.

Use the **PROJECTCPU** option of the IEAOPTxx parmlib member to help determine whether zIIPs can be beneficial to the installation. Setting PROJECTCPU=YES directs z/OS to record the amount of eligible work for zIIPs in SMF record type 72 subtype 3. The field APPL% IIPCP of the Workload Activity Report listing by WLM service class indicates the percentage of a processor that is zIIP eligible. Because of the zIIP's lower price as compared to a CP, even a utilization as low as 10% can provide cost benefits.

Transactional Execution³

Transactional Execution (TX) is known in academia and industry as *hardware transactional memory*. Transactional execution is implemented on IBM zSystems servers.

This feature enables software to indicate to the hardware the beginning and end of a group of instructions that must be treated in an atomic way. All of their results occur or none occur, in true transactional style. The execution is optimistic.

The hardware provides a memory area to record the original contents of affected registers and memory as instruction execution occurs. If the transactional execution group is canceled or must be rolled back, the hardware transactional memory is used to reset the values. Software can implement a fallback capability.

This capability increases the software's efficiency by providing a way to avoid locks (lock elision). This advantage is of special importance for speculative code generation and highly parallelized applications.

TX is used by IBM Java virtual machine (JVM) and might be used by other software. The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248.

System Recovery Boost

System Recovery Boost is a feature that was implemented on the IBM z15 system. The feature provides additional temporary processor capacity and delivers substantially faster system shutdown and restart, critical system operations such as stand-alone dump, short duration Recovery process boost for sysplex events and fast catch-up of an accumulated backlog of critical workload after a planned or unplanned event. There is no additional or increasing IBM software costs by using System Recovery Boost. With the IBM z16 A02 and IBM z16 AGZ system additional Recovery process boost scenarios are supported that allow the customer to define some boost granularity. For more information see *IBM Z System Recovery Boost*, REDP-5563-02.

Table 7-14 Operating system support for System Recovery Boost

Boost type ^a	z/OS V2R5	z/OS V2R4	z/OS V2R3	z/VM V7R3	z/VM V7R2	z/TPF V1R1	z/VSE V6R2	Linux on IBM Z
Subcapacity Boost for IPL, Shutdown ^b	Y	Y	Y ^c	Y	Y ^d	Y ^c	Y ^c	N
zIIP boost ^e for IPL, shutdown and dump events	Y	Y	Y ^c	N	N	N	N	N

³ Statement of Direction: In a future IBM Z hardware system family, the transactional execution and constrained transactional execution facility will no longer be supported. Users of the facility on current servers should always check the facility indications before use.

Boost type ^a	z/OS V2R5	z/OS V2R4	z/OS V2R3	z/VM V7R3	z/VM V7R2	z/TPF V1R1	z/VSE V6R2	Linux on IBM Z
Subcapacity and zIIP Boost for process recovery boost in a sysplex ^f .	Y	Y ^c	Y ^c	N	N	N	N	N
Recovery process boosts (IBM z16 A02 and IBM z16 AGZ).	Y ^c	Y ^c	N	N	N	N	N	N

- a. Boost must be enabled for LPARs to opt in.
- b. Subcapacity Boost is also available for stand-alone dump on z/OS and z/VSE.
- c. With Fixes.
- d. Subcapacity boost might be available during the boost period to guest operating systems except for z/OS.
- e. Allows CP work to be dispatched on zIIPs. zIIP processor capacity boost is only available if customer has at least one active processor characterized as zIIP. For *IBM z16 A02 and IBM z16 AGZ, IBM z16 A01 and IBM z15 T01 only*, more zIIPs can be used if obtained through eBOD (temporary zIIP boost records).
- f. Process recovery boosts support subcapacity CPs speed boost and entitled (purchased) customer zIIPs only - zIIPs provided by FC 9930 and FC 6802 cannot be used for process recovery boosts.

Automation

The client's automation product can be used to automate and control the following System Recovery Boost activities:

- ▶ To activate and deactivate the eBod temporary capacity record to provide more physical zIIPs for an IPL or Shutdown Boost.
- ▶ To dynamically modify LPAR weights, as might be needed to modify the sharing of physical zIIP capacity during a Boost period.
- ▶ To drive the invocation of the PROC that indicates the beginning of a shutdown process (and the start of the shut-down Boost).
- ▶ To take advantage of new composite HW API reconfiguration actions.
- ▶ To control the level of parallelism that is present in the workload at startup (for example, starting middleware regions) and shutdown (for example, performing an orderly shutdown of middleware).

Simultaneous multithreading

SMT is the hardware capability to process up to two simultaneous threads in a single core, sharing the resources of the core such as cache, translation lookaside buffer (TLB), and execution resources. This improves system capacity and efficiency by reducing processor delays, which increases the overall throughput of the system.

SMT⁴ is supported for zIIPs and IFLs.

Note: For zIIPs and IFLs, SMT must be enabled on z/OS, z/VM, or Linux on IBM Z instances. An operating system with SMT support can be configured to dispatch work to a thread on a zIIP (for eligible workloads in z/OS) or an IFL (for z/VM) core in single-thread or SMT mode.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248.

⁴ SMT is also enabled (not user configurable) by default for SAPs.

An operating system that uses SMT controls each core and is responsible for maximizing its throughput and meeting workload goals with the smallest number of cores. In z/OS, consider HiperDispatch cache optimization when you must choose the two threads to be dispatched in the same processor.

HiperDispatch attempts to dispatch guest virtual CPUs on the same logical processor on which they ran. PR/SM attempts to dispatch a vertical low logical processor in the same physical processor. If that process is not possible, it attempts to dispatch it in the same node, or then the same CPC drawer where it was dispatched before to maximize cache reuse.

From the point of view of an application, SMT is transparent and no changes are required in the application for it to run in an SMT environment, as shown in Figure 7-1.

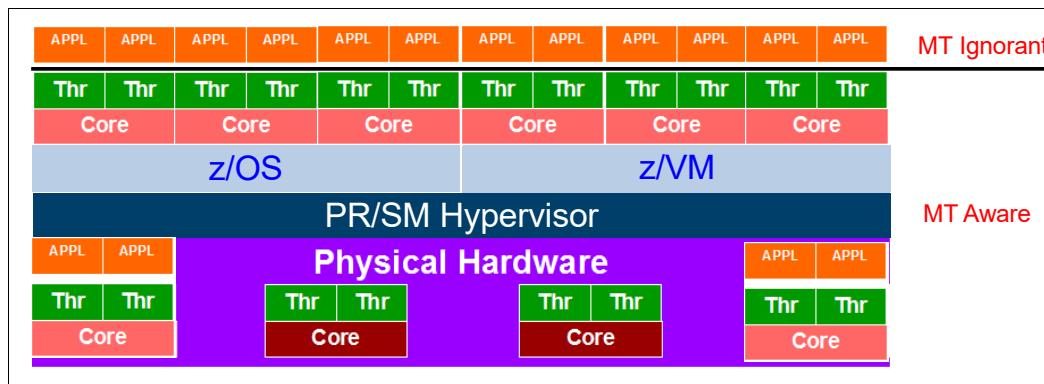


Figure 7-1 Simultaneous multithreading

z/OS

The use of SMT on z/OS requires enabling HiperDispatch, and defining the processor view (**PROCVIEW**) control statement in the LOADxx parmlib member and the **MT_ZIIP_MODE** parameter in the IEAOPTxx parmlib member.

The **PROCVIEW** statement is defined for the life of IPL, and can have the following values:

- **CORE:** This value specifies that z/OS should configure a processor view of core, in which a core can include one or more threads. The number of threads is limited by IBM z16 A02 and IBM z16 AGZ to two threads. If the underlying hardware does not support SMT, a core is limited to one thread.
- **CPU:** This value is the default. It specifies that z/OS should configure a traditional processor view of CPU and not use SMT.
- **CORE,CPU_OK:** This value specifies that z/OS should configure a processor view of core (as with the CORE value) but the CPU parameter is accepted as an alias for applicable commands.

When **PROCVIEW CORE** or **CORE,CPU_OK** are specified in z/OS that is running on an IBM z16 A02 and IBM z16 AGZ, HiperDispatch is forced to run as enabled, and you cannot disable HiperDispatch. The **PROCVIEW** statement cannot be changed dynamically; therefore, you must re-IPL after changing it to make the new setting effective.

The **MT_ZIIP_MODE** parameter in the IEAOPTxx controls zIIP SMT mode. It can be 1 (the default), where only one thread can be running in a core, or 2, where up to two threads can be running in a core. If **PROCVIEW CPU** is specified, the **MT_ZIIP_MODE** is always 1. Otherwise, the use of SMT to dispatch two threads in a single zIIP logical processor (**MT_ZIIP_MODE=2**) can be changed dynamically by using the **SET OPT=xx** setting in the IEAOPTxx parmlib. Changing the MT mode for all cores can take some time to complete.

PROCVIEW CORE requires DISPLAY M=CORE and CONFIG CORE to display the core states and configure an entire core.

With the introduction of Multi-Threading support for SAPs, a maximum of 88 logical SAPs can be used. RMF is updated to support this change by implementing page break support in the I/O Queuing Activity report that is generated by the RMF Post processor.

z/VM V7R3 and V7R2

The use of SMT in z/VM is enabled by using the **MULTITHREADING** statement in the system configuration file. Multithreading is enabled only if z/VM is configured to run with the HiperDispatch vertical polarization mode enabled and with the dispatcher work distribution mode set to reshuffle.

The default in z/VM is multithreading disabled. Dynamic SMT enables dynamically varying the active threads per core. The number of active threads per core can be changed dynamically without a system outage and potential capacity gains going from SMT-1 to SMT-2 (one to two threads per core) can be achieved dynamically.

z/VM V7R3 and V7R2 support up to 40 multithreaded cores (80 threads) for IFLs, and each thread is treated as an independent processor. z/VM dispatches virtual IFLs on the IFL logical processor so that the same or different guests can share a core. Each core has a single dispatch vector, and z/VM attempts to place virtual sibling IFLs on the same dispatch vector to maximize cache reuses.

z/VM guests have no awareness of SMT, and cannot use it directly. z/VM SMT exploitation does not include guest support for multithreading. The value of this support for guests is that the first-level z/VM host of the guests can achieve higher throughput from the multi-threaded IFL cores.

Linux on IBM Z and the KVM hypervisor

The Linux kernel features **SMT functionality** that was developed by the Linux on IBM Z development team. SMT is supported on LPAR only (not as a second-level guest).

The following *minimum* releases of Linux on IBM Z distributions are supported on IBM z16 A02 and IBM z16 AGZ (native SMT support):

- ▶ SUSE:
 - SLES 15 SP4
 - SLES 15 SP3 with service
 - SUSE SLES 12 SP5 with service
- ▶ Red Hat:
 - Red Hat RHEL 9.1
 - Red Hat RHEL 8.4 with service
 - Red Hat RHEL 7.9 with service
- ▶ Ubuntu:
 - Ubuntu 22.04 LTS
 - Ubuntu 20.04.1 LTS with service

The KVM hypervisor is supported on the same Linux on IBM Z distributions in this list.

For most current support, see the [Linux on IBM Z Tested platforms website](#).

Single-instruction multiple-data

The SIMD feature introduces a new set of instructions to enable parallel computing that can accelerate code with string, character, integer, and floating point data types. The SIMD instructions allow a large number of operands to be processed with a single complex instruction.

IBM z16 A02 and IBM z16 AGZ are equipped with a set of instructions to improve the performance of complex mathematical models and analytic workloads through vector processing and complex instructions, which can process numerous data with a single instruction. This set of instructions, which is known as SIMD, enables more consolidation of analytic workloads and business transactions on IBM Z servers.

SIMD on IBM z16 A02 and IBM z16 AGZ has support for enhanced math libraries that provide performance improvements for analytical workloads by processing more information with a single CPU instruction.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248. Operating System support includes the following features⁵:

- ▶ Enablement of vector registers.
- ▶ A math library with an optimized and tuned math function (Mathematical Acceleration Subsystem [MASS]) that can be used in place of some of the C standard math functions. It includes a SIMD vectorized and non-vectorized version.
- ▶ A specialized math library, which is known as Automatically Tuned Linear Algebra Software (ATLAS), that is optimized for the hardware.
- ▶ IBM Language Environment® for C runtime function enablement for ATLAS.
- ▶ DBX to support the disassembly of the new vector instructions, and to display and set vector registers.
- ▶ XML SS exploitation to use new vector processing instructions to improve performance.

MASS and ATLAS can reduce the time and effort for middleware and application developers. IBM provides compiler built-in functions for SIMD that software applications can use as needed, such as for using string instructions.

The use of new hardware instructions require the z/OS V2R5 XL C/C++ compiler with ARCH(14) and TUNE(14) options for targeting IBM z16 A02 and IBM z16 AGZ instructions. The ARCH(14) compiler option allows the compiler to use any new IBM z16 A02 and IBM z16 AGZ instructions where appropriate. The TUNE(14) compiler option allows the compiler to tune for any IBM z16 A02 and IBM z16 AGZ micro-architecture.

Vector programming support is extended for IBM z16 A02 and IBM z16 AGZ to provide access to the new instructions that were introduced by the VEF 2⁶ specification.

Older levels of z/OS XL C/C++ compilers do not provide IBM z16 A02 and IBM z16 AGZ exploitation; however, the z/OS V2R5 XL C/C++ compiler can be used to generate code for the older levels of z/OS running on IBM z16 A02 and IBM z16 AGZ.

The followings compilers include built-in functions for SIMD:

- ▶ IBM Java
- ▶ XL C/C++
- ▶ Enterprise COBOL
- ▶ Enterprise PL/I

⁵ The features that are listed here might not be available on all operating systems that are listed in the tables.

⁶ Hardware-based Vector-extension facility 2.

Code must be developed to take advantage of the SIMD functions. Applications with SIMD instructions abend if they run on a lower hardware level system that do not support SIMD. Some mathematical function replacement can be done without code changes by including the scalar MASS library before the standard math library.

The MASS and standard math library include different accuracies, so assess the accuracy of the functions in the context of the user application before deciding whether to use the MASS and ATLAS libraries.

The SIMD functions can be disabled in z/OS partitions at IPL time by using the **MACHMIG** parameter in the LOADxx member. To disable SIMD code, use the MACHMIG VEF hardware-based vector facility. If you do not specify a **MACHMIG** statement, which is the default, the system not limited in its use of the Vector Facility for z/Architecture (SIMD).

Hardware decimal floating point

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

IBM z16 A02 and IBM z16 AGZ introduce COBOL optimization for Hexadecimal Floating Point (HFP) <--> Binary Coded Decimal (BCD) conversion, and Numeric Editing, and Zoned Decimal operations.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248. For more information, see 7.5.7, “z/OS XL C/C++ considerations” on page 312.

Out-of-order execution

Out-of-order (OOO) execution yields significant performance benefits for compute-intensive applications by reordering instruction execution, which allows later (newer) instructions to be run ahead of a stalled instruction, and reordering storage accesses and parallel storage accesses. OOO maintains good performance growth for traditional applications.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248. For more information, see “3.4.3, “Out-of-Order execution” on page 82.

z/OS DFSORT exploitation of IBM Integrated Accelerator for Z Sort

The Integrated Accelerator for Z Sort, is an “on chip” hardware feature available on IBM z15 and newer servers and which is driven by the SORT LISTS (SORTL) instruction.

z/OS DFSORT takes advantage of Z Sort, providing users with significant performance boosts for their sort workloads. With z/OS DFSORT’s Z Sort algorithm, clients can see batch sort job elapsed time improvements of up to 20–30% depending on record size and CPU time improvements of up to 40% compared to IBM z14.

The function is exploited on z/OS V2.R5 and enabled on z/OS V2.R4 and V2.R3 with PTFs for APAR PH03207.

The sort jobs need to meet certain eligibility criteria - for the full list and other considerations please refer to DFSORT User Guide for PH03207.

CPU Measurement Facility

Also known as Hardware Instrumentation Services (HIS), CPU Measurement Facility (CPUMF) data can be collected by z/OS System Measurement Facility on SMF 113 records. to gain insight into the interaction of workload and hardware it runs on.

CPU MF The supported operating systems are listed in Table 7-3 on page 247.

For more information about this function, see [The Load-Program-Parameter and the CPU-Measurement Facilities](#).

For more information about the CPU Measurement Facility, see the [CPU MF - Update and WSC Experiences page](#) of the IBM Techdocs Library website.

For more information, see “12.2, “IBM z16 A02 and IBM z16 AGZ Large System Performance Reference ratio” on page 458.

z/OS SMF Enhancements for CPACF

SMF 0 records have been enhanced to indicate the number of crypto counters supported by the current IBM Z hardware.

SMF 30 records have been enhanced to include new crypto counter sections that contain counters for CPACF cryptographic instructions utilized by a job in a given period. These sections are produced only for those instructions that are used. These counters are correlated with z/OS jobs and users for the determination of the algorithms, bit lengths and key security utilized by a given workload. This data can aid in compliance, performance and configuration.

The SMF 30 self-defining section indicates the length and number of crypto counter sections

The SMF 30 product / subsystem section indicates if the crypto counters are active

This feature is supported on z/OS 2.4 and later. It requires APAR OA61511.

Large page support

In addition to the existing 1-MB large pages, 4-KB pages, and page frames, IBM z16 A02 and IBM z16 AGZ support pageable 1-MB large pages, large pages that are 2 GB, and large page frames.

z/OS V2.R5 allows 2 GB LFAREA to exceed the prior 4 TB limit, up to 16 TB. All online real storage over 4 TB is part of the 2 GB LFAREA, in addition to what was specified in LFAREA.

Real memory is available only for 2 GB pages.

Applications that make use of 2 GB frames should be reviewed to use more frames if applicable, e.g. Java, Db2.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248.

AI accelerator exploitation

With the IBM z16 A02 and IBM z16 AGZ Integrated Accelerator for AI, clients can benefit from the acceleration of AI operations such as fraud detection, customer behavior predictions and, streamlining of supply chain operations - all in real time. AI accelerator is designed to deliver AI inference in real time, at large scale and rate, with no transaction left behind so clients are able to derive the valuable insights from their data instantly.

The AI capability is applied directly into the running transaction - shifting the traditional paradigm of applying AI to the transactions that have already completed. This innovative technology can be used for intelligent IT workloads placement algorithms, contributing to better overall system performance. The co-processor is driven by the new NNPA (Neural Networks Processing Assist) instruction.

NNPA is a new non-privileged CISC (Complex Instruction Set Computer) memory-to-memory instruction that operates on tensor objects that are in client application program memory. AI functions and macros are abstracted via NNPA.

Virtual Flash Memory

IBM Virtual Flash Memory (FC 0644) offers up to 2.0 TB of memory for IBM z16 A02 and IBM z16 AGZ. VFM is provided for improved application availability and to handle paging workload spikes.

IBM Virtual Flash Memory is designed to help improve availability and handling of paging workload spikes when running z/OS. With this support, z/OS is designed to help improve system availability and responsiveness by using VFM across transitional workload events, such as market openings, and diagnostic data collection. z/OS is also designed to help improve processor performance by supporting middleware exploitation of pageable large (1 MB) pages.

VFM can help organizations meet their most demanding service level agreements and compete more effectively. VFM is easily configurable, and provides rapid time to value.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248.

Guarded Storage Facility

Also known as *less-pausing garbage collection*, Guarded Storage Facility (GSF) is an architecture that was introduced with IBM z14 to enable enterprise scale Java applications to run without periodic pause for garbage collection on larger heaps.

z/OS

GSF support allows an area of storage to be identified such that an Exit routine assumes control if a reference is made to that storage. GSF is managed by new instructions that define Guarded Storage Controls and system code to maintain that control information across undispatch and redispach.

Enabling a less-pausing approach improves Java garbage collection. Function is provided on IBM z14 and subsequent servers that are running z/OS V2.R2 and later with APAR OA51643 installed. **MACHMIG** statement in **LOADxx** of SYS1.PARMLIB provides ability to disable the function.

z/VM

Guarded storage facility is designed to improve the performance of garbage-collection processing by various languages, in particular Java.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248.

Instruction Execution Protection

Instruction Execution Protection (IEP) is a hardware function that enables software, such as Language Environment, to mark certain memory regions (for example, a heap or stack), as non-executable to improve the security of programs running on IBM zSystems against stack-overflow or similar attacks.

Through enhanced hardware features (based on DAT table entry bit) and explicit software requests to obtain memory areas as non-executable, areas of memory can be protected from unauthorized execution. A Protection Exception occurs if an attempt is made to fetch an

instruction from an address in such an element or if an address in such an element is the target of an execute-type instruction.

► z/OS

To use IEP, Real Storage Manager (RSM) is enhanced to request non-executable memory allocation. Use new keyword **EXECUTABLE=YES|NO** on **STORAGE OBTAIN** or **IARV64** to indicate whether memory to be used contains executable code. Recovery Termination Manager (RTM) writes LOGREC record of any program-check that results from IEP.

IEP support is for z/OS V2.R2 and later running on IBM z16 A02 and IBM z16 AGZ with APARs OA51030 and OA51643 installed.

► z/VM

Guest exploitation support for the Instruction Execution Protection Facility is provided with APAR VM65986.

The supported operating systems are listed in Table 7-3 on page 247 and Table 7-4 on page 248.

Secure Boot for Linux and Validated Boot for z/OS

With IBM z16 A02 and IBM z16 AGZ and accompanying z/OS V2.5 operating system support, IBM is providing optional basic support for performing a Validated Boot (IPL) of z/OS systems, using IPL volumes defined and built on ECKD DASD devices. This solution provides digital signatures validation for loaded z/OS executables that have been built and signed as part of the solution. This solution is designed to meet the requirements for achieving the National Information Assurance Partnership (NIAP) OS Protection Profile 4.2.1 Certification.

Secure Boot verification guarantees that the Linux distribution kernel comes from an official provider and has not been compromised. If the signature of the distribution can not be verified, the process of booting the operating system is stopped.

The Secure Boot feature requires OS support (z/OS and Linux on IBM Z).

IBM Integrated Accelerator for zEnterprise Data Compression

The IBM Integrated Accelerator for zEnterprise Data Compression (zEDC) is implemented as on-chip data compression accelerator; that is, Nest Compression Accelerator (NXU) and designed to support Deflate/gzip/zlib algorithms. For more information, see Appendix C, “IBM Integrated Accelerator for zEnterprise Data Compression” on page 475).

Each PU chip has one on-chip compression unit, which is designed to replace the zEnterprise Data Compression (zEDC) Express PCIe feature available on IBM z14 and earlier.

The zEDC Express feature available on older systems is NOT carried forward to IBM z16 A02 and IBM z16 AGZ.

The IBM Integrated Accelerator for zEDC maintains software compatibility with existing zEDC Express use cases. For more information, see [Integrated Accelerator for zEnterprise Data Compression](#).

The z/OS zEDC capability is a software-priced feature that is designed to support compression capable hardware. With IBM z16 A02 and IBM z16 AGZ, the zEDC feature is implemented in the on-chip compression accelerator unit, but the software (z/OS) component is required to maintain the same functionality as previous PCIe based zEDC features.

All data interchange with existing (zEDC) compressed data remains compatible as IBM z16 A02 and IBM z16 AGZ and zEDC capable machines coexist (accessing same data). Data that

is compressed and written with zEDC will be read and decompressed by IBM z16 A02 and IBM z16 AGZ well into the future.

The on-chip compression unit has the following operating modes:

- ▶ Synchronous execution in Problem State, where user application starts instruction in its virtual address space, which provides low latency and high-bandwidth compression/decompression operations). This mode does not require any special hypervisor support, which removes the virtualization layer (sharing the zEDC Express PCIe adapter among LPARs requires virtualization support).
- ▶ Asynchronous optimization for Large Operations under z/OS. The authorized application (for example, BSAM/QSAM) issues I/O for asynchronous execution and SAP (PU) starts instruction (synchronously as described in the previous paragraph) on behalf of application. The on-chip accelerator enables load balancing of high compression loads and low latency and high bandwidth compared to zEDC Express, while maintaining current user experience on compression.

Functionality support for the IBM Integrated Accelerator for zEDC is listed in Table 7-3 on page 247 and Table 7-4 on page 248.

For more information, see Appendix C, “IBM Integrated Accelerator for zEnterprise Data Compression” on page 475.

7.4.3 Coupling and clustering features and functions

In this section, we describe the coupling and cluster features.

Coupling facility and CFCC considerations

Coupling facility (CF) connectivity to an IBM z16 A02 and IBM z16 AGZ is supported on the z15, IBM z14, or another IBM z16 A02 and IBM z16 AGZ. The CFCC levels that are supported on IBM zSystems are listed in Table 7-15.

Table 7-15 IBM zSystems CFCC code-levels

IBM Z server	Code level
IBM z16 A02 and IBM z16 AGZ	CFCC Level 25
IBM z15 T01 and T02	CFCC Level 24
IBM z14M0x and IBM z14ZR1	CFCC Level 23 ^a

a. CFCC Level 22 is not supported when an IBM z16 A02 and IBM z16 AGZ participates in the sysplex.

Consideration: Because coupling link connectivity with an IBM z16 A02 and IBM z16 AGZ, IBM z15, and IBM z14 ZR1 do not support InfiniBand, you must carefully plan your configuration if IBM z14 M0x systems with InfiniBand coupling are present in your configuration. Also, consider the level of CFCC. For more information, see “Migration considerations” on page 179.

CFCC Level 25

CFCC Level 25 is delivered on IBM z16 A02 and IBM z16 AGZ servers with driver level 51. CFCC Level 25 introduces the following enhancements:

- ▶ Scalability Improvements

Processing and dispatching enhancements that result in meaningful scaling of effective throughput up to the limit of 16 ICF processors.

- ▶ Request latency/performance improvements

CFCC and coupling link firmware and hardware improvements to reduce link latency.

- ▶ Elimination of VSAM RLS orphaned castout lock problems and improved VSAM RLS Structure Full recovery processing

Addresses reoccurring problems encountered by installations running VSAM RLS datasharing

Retry Buffer support that is already used on list and lock commands is extended to non-idempotent cache commands and optimized lock commands.

The new support also allows connectors to lock structures to specify a percentage of record data entries to be reserved. These reserved entries are off limits to normal requests to the coupling facility and can only be used if a new keyword is used on lock requests that generate record data entries.

- ▶ Cache residency time metrics

The CF will calculate in microseconds via a moving weighted average the elapsed time a data area or directory entry resides in a storage class before it is reclaimed

XES will return this information on an IXLCACHE REQUEST=READSTGSTATS and IXLMG STRNAME=strname,STGCLASS=stgclass request.

- ▶ Improved exploitation support for handling of lock structure “record data full” conditions, by:

- Thresholding record data structure full conditions to occur when less than 100% full, reserving a special “for emergency use only” pool of record data entries for critical recovery purposes (exploiter-specified threshold).
- Providing new APIs that allow exploiters to make use of this new reserved pool only when needed for recovery actions, but not for normal database locking purposes.

z/OS APAR OA60650 and VSAM RLS APAR OA62059 – z/OS 2.3, 2.4, and 2.5

- ▶ DYNDISP=ONIOFF is deprecated

For CFCC Level 25, DYNDISP=THIN is the only available behavior for shared-engine CF dispatching

Specification of OFF or ON in CF commands and the CF configuration file will be preserved for compatibility, but a warning message will be issued to indicate that these options are no longer supported, and that DYNDISP=THIN behavior will be used.

- ▶ Quantum Safe Signature Checking for CFCC utilizes dual signature scheme

- Further enhancement to the Quantum-Safe security of the IBM z16 A02 and IBM z16 AGZ system, Quantum-safe protection now covers additional CPC firmware, specifically CFCC
 - Both CFCC boot and update processes will be protected
 - Leverages CRYSTALS-Dilithium, recently standardized by NIST for quantum-safe digital signatures
 - Transparent to end users

Before you begin the migration process, install the compatibility and coexistence PTFs. A planned outage is required when you upgrade the CF or CF LPAR to CFCC Level 25.

Upgrades from one CFLEVEL to the next as part of a CPC technology upgrade often cause increases in CF structure sizing requirements, as the coupling facility’s internal data

structures increase in size to accommodate new functions and capabilities, as such, CF structures should always be re-sized on any CFLEVEL increase.

CFCC Level 24

CFCC Level 24 is delivered on IBM z15 servers with driver level 41. CFCC Level 24 introduced the following enhancements:

- ▶ CFCC Fair Latch Manager

This enhancement to the internals of the Coupling Facility (CFCC) dispatcher provides CF work management efficiency and processor scalability improvements, and improve the “fairness” of arbitration for internal CF resource latches across tasks

- ▶ CFCC Message Path Resiliency enhancement

CF Message Paths use a z/OS-provided system identifier (SYID) to uniquely identify which z/OS system image, and instance of that system image, is sending requests over a message path to the CF. With IBM z15, we are providing a new resiliency mechanism that transparently recovers for this “missing” message path deactivate (if and when that deactivation ever occurs).

During path initialization, the CF provides more information to z/OS about every message path that appears active, including the SYID for the path. Whenever z/OS interrogates the state of the message paths to the CF, z/OS checks this SYID information for currency and correctness, and if incorrect, gather diagnostic information and reactivates the path to correct the problem.

- ▶ CF monopolization avoidance

z/OS will take advantage of current CF support in CFLEVEL 24 (IBM z15 T01/T02) to deliver improved z/OS support for handling CF monopolization.

With IBM z15 T01/T02, the CF dispatcher will monitor in real-time the number of CF tasks that have a command assigned to them for a given structure, on a structure-by-structure basis.

When the number of CF tasks being used by any given structure exceeds a model-dependent CF threshold, and a global threshold on the number of active tasks is also exceeded, the structure will be considered to be “monopolizing” the CF, and z/OS will be informed of this monopolization.

New support in z/OS will observe the monopolization state for a structure, and start to selectively queue and throttle incoming requests to the CF, on a structure-specific basis – while other requests, for other “non-monopolizing” structures and workloads, are completely unaffected.

z/OS will dynamically manage the queue of requests for the “monopolizing” structures to limit the number of active CF requests (parallelism) to them, and will monitor the CF’s monopolization state information so as to observe the structure becoming “non-monopolized” again, so that request processing can eventually revert back to a non-throttled mode of operation.

The overall goal of z/OS anti-monopolization support is to protect the ability of ALL well-behaved structures and workloads to access the CF, and get their requests processed in the CF in a timely fashion – while implementing queueing and throttling mechanisms in z/OS to hold back the specific abusive workloads that are causing problems for other workloads.

z/OS XCF/XES exploitation APAR support is required to provide this functionality.

- ▶ CFCC Change Shared-Engine *CF Default* to **DYNDISP=THIN**

Coupling Facility images can run with shared or dedicated processors. Shared processor CFs can operate with different Dynamic Dispatching (DYNDISP) models:

- **DYNDISP=OFF:** LPAR timeslicing completely controls the CF processor.
- **DYNDISP=ON:** an optimization over pure LPAR timeslicing, in which the CFCC code manages timer interrupts to share processors more efficiently.
- **DYNDISP=THIN:** An interrupt-driven model in which the CF processor is dispatched in response to a set of events that generate Thin Interrupts.

Thin Interrupt support was available since zEC12/zBC12, and is proven to be efficient and well-performing in numerous different test and customer shared-engine coupling facility configurations.

Therefore, IBM z15 has made **DYNDISP=THIN** the *default mode* of operation for coupling facility images that use shared processors.

CFCC Level 23

CFCC Level 23 is delivered on IBM z14 servers with driver level 36. In addition to CFCC Level 22 enhancements, it introduces the following enhancements:

- ▶ Asynchronous cross invalidation (XI) for CF cache structures

This enhancement requires z/OS fixes for APARs OA54688 (exploitation) and OA54985 (toleration). It also requires explicit data manager support (Db2 V12 with PTFs).

- ▶ Coupling Facility hang detection

These enhancements provide a significant reduction in failure scope and client disruption (CF-level to structure-level), with no loss of FFDC collection capability. With this support, the CFCC dispatcher significantly reduces the CF hang detection interval to only 2 seconds, which allows more timely detection and recovery from such events.

When a hang is detected, in most cases the CF confines the scope of the failure to “structure damage” for the single CF structure the hung command was processing against, capture diagnostics with a nondisruptive CF dump, and continue operating without stopping or rebooting the CF image.

- ▶ Coupling Facility granular latching

This enhancement eliminates the performance degradation that is caused by structure-wide latching. With this support, most CF list and lock structure ECR processing no longer uses structure-wide latching. It serializes its execution by using the normal structure object latches that all mainline commands use. However, a few “edge conditions” in ECR processing still require structure-wide latching.

For more information about CFCC code levels, see [the Parallel Sysplex page](#) of the IBM IT infrastructure website.

For more information about the latest CFCC code levels, see [the current exception letter](#) that is published on Resource Link website (login is required).

CF structure sizing changes are expected when upgrading from a previous CFCC Level to CFCC Level 25. In fact, CFLEVEL 25 may have more noticeable CF structure size increases associated with it, especially for smaller structures, due to task-related memory increases associated with the increased number of CF tasks in CFLEVEL 25

Review the CF LPAR size by using the available CFSizer tool, which is available for download at: <https://www.ibm.com/support/pages/cfsizer>

Alternatively, the batch SIZER utility is available at:
<https://www.ibm.com/support/pages/cfsizer-alternate-sizing-techniques>

for re-sizing your CF structures as needed. Make sure to update CFRM Policy INITISIZE and/or SIZE values as needed.

Coupling links support

Integrated Coupling Adapter (ICA) Short Reach and Coupling Express2 Long Reach (CE2 LR) coupling link options provide high-speed connectivity at short and longer distances over fiber optic interconnections. For more information, see [4.6.4, “Parallel Sysplex connectivity” on page 179](#).

Integrated Coupling Adapter

PCIe Gen3 coupling fanout, which is also known as Integrated Coupling Adapter Short Reach (ICA SR, ICA SR1.1), supports a maximum distance of 150 meters (492 feet) and is defined as CHPID type CS5 in IOCP.

Coupling Express2 Long Reach⁷

The CE LR link provides point-to-point coupling connectivity at distances of 10 km (6.21 miles) unrepeated and defined as CHPID type CL5 in IOCP. The supported operating systems are listed in Table 7-5 on page 250.

Virtual Flash Memory use by CFCC

VFM can be used in coupling facility images to provide extended capacity and availability for workloads that use WebSphere MQ Shared Queues structures. The use of VFM can help availability by reducing latency from paging delays that can occur at the start of the workday or during other transitional periods. It is also designed to help eliminate delays that can occur when diagnostic data during failures is collected.

Coupling Thin Interrupts (required for IBM z16 A02 and IBM z16 AGZ)

The Coupling Thin Interrupts improves the performance of a CF partition and the dispatching of z/OS LPARs that are awaiting the arrival of returned asynchronous CF requests when used in a shared engine environment.

For more information, see “Coupling Thin Interrupts” on page 103. The supported operating systems are listed in Table 7-5 on page 250.

Asynchronous CF Duplexing for lock structures

Asynchronous CF Duplexing enhancement is a general-purpose interface for any CF Lock structure user. It enables secondary structure updates to be performed asynchronously from primary CF updates. It offers performance advantages for duplexing lock structures and avoids the need for synchronous communication delays during the processing of every duplexed update operation.

Asynchronous CF Duplexing for lock structures requires the following software support:

- ▶ z/OS V2R5, V2R4
- ▶ z/OS V2R3 with PTFs for APAR OA47796 and OA49148
- ▶ z/VM V7R3, V7R2
- ▶ Db2 V12 with PTFs for APAR PI66689 or newer
- ▶ IRLM V2.R3 with PTFs for APAR PI68378

The supported operating systems are listed in Table 7-5 on page 250.

⁷ Coupling Express LR features are NOT carried forward to IBM z16 A02 and IBM z16 AGZ. Coupling Express2 LR is available

Asynchronous cross-invalidate for CF cache structures

Asynchronous cross-invalidate (XI) for CF cache structures enables improved efficiency in CF data sharing by adopting a more transactional behavior for cross-invalidate (XI) processing, which is used to maintain coherency and consistency of data managers' local buffer pools across the sysplex.

Instead of performing XI signals synchronously on every cache update request that causes them, data managers can "opt in" for the CF to perform these XIs asynchronously (and then sync them up with the CF at or before transaction completion). Data integrity is maintained if all XI signals complete by the time transaction locks are released.

The feature enables faster completion of cache update CF requests, especially with cross-site distance involved and provides improved cache structure service times and coupling efficiency. It requires explicit data manager exploitation/participation, which is not transparent to the data manager. No SMF data changes were made for CF monitoring and reporting.

The following requirements must be met:

- ▶ CFCC Level 23 or higher
- ▶ z/OS V2.R5 or V2.R4
- ▶ PTFs on every exploiting system in the sysplex:
 - Fixes for APAR OA54688 - Exploitation support z/OS V2.R3
- ▶ Db2 V12 with PTFs for exploitation

z/VM Dynamic I/O support for ICA CHPIDs

z/VM dynamic I/O configuration support allows you to add, delete, and modify the definitions of channel paths, control units, and I/O devices to the server and z/VM without shutting down the system.

This function refers exclusively to the z/VM dynamic I/O support of InfiniBand⁸ and ICA coupling links. Support is available for the CIB and CS5 CHPID type in the z/VM dynamic commands, including the **change channel path** dynamic I/O command.

Specifying and changing the system name when entering and leaving configuration mode are also supported. z/VM does not use InfiniBand or ICA, and does not support the use of InfiniBand or ICA coupling links by guests. The supported operating systems are listed in Table 7-5 on page 250.

7.4.4 Storage connectivity-related features and functions

In this section, we describe the storage connectivity-related features and functions.

zHyperlink Express

IBM z14 introduced IBM zHyperLink Express as a brand new IBM zSystems input/output (I/O) channel link technology since FICON. The zHyperLink Express 1.1 feature is available with new IBM z16 A02 and IBM z16 AGZ systems and is designed to help bring data close to processing power, increase the scalability of transaction processing, and lower I/O latency.

zHyperLink Express is designed for up to 5x lower latency than High-Performance FICON for IBM Z (zHPF) by directly connecting the IBM Z central processor complex (CPC) to the I/O Bay of the DS8000 (DS8880 or later). This short distance (up to 150 m [492.1 feet]), direct

⁸ InfiniBand coupling is *not* supported on IBM z16 A02 and IBM z16 AGZ.

connection is intended to speed Db2 for z/OS transaction processing and improve active log throughput.

The improved performance of zHyperLink Express allows the Processing Unit (PU) to make a synchronous request for the data that is in the DS8000 cache. This eliminates the undispatch of the running request, the queuing delays to resume the request, and the PU cache disruption.

Support for zHyperLink Writes can accelerate Db2 log writes to help deliver superior service levels by processing high-volume Db2 transactions at speed. IBM zHyperLink Express requires compatible levels of DS8000/F hardware, firmware R8.5.1 or later, and Db2 12 with PTFs and later.

The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

FICON Express32S

FICON Express32S (available for new build IBM z16 A02 and IBM z16 AGZ configurations) supports a link data rate of 32 gigabits per second (Gbps) and auto negotiation to 16 and 8 Gbps for synergy with switches, directors, and storage devices. With support for native FICON, High-Performance FICON for Z (zHPF), and Fibre Channel Protocol (FCP), the IBM z16™ server enables you to position your SAN for even higher performance, which helps you to prepare for an end-to-end 16 Gbps infrastructure to meet the lower latency and increased bandwidth demands of your applications.

The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

IBM Fibre Channel Endpoint Security

IBM z16 A02 and IBM z16 AGZ M/T 3932 supports IBM Fibre Channel Endpoint Security feature (FC 1146). FC 1146 provides FC/FCP link encryption and endpoint authentication. This is an optional priced feature which requires the following:

- ▶ FICON Express32S for both link encryption and endpoint authentication
 - FICON Express16S+ (carry forward) for endpoint authentication only.
- ▶ Select DS8000 storage
- ▶ Supporting infrastructure - IBM Security Guardium Key Lifecycle Manager
- ▶ CPACF enablement (FC 3863)

See the following announcement letter:

<https://www.ibm.com/downloads/cas/US-ENUS120-013-CA/name/US-ENUS120-013-CA.PDF>

FICON Express16S+

FICON Express16S+ (carry forward to IBM z16 A02 and IBM z16 AGZ) supports a link data rate of 16 Gbps and auto negotiation to 4 or 8 Gbps for synergy with switches, directors, and storage devices. With support for native FICON, High-Performance FICON for Z (zHPF), and Fibre Channel Protocol (FCP), the IBM Z systems enable you to position your SAN for even higher performance, which helps you to prepare for an end-to-end 16 Gbps infrastructure to meet the lower latency and increased bandwidth demands of your applications.

The new FICON Express16S+ channel works with your fiber optic cabling environment (single mode and multimode optical cables). The FICON Express16S+ feature running at end-to-end 16 Gbps link speeds provides reduced latency for large read/write operations and increased bandwidth compared to the FICON Express8S feature.

The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

Extended distance FICON

An enhancement to the industry-standard FICON architecture (FC-SB-3) helps avoid degradation of performance at extended distances by implementing a new protocol for persistent IU pacing. Extended distance FICON is transparent to operating systems and applies to all FICON Express32S, FICON Express16SA, and FICON Express16S+ features that carry native FICON traffic (CHPID type FC).

To use this enhancement, the control unit must support the new IU pacing protocol. IBM System Storage™ DS8000 series supports extended distance FICON for IBM zSystems environments. The channel defaults to current pacing values when it operates with control units that cannot use extended distance FICON.

High-performance FICON

High-performance FICON (zHPF) was first provided on System z10®, and is a FICON architecture for protocol simplification and efficiency. It reduces the number of information units (IUs) that are processed. Enhancements were made to the z/Architecture and the FICON interface architecture to provide optimizations for online transaction processing (OLTP) workloads.

zHPF is available on IBM zBC12 and newer IBM zSystems. On IBM z16 A02 and IBM z16 AGZ, the FICON Express32S and FICON Express16S+(CHPID type FC) concurrently support the existing FICON protocol and the zHPF protocol in the server LIC.

When used by the FICON channel, the z/OS operating system, and the DS8000 control unit or other subsystems, the FICON channel processor usage can be reduced and performance improved. Appropriate levels of Licensed Internal Code (LIC) are required.

Also, the changes to the architectures provide end-to-end system enhancements to improve reliability, availability, and serviceability (RAS).

zHPF is compatible with the following standards:

- ▶ Fibre Channel Framing and Signaling standard (FC-FS)
- ▶ Fibre Channel Switch Fabric and Switch Control Requirements (FC-SW)
- ▶ Fibre Channel Single-Byte-4 (FC-SB-4) standards

For example, the zHPF channel programs can be used by the z/OS OLTP I/O workloads, Db2, VSAM, the partitioned data set extended (PDSE), and the z/OS file system (zFS).

At the zHPF announcement, zHPF supported the transfer of small blocks of fixed size data (4 K) from a single track. This capability was extended, first to 64 KB, and then to multitrack operations. The 64 KB data transfer limit on multitrack operations was removed by z196. This improvement allows the channel to fully use the bandwidth of FICON channels, which results in higher throughputs and lower response times.

The multitrack operations extension applies to the FICON Express32S and FICON Express16S+, when configured as CHPID type FC and connecting to z/OS. zHPF requires matching support by the DS8000 series. Otherwise, the extended multitrack support is transparent to the control unit.

zHPF is enhanced to allow all large write operations (greater than 64 KB) at distances up to 100 km (62.13 miles) to be run in a single round trip to the control unit. This process does not elongate the I/O service time for these write operations at extended distances. This

enhancement to zHPF removes a key inhibitor for clients adopting zHPF over extended distances, especially when the IBM HyperSwap capability of z/OS is used.

From the z/OS perspective, the FICON architecture is called *command mode* and the zHPF architecture is called *transport mode*. During link initialization, the channel node and the control unit node indicate whether they support zHPF.

Requirement: All FICON channel path identifiers (CHPIDs) that are defined to the same LCU must support zHPF. The inclusion of any non-compliant zHPF features in the path group causes the entire path group to support command mode only.

The mode that is used for an I/O operation depends on the control unit that supports zHPF and its settings in the z/OS operating system. For z/OS use, a parameter is available in the IECIOSxx member of SYS1.PARMLIB (**ZHPF=YES or NO**) and in the **SETIOS** system command to control whether zHPF is enabled or disabled. The default is ZHPF=NO.

Support is also added for the **D IOS,ZHPF** system command to indicate whether zHPF is enabled, disabled, or not supported on the server.

Similar to the existing FICON channel architecture, the application or access method provides the channel program (CCWs). How zHPF (transport mode) manages channel program operations is different from the CCW operation for the existing FICON architecture (command mode). While in command mode, each CCW is sent to the control unit for execution. In transport mode, multiple channel commands are packaged together and sent over the link to the control unit in a single control block. Fewer processors are used compared to the existing FICON architecture. Certain complex CCW chains are not supported by zHPF.

The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

For more information about FICON channel performance, see the performance technical papers that are available [at the IBM Z I/O connectivity page](#) of the IBM IT infrastructure website.

Modified Indirect Data Address Word facility

The Modified Indirect Data Address Word (MIDAW) facility improves FICON performance. It provides a more efficient channel command word (CCW)/indirect data address word (IDAW) structure for certain categories of data-chaining I/O operations.

The MIDAW facility is a system architecture and software feature that is designed to improve FICON performance. This facility was first made available on System z9 servers, and is used by the Media Manager in z/OS.

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

MIDAW can improve FICON performance for extended format data sets. Non-extended data sets can also benefit from MIDAW.

MIDAW can improve channel utilization and I/O response time. It also reduces FICON channel connect time, director ports, and control unit processor usage.

IBM laboratory tests indicate that applications that use EF data sets, such as Db2, or long chains of small blocks can gain significant performance benefits by using the MIDAW facility.

MIDAW is supported on FICON channels that are configured as CHPID type FC. The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

MIDAW technical description

An IDAW is used to specify data addresses for I/O operations in a virtual environment.⁹ The IDAW design allows the first IDAW in a list to point to any address within a page. Subsequent IDAWs in the same list must point to the first byte in a page. Also, IDAWs (except the first and last IDAW) in a list must manage complete 2 K or 4 K units of data.

Figure 7-2 shows a single CCW that controls the transfer of data that spans non-contiguous 4 K frames in main storage. When the IDAW flag is set, the data address in the CCW points to a list of words (IDAWs). Each IDAW contains an address that designates a data area within real storage.

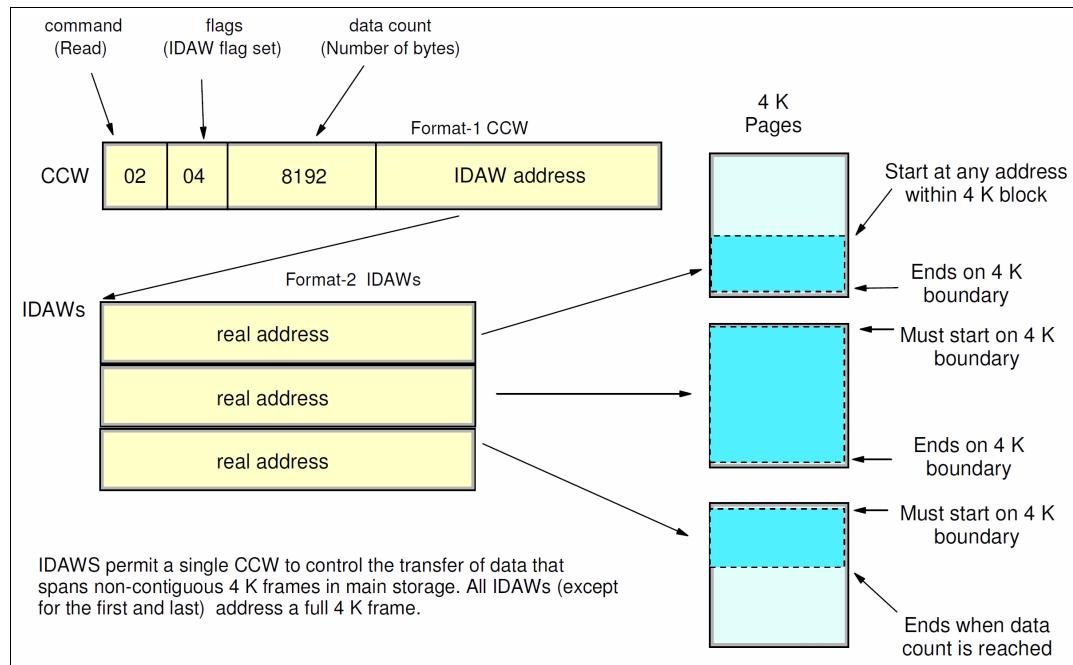


Figure 7-2 IDAW usage

The number of required IDAWs for a CCW is determined by the following factors:

- ▶ IDAW format as specified in the operation request block (ORB)
- ▶ Count field of the CCW
- ▶ Data address in the initial IDAW

For example, three IDAWS are required when the following events occur:

- ▶ The ORB specifies format-2 IDAWs with 4 KB blocks.
- ▶ The CCW count field specifies 8 KB.
- ▶ The first IDAW designates a location in the middle of a 4 KB block.

CCWs with data chaining can be used to process I/O data blocks that have a more complex internal structure, in which portions of the data block are directed into separate buffer areas. This process is sometimes known as *scatter-read* or *scatter-write*. However, as technology evolves and link speed increases, data chaining techniques become less efficient because of switch fabrics, control unit processing and exchanges, and other issues.

⁹ Exceptions are made to this statement, and many details are omitted in this description. In this section, we assume that you can merge this brief description with an understanding of I/O operations in a virtual memory environment.

The MIDAW facility is a method of gathering and scattering data from and into discontinuous storage locations during an I/O operation. The MIDAW format is shown in Figure 7-3. It is 16 bytes long and aligned on a quadword.

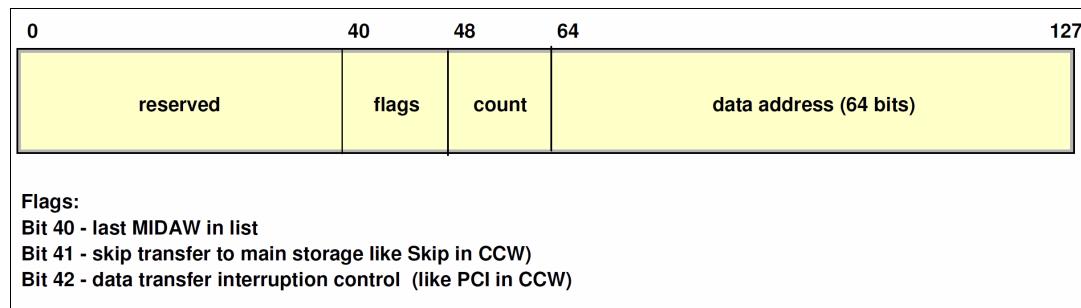


Figure 7-3 MIDAW format

An example of MIDAW usage is shown in Figure 7-4.

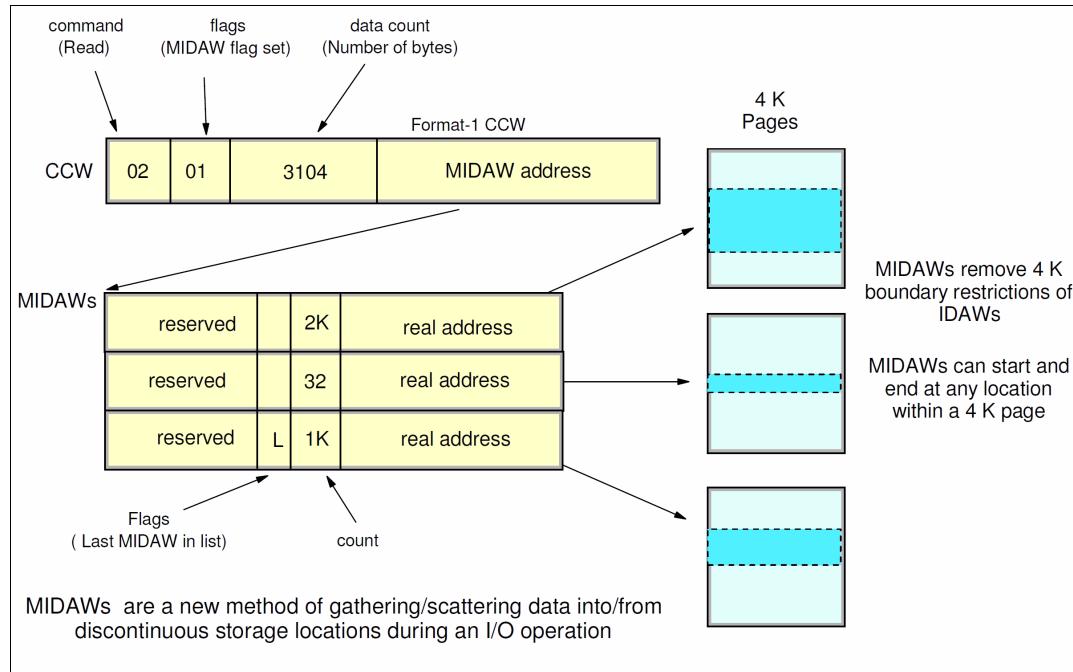


Figure 7-4 MIDAW usage

The use of MIDAWs is indicated by the MIDAW bit in the CCW. If this bit is set, the skip flag cannot be set in the CCW. The skip flag in the MIDAW can be used instead. The data count in the CCW must equal the sum of the data counts in the MIDAWs. The CCW operation ends when the CCW count goes to zero or the last MIDAW (with the last flag) ends.

The combination of the address and count in a MIDAW cannot cross a page boundary. Therefore, the largest possible count is 4 K. The maximum data count of all the MIDAWs in a list cannot exceed 64 K, which is the maximum count of the associated CCW.

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

Extended format data sets

z/OS extended format (EF) data sets use internal structures (often not visible to the application program) that require a scatter-read (or scatter-write) operation. Therefore, CCW data chaining is required, which produces less than optimal I/O performance. Because the most significant performance benefit of MIDAWs is achieved with EF data sets, a brief review of the EF data sets is included here.

VSAM and non-VSAM (DSORG=PS) sets can be defined as EF data sets. For non-VSAM data sets, a 32-byte suffix is appended to the end of every physical record (that is, block) on disk. VSAM appends the suffix to the end of every control interval (CI), which normally corresponds to a physical record.

A 32 K CI is split into two records to span tracks. This suffix is used to improve data reliability, and facilitates other functions that are described next. Therefore, for example, if the DCB BLKSIZE or VSAM CI size is equal to 8192, the actual block on storage consists of 8224 bytes. The control unit does not distinguish between suffixes and user data. The suffix is transparent to the access method and database.

In addition to reliability, EF data sets enable the following functions:

- ▶ DFSMS striping
- ▶ Access method compression
- ▶ Extended addressability (EA)

EA is useful for creating large Db2 partitions (larger than 4 GB). Striping can be used to increase sequential throughput, or to spread random I/Os across multiple logical volumes. DFSMS striping is useful for using multiple channels in parallel for one data set. The Db2 logs are often striped to optimize the performance of Db2 sequential inserts.

Processing an I/O operation to an EF data set normally requires at least two CCWs with data chaining. One CCW is used for the 32-byte suffix of the EF data set. With MIDAW, the additional CCW for the EF data set suffix is eliminated.

MIDAWs benefit EF and non-EF data sets. For example, to read 12 4 K records from a non-EF data set on a 3390 track, Media Manager chains together 12 CCWs by using data chaining. To read 12 4 K records from an EF data set, 24 CCWs are chained (two CCWs per 4 K record). By using Media Manager track-level command operations and MIDAWs, an entire track can be transferred by using a single CCW.

Performance benefits

z/OS Media Manager includes I/O channel program support for implementing EF data sets, and automatically uses MIDAWs when appropriate. Most disk I/Os in the system are generated by using Media Manager.

Users of the Executing Fixed Channel Programs in Real Storage (EXCPVR) instruction can construct channel programs that contain MIDAWs. However, doing so requires that they construct an IOBE with the IOBEMIDA bit set. Users of the EXCP instruction cannot construct channel programs that contain MIDAWs.

The MIDAW facility removes the 4 K boundary restrictions of IDAWs and for EF data sets, reduces the number of CCWs. Decreasing the number of CCWs helps to reduce the FICON channel processor utilization. Media Manager and MIDAWs do not cause the bits to move any faster across the FICON link. However, they reduce the number of frames and sequences that flow across the link, and therefore use the channel resources more efficiently.

The performance of a specific workload can vary based on the conditions and hardware configuration of the environment. IBM laboratory tests found that Db2 gains significant performance benefits by using the MIDAW facility in the following areas:

- ▶ Table scans
- ▶ Logging
- ▶ Utilities
- ▶ Use of DFSMS striping for Db2 data sets

Media Manager with the MIDAW facility can provide significant performance benefits when used in combination applications that use EF data sets (such as Db2) or long chains of small blocks.

For more information about FICON and MIDAW, see the following resources:

- ▶ The [I/O Connectivity page](#) of the IBM IT infrastructure website includes information about FICON channel performance
- ▶ *DS8000 Performance Monitoring and Tuning*, SG24-7146

ICKDSF

Device Support Facilities, ICKDSF, Release 17 is required on all systems that share disk subsystems with an IBM z16 A02 or IBM z16 AGZ processor.

ICKDSF supports a modified format of the CPU information field that contains a two-digit LPAR 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 data corruption, ICKDSF must determine all sharing systems that might run ICKDSF. Therefore, this support is required for IBM z16 A02 and IBM z16 AGZ.

Remember: The need for ICKDSF Release 17 also applies to systems that are not part of the same sysplex, or are running an operating system other than z/OS, such as z/VM.

z/OS Discovery and Auto-Configuration

z/OS Discovery and Auto Configuration (zDAC) is designed to automatically run several I/O configuration definition tasks for new and changed disk and tape controllers that are connected to a switch or director, when attached to a FICON channel.

The zDAC function is integrated into the hardware configuration definition (HCD). Clients can define a policy that can include preferences for availability and bandwidth that include parallel access volume (PAV) definitions, control unit numbers, and device number ranges. When new controllers are added to an I/O configuration or changes are made to existing controllers, the system discovers them and proposes configuration changes that are based on that policy.

zDAC provides real-time discovery for the FICON fabric, subsystem, and I/O device resource changes from z/OS. By exploring the discovered control units for defined logical control units (LCUs) and devices, zDAC compares the discovered controller information with the current system configuration. It then determines delta changes to the configuration for a proposed configuration.

All added or changed logical control units and devices are added into the proposed configuration. They are assigned proposed control unit and device numbers, and channel paths that are based on the defined policy. zDAC uses channel path chosen algorithms to minimize single points of failure. The zDAC proposed configurations are created as work I/O definition files (IODFs) that can be converted to production IODFs and activated.

zDAC is designed to run discovery for all systems in a sysplex that support the function. Therefore, zDAC helps to simplify I/O configuration on IBM z16 A02 and IBM z16 AGZ that run z/OS, and reduces complexity and setup time.

zDAC applies to all FICON features that are supported on IBM z16 A02 and IBM z16 AGZ when configured as CHPID type FC. The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

Platform and name server registration in FICON channel

The FICON Express32S, FICON Express16SA¹⁰, and FICON Express16S+ features support platform and name server registration to the fabric for CHPID types FC and FCP.

Information about the channels that are connected to a fabric (if registered) allows other nodes or storage area network (SAN) managers to query the name server to determine what is connected to the fabric.

The following attributes are registered for the IBM z16 A02 and IBM z16 AGZ:

- ▶ Platform information
- ▶ Channel information
- ▶ Worldwide port name (WWPN)
- ▶ Port type (N_Port_ID)
- ▶ FC-4 types that are supported
- ▶ Classes of service that are supported by the channel

The platform and name server registration service are defined in the Fibre Channel Generic Services 4 (FC-GS-4) standard.

The 63.75-K subchannels

Servers before IBM z9 EC reserved 1024 subchannels for internal system use, out of a maximum of 64 K subchannels. Starting with IBM z9 EC, the number of reserved subchannels was reduced to 256, which increased the number of subchannels that are available. Reserved subchannels exist in subchannel set 0 only. One subchannel is reserved in each of subchannel sets 1, 2, and 3.

The informal name, 63.75-K subchannels, represents 65280 subchannels, as shown in the following equation:

$$63 \times 1024 + 0.75 \times 1024 = 65280$$

This equation is applicable for subchannel set 0. For subchannel sets 1, 2 and 3, the available subchannels are derived by using the following equation:

$$(64 \times 1024) - 1 = 65535$$

The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

Multiple subchannel sets

First introduced in IBM z9 EC, multiple subchannel sets (MSS) provide a mechanism for addressing more than 63.75-K I/O devices and aliases for FICON (CHPID types FC) on the IBM z16 A02 and IBM z16 AGZ, IBM z15, IBM z14, IBM z13, IBM z13s, IBM zEC12, and IBM zBC12. IBM z196 introduced the third subchannel set (SS2). With IBM z13, one more subchannel set (SS3) was introduced, which expands the alias addressing by 64-K more I/O devices.

¹⁰ FICON Express16 SA is not supported on IBM z16 A02 or IBM z16 AGZ.

Current z/VM versions MSS support for mirrored direct access storage device (DASD) provides a subset of host support for the MSS facility to allow the use of an alternative subchannel set for Peer-to-Peer Remote Copy (PPRC) secondary volumes.

The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251. For more information about channel subsystem, see Chapter 5, “Logical I/O - Channel Subsystem” on page 185.

Subchannel sets

IBM z16 A02 and IBM z16 AGZ support four subchannel sets (SS0, SS1, SS2, SS3).

Subchannel sets SS1, SS2, and SS3 can be used for disk alias devices of primary and secondary devices, and as Metro Mirror secondary devices. This set helps facilitate storage growth and complements other functions, such as extended address volume (EAV) and Hyper Parallel Access Volumes (Hyperplane).

See Table 7-6 on page 250 and Table 7-7 on page 251 for list of supported operating systems.

IPL from an alternative subchannel set

IBM z16 A02 and IBM z16 AGZ support IPL from subchannel set 1 (SS1), subchannel set 2 (SS2), or subchannel set 3 (SS3), in addition to subchannel set 0.

See Table 7-6 on page 250 and Table 7-7 on page 251 for list of supported operating systems. For more information, refer to “IPL from an alternative subchannel set” on page 288.

32 K subchannels

To help facilitate growth and continue to enable server consolidation, the IBM z16 A02 and IBM z16 AGZ support up to 32 K subchannels per FICON Express32S, FICON Express16SA and FICON Express16S+ channels (CHPID). More devices can be defined per FICON channel, which includes primary, secondary, and alias devices. The maximum number of subchannels across all device types that are addressable within an LPAR remains at 63.75 K for subchannel set 0 and 64 K (64 X 1024)-1 for subchannel sets 1, 2, and 3.

This support is available to the IBM z16 A02 and IBM z16 AGZ, IBM z15, z14, IBM z13, and IBM z13s servers and applies to the FICON Express32S, FICON Express16SA¹¹, and FICON Express16S+ features (defined as CHPID type FC).

The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

Request node identification data

The request node identification data (RNID) function for native FICON CHPID type FC allows isolation of cabling-detected errors. The supported operating systems are listed in Table 7-6 on page 250.

FICON link incident reporting

FICON link incident reporting allows an operating system image (without operator intervention) to register link incident reports. The supported operating systems are listed in Table 7-6 on page 250.

¹¹ FICON Express 16SA is not supported on IBM z16A02 or IBM z16 AGZ.

Health Check for FICON Dynamic routing

Starting with IBM z13, the channel microcode was changed to support FICON dynamic routing. Although change is required in z/OS to support dynamic routing, I/O errors can occur if the FICON switches are configured for dynamic routing despite the missing support in the processor or storage controllers. Therefore, a health check is provided that interrogates the switch to determine whether dynamic routing is enabled in the switch fabric.

No action is required on z/OS to enable the health check; it is automatically enabled at IPL and reacts to changes that might cause problems. The health check can be disabled by using the **PARMLIB** or **SDSF** modify commands.

The supported operating systems are listed in Table 7-6 on page 250. For more information about FICON Dynamic Routing (FIDR), see Chapter 4, “I/O structure” on page 137.

Global resource serialization FICON CTC toleration

For some configurations that depend on ESCON CTC definitions, global resource serialization (GRS) FICON CTC toleration that is provided with APAR OA38230 is essential, especially after ESCON channel support was removed from IBM zSystems starting with IBM zEC12.

The supported operating systems are listed in Table 7-6 on page 250.

Increased performance for the FCP protocol

The FCP LIC is modified to help increase I/O operations per second for small and large block sizes, and to support 32-Gbps link speeds.

For more information about FCP channel performance, see [the performance technical papers that are available](#) at the IBM zSystems I/O connectivity page of the IBM IT infrastructure website.

The FCP protocol is supported by z/VM, z/VSE, and Linux on IBM Z. The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

T10-DIF support

American National Standards Institute (ANSI) T10 Data Integrity Field (DIF) standard is supported on IBM Z for SCSI end-to-end data protection on fixed block (FB) LUN volumes. IBM Z provides added end-to-end data protection between the operating system and the DS8870 unit. This support adds protection information that consists of Cyclic Redundancy Checking (CRC), Logical Block Address (LBA), and host application tags to each sector of FB data on a logical volume.

IBM Z support applies to FCP channels only. The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

N_Port ID Virtualization

N_Port ID Virtualization (NPIV) allows multiple system images (in LPARs or z/VM guests) to use a single FCP channel as though each were the sole user of the channel. First introduced with z9 EC, this feature can be used with supported FICON features on IBM z16 A02 and IBM z16 AGZ. The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

Worldwide port name tool

Part of the IBM z16 A02 and IBM z16 AGZ system installation is the pre-planning of the SAN environment. IBM includes a stand-alone tool to assist with this planning before the installation.

The capabilities of the WWPN are extended to calculate and show WWPNs for virtual and physical ports ahead of system installation.

The tool assigns WWPNs to each virtual FCP channel or port by using the same WWPN assignment algorithms that a system uses when assigning WWPNs for channels that use NPIV. Therefore, the SAN can be set up in advance, which allows operations to proceed much faster after the server is installed. In addition, the SAN configuration can be retained instead of altered by assigning the WWPN to physical FCP ports when a FICON feature is replaced.

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 be created manually or exported from the HCD/HCM. The supported operating systems are listed in Table 7-6 on page 250 and Table 7-7 on page 251.

The WWPN tool is applicable to all FICON channels that are defined as CHPID type FCP (for communication with SCSI devices) on IBM z16 A02 and IBM z16 AGZ. It is available [for download at the Resource Link](#) at the following website (log in is required).

Note: An optional feature can be ordered for WWPN persistency before shipment to keep the same I/O serial number on the new CPC. Current information must be provided during the ordering process.

7.4.5 Networking features and functions

In this section, we describe the networking features and functions supported on IBM z16 A02 and IBM z16 AGZ M/T 3932.

25GbE RoCE Express3 SR and 25GbE RoCE Express3 LR

25GbE RoCE Express3 features (FC 0452 and FC 0453) are RDMA/Ethernet technology refresh. The features support now also LR optics and 25 Gbit Ethernet switches with relaxed requirements. The features are installed in the PCIe+ I/O drawer and are supported only on IBM z16 A02 and IBM z16 AGZ.

The 25GbE RoCE Express3 are native PCIe features which not use a CHPID and are defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

The virtualization capabilities for IBM z16 A02 and IBM z16 AGZ are 63 Virtual Functions per port (126 VFs per feature/PCHID). One RoCE Express feature can be shared by up to 126 partitions (LPARs) (one adapter is one PCHID).

The 25GbE RoCE Express3 SR (FC 0452) feature connects using SR optics and multimode fiber terminated with LC connector, while 25GbE RoCE Express3 LR (FC 0453) connects using LR optics and single mode fiber terminated with LC connector.

Support for select Linux on IBM Z distributions is provided for Shared Memory Communications over Remote Direct Memory Access (SMC-R) by using RoCE Express features. For more information, see [this Linux on IBM Z Blogspot web page](#).

The RoCE Express3 features can also provide local area network (LAN) connectivity for Linux on IBM zSystems, and comply with IEEE standards. In addition, RoCE Express features assume several functions of the TCP/IP stack that normally are performed by the PU, which allows significant performance benefits by offloading processing from the operating system.

10GbE RoCE Express3 SR and 10GbE RoCE Express3 LR

10GbE RoCE Express3 features (FC 0440 and FC 0441) are a RDMA/Ethernet technology refresh. The features support now also LR optics and 10 Gbit Ethernet switches with relaxed requirements. The features are installed in the PCIe+ I/O drawer and are supported only on IBM z16 A02 and IBM z16 AGZ configurations.

The 10GbE RoCE Express3 are native PCIe features which not use a CHPID and are defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

The virtualization capabilities for IBM z16 A02 and IBM z16 AGZ are 63 Virtual Functions per port (126 VFs per feature/PCHID). One RoCE Express feature can be shared by up to 126 partitions (LPARs) (one adapter is one PCHID).

The 25GbE RoCE Express3 SR (FC 0452) feature connects using SR optics and multimode fiber terminated with LC connector, while 25GbE RoCE Express3 LR (FC 0453) connects using LR optics and single mode fiber terminated with LC connector.

25GbE RoCE Express2.1 and 25GbE RoCE Express2

Based on the RoCE Express2 generation hardware, the 25GbE RoCE Express2 (FC 0430 and 0450) features provide two 25GbE physical ports and requires 25GbE optics and Ethernet switch 25GbE support. The switch port must support 25GbE (negotiation down to 10GbE is not supported).

The 25GbE RoCE Express2 has one PCHID and the same virtualization characteristics and the 10GbE RoCE Express2 (FC 0412 and FC 0432); that is, 126 Virtual Functions per PCHID.

z/OS requires fixes for APAR OA55686. RMF 2.2 and later is also enhanced to recognize the CX4 card type and properly display CX4 cards in the PCIe Activity reports.

25GbE RoCE Express2 feature also are used by Linux on IBM Z for applications that are coded to the native RoCE verb interface or use Ethernet (such as TCP/IP). This native exploitation does not require a peer OSA.

Support for select Linux on IBM Z distributions is now provided for Shared Memory Communications over Remote Direct Memory Access (SMC-R) by using RoCE Express features. For more information, see [this Linux on IBM Z Blogspot web page](#).

The RoCE Express3 features can also provide local area network (LAN) connectivity for Linux on IBM Z, and comply with IEEE standards. In addition, RoCE Express features assume several functions of the TCP/IP stack that normally are performed by the PU, which allows significant performance benefits by offloading processing from the operating system.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

10GbE RoCE Express2.1 and 10GbE RoCE Express2

IBM 10GbE RoCE Express2 provides a natively attached PCIe I/O Drawer-based Ethernet feature that supports 10 Gbps Converged Enhanced Ethernet (CEE) and RDMA over CEE

(RoCE). The RoCE feature, with an OSA feature, enables shared memory communications between two CPCs by using a shared switch.

RoCE Express2 provides increased virtualization (sharing capability) by supporting 63 Virtual Functions (VFs) per physical port for a total of 126 VFs per PCHID. This configuration allows RoCE to be extended to more workloads.

z/OS Communications Server (CS) provides a new software device driver ConnectX4 (CX4) for RoCE Express2. The device driver is not apparent to both upper layers of the CS (the SMC-R and TCP/IP stack) and application software (using TCP sockets). RoCE Express2 introduces a minor change in how the physical port is configured.

RMF 2.2 and later is also enhanced to recognize the new CX4 card type and properly display CX4 cards in the PCIE Activity reports.

Support in select Linux on IBM Z distributions is now provided for Shared Memory Communications over Remote Direct Memory Access (SMC-R) using the supported RoCE Express features. For more information, see [this Linux on IBM Z Blogspot web page](#).

The RoCE Express3 features can also provide local area network (LAN) connectivity for Linux on IBM Z, and comply with IEEE standards. In addition, RoCE Express features assume several functions of the TCP/IP stack that normally are performed by the PU, which allows significant performance benefits by offloading processing from the operating system.

The 10GbE RoCE Express2 feature also is used by Linux on IBM Z for applications that are coded to the native RoCE verb interface or use Ethernet (such as TCP/IP). This native use does not require a peer OSA.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

Shared Memory Communication - Direct Memory Access

First introduced with IBM z13 servers, the Shared Memory Communication - Direct Memory Access (SMC-D) feature maintains the socket-API transparency aspect of SMC-R so that applications that use TCPI/IP communications can benefit immediately without requiring application software to undergo IP topology changes.

Similar to SMC-R, this protocol uses shared memory architectural concepts that eliminate TCP/IP processing in the data path, yet preserve TCP/IP Qualities of Service for connection management purposes.

Support in select Linux on IBM Z distributions is now provided for Shared Memory Communications over Direct Memory Access (SMC-D). For more information, see [this Linux on IBM Z Blogspot web page](#).

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

Shared Memory Communications Version 2

Shared Memory Communications v2 is available in z/OS V2R4 (with PTFs) and z/OS V2R5.

The initial version of SMC was limited to TCP/IP connections over the same layer 2 network and therefore was not routable across multiple IP subnets. The associated TCP/IP connection was limited to hosts within a single IP subnet requiring the hosts to have direct access to the same physical layer 2 network (i.e. same Ethernet LAN over a single VLAN ID). The scope of eligible TCP/IP connections for SMC was limited to and defined by the single IP subnet.

SMC Version 2 (SMCv2) provides support for SMC over multiple IP subnets for both SMC-D and SMC-R and is referred to as SMC-Dv2 and SMC-Rv2. SMCv2 requires updates to the underlying network technology. SMC-Dv2 requires ISMv2 and SMC-Rv2 requires RoCEv2.

The SMCv2 protocol is downward compatible allowing SMCv2 hosts to continue to communicate with SMCv1 down-level hosts.

While SMCv2 changes the SMC connection protocol enabling multiple IP subnet support, SMCv2 does not change how actual user TCP socket data is transferred, which preserves the benefits of SMC to TCP workloads.

TCP/IP connections that require IPSec are not eligible for any form of SMC.

HiperSockets Completion Queue

The HiperSockets Completion Queue function is implemented on IBM z16 A01, IBM z16 A02, IBM z16 AGZ, IBM z15, and IBM z14. This function is designed to allow HiperSockets to transfer data synchronously (if possible) and asynchronously, if necessary. Therefore, it combines ultra-low latency with more tolerance for traffic peaks. HiperSockets Completion Queue can be especially helpful in burst situations. The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

HiperSockets Virtual Switch Bridge

The HiperSockets Virtual Switch Bridge is supported on IBM Z servers. With the HiperSockets Virtual Switch Bridge, 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 components:

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

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

HiperSockets Multiple Write Facility

The HiperSockets Multiple Write Facility allows the streaming of bulk data over a HiperSockets link between two LPARs. Multiple output buffers are supported on a single Signal Adapter (SIGA) write instruction. The key advantage of this enhancement is that it allows the receiving LPAR to process a much larger amount of data per I/O interrupt. This process is transparent to the operating system in the receiving partition. HiperSockets Multiple Write Facility with fewer I/O interrupts is designed to reduce processor utilization of the sending and receiving partitions.

Support for this function is required by the sending operating system. For more information, see “HiperSockets” on page 173. The supported operating systems are listed in Table 7-8 on page 253.

HiperSockets support of IPv6

IPv6 is a key element in the future of networking. The IPv6 support for HiperSockets allows compatible implementations between external networks and internal HiperSockets networks. The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

HiperSockets Layer 2 support

For flexible and efficient data transfer for IP and non-IP workloads, the HiperSockets internal networks on IBM Z servers can support two transport modes: Layer 2 (Link Layer) and the current Layer 3 (Network or IP Layer). Traffic can be Internet Protocol (IP) Version 4 or Version 6 (IPv4, IPv6) or non-IP (AppleTalk, DECnet, IPX, NetBIOS, or SNA).

HiperSockets devices are protocol-independent and Layer 3-independent. Each HiperSockets device features its own Layer 2 Media Access Control (MAC) address. This MAC address allows 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 a non-mainframe environment.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

HiperSockets network traffic analyzer for Linux on IBM Z

HiperSockets network traffic analyzer (HS NTA) provides support for tracing Layer2 and Layer3 HiperSockets network traffic in Linux on IBM Z. This support allows Linux on IBM Z to control the trace for the internal virtual LAN to capture the records into host memory and storage (file systems).

Linux on IBM Z tools can be used to format, edit, and process the trace records for analysis by system programmers and network administrators.

OSA-Express7S 1.2 25 GbE LR and SR features

OSA-Express7S 1.2 features are an Ethernet technology refresh introduced with IBM z16 A02 and IBM z16 AGZ.

OSA-Express7S 1.2 25 GbE SR (FC 0459) and OSA-Express7S 1.2 25 GbE LR (FC 0460) are installed in the PCIe+ I/O Drawer and have 25 GbE physical port. New with the generation is the Long Reach version which uses single mode fiber and can be point to point connected to a distance of up to 10 km (6.2 miles). The features connect to a 25 GbE switch and do not support auto-negotiation to a different speed.

Consider the following points regarding operating system support:

- ▶ z/OS V2R3 requires fixes for the following APARs: OA55256 (VTAM) and PI95703 (TCP/IP).

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

OSA-Express7S 1.2 10 GbE LR and SR features

OSA-Express7S 1.2 features are an Ethernet technology refresh introduced with IBM z16 A02 and IBM z16 AGZ.

OSA-Express7S 1.2 10 GbE SR (FC 0457) and OSA-Express7S 1.2 10 GbE LR (FC 0456) are installed in the PCIe+ I/O Drawer and have 10 GbE physical port. The features connect to a 10 GbE switch and do not support auto-negotiation to a different speed.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

OSA-Express7S 1.2 GbE LX and SX features

OSA-Express7S 1.2 features are an Ethernet technology refresh introduced with IBM z16 A02 and IBM z16 AGZ.

OSA-Express7S 1.2 GbE SX (FC 0455) and OSA-Express7S 1.2 10 GbE LX (FC 0454) are installed in the PCIe+ I/O Drawer and have two GbE physical ports. The features connect to a GbE switch and do not support auto-negotiation to a different speed.

Each adapter can be configured in the following modes:

- ▶ QDIO mode, with CHPID types OSD
- ▶ Local 3270 emulation mode, including OSA-ICC, with CHPID type OSC

Note: Operating system support is required to recognize and use the second port on the OSA-Express6S 1000BASE-T Ethernet feature.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

OSA-Express7S 1.2 1000BASE-T features (FC 0458)

OSA-Express7S 1.2 features are an Ethernet technology refresh introduced with IBM z16 A02 and IBM z16 AGZ. The performance characteristics are comparable to the OSA-Express7S and OSA-Express6S features and they also retain the same form factor and port granularity.

Removal of support for OSA-Express 1000BASE-T hardware adapters^a: IBM z16 A02 and IBM z16 AGZ is planned to be the last IBM zSystems to support OSA-Express 1000BASE-T hardware adapters (FC 0426, 0446, and 0458) on new build servers. Definition of all valid OSA CHPID types will be allowed only on OSA-Express GbE adapters. IBM plans to continue moving forward with OSA CHPID types on higher bandwidth fiber Ethernet adapters on future servers.

- a. Statements by IBM regarding its plans, directions, and intent are subject to change or withdrawal without notice at the sole discretion of IBM.

Each adapter can be configured in the following modes:

- ▶ QDIO mode, with CHPID types OSD
- ▶ Non-QDIO mode, with CHPID type OSE
- ▶ Local 3270 emulation mode, including OSA-ICC, with CHPID type OSC

Notes: Consider the following points:

- ▶ Operating system support is required to recognize and use the second port on the OSA-Express7S 1.2 1000BASE-T Ethernet feature.
- ▶ OSA-Express7S 1.2 1000BASE-T Ethernet feature supports only 1000 Mbps duplex mode (no auto-negotiation to 100 or 10 Mbps)

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

OSA-Express6S 10-Gigabit Ethernet LR and SR (carry forward)

OSA-Express6S 10-Gigabit Ethernet features are installed in the PCIe I/O drawer, which is supported by the 16 GBps PCIe Gen3 host bus. The performance characteristics are comparable to the OSA-Express5S features and they also retain the same form factor and

port granularity. OSA-Express6S features have been introduced with IBM z14 and can be carried forward to an IBM z15 (T01 and T02), as well as ordered with a new IBM z16 A02 and IBM z16 AGZ.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

OSA-Express6S 1000BASE-T Ethernet (carry forward)

IBM z14 has introduced an Ethernet technology refresh with OSA-Express6S 1000BASE-T Ethernet features to be installed in the PCIe I/O drawer, which is supported by the 16 GBps PCIe Gen3 host bus. The performance characteristics are comparable to the OSA-Express5S features and they also retain the same form factor and port granularity.

Each adapter can be configured in the following modes:

- ▶ QDIO mode, with CHPID types OSD
- ▶ Non-QDIO mode, with CHPID type OSE
- ▶ Local 3270 emulation mode, including OSA-ICC, with CHPID type OSC

Note: Operating system support is required to recognize and use the second port on the OSA-Express6S 1000BASE-T Ethernet feature.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

OSA-Integrated Console Controller

The OSA-Express 1000BASE-T Ethernet features provide the Integrated Console Controller (OSA-ICC) function, which supports TN3270E (RFC 2355) and non-SNA DFT 3270 emulation. The OSA-ICC function is defined as CHPID type OSC and console controller, and includes multiple LPAR support as shared or spanned channels.

Removal of support for OSA-Express 1000BASE-T hardware adapters^a: IBM z16 A02 and IBM z16 AGZ is planned to be the last IBM zSystems to support OSA-Express 1000BASE-T hardware adapters (FC 0426, 0446, and 0458) on new build servers. Definition of all valid OSA CHPID types will be allowed only on OSA-Express GbE adapters. IBM plans to continue moving forward with OSA CHPID types on higher bandwidth fiber Ethernet adapters on future servers.

- a. Statements by IBM regarding its plans, directions, and intent are subject to change or withdrawal without notice at the sole discretion of IBM.

With the OSA-ICC function, 3270 emulation for console session connections is integrated in the IBM z16 A02 and IBM z16 AGZ through a port on the OSA-Express7S 1000BASE-T, OSA-Express7S GbE, OSA-Express6S 1000BASE-T, or OSA-Express5S 1000BASE-T.

OSA-ICC can be configured on a PCHID-by-PCHID basis, and is supported at any of the feature settings. Each port can support up to 120 console session connections.

To improve security of console operations and to provide a secure, validated connectivity, OSA-ICC supports Transport Layer Security/Secure Sockets Layer (TLS/SSL) with Certificate Authentication starting with IBM z13 GA2 (Driver level 27).

Note: OSA-ICC supports up to 48 *secure* sessions per CHPID (the overall maximum of 120 connections is unchanged).

OSA-ICC Enhancements

With HMC 2.14.1 and newer the following enhancements are available:

- ▶ The IPv6 communications protocol is supported by OSA-ICC 3270 so that clients can comply with regulations that require all computer purchases to support IPv6.
- ▶ TLS negotiation levels (the supported TLS protocol levels) for the OSA-ICC 3270 client connection can now be specified:
 - TLS 1.0 OSA-ICC 3270 server permits TLS 1.0, TLS 1.1, and TLS 1.2 client connections.
 - TLS 1.1 OSA-ICC 3270 server permits TLS 1.1 and TLS 1.2 client connections.
 - TLS 1.2 OSA-ICC 3270 server permits only TLS 1.2 client connections.
- ▶ Separate and unique OSA-ICC 3270 certificates are supported (for each PCHID), for the benefit of customers who host workloads across multiple business units or data centers, where cross-site coordination is required. Customers can avoid interruption of all the TLS connections at the same time when having to renew expired certificates.
OSA-ICC continues to also support a single certificate for all OSA-ICC PCHIDs in the system.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

Checksum offload for in QDIO mode (CHPID type OSD)

Checksum offload provides the capability of calculating the Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and IP header checksum. Checksum verifies the accuracy of files. By moving the checksum calculations to a Gigabit or 1000BASE-T Ethernet feature, host processor cycles are reduced and performance is improved.

Checksum offload provides checksum offload for several types of traffic and is supported by the following features when configured as CHPID type OSD (QDIO mode only):

- ▶ OSA-Express7S 1.2 25GbE
- ▶ OSA-Express7S 1.2 10GbE
- ▶ OSA-Express7S 1.2 GbE
- ▶ OSA-Express7S 1.2 1000BASE-T Ethernet
- ▶ OSA-Express6S 10GbE
- ▶ OSA-Express6S GbE
- ▶ OSA-Express6S 1000BASE-T Ethernet

When checksum is offloaded, the OSA-Express feature runs the checksum calculations for Internet Protocol version 4 (IPv4) and Internet Protocol version 6 (IPv6) packets. The checksum offload function applies to packets that go to or come from the LAN.

When multiple IP stacks share an OSA-Express, and an IP stack sends a packet to a next hop address that is owned by another IP stack that is sharing the OSA-Express, OSA-Express sends the IP packet directly to the other IP stack. The packet does not have to be placed out on the LAN, which is termed LPAR-to-LPAR traffic. Checksum offload is enhanced to support the LPAR-to-LPAR traffic, which was not originally available.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

Querying and displaying an OSA configuration

OSA-Express3 introduced the capability for the operating system to query and display directly the current OSA configuration information (similar to OSA/SF). z/OS uses this OSA capability

by introducing the TCP/IP operator command **display OSAINFO**. z/VM provides this function with the **NETSTAT OSAINFO TCP/IP** command.

The use of **display OSAINFO** (z/OS) or **NETSTAT OSAINFO** (z/VM) allows the operator to monitor and verify the current OSA configuration and helps improve the overall management, serviceability, and usability of OSA-Express cards.

These commands apply to CHPID type OSD. The supported operating systems are listed in Table 7-8 on page 253.

QDIO data connection isolation for z/VM

The QDIO data connection isolation function provides a higher level of security when sharing an OSA connection in z/VM environments that use VSWITCH. The VSWITCH is a virtual network device that provides switching between OSA connections and the connected guest systems.

QDIO data connection isolation allows disabling internal routing for each QDIO connected. It also provides a means for creating security zones and preventing network traffic between the zones.

QDIO data connection isolation is supported by all OSA-Express features on IBM z16 A02 and IBM z16 AGZ. The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

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 connection 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 OSA Address Table (OAT) as isolated.

QDIO interface isolation is supported on all OSA-Express features on IBM z16 A02 and IBM z16 AGZ. The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

QDIO optimized latency mode

QDIO optimized latency mode (OLM) can help improve performance for applications that feature a critical requirement to minimize response times for inbound and outbound data.

OLM optimizes the interrupt processing in the following manner:

- ▶ For inbound processing, the TCP/IP stack looks more frequently for available data to process. This process ensures that any new data is read from the OSA-Express features without needing more program controlled interrupts (PCIs).
- ▶ For outbound processing, the OSA-Express cards also look more frequently for available data to process from the TCP/IP stack. Therefore, the process does not require a Signal Adapter (SIGA) instruction to determine whether more data is available.

The supported operating systems are listed in Table 7-8 on page 253.

QDIO Diagnostic Synchronization

QDIO Diagnostic Synchronization enables system programmers and network administrators to coordinate and simultaneously capture software and hardware traces. It allows z/OS to signal OSA-Express features (by using a diagnostic assist function) to stop traces and capture the current trace records.

QDIO Diagnostic Synchronization is supported by the OSA-Express features on IBM z16 A02 and IBM z16 AGZ when in QDIO mode (CHPID type OSD). The supported operating systems are listed in Table 7-8 on page 253.

Inbound workload queuing (IWQ) for OSA

Inbound workload queuing (IWQ) creates multiple input queues and allows OSA to differentiate workloads “off the wire.” It then assigns work to a specific input queue (per device) to z/OS.

Each input queue is a unique type of workload, and has unique service and processing requirements. The IWQ function allows z/OS to preassign the appropriate processing resources for each input queue. This approach allows multiple concurrent z/OS processing threads to process each unique input queue (workload), which avoids traditional resource contention.

IWQ reduces the conventional z/OS processing that is required to identify and separate unique workloads. This advantage results in improved overall system performance and scalability.

A primary objective of IWQ is to provide improved performance for business-critical interactive workloads by reducing contention that is created by other types of workloads. In a heavily mixed workload environment, this “off the wire” network traffic separation is provided by OSA-Express7S 1.2, OSA-Express7S and OSA-Express6S¹² features that are defined as CHPID type OSD. OSA IWQ is shown in Figure 7-5.

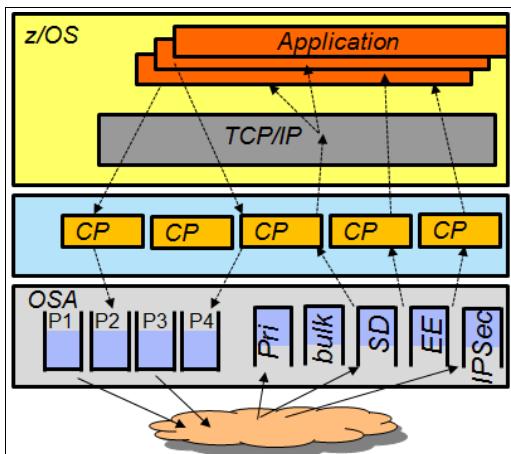


Figure 7-5 OSA inbound workload queuing

The following types of z/OS workloads are identified and assigned to unique input queues:

- ▶ z/OS Sysplex Distributor traffic

Network traffic that is associated with a distributed virtual Internet Protocol address (VIPA) is assigned to a unique input queue. This configuration allows the Sysplex Distributor traffic to be immediately distributed to the target host.

- ▶ z/OS bulk data traffic

Network traffic that is dynamically associated with a streaming (bulk data) TCP connection is assigned to a unique input queue. This configuration allows the bulk data processing to be assigned the appropriate resources and isolated from critical interactive workloads.

¹² Only OSA-Express6S and OSA-Express7S cards are supported on IBM z16 A02 and IBM z16 AGZ as carry forward.

- ▶ EE (Enterprise Extender / SNA traffic)

IWQ for the OSA-Express features is enhanced to differentiate and separate inbound Enterprise Extender traffic to a dedicated input queue.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

GARP VLAN Registration Protocol

All OSA-Express features support VLAN prioritization, which is a component of the IEEE 802.1 standard. GARP VLAN Registration Protocol (GVRP) support allows an OSA-Express port to register or unregister its VLAN IDs with a GVRP-capable switch and dynamically update its table as the VLANs change. This process simplifies the network administration and management of VLANs because manually entering VLAN IDs at the switch is no longer necessary. The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

Link aggregation support for z/VM

Link aggregation (IEEE 802.3ad) that is controlled by the z/VM Virtual Switch (VSWITCH) allows the dedication of an OSA-Express port to the z/VM operating system. The port must be participating in an aggregated group that is configured in Layer 2 mode. Link aggregation (trunking) combines multiple physical OSA-Express ports into a single logical link. This configuration increases throughput, and provides nondisruptive failover if a port becomes unavailable. The target links for aggregation must be of the same type.

Link aggregation is applicable to CHPID type OSD (QDIO). The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

Multi-VSwitch Link Aggregation

Multi-VSwitch Link Aggregation support allows a port group of OSA-Express features to span multiple virtual switches within a single z/VM system or between multiple z/VM systems. Sharing a Link Aggregation Port Group (LAG) with multiple virtual switches increases optimization and utilization of the OSA-Express features when handling larger traffic loads.

Higher adapter utilization protects customer investments, which is increasingly important as 10 GbE deployments become more prevalent. The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

Large send for IPv6 packets

Large send for IPv6 packets improves performance by offloading outbound TCP segmentation processing from the host to an OSA-Express feature by using a more efficient memory transfer into it.

Large send support for IPv6 packets applies to the OSA-Express7S 1.2, OSA-Express7S, and OSA-Express6S¹² features (CHPID type OSD) on IBM z16 A01, IBM z16 A02, IBM z16 AGZ, IBM z15, and IBM z14.

OSA-Express6S added TCP checksum on large send, which reduces the cost (CPU time) of error detection for large send.

The supported operating systems are listed in Table 7-8 on page 253 and Table 7-9 on page 255.

OSA Dynamic LAN idle

The OSA Dynamic LAN idle parameter helps reduce latency and improve performance by dynamically adjusting the inbound blocking algorithm. System administrators can authorize the TCP/IP stack to enable a dynamic setting that previously was static.

The blocking algorithm is modified based on the following application requirements:

- ▶ For latency-sensitive applications, the blocking algorithm is modified considering latency.
- ▶ For streaming (throughput-sensitive) applications, the blocking algorithm is adjusted to maximize throughput.

In all cases, the TCP/IP stack determines the best setting based on the current system and environmental conditions, such as inbound workload volume, processor utilization, and traffic patterns. It can then dynamically update the settings.

Supported OSA-Express features adapt to the changes, which avoids thrashing and frequent updates to the OAT. Based on the TCP/IP settings, OSA holds the packets before presenting them to the host. A dynamic setting is designed to avoid or minimize host interrupts.

OSA Dynamic LAN idle is supported by all OSA-Express features on IBM z16 A02 and IBM z16 AGZ when in QDIO mode (CHPID type OSD). The supported operating systems are listed in Table 7-8 on page 253.

OSA Layer 3 virtual MAC for z/OS environments

To help simplify the infrastructure and facilitate load balancing when an LPAR is sharing an OSA MAC address with another LPAR, each operating system instance can have its own unique logical or virtual MAC (VMAC) address. All IP addresses that are associated with a TCP/IP stack are accessible by using their own VMAC address instead of sharing the MAC address of an OSA port. This situation also applies to Layer 3 mode and to an OSA port spanned among channel subsystems.

OSA Layer 3 VMAC is supported by all OSA-Express features on IBM z16 A02 and IBM z16 AGZ when in QDIO mode (CHPID type OSD). The supported operating systems are listed in Table 7-8 on page 253.

Network Traffic Analyzer

IBM zSystems offer systems programmers and network administrators the ability to more easily solve network problems despite high traffic. With the OSA-Express Network Traffic Analyzer and QDIO Diagnostic Synchronization on the server, you can capture trace and trap data. This data can then be forwarded to z/OS tools for easier problem determination and resolution.

The Network Traffic Analyzer is supported by all OSA-Express features on IBM z16 A02 and IBM z16 AGZ when in QDIO mode (CHPID type OSD). The supported operating systems are listed in Table 7-8 on page 253.

7.4.6 Cryptography Features and Functions Support

IBM z16™ provides the following major groups of cryptographic functions:

- ▶ Synchronous cryptographic functions, which are provided by CPACF
- ▶ Asynchronous cryptographic functions, which are provided by the Crypto Express8S feature

The minimum software support levels are described in the following sections. Review the current PSP buckets to ensure that the latest support levels are known and included as part of the implementation plan.

CP Assist for Cryptographic Function

Central Processor Assist for Cryptographic Function (CPACF), which is standard¹³ on every IBM z16 A02 or IBM z16 AGZ core, now supports pervasive encryption. Simple policy controls allow business to enable encryption to protect data in mission-critical databases without needing to stop the database or re-create database objects. Database administrators can use z/OS Dataset Encryption, z/OS Coupling Facility Encryption, z/VM encrypted hypervisor paging, and z/TPF transparent database encryption, which use the performance enhancements in the hardware.

CPACF supports the following features in IBM z16 A02 and IBM z16 AGZ:

- ▶ Processor Activity Instrumentation to count cryptographic operations
- ▶ Advanced Encryption Standard (AES, symmetric encryption)
- ▶ Data Encryption Standard (DES, symmetric encryption)
- ▶ Secure Hash Algorithm (SHA, hashing)
- ▶ SHAKE Algorithms
- ▶ True Random Number Generation (TRNG)
- ▶ Improved GCM (Galois Counter Mode) encryption (enabled by a single hardware instruction)

In addition, the IBM z16 A02 and IBM z16 AGZ core implements a Modulo Arithmetic unit in support of Elliptic Curve Cryptography.

CPACF is used by several IBM software product offerings for z/OS, such as IBM WebSphere Application Server for z/OS. For more information, see 6.4, “CP Assist for cryptographic functions” on page 209.

The supported operating systems are listed in Table 7-10 on page 257 and Table 7-11 on page 258.

Crypto Express8S (new on IBM z16 A02 and IBM z16 AGZ)

Crypto Express8S includes a single- or dual- HSM adapter (single or dual IBM 4770 PCIe Cryptographic Co-processor [PCIeCC]) and complies with the following Physical Security Standards:

- ▶ FIPS 140-3 level 4
- ▶ Common Criteria EP11 EAL4+
- ▶ Payment Card Industry (PCI) HSM
- ▶ German Banking Industry Commission (GBIC, formerly DK)
- ▶ AusPayNet (APN)

Support of Crypto Express8S functions varies by operating system and release and by the way the card is configured as a coprocessor or an accelerator. The supported operating systems are listed in Table 7-10 on page 257 and Table 7-11 on page 258.

¹³ CPACF hardware is implemented on each IBM z15 core. CPACF functionality is enabled with FC 3863.

Crypto Express7S (carry forward on IBM z16 A02 and IBM z16 AGZ)

Introduced with IBM z15, Crypto Express7S includes a single- or dual-port adapter (single or dual IBM 4769 PCIe Cryptographic Co-processor [PCIeCC]) and complies with the following Physical Security Standards:

- ▶ FIPS 140-2 level 4
- ▶ Common Criteria EP11 EAL4
- ▶ Payment Card Industry (PCI) HSM
- ▶ German Banking Industry Commission (GBIC, formerly DK)

The supported operating systems are listed in Table 7-10 on page 257 and Table 7-11 on page 258.

Crypto Express6S (carry forward on IBM z16 A02 and IBM z16 AGZ)

Introduced with IBM z14, Crypto Express6S includes one IBM 4768 PCIe Cryptographic Co-processor (PCIeCC) and complies with the following Physical Security Standards:

- ▶ FIPS 140-2 level 4
- ▶ Common Criteria EP11 EAL4
- ▶ Payment Card Industry (PCI) HSM
- ▶ German Banking Industry Commission (GBIC, formerly DK)

The supported operating systems are listed in Table 7-10 on page 257 and Table 7-11 on page 258.

Web deliverables

For more information about web-deliverable code on z/OS, see [the z/OS downloads website](#).

For Linux on IBM Z, support is delivered through IBM and the distribution partners. For more information, see [Linux on IBM Z on the IBM developerWorks website](#).

z/OS Integrated Cryptographic Service Facility

To achieve security in a distributed computing environment, a combination of elements must work together. A security policy should be based on an appraisal of the value of data and the potential threats to that data. This provides the foundation for a secure environment.

IBM has categorized these security functions according to International Organization for Standardization (ISO) standard 7498-2:

- ▶ Identification and authentication - includes the ability to identify users to the system and provide proof that they are who they claim to be.
- ▶ Access control - determines which users can access which resources.
- ▶ Data confidentiality - protects an organization's sensitive data from being disclosed to unauthorized persons.
- ▶ Data integrity - ensures that data is in its original form and that nothing has altered it.
- ▶ Security management - administers, controls, and reviews a business security policy.
- ▶ Nonrepudiation - assures that the appropriate individual sent the message.

Only cryptographic services can provide the data confidentiality and the identity authentication that is required to protect business commerce on the Internet¹⁴.

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, to balance the workload and help address the bandwidth requirements of the applications.

¹⁴ Quoted from z/OS V2.R5 publications.

ICSF support for IBM z16 A02 and IBM z16 AGZ is provided with PTFs, not as previously was the case, through Web deliverables.

Supported levels of ICSF automatically detect what HW cryptographic capabilities are available where it is running, then enables functions accordingly. No toleration of new HW is necessary, it is “just there”. ICSF maintenance is necessary if you want to exploit new capabilities.

Exploitation of new function is supplied in ICSF PTFs on:

z/OS V2.R2-V2.R4 (Web deliverable HCR77D1) or V2.R5 (base, which is HCR77D2)

New function exploitation includes:

- ▶ CCA and EP11 CEX8 Coprocessor support
- ▶ CCA and EP11 Quantum Safe Algorithms (Kyber & Dilithium 8,7)
- ▶ EP11 mechanism for data re-encryption and new ECC curve support
- ▶ Fully homomorphic encryption
- ▶ Usage counters that count classes of crypto operations (to meet audit requirements)

When exploiting new Quantum Safe Algorithms and sharing a KDS in a sysplex, ensure all ICSF PTFs are installed on all systems.

For more information about ICSF versions and FMID cross-references, see the [z/OS: ICSF Version and FMID Cross Reference](#), TD103782, abstract that is available at the IBM Techdoc website.

RMF Support for Crypto Express

RMF enhances the Monitor I Crypto Activity data gatherer to recognize and use performance data for the new Crypto Express8S (CEX8), Crypto Express7S (CEX7) and CryptoExpress6S (CEX6) cards. RMF supports all valid card configurations on IBM z16 A02 and IBM z16 AGZ and provides CEX7 and CEX6 crypto activity data in the SMF type 70 subtype 2 records and RMF Postprocessor Crypto Activity Report.

Reporting can be done at an LPAR/domain level to provide more granular reports for capacity planning and diagnosing problems. This feature requires fix for APAR OA54952.

The supported operating systems are listed in Table 7-10 on page 257.

z/OS Data Set Encryption

Aligned with IBM Z Pervasive Encryption initiative, IBM provides application-transparent, policy-controlled dataset encryption in IBM z/OS.

Policy-driven z/OS Data Set Encryption enables users to perform the following tasks:

- ▶ De-couple encryption from data classification; encrypt data automatically independent of labor-intensive data classification work.
- ▶ Encrypt data immediately and efficiently at the time it is written.
- ▶ Reduce risks that are associated with mis-classified or undiscovered sensitive data.
- ▶ Help protect digital assets automatically.
- ▶ Achieve application transparent encryption.

IBM Db2® for z/OS and IBM Information Management System (IMS) intend to use z/OS Data Set Encryption.

With z/OS, Data Set Encryption DFSMS enhances data security with support for data set level encryption by using DFSMS access methods. This function is designed to give users the ability to encrypt their data sets without changing their application programs.

DFSMS users can identify which data sets require encryption by using JCL, Data Class, or the RACF data set profile. Data set level encryption can allow the data to remain encrypted during functions, such as backup and restore, migration and recall, and replication.

z/OS Data Set Encryption requires CP Assist for Cryptographic Functions (CPACF).

Considering the significant enhancements that were introduced with z14, the encryption mode of XTS is used by access method encryption to obtain the best performance possible. It is not recommended to enable z/OS data set encryption until all sharing systems, fallback, backup, and DR systems support encryption.

In addition to applying PTFs enabling the support, ICSF configuration is required. The supported operating systems are listed in Table 7-10 on page 257.

Quantum-safe encryption with Crypto Express8S

Quantum-safe cryptography strengthens the portfolio of pervasive encryption capabilities on IBM z16 A02 and IBM z16 AGZ, allowing clients not only to encrypt the data with a quantum-safe cryptographic algorithm (AES with 256-bit keys) as is the case with prior IBM Z systems, but with the use of quantum-safe algorithms for internal system protection of encryption keys. The enhancements introduced with IBM z16 A02 and IBM z16 AGZ to accomplish this are:

- ▶ Quantum-safe key generation
- ▶ Quantum-safe hybrid key exchange schemes
- ▶ Quantum-safe dual digital signature schemes

Crypto Analytics Tool for IBM Z

The IBM Crypto Analytics Tool (CAT) for IBM Z is an analytics solution that collects data on your z/OS cryptographic infrastructure, presents reports, and analyzes if any vulnerabilities exist. CAT collects cryptographic information from across the enterprise and provides reports to help users better manage the crypto infrastructure and ensure it follows best practices. The use of CAT can help you deal with managing complex cryptography resources across your organization.

z/VM encrypted hypervisor paging (encrypted paging support)

With the PTF for APAR VM65993, z/VM V6.4 provides support for encrypted paging in support of the IBM z16 A02 and IBM z16 AGZ pervasive encryption philosophy of encrypting all data in flight and at rest. Ciphering occurs as data moves between active memory and a paging volume that is owned by z/VM.

Included in this support is the ability to dynamically control whether a running z/VM system is encrypting this data. This support protects guest paging data from administrators or users with access to volumes. Enabled with AES encryption, z/VM Encrypted Paging includes low overhead by using CPACF.

The supported operating systems are listed in Table 7-10 on page 257.

z/TPF transparent database encryption

Shipped in August 2016, z/TPF at-rest Data Encryption provides following features and benefits:

- ▶ Automatic encryption of at-rest data by using AES CBC (128 or 256).

- ▶ No application changes required.
- ▶ Database level encryption by using highly efficient CPACF.
- ▶ Inclusion of data on disk and cached in memory.
- ▶ Ability to include data integrity checking (optionally by using SHA-256) to detect accidental or malicious data corruption.
- ▶ Tools to migrate a database from unencrypted to encrypted state or change the encryption key/algorithm for a specific DB while transactions are flowing (no database downtime).

Pervasive encryption for Linux on IBM Z

Pervasive encryption for Linux on IBM Z combines the full power of Linux with IBM z16 A02 and IBM z16 AGZ capabilities by using the support of the following features:

- ▶ Kernel Crypto: IBM z16 A02 and IBM z16 AGZ CPACF
- ▶ LUKS dm-crypt Protected-Key CPACF
- ▶ Libica and openssl: IBM z16 A02 and IBM z16 AGZ CPACF and acceleration of RSA handshakes by using SIMD
- ▶ Secure Service Container: High security virtual appliance deployment infrastructure

Protection of data at-rest

By using the integration of industry-unique hardware accelerated CPACF encryption into the standard Linux components, users can achieve optimized encryption transparently to prevent raw key material from being visible to operating systems and applications.

Because of the potential costs and overheads, most of the organizations avoid the use of host-based network encryption today. By using enhanced CPACF and SIMD on IBM z16 A02 and IBM z16 AGZ, TLS and IPSec can use hardware performance gains while benefitting from transparent enablement. Reduced cost of encryption enables broad use of network encryption.

IBM Z Security and Compliance Center

The IBM Z Security and Compliance Center is a modern, browser-based application providing compliance capability mapping, fact collection, and validation. Designed for use with minimal technical skills, this solution can automate evidence collection of compliant-relevant facts from IBM Z platforms including the new CPACF usage counters which demonstrate crypto algorithm strength and key protection. The IBM Z Security and Compliance Center enables clients to:

- ▶ Generate detailed reports to enable executives, administrators, and auditors to understand compliance metrics with ease
- ▶ Track compliance drift over time with dashboard visualizations that include historical compliance information
- ▶ Utilize compliance evidence generation facilities from IBM Z software stack (for example, z/OS, z/OS Middleware, z/OS Compliance Integration Manager, Oracle on Linux on IBM Z, and PostgreSQL on Linux on IBM Z)
- ▶ Provide an interactive view of the compliance posture and details around the severity of control deviations from regulations, such as PCI-DSS v3.2.1, NIST SP800-53, and CIS Benchmarks

z/OS support for compliance

z/OS has been enhanced to enable the collection of compliance data from IBM z16 A02 and IBM z16 AGZ CPACF counters and several z/OS products and components.

A new z/OSMF Compliance fact collection REST API sends an ENF86 signal to all systems. Participating products and components will collect and write compliance data to new SMF

1154 records associated with its unique subtype. These new SMF 1154 records may be integrated into solutions such as the IBM z16 A02 and IBM z16 AGZ IBM Z Security and Compliance Center.

This support requires PTFs for z/OS 2.4 and z/OS 2.5. The PTFs will be identified by a fix category designated specifically for Compliance data collection support named

IBM.Function.Compliance.DataCollection. See “IBM Fix Category Values and Descriptions” for information on how to use this fix category to identify and install the specific PTFs that enable compliance data collection.

For additional information about z/OS collection sources and enablement refer to:

- Software Announcement 222-005, IBM Z Security and Compliance Center.
- Software Announcement 222-092, CICS Transaction Server for z/OS 6.1.
- Software Announcement 222-003, Db2 13 for z/OS powered by AI innovations provides industry scalability, business resiliency and intelligence.

Linux support for compliance

Linux on IBM Z supports the collection of compliance data from the Linux environment.

Pre-requisite operating systems:

- Red Hat Enterprise Linux 8.0 (RHEL) on IBM Z, or later
- SUSE Linux Enterprise Server (SLES) on IBM Z 15
- Ubuntu Server LTS for IBM Z 22.04

Optional middleware and software:

- Oracle 19c
- PostgreSQL 13.x, 14.x

7.5 z/OS migration considerations

Except for base processor support, z/OS releases do not require any of the functions that are introduced with the IBM z16 A02 and IBM z16 AGZ. Minimal toleration support that is needed depends on z/OS release.

Although IBM z16 A02 and IBM z16 AGZ servers *do not* require any “functional” software, it is recommended to install all IBM z16 A02 and IBM z16 AGZ service before upgrading to the new server. The support matrix for z/OS releases and the IBM Z servers that support them are listed in Table 7-16. X means hardware is supported.

Table 7-16 z/OS support summary

z/OS Release	IBM zEC12 IBM zBC12 WDFM ^a	IBM z13 IBM z13s WDFM ^a	IBM z14 ^a	IBM z15	IBM z16 A02 and IBM z16 AGZ	End of Service	Extended Defect Support ^b
V2R2	X	X	X	X	X	09/2020 ^b	09/2023 ^b
V2R3	X	X	X	X	X	09/2022	09/2025 ^b
V2R4	-	X	X	X	X	09/2024	09/2027 ^b

z/OS Release	IBM zEC12 IBM zBC12 WDFM^a	IBM z13 IBM z13s WDFM^a	IBM z14^a	IBM z15	IBM z16 A02 and IBM z16 AGZ	End of Service	Extended Defect Support^b
V2R5	-	-	X	X	X	09/2026	09/2029 ^b

a. Server is withdrawn from marketing.

b. The IBM Software Support Services provides the ability for customers to purchase extended defect support service for z/OS.

7.5.1 General guidelines

The IBM z16™ introduces the latest IBM zSystems technology. Although support is provided by z/OS starting with z/OS V2.R2, the capabilities and use of IBM z16 A02 or IBM z16 AGZ depends on the z/OS release. Optional web deliverables¹⁵ are needed for some functions on some releases.

New: ICSF support for IBM z16 A02 and IBM z16 AGZ is provided with PTFs, not Web deliverables

In general, consider the following guidelines:

- ▶ Do not change software releases and hardware at the same time.
- ▶ At minimum apply maintenance from the following FIXCAT to all systems that will participate in a sysplex with IBM z16 A02 and IBM z16 AGZ regardless of whether the systems will be migrated to the current hardware:
 - IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ -3931.RequiredService
 - IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ A02-3932.RequiredService
- ▶ Keep members of the sysplex at the same software level, except during brief migration periods.
- ▶ Upgrade Coupling Facility LPARs to current levels (prior to the CF upgrade you should review all structure sizes using the CFSIZER tool).
- ▶ Review any restrictions and migration considerations before creating an upgrade plan.
- ▶ Acknowledge that some hardware features cannot be ordered or carried forward for an upgrade from an earlier server to IBM z16 A02 and IBM z16 AGZ and plan accordingly.
- ▶ Determine the changes in IOCP, HCD, and HCM to support defining IBM z16 A02 and IBM z16 AGZ configuration and the new features and functions it introduces.
- ▶ Ensure that none of the new z/Architecture Machine Instructions (mnemonics) that were introduced with IBM z16 A02 and IBM z16 AGZ are colliding with the names of Assembler macro instructions you use¹⁶.
- ▶ Check the use of **MACHMIG** statements in **LOADxx PARMLIB** commands.
- ▶ Contact software vendors to inform them of new machine model and request new license keys if applicable.

¹⁵ For example, the use of Crypto Express7S requires the Cryptographic Support for z/OS V2R2 - z/OS V2R3 web deliverable.

¹⁶ For more information, see the [Tool to Compare IBM z16 A02 and IBM z16 AGZ Instruction Mnemonics with Macro Libraries](#) IBM Technote.

- ▶ Review the z/OS Upgrade Workflow for z/OSMF provided with APAR OA62703 for z/OS V2R2 and higher. This Workflow is also available in the IBM Documentation library.

7.5.2 Hardware Fix Categories (FIXCATs)

Important: z16 M/T 3932 users must use the FIXCATs for both the 3932 and the 3931. The Required FIXCAT is absolutely critical, and cannot be understated. Only unique capabilities for the 3932 will be identified with the 3932 FIXCATs.

Base support includes fixes that are required to run z/OS on the IBM z16™ server. They are identified by:

IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ -3931.RequiredService
IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ A02-3932.RequiredService

The use of many functions covers fixes that are required to use the capabilities of the IBM z16™ servers. They are identified by:

IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ -3931.Exploitation
IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ A02-3932.Exploitation

Recommended service is identified by:

IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ -3931.Exploitation
IBM.Device.Server.IBM z16 A02 and IBM z16 AGZ A02-3932.Exploitation

General z/OS support documentation provided in the PSP Bucket:

- For IBM z16 A02 and IBM z16 AGZ: PSP Bucket:
Upgrade = 3932DEVICE, Subset = 3932/ZOS

Attention: PSP Bucket should not be used to find PTF support for IBM z16 A02 and IBM z16 AGZ. Use SMP/E FIXCATs.

Consider the following other Fix Categories of Interest:

- ▶ Fixes that are required to use the Server Time Protocol function:
IBM.Function.ServerTimeProtocol
- ▶ Fixes that are required to use the High-Performance FICON function:
IBM.Function.zHighPerformanceFICON
- ▶ Fixes that are required for IBM Z System Recovery Boost:
IBM.Function.SystemRecoveryBoost
- ▶ PTFs that allow previous levels of ICSF to coexist with the latest Cryptographic Support for z/OS V2R2 - z/OS V2R4 (HCR77D1) web deliverable:
IBM.Coexistence.ICSF.z/OS_V2R2-V2R4-HCR77D1

Use the SMP/E REPORT MISSINGFIX command to determine whether any FIXCAT APARs exist that are applicable and are not yet installed, and whether any SYSMODs are available to satisfy the missing FIXCAT APARs.

For more information about IBM Fix Category Values and Descriptions, see the [IBM Fix Category Values and Descriptions page](#) of the IBM IT infrastructure website.

7.5.3 z/OS V2.R5

IBM z/OS, Version 2 Release 5 has been announced July 27, 2021. One of the highlights of this release is the support of 16TB real memory per z/OS image. This allows new workloads which require more storage than is currently available.

Further details of this release are available in the [announcement letter](#).

7.5.4 z/OS V2.R4

IBM z/OS, Version 2 Release 4, was made generally available on September 30, 2019. This release delivers innovation through an agile, optimized, and resilient platform that helps companies build applications and services based on a highly scalable and secure infrastructure that provides the performance and availability for on-premise or provisioned as-a-service workloads.

z/OS V2.R4 delivers the following capabilities (list is not exclusive):

- ▶ IBM z/OS Container Extensions (zCX), which enables the ability to run almost any Linux on IBM Z Docker container in z/OS alongside existing z/OS applications and data without a separate provisioned Linux server
- ▶ Easier integration of z/OS into private and multi-cloud environments with improvements that deliver a more robust, easy to use, and highly available implementation using IBM Cloud™ Provisioning and Management for z/OS, IBM z/OS Cloud Broker and IBM Cloud Storage Access for z/OS Data,
- ▶ Enhancements that continue to simplify and modernize the z/OS environment for a better user experience and improved productivity by reducing the level of IBM Z specific skills that are required to maintain z/OS,
- ▶ Ongoing industry-wide simplification improvements to help companies install and configure software using a common and modern method. These installation improvements range from the packaging of software through the configuration so that faster time to value can be realized throughout the enterprise,
- ▶ IBM Open Data Analytics for z/OS provides enhancements to simplify data analysis by combining open source runtimes and libraries with analysis of z/OS data at its source,
- ▶ Enhancements to security and data protection on the system with support for new industry cryptography and continued enhancements driving pervasive encryption through the ability to encrypt data without application changes. A new RACF capability improves management of access and privileges
- ▶ Leveraging IBM z16 A02 and IBM z16 AGZ capabilities - System Recovery Boost which reduces the time that z/OS is offline when the operating system is offline for any reason. The use of IBM System Recovery Boost expedites planned operating system shutdown processing, operating system IPL (Initial Program Load), middleware/workload restart and recovery, and the client workload execution that follows. It will let businesses return their systems to work faster, not just from catastrophes, but after all kinds of disruptions, both planned and unplanned. Another aspect of System Recovery Boost is to expedite and streamline the execution of GDPS recovery scripts which perform reconfiguration actions during various planned and unplanned operational scenarios.
- ▶ Dynamic activation of I/O configurations for stand-alone Coupling Facilities.

Stand-alone CFs (Coupling Facility images that reside on a server without a co-resident z/OS image), are now able to participate in dynamic I/O configuration changes that affect the stand-alone CF and no longer require the server to be restarted to activate such changes

7.5.5 z/OS V2.R3

IBM announced z/OS Version 2 Release 3 - Engine for digital transformation through Announcement letter 217-246 on July 17, 2017. Focusing on three critical areas (Security, Simplification, and Cloud), z/OS V2.R3 provides a simple and transparent approach to enable extensive encryption of data and to simplify the overall management of the z/OS system to increase productivity. Focus is also given to providing a simple approach for self-service provisioning and rapid delivery of software as a service, while enabling for the API economy.

Consider the following points before migrating z/OS V2.R3 to IBM z16™:

- ▶ IBM z/OS V2.R3 with IBM z16 A02 and IBM z16 AGZ requires a minimum of 8 GB of memory. When running as a z/VM guest or on an IBM System z® Personal Development Tool, a minimum of 2 GB is required for z/OS V2.R3. If the minimum is not met, a warning WTO is issued at IPL.
Continuing with less than the minimum memory might affect availability. A migration health check will warn if the system is configured with less than 8 GB.
- ▶ Dynamic splitting and merging of Coordinated Timing Network (CTN) is available with IBM z16 A02 and IBM z16 AGZ.
- ▶ RMF support is provided to collect SMC-D related performance measurements in SMF 73 Channel Path Activity and SMF 74 subtype 9 PCIE Activity records. It also provides these measurements in the RMF Postprocessor and Monitor III PCIE and Channel Activity reports.

HyperSwap support is enhanced to allow RESERVE processing. When a system runs a request to swap to secondary devices that are managed by HyperSwap, z/OS detects when RESERVEs are held and ensures that the devices that are swapped also hold the RESERVE. This enhancement is provided with collaboration from z/OS, GDPS HyperSwap, and CSM HyperSwap.

7.5.6 Coupling links¹⁷

IBM z16 A02 and IBM z16 AGZ servers support only active participation in the same Parallel Sysplex with IBM z15 and IBM z14. Configurations with z/OS on one of these servers can add an IBM z16 A02 and IBM z16 AGZ to their Sysplex for a z/OS or a Coupling Facility image.

Configurations with a Coupling Facility on one of these servers can add an IBM z16 A02 and IBM z16 AGZ to their Sysplex for a z/OS or a Coupling Facility image. IBM z16 A02 and IBM z16 AGZ does not support participating in a Parallel Sysplex with System IBM z13/IBM z13s and earlier systems.

Each system can use, or not use, internal coupling links, CE LR links, or ICA SR coupling links independently of what other systems are using.

Coupling connectivity is available only when other systems also support the same type of coupling. For more information about supported coupling link technologies on IBM z16 A02

¹⁷ IBM z16 A02 and IBM z16 AGZ does not support InfiniBand coupling links. More planning might be required to integrate the IBM z16 A02 and IBM z16 AGZ in a Parallel Sysplex with IBM z14 servers.

and IBM z16 AGZ, see 4.6.4, “Parallel Sysplex connectivity” on page 175, and the *Coupling Facility Configuration Options* white paper.

7.5.7 z/OS XL C/C++ considerations

IBM z/OS V2.R4 XL C/C++ is an optional feature of z/OS that will continue to ship with IBM z16 A02 and IBM z16 AGZ.

The following new functions provide performance improvements for applications by using new IBM z16 A02 and IBM z16 AGZ instructions:

- ▶ High performance math libraries
- ▶ MASS
- ▶ Replace Atlas with OpenBLAS
- ▶ Metal C for modernizing HLLASM applications and systems programming

To enable the use of new functions, specify **ARCH(14)** and **VECTOR** for compilation. The binaries that are produced by the compiler on IBM z16 A02 or IBM z16 AGZ can be run only on IBM z16 A02 and IBM z16 AGZ, because it uses the vector facility available on them for new functions. The use of older versions of the compiler on IBM z16 A02 or IBM z16 AGZ does not enable new functions.

z/OS V2R4 is able to use the latest level (14) of the following C/C++ compiler options:

- ▶ **ARCHITECTURE**: This option selects the minimum level of system architecture on which the program can run. Certain features that are provided by the compiler require a minimum architecture level. ARCH(14) uses instructions that are available on the IBM z16 A02 and IBM z16 AGZ.
- ▶ **TUNE**: This option allows optimization of the application for a specific system 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.

Important: Use the previous IBM Z **ARCHITECTURE** or **TUNE** options for C/C++ programs if the same applications run on the previous IBM Z servers. However, if C/C++ applications run on IBM z16 A02 or IBM z16 AGZ servers only, use the latest **ARCHITECTURE** and **TUNE** options to ensure that the best performance possible is delivered through the latest instruction set additions.

For more information about the **ARCHITECTURE**, **TUNE**, and **VECTOR** compiler options, see [z/OS XL C/C++ User's Guide, SC14-7307-40](#).

z/OS XL C/C++ Web deliverables are available at no charge to z/OS XL C/C++ customers

- ▶ Based on Open source LLVM infrastructure – supports up to date C++ language standards
- ▶ 64-bit, USS only

Statement of Direction: IBM will continue to adopt the LLVM and Clang compiler infrastructure in future C/C++ offerings on IBM Z^a

- a. Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

7.6 IBM z/VM migration considerations

IBM z16 A02 and IBM z16 AGZ M/T 3932 supports z/VM 7.3, and z/VM 7.2. z/VM is moving to continuous delivery model. For more information, see [this web page](#).

7.6.1 IBM z/VM 7.3

z/VM 7.3 has been available on September 16, 2022. It features the following:

- ▶ 8-Member SSI: increases the maximum size of a Single System Image (SSI) cluster from four members to eight, enabling clients to grow their SSI clusters to allow for more workload, and providing more flexibility to use live guest relocation (LGR) for non-disruptive upgrades and workload balancing.
- ▶ New Architecture Level Set of IBM z14, IBM z14 ZR1.
- ▶ It also features all new functions made available in z/VM 7.2 throughout the continuous delivery process.

7.6.2 IBM z/VM 7.2

z/VM 7.2 has been available since September, 2020. It features the following (excerpt):

- ▶ Centralized Service Management for non-SSI environments to deploy service to multiple systems, regardless of geographic location, from a centralized primary location.
- ▶ Multiple Subchannel Set (MSS) Multi-Target Peer-To-Peer Remote Copy (MT-PPRC) z/VM support for the GDPS environment, allowing a device to be the primary to up to three secondary devices, each defined in a separate alternate subchannel set (supporting up to 3 alternate subchannel sets). Also provides the CP updates necessary for VM/HCD support of alternate subchannel sets.
- ▶ New Architecture Level Set of IBM z13, IBM z13s (LinuxONE Emperor / Rockhopper), or newer processor families
- ▶ z/VM 7.2 includes New Function APARs shipped for z/VM 7.1, such as:
VSwitch Priority Queuing, EAV Paging, 80 Logical Processors, Dynamic Crypto, System Recovery Boost support (subcapacity CPs speed boost only), and so on.

7.6.3 Capacity

For the capacity of any z/VM logical partition (LPAR) and any z/VM guest, in terms of the number of Integrated Facility for Linux (IFL) processors and central processors (CPs), real or virtual, you might want to adjust the number to accommodate the PU capacity of IBM z16 A02 and IBM z16 AGZ configurations.

7.7 z/VSE migration considerations

As described in 7.2.3, “z/VSE” on page 244, IBM z16 A02 and IBM z16 AGZ support z/VSE 6.2.

Consider the following general guidelines when you are migrating z/VSE environment to IBM z16 A02 and IBM z16 AGZ servers:

- ▶ Collect reference information before migration

This information includes baseline data that reflects the status of, for example, performance data, CPU utilization of reference workload, I/O activity, and elapsed times.

This information is required to size IBM z16 A02 and IBM z16 AGZ and is the only way to compare workload characteristics after migration.

For more information, see the *z/VSE Release and Hardware Upgrade* document.

- ▶ Apply required maintenance for IBM z16 A02 and IBM z16 AGZ

Review the Preventive Service Planning (PSP) bucket 3932DEVICE for IBM z16 A02 and IBM z16 AGZ M/T 3932 and apply the required PTFs for IBM and independent software vendor (ISV) products.

Note: IBM z16™ supports z/Architecture mode only.

7.8 Software licensing

The IBM z16™ software portfolio includes operating system software (that is, z/OS, z/VM, z/VSE, and z/TPF) and middleware that runs on these operating systems. The portfolio also includes middleware for Linux on IBM Z environments.

This section provides an overview of IBM Z® software licensing options that are available for IBM z16 A02 and IBM z16 AGZ software, including MLC, zIPLA, sub-capacity, sysplex, and Taylor Fit Pricing.

- ▶ Monthly license charge (MLC)

MLC is a recurring charge that is applied monthly. It includes the right to use the product and provides access to IBM product support during the support period. Select an MLC pricing metric based on your goals and environment. The selected metric will be used to price MLC products, such as z/OS®, z/TPF, z/VSE®, middleware, compilers and selected systems management tools and utilities.

- Key MLC Metrics and Offerings

MLC metrics include various offerings. The metrics and pricing schemes available on IBM z14, IBM z15, IBM z16 A01 and IBM z16 A02 and IBM z16 AGZ are shown in Table 7-17:

Table 7-17 MLC metrics offerings

Key MLC Metric	Sub-Capacity	Sysplex Aggregation	Contract No.
Advanced Workload License Charges (AWLC)	Y	Y	Z125-8538
Country Multiplex License Charges (CMLC) ^a	Y		Z126-6965
Flat Workload License Charges (FWLC) ^b			
System z New Application License Charges (zNALC)	Y	Y	Z125-7454
Parallel Sysplex License Charges (PSLC)		Y	Z125-5205
Midrange Workload License Charges (MWLC)	Y		Z125-7452

a. The Country Multiplex offering was withdrawn as of January 1, 2021. For existing CMP clients, machines currently eligible to be included in an existing multiplex cannot be older than two generations prior to the most recently available server.

b. Metric available only in conjunction with AWLC or CMLC.

- ▶ zIPLA licensing

International Program License Agreement (IPLA) programs have a one-time-charge (OTC) and an optional annual maintenance charge, called Subscription & Support. This annual charge includes access to IBM technical support and enables you to obtain version upgrades at no charge for products that generally fall under the zIPLA such as appl. development tools, CICS tools, data management tools, WebSphere® for IBM Z products, Linux® on IBM Z middleware and z/VM®.

There are three pricing metrics that apply to IBM Z IPLA products:

- Value Unit pricing applies to most IPLA products that run on z/OS. Value Unit pricing is typically based upon a number of MSUs and allows for lower cost of incremental growth.
- z/VM V5, V6, V7 and certain z/VM middleware have pricing based on the number of engines. Engine based Value Unit pricing allows for a lower cost of incremental growth with additional engine-based licenses purchased.
- Most Linux middleware is also priced based on the number of engines. The number of engines is converted into Processor Value Units under the Passport Advantage terms and conditions.

For more information see the zIPLA section under the licensing tab at:

<https://www.ibm.com/it-infrastructure/z/pricing-licensing>

- ▶ Sub-capacity licensing

Sub-capacity licensing includes software charges for certain IBM products based on the utilization capacity of the logical partitions (LPARs) on which the product runs.

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.

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 PR/SM, which limits 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 limit the maximum instantaneous partition use.

You can also use an LPAR group capacity limit, which sets soft capping by PR/SM for a group of logical partitions that are running z/OS. Only the 4-hour rolling average utilization of the LPAR group is tracked, which allows utilization peaks above the group capacity value.

- ▶ Sysplex licensing

Sysplex licensing allows monthly software licenses to be aggregated across a qualified Parallel Sysplex®. To be eligible for Sysplex pricing aggregation the customer environment must meet hardware, software, operation and verification criteria to be considered “actively coupled”. For more information about Sysplex licensing check the licensing tab on the IBM Software Pricing website:

<https://www.ibm.com/it-infrastructure/z/pricing-licensing>

- ▶ Taylor Fit Software Consumption

Taylor Fit Software Consumption Solution is a cloud-like usage-based licensing model. Usage is measured on the basis of MSUs consumed, which removes the need for manual or automated capping and allows customers to configure their systems to support optimal response times and service level agreements.

Tailored Fit Pricing requires z/OS V2.R2, or later, with the applicable PTFs applied.

The requirements for Tailored Fit Pricing (TFP) vary with the solution. The specific requirements for a solution must be met before IBM can accept and process sub-capacity reports in which Tailored Fit Pricing solutions are reported. Further information about TFP can be found on the IBM Infrastructure website:

<https://www.ibm.com/it-infrastructure/z/pricing-tailored-fit>

Technology Transition Offerings with IBM z16 A02 and IBM z16 AGZ

Complementing the announcement of the IBM z16 A02 and IBM z16 AGZ, IBM introduced the following Technology Transition Offerings (TTOs):

- ▶ Technology Update Pricing for the IBM z16™.
- ▶ New and revised Transition Charges for Sysplexes or Multiplexes TTOs for actively coupled Parallel Sysplexes (z/OS), Loosely Coupled Complexes (z/TPF), and Multiplexes (z/OS and z/TPF).

Technology Update Pricing for the IBM z16™ extends the software price and performance that is provided by AWLC for IBM z16 A02 and IBM z16 AGZ servers. The new and revised Transition Charges for Sysplexes offerings provide a transition to Technology Update Pricing for the IBM z16™ for customers who have not fully migrated to IBM z16 A02 or IBM z16 AGZ. This transition ensures that aggregation benefits are maintained and also phases in the benefits of Technology Update Pricing for the IBM z16™ pricing as customers migrate.

When an IBM z16 A02 or IBM z16 AGZ server is in an actively coupled Parallel Sysplex or a Loosely Coupled Complex, you might choose aggregated Advanced Workload License Charges (AWLC) pricing or aggregated Parallel Sysplex License Charges (PSLC) pricing (subject to all applicable terms and conditions).

When an IBM z16 A02 or IBM z16 AGZ is part of a Multiplex under Country Multiplex Pricing (CMP) terms, Country Multiplex License Charges (CMLC), Multiplex zNALC (MzNALC), and Flat Workload License Charges (FWLC) are the only pricing metrics available (subject to all applicable terms and conditions).

When an IBM z16 A02 or IBM z16 AGZ is running z/VSE, you can choose Mid-Range Workload License Charges (MWLC), which are subject to all applicable terms and conditions.

For more information about AWLC, CMLC, MzNALC, PSLC, MWLC, or the Technology Update Pricing and Transition Charges for Sysplexes or Multiplexes TTO offerings, see the [IBM Z Software Pricing page](#) of the IBM IT infrastructure website.

7.9 References

For more information about planning, see the home pages for the following operating systems:

- ▶ [z/OS](#)
- ▶ [z/VM](#)
- ▶ [z/VSE](#)
- ▶ [z/TPF](#)
- ▶ [Linux on IBM Z](#)
- ▶ [KVM for IBM Z](#)



System upgrades

This chapter provides an overview of the IBM zSystems upgrade process and how, in many cases, customers can manage capacity upgrades by using online tools and automation. The chapter also includes a detailed description of capacity on demand (CoD) offerings available on the [IBM z16 A02](#) and [IBM z16 AGZ](#).

[IBM z16 A02](#) and [IBM z16 AGZ](#) support dynamic provisioning features to give clients exceptional flexibility and control over system capacity and costs.

A key resource for managing client IBM zSystems is the [IBM Resource Link website](#). Once registered, a client can view product information by clicking **Resource Link → Client Initiated Upgrade Information**, and selecting **Education**. Select your particular product from the list of available systems.

This chapter includes the following topics:

- ▶ 8.1, “Permanent and Temporary Upgrades” on page 318
- ▶ 8.2, “PU upgrades” on page 323
- ▶ 8.3, “Miscellaneous equipment specification (MES) upgrades” on page 330
- ▶ 8.4, “Permanent upgrade by using the CIU facility” on page 334
- ▶ 8.5, “On/Off Capacity on Demand” on page 338
- ▶ 8.6, “z/OS Capacity Provisioning” on page 344
- ▶ 8.7, “Capacity for Planned Event” on page 349
- ▶ 8.8, “Capacity Backup” on page 349
- ▶ 8.9, “Flexible Capacity Cyber Resiliency” on page 353
- ▶ 8.10, “Planning for nondisruptive upgrades” on page 355
- ▶ 8.11, “Summary of Capacity on-Demand offerings” on page 361

8.1 Permanent and Temporary Upgrades

The terminology for CoD and the types of upgrades for an [IBM z16 A02](#) and [IBM z16 AGZ](#) are described in this section.

8.1.1 Overview

Upgrades can be categorized as described in this section.

Permanent versus temporary upgrades

Deciding whether to perform a Permanent or a Temporary upgrade depends on the situation. For example, a growing workload might require more memory, I/O cards, or processor capacity. However, to handle a peak workload, or to temporarily replace a system that is down during a disaster or data center maintenance, might require only a temporary upgrade. [IBM z16 A02](#) and [IBM z16 AGZ](#) offer the following solutions:

- ▶ Permanent upgrades
 - Miscellaneous equipment specification (MES)
An MES upgrade might involve the addition of physical hardware or the installation of Licensed Internal Code Configuration Control (LICCC). In both cases, the hardware installation is performed by IBM personnel.
 - Customer Initiated Upgrade (CIU)
The use of the CIU facility for a system requires that the online CoD buying feature (FC 9900) is installed on the system and the relevant CIU contract agreements are in place. The CIU facility supports only LICCC upgrades.

For more information, see 8.1.4, “Permanent upgrades” on page 321.

Tip: An MES provides system upgrades that can result in more enabled processors, a different central processor (CP) capacity level, more CPC drawers, memory, PCIe+ I/O drawers, and I/O features (physical upgrade). Extra planning tasks are required for nondisruptive logical upgrades. An MES is ordered through your IBM representative and installed by IBM service support representatives (IBM SSRs).

- ▶ Temporary upgrades

All temporary upgrades are LICCC-based. The one billable capacity offering is On/Off Capacity on Demand (On/Off CoD), which can be used for short-term capacity requirements and are pre-paid or post-paid.

The two replacement capacity offerings available are Capacity Backup (CBU) and Capacity for Planned Event (CPE).

Note: CPE is only available when carried forward to [IBM z16 A02](#) and [IBM z16 AGZ](#).

Flexible Capacity for Cyber Resiliency is new type of temporary record that is introduced with IBM z16 A01. This record holds the Flexible Capacity Entitlements for IBM z16 A01, IBM z16 A02 and IBM z16 AGZ machines across two or more sites.

8.1.2 CoD for IBM z16 A02 and IBM z16 AGZ systems-related terminology

The most frequently used terms that are related to CoD for IBM z16 A02 and IBM z16 AGZ are listed in Table 8-1.

Table 8-1 CoD terminology

Term	Description
Activated capacity	Capacity that is purchased and activated. Purchased capacity can be greater than the activated capacity.
Billable capacity	Capacity that helps handle workload peaks (expected or unexpected). The one billable offering that is available is On/Off Capacity on Demand (OOCoD).
Capacity	Hardware resources (processor and memory) that can process the workload can be added to the system through various capacity offerings.
Capacity Backup (CBU)	Capacity Backup allows you to place model capacity or specialty engines in a backup system. CBU is used in an unforeseen loss of system capacity because of an emergency or for Disaster Recovery testing.
Capacity for Planned Event (CPE)	Used when temporary replacement capacity is needed for a short-term event. CPE activates processor capacity temporarily to facilitate moving systems between data centers, upgrades, and other routine management tasks. CPE is an offering of CoD. CPE is only available as carry forward on IBM z16 A02 and IBM z16 AGZ.
Capacity levels	Can be full capacity or subcapacity. For an IBM IBM z16 A02 and IBM z16 AGZ system, capacity levels for the CP engine are A-Z (26 subcapacity levels): <ul style="list-style-type: none"> ▶ A full capacity CP engine is indicated by Z. ▶ A subcapacity CP engine is indicated by A - Y. ▶ Each capacity level can have 1 to 6 CPs, which will result in a total of 156 different options.
Capacity setting	Derived from the capacity level and the number of processors. For the IBM z16 A02 and IBM z16 AGZ , the capacity levels are A01 - Z06, where the last digit indicates the number of active CPs, and the letter A - Z indicates the processor capacity level. An all IFL or all ICF system has a capacity setting of A00.
Customer Initiated Upgrade (CIU)	A web-based facility where you can request processor and memory upgrades by using the IBM Resource Link and the system's Remote Support Facility (RSF) connection.
Capacity on Demand (CoD)	The ability of a system to increase or decrease its performance capacity as needed to meet fluctuations in demand.
Capacity Provisioning Manager (CPM)	As a component of z/OS Capacity Provisioning, CPM monitors business-critical workloads that are running z/OS on IBM z16 A02 and IBM z16 AGZ .
Customer profile	This information is on Resource Link and contains client and system information. A customer profile can contain information about systems that are related to their IBM customer numbers.
Flexible Capacity for Cyber Resiliency	Available on IBM z16 A02 and IBM z16 AGZ, the optional Flexible Capacity Record is an orderable feature that entitles a customer to active MIPS flexibility for all engine types between IBM z16 A02 and IBM z16 AGZ across two or more sites. It allows capacity swaps for an extended term.
Full capacity CP feature	For IBM z16 A02 and IBM z16 AGZ , capacity settings "Z" (Z0x, x=1-6) are full capacity settings.
High-water mark	Capacity that is purchased and owned by the client.

Term	Description
Installed record	The LICCC record is downloaded, staged to the Support Element (SE), and is installed on the central processor complex (CPC). A maximum of eight different records can be concurrently installed.
Model capacity identifier (MCI)	Shows the current active capacity on the server, including all replacement and billable capacity. For IBM z16 A02 and IBM z16 AGZ , the model capacity identifier is in the form of A0x - Z0x, where x indicates the number of active CPs (x can have a range of 1 - 6).
Model Permanent Capacity Identifier (MPCI)	Keeps information about the capacity settings that are active before any temporary capacity is activated.
Model Temporary Capacity Identifier (MTCI)	Reflects the permanent capacity with billable capacity only, without replacement capacity. If no billable temporary capacity is active, MTCI equals the MPCI.
On/Off Capacity on Demand (CoD)	Represents a function that allows spare capacity in a CPC to be made available to increase the total capacity of a CPC. For example, On/Off CoD can be used to acquire more capacity for handling a workload peak.
Permanent capacity	The capacity that a client purchases and activates. This amount might be less capacity than the total capacity purchased.
Permanent upgrade	LICC that is licensed by IBM to enable the activation of applicable computing resources, such as processors or memory, for a specific CIU-eligible system on a permanent basis.
Purchased capacity	Capacity that is delivered to and owned by the client. It can be higher than the permanent capacity.
Permanent/Temporary entitlement record	The internal representation of a temporary (TER) or permanent (PER) capacity upgrade that is processed by the CIU facility. An <i>entitlement record</i> contains the encrypted representation of the upgrade configuration with the associated time limit conditions.
Replacement capacity	A temporary capacity that is used for situations in which processing capacity in other parts of the enterprise is lost. This loss can be a planned event or an unexpected disaster. The two replacement offerings available are Capacity for Planned Events and Capacity Backup.
Resource Link	The IBM Resource Link is a technical support website that provides a comprehensive set of tools and resources (log in required).
Secondary approval	An option that is selected by the client that requires second approver control for each CoD order. When a secondary approval is required, the request is sent for approval or cancellation to the Resource Link secondary user ID.
Staged record	The point when a record that represents a temporary or permanent capacity upgrade is retrieved and loaded on the SE disk.
Subcapacity	For IBM z16 A02 and IBM z16 AGZ , CP features A01 to Y06 represent subcapacity configurations, and CP features Z01 to Z06 represent full capacity configurations.
Temporary capacity	An optional capacity that is added to the current system capacity for a limited amount of time. It can be capacity that is owned or not owned by the client.
Vital product data (VPD)	Information that uniquely defines system, hardware, and microcode elements of a processing system.

8.1.3 Concurrent and nondisruptive upgrades

Depending on the effect on the system and application availability, upgrades can be classified in the following manner:

- ▶ Concurrent

In general, *concurrency* addresses the continuity of operations of the *hardware* during an upgrade; for example, whether a system (hardware) must be turned off during the upgrade. For more information, see 8.2, “PU upgrades” on page 323.

- ▶ Non-concurrent

This type of upgrade requires turning off the hardware that is being upgraded. Example is a physical memory upgrade to IBM z16 A02 or IBM z16 AGZ.

- ▶ Nondisruptive

Nondisruptive upgrades do not require the software or operating system to be restarted for the upgrade to take effect.

- ▶ Disruptive

An upgrade is considered *disruptive* when resources that are modified or added to an operating system image require that the operating system be restarted to configure the newly added resources.

A Concurrent upgrade might be disruptive to operating systems or programs that do not support the upgrades while being nondisruptive to others. For more information, see 8.10, “Planning for nondisruptive upgrades” on page 355.

8.1.4 Permanent upgrades

Permanent upgrades can be obtained by using the following processes:

- ▶ Ordered through an IBM marketing representative
- ▶ Initiated by the client with the CIU on the IBM Resource Link

Tip: The use of the CIU facility for a system requires that the online CoD buying feature (FC 9900) is installed on the system. The CIU facility is enabled through the permanent upgrade authorization feature code (FC 9898). A prerequisite to FC 9898 is the online CoD buying feature code (FC 9900).

Permanent upgrades that are ordered through an IBM representative

Through a permanent upgrade, you can accomplish the following tasks:

- ▶ Add CPC drawers
- ▶ Add Peripheral Component Interconnect Express (PCIe) drawers and features
- ▶ Add model capacity
- ▶ Add specialty engines
- ▶ Add memory
- ▶ Activate unassigned model capacity or IFLs, ICFs, or zIIPs.
- ▶ Deactivate activated model capacity or IFLs, ICFs, or zIIPs.
- ▶ Add IO channels
- ▶ Add Crypto Express cards
- ▶ Change specialty engines (recharacterization)

Considerations: Most of the MESs can be concurrently applied without disrupting the workload. For more information, see 8.2, “PU upgrades” on page 323. However, certain MES changes are non-concurrent; for example, CPC feature upgrades such as from IBM z16 A02 and IBM z16 AGZ Max5 or Max16 to a Max32 feature.

Memory upgrades are only concurrent when the required memory capacity is already physical available and can be activated through LICCC.

Permanent upgrades by using CIU on the IBM Resource Link

Ordering the following permanent upgrades by using the CIU application through Resource Link allows you to add capacity to fit within your hardware:

- ▶ Add model capacity
- ▶ Add specialty engines
- ▶ Add memory
- ▶ Activate unassigned model capacity or IFLs, ICFs, or zIIPs.
- ▶ Deactivate activated model capacity or IFLs, ICFs, or zIIPs.

8.1.5 Temporary upgrades

IBM z16 A02 and IBM z16 AGZ offer the following types of temporary upgrades:

- ▶ On/Off Capacity on Demand (On/Off CoD)

This offering allows you to temporarily add capacity or specialty engines to cover seasonal activities, period-end requirements, peaks in workload, or application testing. This temporary upgrade can be ordered by using the CIU application through Resource Link only.

Prepaid OOCoD tokens: Beginning with IBM z16 A02 and IBM z16 AGZ, new prepaid OOCoD tokens that are purchased will not carry forward to future systems.

- ▶ CBU

This offering allows you to replace model capacity or specialty engines in a backup system that is used in an unforeseen loss of system capacity because of a disaster.

- ▶ CPE¹

This offering allows you to replace model capacity or specialty engines because of a relocation of workload during system migrations or a data center move.

- ▶ Flexible Capacity Record

This offering allows you to move CPU capacity between machines across two or more sites. Capacity can be moved between sites a maximum of 12 times per year for a maximum of 12 months per move.

Consider the following points:

- ▶ CBU can be ordered by using the CIU application through Resource Link or by contacting your IBM marketing representative.
- ▶ Flexible Capacity can be ordered by contacting your IBM representative.
- ▶ Temporary upgrade capacity changes might be billable or a replacement.

¹ CPE is only available as carry forward on IBM z16 A02 and IBM z16 AGZ

Billable capacity

To handle a peak workload, you can activate up to double the purchased capacity of any processor unit (PU) type temporarily. You are charged daily.

This billable capacity offering is On/Off Capacity on Demand (On/Off CoD).

Replacement capacity

When processing capacity is lost in part of an enterprise, replacement capacity can be activated. It allows you to activate any PU type up to your authorized limit.

The following replacement capacity offerings are available:

- ▶ Capacity Backup
- ▶ Capacity for Planned Event²
- ▶ Flexible Capacity for Cyber Resiliency

8.2 PU upgrades

Concurrent upgrades on [IBM z16 A02](#) and [IBM z16 AGZ](#) can provide more capacity with no system outage. In most cases, a concurrent upgrade can be nondisruptive to the operating system with planning and operating system support.

The concurrent capacity growth capabilities that are provided by [IBM z16 A02](#) and [IBM z16 AGZ](#) include, but are not limited to, the following benefits:

- ▶ Enabling the meeting of new business opportunities
- ▶ Supporting the growth of smart and cloud environments
- ▶ Managing the risk of volatile, high-growth, and high-volume applications
- ▶ Supporting 24 x 7 application availability
- ▶ Enabling capacity growth during lockdown or frozen periods
- ▶ Enabling planned-downtime changes without affecting availability

This capability is based on the flexibility of the design and structure, which allows concurrent hardware installation and Licensed Internal Control Code (LICC) configuration changes.

Subcapacity models provide for CP capacity increase in two dimensions that can be used together to deliver configuration granularity. The first dimension is adding CPs to the configuration. The second is changing the capacity setting of the CPs currently installed to a higher model capacity identifier. In addition, a capacity increase can be delivered by increasing the CP capacity setting, and at the same time decreasing the number of active CPs.

Consideration: An [IBM z16 A02](#) and [IBM z16 AGZ](#) Max5 has a maximum of five PUs available, so it can concurrently be upgraded to models A05 - Z05. If a capacity setting with more than five CPs is required or the combination of CPs and specialty engines exceeds five, a concurrent upgrade to a Max16 feature is required.

[IBM z16 A02](#) and [IBM z16 AGZ](#) allows the concurrent and nondisruptive addition of processors to a running logical partition (LPAR). As a result, you can have a flexible

² CPE is only available as carry forward on IBM z16 A02 and IBM z16 AGZ

infrastructure to which you can add capacity. This function is supported by z/OS, z/VM, and z/VSE. This addition is made by using one of the following methods:

- ▶ With planning ahead for the future need of extra processors. Reserved processors can be specified in the LPAR's profile. When the extra processors are installed, the number of active processors for that LPAR can be increased without the need for a partition reactivation and initial program load (IPL).
- ▶ Another (easier) way is to enable the dynamic addition of processors through the z/OS LOADxx member. Set the **DYNCPADD** parameter in member LOADxx to ENABLE.

8.2.1 CPC drawer feature and PU capacity upgrades

IBM z16 A02 and IBM z16 AGZ have a machine type 3932 and a model capacity identifier.

The 3932 is available with the following four CPC drawer features:

- ▶ Feature Max5 (single CPC drawer installed) can have a maximum of five PUs for client characterization.
- ▶ Feature Max16 (single CPC drawer) can have a maximum of 16 client PUs.
- ▶ Feature Max32 (single CPC drawer) can have a maximum of 32 client PUs.
- ▶ Feature Max68 (two CPC drawers) can have a maximum of 68 client PUs.

Model capacity identifiers (MCI) are A01 to Z06. The MCI described how many CPs are characterized (01 - 06) and the capacity setting (A to Z) of the CPs.

A hardware configuration upgrade always requires more physical hardware (CPC drawers, PCIe+ I/O drawers, or both). A system upgrade can change the feature description (Max) and /or the MCI.

Consider the following points regarding model upgrades:

- ▶ LICCC only upgrade:
 - Can add memory or Virtual Flash Memory (VFM) up to the amount that is physically installed
 - Can change the model capacity identifier, the capacity setting, or both up to the amount that is physically installed
- ▶ Hardware installation and according LICCC upgrade:
 - Can change the CPC drawer feature by adding more PU DCMs
 - Can change the model capacity identifier, the capacity setting, or both
 - Can add physical memory, PCIe+ I/O drawers, and other hardware features

The model capacity identifier can be concurrently changed. Concurrent upgrades can be performed for permanent and temporary upgrades (if the according physical hardware is installed).

CPC drawer feature upgrades:

Consider the following points:

- ▶ Upgrades from CPC feature Max5 to Max16 is concurrent.
- ▶ Upgrades from CPC feature Max5 and Max16 to Max32 and Max68 is disruptive.
- ▶ Upgrades from IBM z16 A02 and IBM z16 AGZ Max32 to Max68 is concurrent.
- ▶ IBM z16 AGZ Max32 to Max68 is concurrent if plan ahead feature (2332) CPC1 Reserve was ordered with the initial system configuration.

Licensed Internal Code upgrades (MES ordered)

The LICCC provides for system upgrades without hardware changes by activating extra (physically installed) unused capacity. Concurrent upgrades through LICCC can be performed for the following resources:

- ▶ Processors, such as CPs, ICFs, z Integrated Information Processors (zIIPs), and IFLs, if unused PUs are available on the installed CPC drawers, or if the model capacity identifier for the CPs can be increased.
- ▶ Memory and VFM, when unused capacity is available on the installed memory cards.

Concurrent hardware installation upgrades (MES ordered)

Configuration upgrades can be concurrent when installing the following resources:

- ▶ The second CPC drawer if coming from a Max32
- ▶ PCIe+ fanouts
- ▶ I/O cards
- ▶ PCIe+ I/O drawers

The concurrent I/O upgrade capability can be better used if a future target configuration is considered during the initial configuration.

Concurrent PU conversions (MES ordered)

[IBM z16 A02](#) and [IBM z16 AGZ](#) supports concurrent conversion between all PU types, to provide flexibility and meet changing business requirements.

Important: The LICCC-based PU conversions require that at least one PU (CP, ICF, or IFL), remains unchanged. Otherwise, the conversion is disruptive. The PU conversion generates a LICCC that can be installed concurrently in two steps:

1. Remove the assigned PU from the configuration.
2. Activate the newly available PU as the new PU type.

LPARs also might have to free the PUs to be converted. The operating systems must include support to configure processors offline or online so that the PU conversion can be done nondisruptively.

Considerations: Client planning and operator action are required to use concurrent PU conversion. Consider the following points about PU conversion:

- ▶ It is disruptive if *all* current PUs are converted to different types.
- ▶ It might require individual LPAR outages if dedicated PUs are converted.

Unassigned CP capacity is recorded by a model capacity identifier. CP feature conversions change (increase or decrease) the model capacity identifier.

8.2.2 Customer Initiated Upgrade facility

The CIU facility is an IBM online system through which you can order, download, and install permanent and temporary upgrades for IBM zSystems. Access to and use of the CIU facility requires a contract between the client and IBM through which the terms and conditions for use of the CIU facility are accepted.

The CIU facility is controlled through the permanent upgrade authorization FC 9898. A prerequisite to FC 9898 is the online CoD buying feature code (FC 9900). Although FC 9898 can be installed on your [IBM z16 A02 and IBM z16 AGZ](#) at any time, often it is added when ordering an [IBM z16 A02 and IBM z16 AGZ](#).

After you place an order through the CIU facility, you receive a notice via e-mail that the order is ready for download. You can then download and apply the upgrade by using functions that are available through the Hardware Management Console (HMC) in the task **Perform Model Conversion** and if Remote Support Facility (RSF) is available. After all of the prerequisites are met, the entire process (from ordering to activation of the upgrade) is performed by the client and does not require any onsite presence of IBM System Service Representative (SSR).

CIU prerequisites

The CIU facility supports LICCC upgrades only. It does not support I/O upgrades. All other capacity that is required for an upgrade must be previously installed. Extra CPC drawers or I/O cards cannot be installed as part of an order that is placed through the CIU facility.

To place an CIU enough un-characterized PUs must be available according your CPC Max feature and already characterized PU (including unassigned -ICFs, -zIIPs, and -IFLs).

CIU registration and contract for CIU

To use the CIU facility, a client must be registered and the system must be set up. After you complete the CIU registration, access to the CIU application is available through the [IBM Resource Link website](#).

As part of the setup, provide one resource link ID for configuring and placing CIU orders and, if required, a second ID as an approver. The IDs are then set up for access to the CIU support. The CIU facility allows upgrades to be ordered and delivered much faster than through the regular MES process.

To order and activate the upgrade, log on to the [IBM Resource Link website](#) and start the CIU application to upgrade a system for processors or memory. You can request a client order approval to conform to your operational policies. You also can allow the definition of more IDs to be authorized to access the CIU. More IDs can be authorized to enter or approve CIU orders, or only view orders.

Permanent upgrades

Permanent upgrades can be ordered by using the CIU facility. Through the CIU facility, you can generate online permanent upgrade orders to concurrently add processors (CPs, ICFs, zIIPs, and IFLs) and memory, or change the model capacity identifier. You can do so up to the limits of the installed CPC drawers on a system.

Temporary upgrades

The [IBM z16 A02](#) and [IBM z16 AGZ](#) base model describes permanent and dormant capacity by using the capacity marker and the number of PU features that are installed on the system. Up to eight temporary offerings can be present. Each offering includes its own policies and controls, and each can be activated or deactivated independently in any sequence and combination. Although multiple offerings can be active at any time, *only one On/Off CoD offering can be active at any time* if enough resources are available to fulfill the offering specifications.

Temporary upgrades are represented in the system by a *record*. All temporary upgrade records are on the SE hard disk drive (HDD). The records can be downloaded from the RSF or installed from portable media. At the time of activation, you can control everything locally.

The provisioning architecture is shown in Figure 8-1.

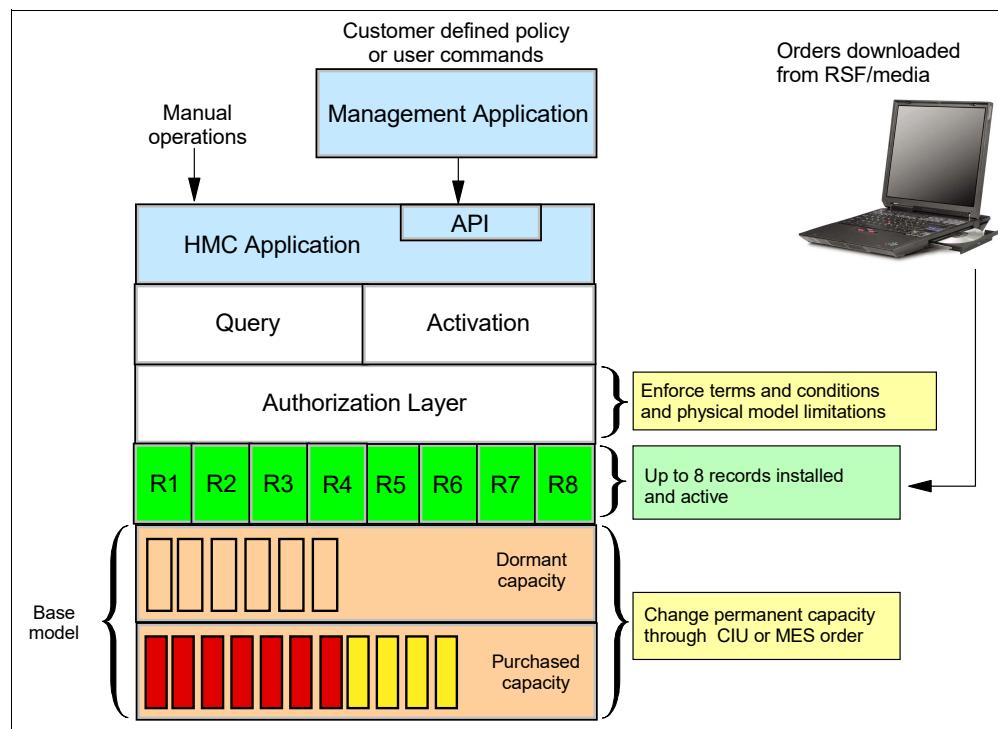


Figure 8-1 Provisioning architecture

The authorization layer enables administrative control over the temporary offerings. The activation and deactivation can be driven manually in task **Perform Model Conversion** at HMC or SE or with using Web Services API / BCPii.

By using the API approach, you can customize at activation time the resources that are necessary to respond to the current situation up to the maximum that is specified in the order record. If the situation changes, you can add or remove resources without having to go back to the base configuration. This process eliminates the need for temporary upgrade specifications for all possible scenarios.

For a CPE record, only the ordered configuration can be activated.

This approach also enables you to update and replenish temporary upgrades, even in situations where the upgrades are active. Likewise, depending on the configuration, permanent upgrades can be performed while temporary upgrades are active. Examples of the activation sequence of multiple temporary upgrades are shown in Figure 8-2.

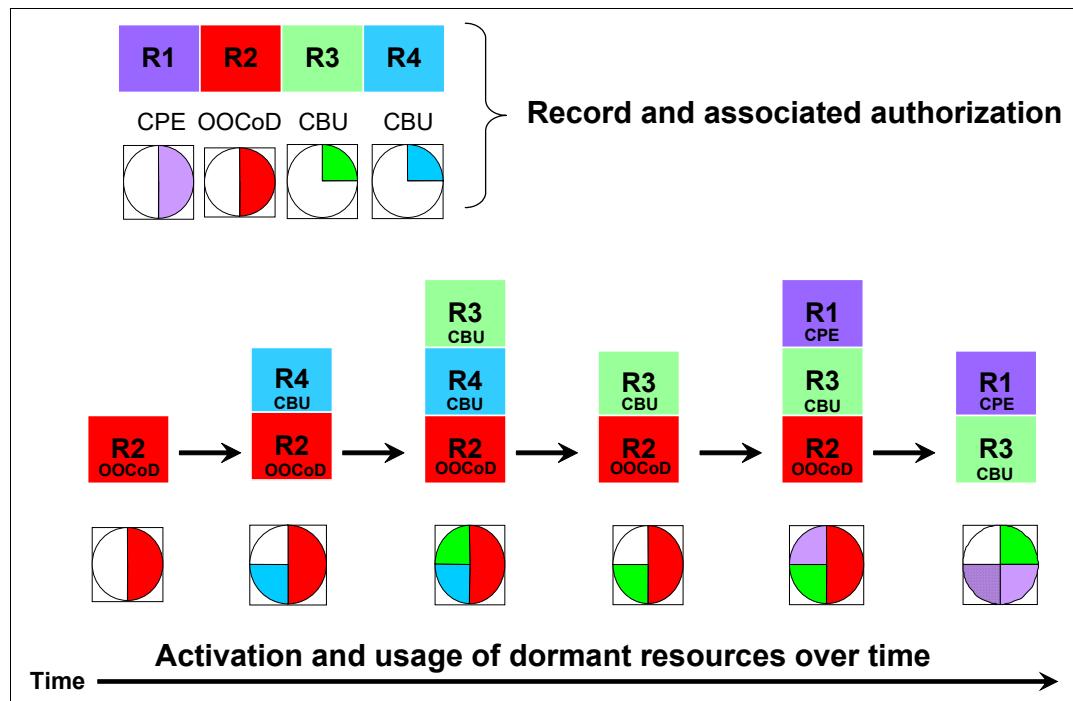


Figure 8-2 Example of temporary upgrade activation sequence

As shown in Figure 8-2, if R2, R3, and R1 are active at the same time, only parts of R1 can be activated because not enough resources are available to fulfill all of R1. When R2 is deactivated, the remaining parts of R1 can be activated as shown.

Temporary capacity can be billable as On/Off CoD, or replacement capacity as CBU, or CPE. Consider the following points:

- ▶ On/Off CoD is a function that enables *concurrent* and *temporary* capacity growth of the system.
On/Off CoD can be used for client peak workload requirements, for any length of time, and includes a daily hardware and maintenance charge. The software charges can vary according to the license agreement for the individual products. For more information, contact your IBM representative.

On/Off CoD can concurrently add processors (CPs, ICFs, zIIPs, and IFLs), increase the model capacity identifier, or both. It can do so up to the limit of the installed CPC drawers of a system. It is restricted to twice the installed capacity. On/Off CoD requires a contractual agreement between you and IBM.

You decide whether to pre-pay or post-pay On/Off CoD. Capacity tokens that are inside the records are used to control activation time and resources.

- ▶ CBU is a concurrent and temporary activation of more CPs, ICFs, zIIPs, and IFLs; or an increase of the model capacity identifier; or both.

Note: CBU cannot be used for peak workload management in any form.

On/Off CoD is the correct method to use for workload management. A CBU activation can last up to 90 days when a disaster or recovery situation occurs.

CBU features are optional, and require unused capacity to be available on installed CPC drawers of the backup system. They can be available as unused PUs, an increase in the model capacity identifier, or both.

A CBU contract must be in place before the special code that enables this capability can be loaded on the system. The standard CBU contract provides for five 10-day tests (the *CBU test activation*) and one 90-day activation over a five-year period. For more information, contact your IBM representative.

You can run production workload on a CBU upgrade during a CBU test. At least an *equivalent amount* of production capacity must be shut down during the CBU test. If you signed CBU contracts, you also must sign an Amendment with IBM to allow you to run production workload on a CBU upgrade during your CBU tests. More 10-day tests can be purchased with the CBU record.

- ▶ CPE is a concurrent and temporary activation of extra CPs, ICFs, zIIPs, and IFLs; or an increase of the model capacity identifier; or both.

The CPE offering is used to replace temporary lost capacity within a client's enterprise for planned downtime events, such as data center changes.

Notes:

1. CPE cannot be used for peak load management of client workload or for a disaster situation.
2. CPE is only available as carry forward on IBM z16 A02 and IBM z16 AGZ.

The CPE feature requires unused capacity to be available on installed CPC drawers of the backup system. The capacity must be available as unused PUs, as a possibility to increase the model capacity identifier on a subcapacity system, or as both.

A CPE contract must be in place before the special code that enables this capability can be loaded on the system. The standard CPE contract provides for one 3-day planned activation at a specific date. For more information, contact your IBM representative.

8.2.3 Concurrent upgrade functions summary

The possible concurrent upgrades combinations are listed in Table 8-2.

Table 8-2 Concurrent upgrade summary

Type	Name	Upgrade	Process
Permanent	MES	CPs, ICFs, zIIPs, IFLs, CPC drawer, memory, and I/Os	Installed by IBM SSRs
	Online permanent upgrade	CPs, ICFs, zIIPs, IFLs and memory	Performed through the CIU facility
Temporary	On/Off CoD	CPs, ICFs, zIIPs, and IFLs	Performed through the OOCoD facility
	CBU	CPs, ICFs, zIIPs, and IFLs	Activated through model conversion
	CPE	CPs, ICFs, zIIPs, and IFLs	Activated through model conversion
	Flexible Capacity Record	CPs, ICFs, zIIPs, and IFLs	Activated through model conversion

8.3 Miscellaneous equipment specification (MES) upgrades

MES upgrades enable concurrent and permanent capacity growth. MES upgrades allow the concurrent adding of processors (CPs, ICFs, zIIPs, and IFLs), memory capacity, and I/O ports. For subcapacity models, MES upgrades allow the concurrent adjustment of both the number of processors and the capacity level. The MES upgrade can be performed by using LICCC only, installing the second CPC drawer, adding I/O drawers, adding I/O features, or using the following combinations:

- ▶ MES upgrades for processors are done by any of the following methods:
 - LICCC assigning and activating unassigned PUs up to the limit of the installed CPC drawers.
 - LICCC to adjust the number and types of PUs to change the capacity setting, or both.
 - Installing the second CPC drawer and LICCC assigning and activating unassigned PUs on the installed CPC drawers.
- ▶ MES upgrades for memory are done by one of the following methods:
 - By using LICCC to activate more memory capacity up to the limit of the memory cards on the currently installed CPC drawers.
 - Installing the second CPC drawer and the use of LICCC to activate more memory capacity on installed CPC drawers.
 - By using the CPC Enhanced Drawer Availability (EDA), where possible, on multi-drawer systems to add or change the memory cards.
- ▶ MES upgrades for I/O are done by installing more I/O features and supporting infrastructure (if required) on I/O drawers that are installed, or installing more I/O drawers to hold the new cards.

A physical MES upgrade requires an IBM SSRs for the installation.

To better use the MES upgrade function, carefully plan the initial configuration to allow a concurrent upgrade to a target configuration. The availability of I/O drawers improves the flexibility to perform unplanned I/O configuration changes concurrently.

The Store System Information (STSI) instruction gives more useful and detailed information about the base configuration and temporary upgrades.

The model and model capacity identifiers that are returned by the STSI instruction are updated to coincide with the upgrade. For more information, see “Store System Information instruction” on page 358.

Upgrades: An MES provides the physical upgrade, which results in more enabled processors, different capacity settings for the CPs, and more memory, I/O ports, I/O adapters, and I/O drawers. Extra planning tasks are required for nondisruptive logical upgrades. For more information, see “Guidelines to avoid disruptive upgrades” on page 360.

8.3.1 MES upgrade for processors

An MES upgrade for processors can concurrently add CPs, ICFs, zIIPs, and IFLs to an [IBM z16 A02](#) and [IBM z16 AGZ](#) by assigning available PUs on the CPC drawers through LICCC. Depending on the quantity of the extra processors in the upgrade, more PU DCMs and one or the second CPC drawer might be required, and can be concurrently installed before the LICCC is enabled. With the subcapacity models, more capacity can be provided by adding CPs, changing the capacity identifier on the current CPs, or both.

Limits: The sum of CPs, inactive CPs, ICFs, unassigned ICFs, zIIPs, unassigned zIIPs, IFLs, unassigned and IFLs cannot exceed the maximum limit of PUs available for client use. The number of zIIPs cannot exceed twice the number of purchased CPs.

An example of an MES upgrade for processors (with two upgrade steps) is shown in Figure 8-3.

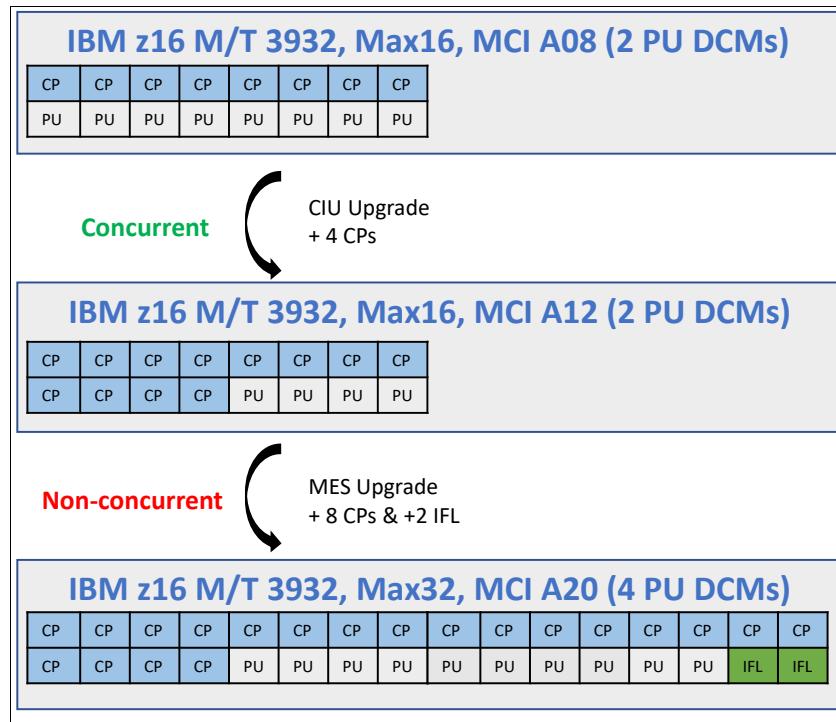


Figure 8-3 MES for processor example

In the example that is shown in Figure 8-3 on page 331 it is an IBM z16 A02 with an MCI A08. An upgrade to an MCI A12 is current, as with a Max16 you have in total 16 PUs to characterize.

The next upgrade step from A12 to A20 and additional 2 IFLs is non-concurrent, as you need additional PUs to fulfil the 20 CPs and 2 IFLs. The next possible Max feature is Max32. To reach the 32 PUs, additional physical HW (two additional DCMs) has to be non-concurrently installed.

Consideration: All available processors on a server (including reserved processors) can be defined to an LPAR. However, do not define more processors to an LPAR than the target operating system supports.

The number of processors that are supported by various operating systems releases are listed in Table 8-3.

Table 8-3 Number of processors that are supported by operating system

Operating system	Number of processors that are supported
z/OS V2R3 and later	200 PUs per z/OS LPAR in non-SMT mode and 128 PUs per z/OS LPAR in SMT mode. For both, the PU total is the sum of CPs and zIIPs.
z/VM V7R2 and later	80 (or 40 in SMT mode).
z/VSE	z/VSE Turbo Dispatcher can use up to 4CPs, and tolerates up to 10-way LPARs.
z/TPF	86 CPs.
Linux on IBM Z	The IBM z16 limit is 200 CPs although Linux ^a supports 256 cores without SMT and 128 cores with SMT (256 threads).

a. Supported Linux on IBM Z distributions (for more information, see Chapter 7, “Operating system support” on page 241).

Software charges, which are based on the total capacity of the system on which the software is installed, are adjusted to the new capacity after the MES upgrade.

Software products that use Workload License Charges (WLC) or Taylor Fit Pricing (TFP) might not be affected by the system upgrade. Their charges are based on partition usage, not on the system total capacity. For more information about WLC, see 7.8, “Software licensing” on page 314.

8.3.2 MES upgrades for memory

MES upgrades for memory can concurrently add more memory in the following ways:

- ▶ Through LICCC, which enables more capacity up to the limit of the currently installed DIMM memory cards
- ▶ Concurrently installing the second CPC drawer and LICCC-enabling memory capacity on the new CPC drawers.

If the IBM z16 A02 or IBM z16 AGZ have the second CPC drawer, you can use the Concurrent Drawer Replacement (CDR) feature to remove a CPC drawer and add DIMM memory cards concurrent. It can also be used to upgrade the installed memory cards to a

larger capacity size. After physical installation the LICCC is then installed to enable the extra memory.

With proper planning, memory can be added nondisruptive to z/OS partitions and z/VM partitions. If necessary, new LPARs can be created nondisruptive to use the newly added memory.

An LPAR can dynamically take advantage of a memory upgrade if reserved storage is defined to that LPAR. The reserved storage is defined to the LPAR as part of the image profile.

Reserved memory can be configured online to the LPAR by using the LPAR dynamic storage reconfiguration (DSR) function. DSR allows a z/OS operating system image and z/VM partitions to add reserved storage to their configuration if any unused storage exists.

The nondisruptive addition of storage to a z/OS and z/VM partition requires the correct operating system parameters to be set. If reserved storage is not defined to the LPAR, the LPAR must be deactivated, the image profile changed, and the LPAR reactivated. This process allows the extra storage resources to be available to the operating system image.

Adding or changing physical memory for a single CPC drawer to an IBM z16 A02 or IBM z16 AGZ is disruptive.

8.3.3 MES upgrades for I/O and CPC drawers

MES upgrades for I/O can concurrently add more I/O features by using one of the following methods:

- ▶ Installing more I/O features on an installed PCIe+ I/O drawer.
- ▶ Adding a PCIe+ I/O drawer to hold the new I/O features.

You can not order empty PCIe+ I/O drawers. Depending on the number of I/O features, the configurator determines the number of PCIe+ I/O drawers required.

For more information about PCIe+ I/O drawers, see 4.2, “I/O system overview” on page 140.

IBM z16 A02

For IBM z16 A02 you can have 1-2 CPC drawers and 0-3 PCIe+ I/O drawers. To add more CPC drawers or PCIe+ I/O drawers, reserved space (as needed in previous zSystems) are no longer required for CPC in IBM z16 A02.

IBM z16 AGZ

For IBM z16 AGZ you can have the same amount of CPC drawers and I/O drawers than in IBM z16 A02.

Important: To upgrade CPC drawers and PCIe+ I/O drawers concurrent, you need to have plan-ahead reserved features:

- ▶ FC 2332 - CPC1 Reserve
- ▶ FC 2333 - IO1 Reserve
- ▶ FC 2334 - IO2 Reserve
- ▶ FC 2335 - IO3 Reserve

The number of PCIe+ I/O drawers that can be present in an [IBM z16 A02 or IBM z16 AGZ](#) configurations is listed in Table 8-4 on page 334.

Table 8-4 PCIe+ I/O drawer summary

Description	New build	MES add
PCIe+ I/O drawers for IBM z16 A02	0-3	1-3
PCIe+ I/O drawers for IBM z16 AGZ	0-3	1-3

To better use the MES for I/O capability, carefully plan the initial configuration to allow concurrent upgrades up to the target configuration.

If a PCIe+ I/O drawer is added to an IBM z16 A02 or IBM z16 AGZ and original features must be physically moved to another PCIe+ I/O drawer, original card moves are disruptive.

z/VSE, z/TPF, and Linux on IBM Z do *not* provide dynamic I/O configuration support. Although installing the new hardware is done concurrently, defining the new hardware to these operating systems requires an IPL.

Tip: IBM z16 A02 and IBM z16 AGZ feature a hardware system area (HSA) of 160 GB (same as IBM z15 T02). HSA is *not* part of the client-purchased memory.

8.4 Permanent upgrade by using the CIU facility

By using the CIU facility (through [the IBM Resource Link](#)), you can start a permanent upgrade for CPs, ICFs, zIIPs, and IFLs, or memory. When performed through the CIU facility, you add the resources without IBM personnel present at your location. You can also unassign previously purchased CPs and IFL processors through the CIU facility.

Adding permanent upgrades to a system through the CIU facility requires that Online CoD Buying feature (FC 9900) and the permanent upgrade enablement feature (FC 9898) is installed on the system. A permanent upgrade might change the system model capacity identifier ($A0x - Z0x$) if more CPs are requested, or if the capacity identifier is changed as part of the permanent upgrade. If necessary, more LPARs can be created concurrently to use the newly added processors. Maintenance charges are automatically adjusted as a result of a permanent upgrade.

Software charges that are based on the total capacity of the system on which the software is installed are adjusted to the new capacity after the permanent upgrade is installed. Software products that use WLC or customers with TFP might not be affected by the system upgrade because their charges are based on LPAR usage rather than system total capacity. For more information about WLC, see 7.8, “Software licensing” on page 314.

The CIU facility process on IBM Resource Link is shown in Figure 8-4.

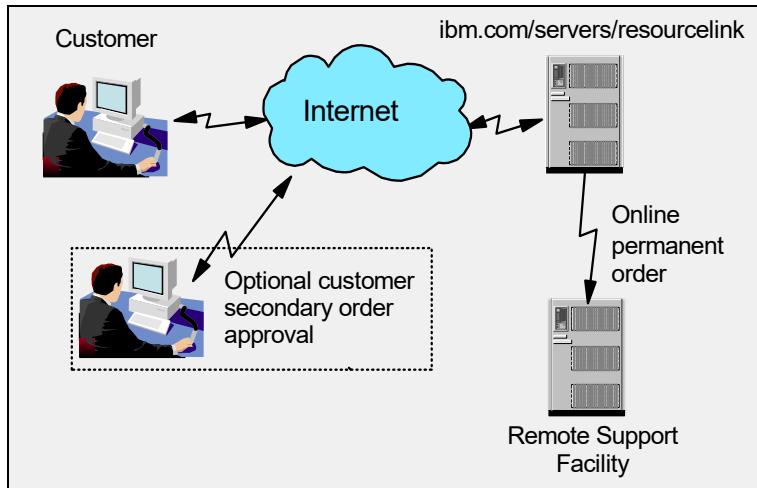


Figure 8-4 Permanent upgrade order example

The following sample sequence shows how to start an order on the IBM Resource Link:

1. Sign on to Resource Link.
2. Select **Customer Initiated Upgrade** from the main Resource Link page. Client and system information that is associated with the user ID are displayed.
3. Select the system to receive the upgrade. The current configuration (PU allocation and memory) is shown for the selected system.
4. Select **Order Permanent Upgrade**. The Resource Link limits the options to those options that are valid or possible for the selected configuration (system).
5. After the target configuration is verified by the system, accept or cancel the order. An order is created and verified against the pre-established agreement.
6. Accept or reject the price that is quoted. A secondary order approval is optional. Upon confirmation, the order is processed. The LICCC for the upgrade is available within hours.

The order activation process for a permanent upgrade is shown in Figure 8-5 on page 336. When the LICCC is passed to the Remote Support Facility, you are notified through an e-mail that the upgrade is ready to be downloaded.

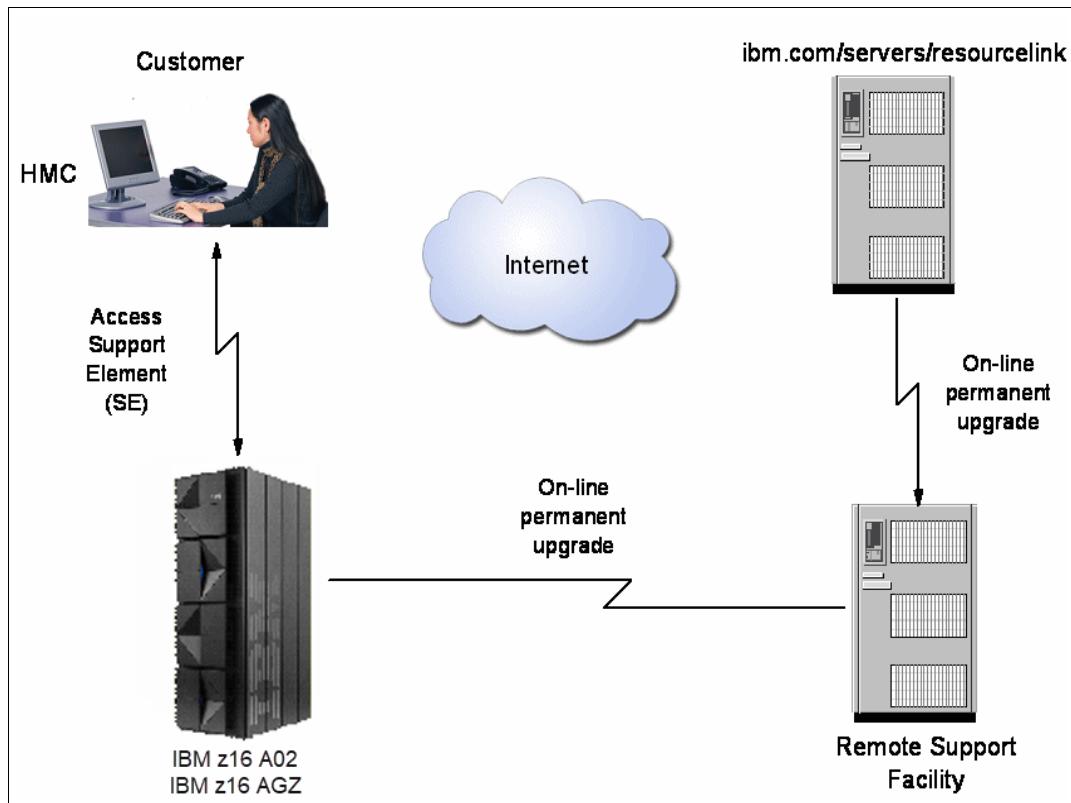


Figure 8-5 CIU-eligible order activation example

8.4.1 Ordering

IBM Resource Link provides the interface that enables you to order a concurrent upgrade for a system. You can create, cancel, or view the order, and view the history of orders that were placed through this interface.

Configuration rules enforce that only valid configurations are generated within the limits of the individual system. Warning messages are issued if you select invalid upgrade options. The process allows only one permanent CIU-eligible order for each system to be placed at a time.

For more information, see the [IBM Resource Link website](#) (log in required).

The initial view of the Machine profile on Resource Link is shown in Figure 8-6.

Current configuration		Machine summary	Ordering options
Model Capacity:	729 (29 CPs)	Type, model, serial: 8561 - T01 - AFPS06SE	Order permanent upgrade Order On/Off CoD record Order On/Off CoD test record Order On/Off CoD record with prepaid upgrades
ICF:	3	System name: AFPS06SE	Order On/Off CoD record with spending limits
zIIP:	11	Customer summary	
IFL:	28	Company name: [REDACTED]	Order administrative On/Off CoD test record
SAP:	8	Customer number: [REDACTED]	Order Capacity Backup (CBU) record
Memory:	4864	GEO, country: Americas - zDutchy of Merwyn	Order Capacity for Planned Events (CPE) record Order System Recovery Boost Upgrade record
Unassigned IFLs: 0		Display upgrade matrix	
Current configuration as of 26 Feb 2022 18:49:18			
About ordering		To update profile	
Authorization to create orders User ID: brunofarrugia@fr.ibm.com and 3 more		Upload VPD	
Authorization to approve orders		Upload upgrade billing XML data Disable machine profile...	
Order CIU permanent Enabled Order On/Off CoD Enabled			

Figure 8-6 Machine profile window

The number of CPs, ICFs, zIIPs, IFLs, memory size, unassigned IFLs, unassigned ICFs, and unassigned zIIPs on the current configuration are displayed on the left side of the page.

Resource Link retrieves and stores relevant data that is associated with the processor configuration, such as the number of CPs and installed memory cards. It allows you to select only those upgrade options that are deemed valid by the order process. It also allows upgrades only within the bounds of the currently installed hardware.

8.4.2 Retrieval and activation

After an order is placed and processed, the upgrade record is passed to the IBM support system for download.

When the order is available for download, you receive an email that contains an activation number. You can then retrieve the order by using the Perform Model Conversion task from the SE, or through the Single Object Operation to the SE from an HMC.

In the **Perform Model Conversion** window, select **Permanent upgrades** to start the process, as shown in Figure 8-7.

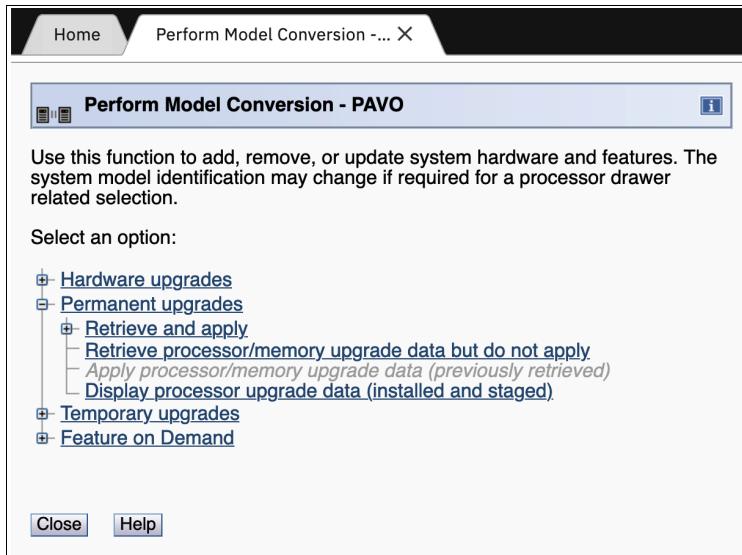


Figure 8-7 IBM z16 Perform Model Conversion window

The window provides several possible options. If you select the **Retrieve and apply** data option, you are prompted to enter the order activation number to start the permanent upgrade, as shown in Figure 8-8.

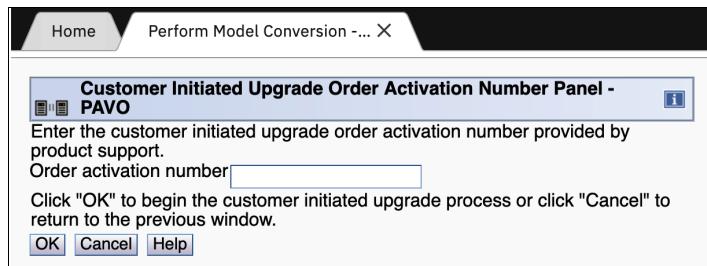


Figure 8-8 Customer Initiated Upgrade Order Activation Number window

8.5 On/Off Capacity on Demand

On/Off CoD allows you to temporarily enable processors that are available within the current hardware model. You can also use it to change capacity settings for CPs to help meet your peak workload requirements.

8.5.1 Overview

The capacity for CPs is expressed in millions of service units (MSUs). Capacity for speciality engines is expressed in number of speciality engines. *Capacity tokens* are used to limit the resource consumption for all types of processor capacity.

Capacity tokens were introduced to provide better control over resource consumption when On/Off CoD offerings are activated. Tokens represent the following resource consumptions:

- ▶ For CP capacity, each token represents the amount of CP capacity that results in one MSU of software cost for one day (*an MSU-day token*).
- ▶ For speciality engines, each token is equivalent to one speciality engine capacity for one day (*an engine-day token*).

Each speciality engine type features its own tokens, and each On/Off CoD record includes separate token pools for each capacity type. During the ordering sessions on Resource Link, select how many tokens of each type to create for an offering record. Each engine type must include tokens for that engine type to be activated. Capacity that has no tokens cannot be activated.

When resources from an On/Off CoD offering record that contains capacity tokens are activated, a *billing window* is started. A billing window is always 24 hours. Billing occurs at the end of each billing window.

The resources that are billed are the highest resource usage inside each billing window for each capacity type. An activation period is one or more complete billing windows. The activation period is the time from the first activation of resources in a record until the end of the billing window in which the last resource in a record is deactivated.

At the end of each billing window, the tokens are decremented by the highest usage of each resource during the billing window. If any resource in a record does not have enough tokens to cover usage for the next billing window, the entire record is deactivated.

Note: On/Off CoD requires that the Online CoD Buying features (FC 9900) and (FC 9896) are installed on the system that you want to upgrade.

The On/Off CoD to Permanent Upgrade Option gives customers a window of opportunity to assess capacity additions to your permanent configurations by using On/Off CoD. If a purchase is made, the hardware On/Off CoD charges during this window (three days or less) are waived. If no purchase is made, you are charged for the temporary use.

The resources eligible for temporary use are CPs, ICFs, zIIPs, and IFLs. The temporary addition of memory and I/O ports or adapters is not supported.

Unassigned PUs that are on the installed CPC drawers can be temporarily and concurrently activated as CPs, ICFs, zIIPs, and IFLs through LICCC. You can assign PUs up to twice the currently installed CP capacity, and up to twice the number of ICFs, zIIPs, or IFLs.

An On/Off CoD upgrade cannot change the system capacity feature. The addition of new CPC drawers is not supported. However, the activation of an On/Off CoD upgrade can increase the model capacity identifier (*A0x-Z0x*).

8.5.2 Capacity Provisioning Manager

The installation of the capacity provision function on z/OS requires the following prerequisites:

- ▶ Setting up and customizing z/OS RMF, including the Distributed Data Server (DDS).
- ▶ Setting up the z/OS CIM Server (included in z/OS base).
- ▶ Performing capacity provisioning customization. For more information, see *z/OS MVS Capacity Provisioning User's Guide*, SC34-2661.

Using the capacity provisioning function requires the following prerequisites:

- ▶ TCP/IP connectivity to observed systems.
- ▶ RMF Distributed Data Server must be active.
- ▶ CIM server must be active.
- ▶ Security and CIM customization.
- ▶ Capacity Provisioning Manager customization.

The Capacity Provisioning Manager Console is provided as part of z/OSMF, which provides a browser-based interface for managing z/OS systems.

Customizing the capacity provisioning function is required on the following systems:

- ▶ Observed z/OS systems

These systems are in one or multiple sysplexes that are to be monitored. For more information about the capacity provisioning domain, see 8.10, “Planning for nondisruptive upgrades” on page 355.

- ▶ Runtime systems

These are systems where the Capacity Provisioning Manager is running, or to which the server can fail over after a system failure.

8.5.3 Ordering

On/Off CoD allows you to temporarily turn on unowned PUs, unassigned CPs (or unassigned CP capacity), and unassigned specialty engines (IFLs, ICFs and zIIPs) available within the current CPC drawer feature (Max) with the following limitations:

- ▶ Temporary model capacity with CPs and capacity level equal to or greater than the active model capacity, up to 100% of the purchased capacity (active permanent capacity plus unassigned permanent capacity).
- ▶ As many temporary specialty engines up to the total of purchased specialty engines (permanently active specialty engines plus unassigned specialty engines).
- ▶ As many additional specialty engines of each type up to the total purchased specialty engines of each type.

On/Off CoD can be ordered as prepaid or postpaid. A prepaid On/Off CoD offering record contains resource descriptions, MSUs, specialty engines, and tokens that describe the total capacity that can be used. For CP capacity, the token contains MSU-days. For specialty engines, the token contains specialty engine-days.

When resources on a prepaid offering are activated, they must have enough capacity tokens to allow the activation for an entire billing window, which is 24 hours. The resources remain active until you deactivate them or until one resource uses all of its capacity tokens. Then, all activated resources from the record are deactivated.

A postpaid On/Off CoD offering record contains resource descriptions, MSUs, specialty engines, and can contain capacity tokens that denote MSU-days and specialty engine-days.

When resources in a postpaid offering record *without* capacity tokens are activated, those resources remain active until they are deactivated, or until the offering record expires. The record normally expires 180 days after its installation.

When resources in a postpaid offering record *with* capacity tokens are activated, those resources must include enough capacity tokens to allow the activation for an entire billing window (24 hours). The resources remain active until they are deactivated, until all of the

resource tokens are used, or until the record expires. The record usually expires 180 days after its installation. If one capacity token type is used, resources from the entire record are deactivated.

For example, for an [IBM z16 A02](#) and [IBM z16 AGZ](#) with capacity identifier D02 (two CPs), a capacity upgrade through On/Off CoD can be delivered in the following ways:

- ▶ Add CPs of the same capacity setting. With this option, the model capacity identifier can be changed to a D03, which adds another CP to make it a three-way CP. It can also be changed to a D04, which adds two CPs, making it a four-way CP.
- ▶ Change to a different capacity level of the current CPs and change the model capacity identifier to a E02 or F02. The capacity level of the CPs is increased, but no other CPs are added. The D02 also can be temporarily upgraded to a E03, which increases the capacity level and adds another processor.

Use the Large System Performance Reference (LSPR) information to evaluate the capacity requirements according to your workload type. For more information about LSPR data for current IBM processors, see the [Large Systems Performance Reference for IBM zSystems page](#) of the IBM Systems website.

The On/Off CoD hardware capacity is charged on a 24-hour basis. A grace period is granted at the end of the On/Off CoD day. This grace period allows up to an hour after the 24-hour billing period to change the On/Off CoD configuration for the next 24-hour billing period or deactivate the current On/Off CoD configuration. The times when the capacity is activated and deactivated are maintained in the BM z16 A02 or IBM z16 AGZ and sent back to the IBM support systems.

If On/Off capacity is active, On/Off capacity can be added without having to return the system to its original capacity. If the capacity is increased multiple times within a 24-hour period, the charges apply to the highest amount of capacity active in that period.

If more capacity is added from an active record that contains capacity tokens, the system checks whether the resource has enough capacity to be active for an entire billing window (24 hours). If that criteria is not met, no extra resources are activated from the record.

If necessary, more LPARs can be activated concurrently to use the newly added processor resources.

Consideration: On/Off CoD provides a concurrent hardware upgrade that results in more capacity being made available to a system configuration. Extra planning tasks are required for nondisruptive upgrades. For more information, see “Guidelines to avoid disruptive upgrades” on page 360.

To participate in this offering, you must accept contractual terms for purchasing capacity through the Resource Link, establish a profile, and install an On/Off CoD enablement feature on the system. Later, you can concurrently install temporary capacity up to the limits in On/Off CoD and use it for up to 180 days.

Monitoring occurs through the system call-home facility. An invoice is generated if the capacity is enabled during the calendar month. You are billed for the use of temporary capacity until the system is returned to the original configuration. Remove the enablement code if the On/Off CoD support is no longer needed.

On/Off CoD orders can be pre-staged in Resource Link to allow multiple optional configurations. The pricing of the orders is done at the time that you order them, and the pricing can vary from quarter to quarter. Staged orders can have different pricing.

When the order is downloaded and activated, the daily costs are based on the pricing at the time of the order. The staged orders do not have to be installed in the order sequence. If a staged order is installed out of sequence and later a higher-priced order is staged, the daily cost is based on the lower price.

Another possibility is to store multiple On/Off CoD LICCC records on the SE with the same or different capacities, which gives you greater flexibility to enable quickly needed temporary capacity. Each record is easily identified with descriptive names, and you can select from a list of records that can be activated.

Resource Link provides the interface to order a dynamic upgrade for a specific system. You can create, cancel, and view the order. Configuration rules are enforced, and only valid configurations are generated based on the configuration of the individual system. After you complete the prerequisites, orders for the On/Off CoD can be placed. The order process uses the CIU facility on Resource Link.

Memory and channels are not supported on On/Off CoD.

An individual record can be activated only once. Subsequent sessions require a new order to be generated, which produces a new LICCC record for that specific order. Alternatively, you can use an *auto-renewal* feature to eliminate the need for a manual replenishment of the On/Off CoD order. This feature is implemented in Resource Link, and you must also select this feature in the machine profile, as shown in Figure 8-9.

Figure 8-9 Order On/Off CoD record window

8.5.4 On/Off CoD testing

Each On/Off CoD-enabled system is entitled to one no-charge 24-hour test. No IBM charges are assessed for the test, including charges that are associated with temporary hardware capacity, IBM software, and IBM maintenance. The test can be used to validate the processes to download, stage, install, activate, and deactivate On/Off CoD capacity.

This test can have a maximum duration of 24 hours, which commences upon the activation of any capacity resource that is contained in the On/Off CoD record. Activation levels of capacity

can change during the 24-hour test period. The On/Off CoD test automatically stops at the end of the 24-hour period.

You also can perform administrative testing. No capacity is added to the system, but you can test all the procedures and automation for the management of the On/Off CoD facility.

8.5.5 Activation and deactivation

When a previously ordered On/Off CoD is retrieved from Resource Link, it is downloaded and stored on the Support Element (SE). You can activate the order manually or through automation when the capacity is needed.

If the On/Off CoD offering record does not contain resource tokens, you must deactivate the temporary capacity manually. Deactivation is done from the SE and is nondisruptive. Depending on how the capacity was added to the LPARs, you might be required to perform tasks at the LPAR level to remove it. For example, you might have to configure offline any CPs that were added to the partition, deactivate LPARs that were created to use the temporary capacity, or both.

On/Off CoD orders can be staged in Resource Link so that multiple orders are available. An order can be downloaded and activated only once. If a different On/Off CoD order is required or a permanent upgrade is needed, it can be downloaded and activated without having to restore the system to its original purchased capacity.

In support of automation, an API is available that allows the activation of the On/Off CoD records. The activation is performed from the HMC and requires specifying the order number. With this API, automation code can be used to send an activation command along with the order number to the HMC to enable the order.

8.5.6 Termination

A client is contractually obligated to end the On/Off CoD right-to-use feature when a transfer in asset ownership occurs. A client also can choose to end the On/Off CoD right-to-use feature without transferring ownership.

Removing FC 9898 ends the right to use the On/Off CoD. This feature cannot be ordered if a temporary session is active. Similarly, the CIU enablement feature cannot be removed if a temporary session is active. When the CIU enablement feature is removed, the On/Off CoD right-to-use feature is simultaneously removed. Reactivating the right-to-use feature subjects the client to the terms and fees that apply then.

Upgrade capability during On/Off CoD

Upgrades that involve physical hardware are supported while an On/Off CoD upgrade is active on a particular [IBM z16 A02](#) and [IBM z16 AGZ](#) configuration. LICCC-only upgrades can be ordered and retrieved from Resource Link, and can be applied while an On/Off CoD upgrade is active. LICCC-only memory upgrades can be retrieved and applied while an On/Off CoD upgrade is active.

Repair capability during On/Off CoD

If the BM z16 A02 or IBM z16 AGZ require service while an On/Off CoD upgrade is active, the repair can take place without affecting the temporary capacity.

Monitoring

When you activate an On/Off CoD upgrade, an indicator is set in vital product data. This indicator is part of the call-home data transmission, which is sent on a scheduled basis. A time stamp is placed into the call-home data when the facility is deactivated. At the end of each calendar month, the data is used to generate an invoice for the On/Off CoD that was used during that month.

Maintenance

The maintenance price is adjusted as a result of an On/Off CoD activation.

Software

Software Parallel Sysplex license charge (PSLC) clients are billed at the MSU level that is represented by the combined permanent and temporary capacity. All PSLC products are billed at the peak MSUs that are enabled during the month, regardless of usage. Clients with WLC licenses are billed by product at the highest four-hour rolling average for the month. In this instance, temporary capacity does not increase the software bill until that capacity is allocated to LPARs and used.

Results from the STSI instruction reflect the current permanent and temporary CPs. For more information, see “Store System Information instruction” on page 358.

8.6 z/OS Capacity Provisioning

This section describes how z/OS Capacity Provisioning can help you manage the addition of capacity to a server to handle workload peaks.

z/OS Capacity Provisioning is delivered as part of the z/OS MVST™ Base Control Program (BCP).

Capacity Provisioning includes the following components:

- ▶ Capacity Provisioning Manager (Provisioning Manager)
- ▶ Capacity Provisioning Management Console, available in the IBM z/OS Management Facility
- ▶ Sample data sets and files

The Provisioning Manager monitors the workload on a set of z/OS systems and organizes the provisioning of extra capacity to these systems when required. You define the systems to be observed in a domain configuration file.

The details of extra capacity and the rules for its provisioning are stored in a policy file. These two files are created and maintained through the Capacity Provisioning Management Console (CPMC).The operational flow of Capacity Provisioning is shown in Figure 8-10 on page 345.

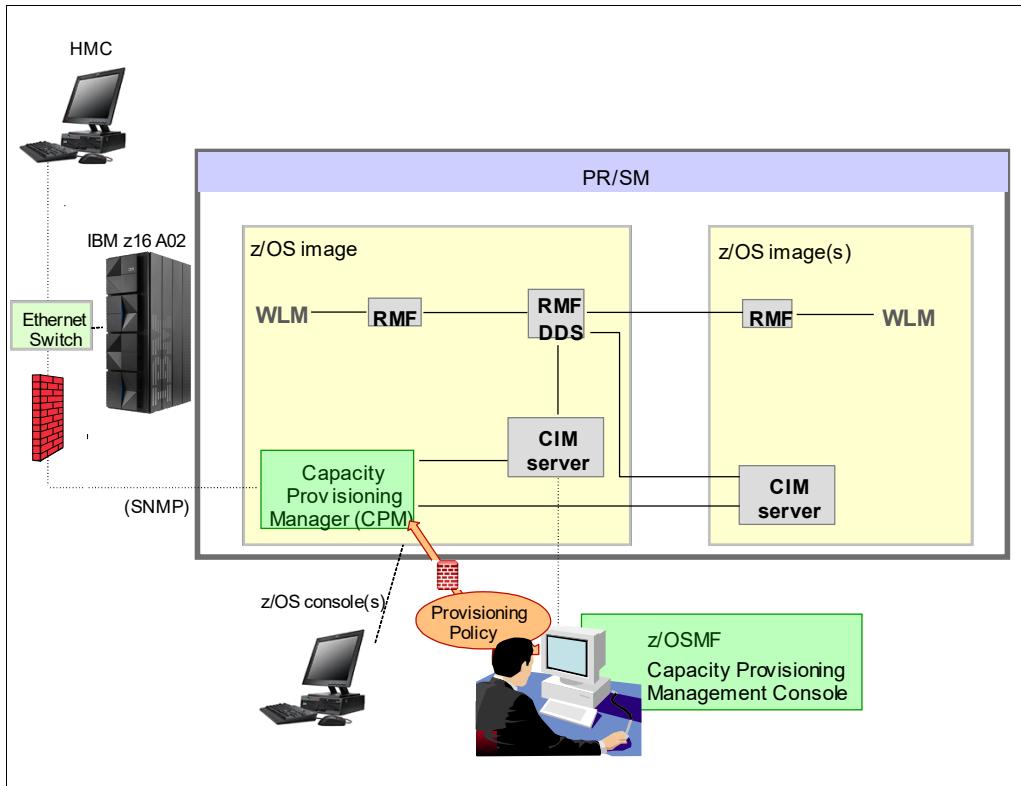


Figure 8-10 The capacity provisioning process and infrastructure

The z/OS WLM manages the workload by goals and business importance on each z/OS system. WLM metrics are available through existing interfaces, and are reported through IBM Resource Measurement Facility (RMF) Monitor III, with one RMF gatherer for each z/OS system.

Sysplex-wide data aggregation and propagation occur in the RMF Distributed Data Server (DDS). The RMF Common Information Model (CIM) providers and associated CIM models publish the RMF Monitor III data.

CPM retrieves critical metrics from one or more z/OS systems' CIM structures and protocols. CPM communicates to local and remote SEs and HMCs by using the Simple Network Management Protocol (SNMP).

CPM can see the resources in the individual offering records and the capacity tokens. When CPM activates resources, a check is run to determine whether enough capacity tokens remain for the specified resource to be activated for at least 24 hours. If insufficient tokens remain, no resource from the On/Off CoD record is activated.

If a capacity token is used during an activation that is driven by the CPM, the corresponding On/Off CoD record is deactivated prematurely by the system. This process occurs even if the CPM activates this record, or parts of it. However, you do receive warning messages if capacity tokens are close to being fully used.

You receive the messages five days before a capacity token is fully used. The five days are based on the assumption that the consumption is constant for the five days. You must put operational procedures in place to handle these situations. You can deactivate the record manually, allow it occur automatically, or replenish the specified capacity token by using the Resource Link application.

The Capacity Provisioning Management Console (CPMC) is a console that administrators use to work with provisioning policies and domain configurations and to monitor the status of a Provisioning Manager. The management console is implemented by the Capacity Provisioning task in the IBM z/OS Management Facility (z/OSMF). z/OSMF provides a framework for managing various aspects of a z/OS system through a web browser interface.

Capacity Provisioning Domain

The provisioning infrastructure is managed by the CPM through the Capacity Provisioning Domain (CPD), which is controlled by the Capacity Provisioning Policy (CPP). The CPD is shown in Figure 8-11.

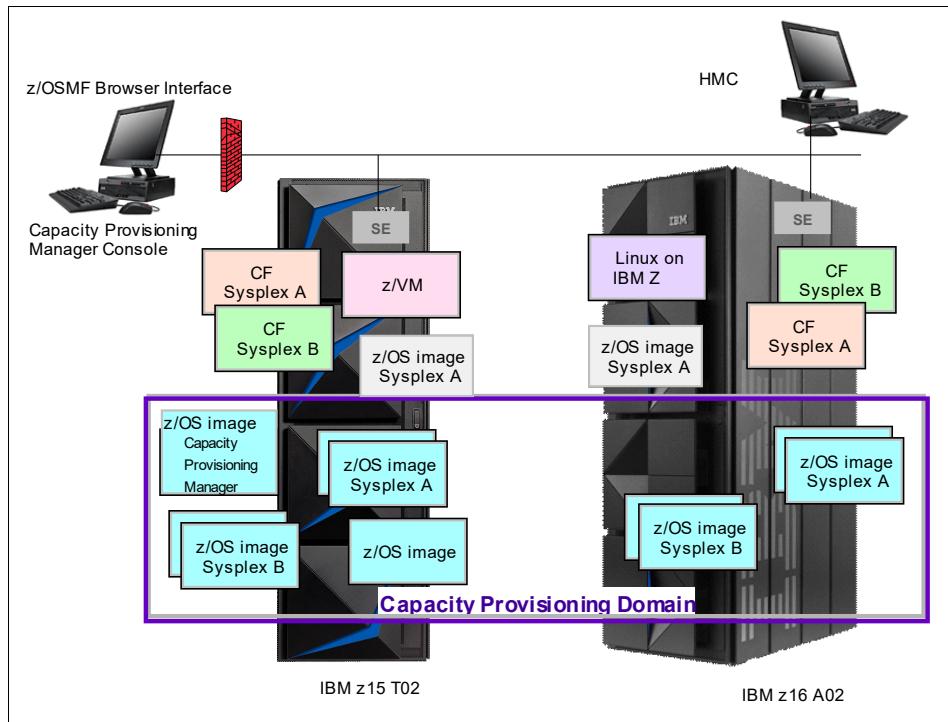


Figure 8-11 Capacity Provisioning Domain

The CPD configuration defines the CPCs and z/OS systems that are controlled by an instance of the CPM. One or more CPCs, sysplexes, and z/OS systems can be defined into a domain. Although sysplexes and CPCs do not have to be contained in a domain, they must not belong to more than one domain.

Each domain has one active capacity provisioning policy.

CPM operates in the following modes, which allows four different levels of automation:

- ▶ Manual mode

Use this command-driven mode when no CPM policy is active.

- ▶ Analysis mode

In analysis mode, CPM processes capacity-provisioning policies and informs the operator when a provisioning or deprovisioning action is required according to policy criteria.

Also, the operator determines whether to ignore the information or to manually upgrade or downgrade the system by using the HMC, SE, or available CPM commands.

- ▶ Confirmation mode

In this mode, CPM processes capacity provisioning policies and interrogates the installed temporary offering records. Every action that is proposed by the CPM must be confirmed by the operator.

- ▶ Autonomic mode

This mode is similar to the confirmation mode, but no operator confirmation is required.

Several reports are available in all modes that contain information about the workload, provisioning status, and the rationale for provisioning guidelines. User interfaces are provided through the z/OS console and the CPMC application.

The provisioning policy defines the circumstances under which more capacity can be provisioned (when, which, and how). The criteria features the following elements:

- ▶ A time condition is when provisioning is allowed:
 - Start time indicates when provisioning can begin.
 - Deadline indicates that provisioning of more capacity is no longer allowed.
 - End time indicates that deactivation of capacity must begin.
- ▶ A workload condition is which work qualifies for provisioning. It can have the following parameters:
 - The z/OS systems that can run eligible work.
 - The importance filter indicates eligible service class periods, which are identified by WLM importance.
 - Performance Index (PI) criteria:
 - Activation threshold: PI of service class periods must exceed the activation threshold for a specified duration before the work is considered to be suffering.
 - Deactivation threshold: PI of service class periods must fall below the deactivation threshold for a specified duration before the work is considered to no longer be suffering.
 - Included service classes are eligible service class periods.
 - Excluded service classes are service class periods that must not be considered.

Tip: If no workload condition is specified, the full capacity that is described in the policy is activated and deactivated at the start and end times that are specified in the policy.

- ▶ Provisioning scope is how much more capacity can be activated and is expressed in MSUs.

The number of zIIPs must be one specification per CPC that is part of the CPD and are specified in MSUs.

The maximum provisioning scope is the maximum extra capacity that can be activated for all the rules in the CPD.

In the specified time interval, the provisioning rule is that up to the defined extra capacity can be activated if the specified workload is behind its objective.

The rules and conditions are named and stored in the Capacity Provisioning Policy.

For more information about z/OS Capacity Provisioning functions, see *z/OS MVS Capacity Provisioning User's Guide*, SC34-2661.

Planning considerations for using automatic provisioning

Although only one On/Off CoD offering can be active at any one time, several On/Off CoD offerings can be present on the system. Changing from one to another requires stopping the active one before the inactive one can be activated. This operation decreases the current capacity during the change.

The provisioning management routines can interrogate the installed offerings, their content, and the status of the content of the offering. To avoid the decrease in capacity, create only one On/Off CoD offering on the system by specifying the maximum allowable capacity. The CPM can then, when an activation is needed, activate a subset of the contents of the offering sufficient to satisfy the demand. If more capacity is needed later, the Provisioning Manager can activate more capacity up to the maximum allowed increase.

Multiple offering records can be pre-staged on the SE hard disk. Changing the content of the offerings (if necessary) is also possible.

Remember: CPM controls capacity tokens for the On/Off CoD records. In a situation where a capacity token is used, the system deactivates the corresponding offering record. Therefore, you must prepare routines for catching the warning messages about capacity tokens being used, and have administrative procedures in place for such a situation.

The messages from the system begin five days before a capacity token is fully used. To avoid capacity records being deactivated in this situation, replenish the necessary capacity tokens before they are used.

The Capacity Provisioning Manager operates based on Workload Manager (WLM) indications, and the construct that is used is the Performance Index (PI) of a service class period. It is important to select service class periods that are appropriate for the business application that needs more capacity. For example, the application in question might be running through several service class periods, where the first period is the important one. The application might be defined as importance level 2 or 3, but might depend on other work that is running with importance level 1. Therefore, it is important to consider which workloads to control and which service class periods to specify.

8.7 Capacity for Planned Event

Note: The Capacity Planned Event feature (6833) can no longer be ordered for a new IBM BM z16 A02 or IBM z16 AGZ, but, if installed on the base system, the record will be brought forward during an upgrade into [IBM z16 A02](#) and [IBM z16 AGZ](#) (via the Support Element Save/Restore process). Also, the CPE record cannot be replenished through e-config or Resource Link.

Flexible Capacity for Cyber Resiliency is a new Capacity on Demand (CoD) offering available on IBM z16 A01 [IBM z16 A02](#) and [IBM z16 AGZ](#) machines that allows processing capacity flexibility between an organization's primary site and alternate data centers.

Flexible Capacity for Cyber Resiliency is designed to provide increased flexibility and control to organizations that want to shift production capacity between participating IBM IBM z16 A01, IBM z16 A02 and IBM z16 AGZ at different sites. The capacity of any engine type can be shifted up to 12 times a year and stay at the target machine for up to 12 months after the flexible capacity record activation on the target machine. Capacity shifts can be done under full client control without IBM intervention and can be fully automated using IBM GDPS automation tools.

Flexible Capacity for Cyber Resiliency supports a broad set of scenarios and can be combined with other IBM On-Demand offerings. For additional informations see the Redpaper *IBM Z Flexible Capacity for Cyber Resiliency*, REDP-5702.

8.8 Capacity Backup

CBU provides reserved emergency backup processor capacity for unplanned situations in which capacity is lost in another part of your enterprise. It allows you to recover by adding the reserved capacity on a designated IBM zSystems.

CBU is the quick, temporary activation of PUs:

- ▶ For up to 90 contiguous days, for a loss of processing capacity as a result of an emergency or disaster recovery situation.
- ▶ For 10 days, for testing your disaster recovery procedures or running the production workload. This option requires that IBM zSystems workload capacity that is equivalent to the CBU upgrade capacity is shut down or otherwise made unusable during the CBU test.³

Important: CBU is for disaster and recovery purposes only. It *cannot* be used for peak workload management or for a planned event.

8.8.1 Ordering

The CBU process allows for CBU to activate CPs, ICFs, zIIPs, and IFLs. To use the CBU process, a CBU enablement feature (FC 9910) must be ordered and installed. You must order the quantity and type of PU that you require by using the following feature codes:

- ▶ FC 6805: More CBU test activations

³ All new CBU contract documents contain new CBU test terms to allow execution of production workload during CBU test. CBU clients must sign the IBM client Agreement Amendment for IBM zSystems Capacity Backup Upgrade Tests (US form #Z125-8145).

- ▶ FC 6817: Total CBU years ordered
- ▶ FC 6818: CBU records that are ordered
- ▶ FC 6820: Single CBU CP-year
- ▶ FC 6821: 25 CBU CP-year
- ▶ FC 6822: Single CBU IFL-year
- ▶ FC 6823: 25 CBU IFL-year
- ▶ FC 6824: Single CBU ICF-year
- ▶ FC 6825: 25 CBU ICF-year
- ▶ FC 6828: Single CBU zIIP-year
- ▶ FC 6829: 25 CBU zIIP-year
- ▶ FC 6832: CBU replenishment

The CBU entitlement record (FC 6818) contains an expiration date that is established at the time of the order. This date depends on the quantity of CBU years (FC 6817). You can extend your CBU entitlements through the purchase of more CBU years.

The number of FC 6817 per instance of FC 6818 remains limited to five. Fractional years are rounded up to the nearest whole integer when calculating this limit.

If two years and eight months exist before the expiration date at the time of the order, the expiration date can be extended by no more than two years. One test activation is provided for each CBU year that is added to the CBU entitlement record.

FC 6805 allows for ordering more tests in increments of one. The maximum number of tests that is allowed is 15 for each FC 6818.

The PUs that can be activated by CBU come from the available unassigned PUs on any installed CPC drawer. The maximum number of CBU features that can be *ordered* is 68. The number of features that can be *activated* is limited by the number of unused PUs on the system.

The ordering system allows for over-configuration in the order. You can order up to 68 CBU features, regardless of the current configuration. However, at activation, only the capacity that is installed can be activated. At activation, you can decide to activate only a subset of the CBU features that are ordered for the system.

Subcapacity makes a difference in the way that the CBU features are completed. On the full-capacity models, the CBU features indicate the amount of extra capacity that is needed. If the amount of necessary CBU capacity is equal to four CPs, the CBU configuration is four CBU CPs.

The number of CBU CPs must be equal to or greater than the number of CPs in the base configuration. Also, all of the CPs in the CBU configuration must have the same capacity setting. For example, if the base configuration is a two-way D02, providing a CBU configuration of a four-way of the same capacity setting requires two CBU feature codes.

If the required CBU capacity changes the capacity setting of the CPs, going from model capacity identifier D02 to a CBU configuration of a four-way E04 requires four CBU feature codes: two to upgrade from a D02 to a E02 and two to upgrade from an E02 to a E04.

If the capacity setting of the CPs is changed, more CBU features are required, not more physical PUs. Therefore, your CBU contract requires more CBU features when the capacity setting of the CPs is changed.

CBU can add CPs through LICCC only, and the IBM z16 A02 and IBM z16 AGZ A01 must have the correct number of installed CPC drawers to allow the required upgrade. CBU can

change the model capacity identifier to a *higher* value than the base setting, but does not change the system model. The CBU feature cannot *decrease* the capacity setting.

A CBU contract must be in place before the special code that enables this capability can be installed on the system. CBU features can be added to an IBM z16 A02 or IBM z16 AGZ non disruptively. For each system enabled for CBU, the authorization to use CBU is available for 1 - 5 years.

The alternative configuration is activated *temporarily*, and provides more capacity than the system's original, *permanent* configuration. At activation time, determine the capacity that you require for that situation. You can decide to activate only a subset of the capacity that is specified in the CBU contract.

The base system configuration must have sufficient memory and channels to accommodate the potential requirements of the large CBU target system. Ensure that all required functions and resources are available on the backup systems. These functions include CF LEVELs for coupling facility partitions, memory, and cryptographic functions, and connectivity capabilities.

When the emergency is over (or the CBU test is complete), the system must be returned to its original configuration. The CBU features can be deactivated at any time before the expiration date. Failure to deactivate the CBU feature before the expiration date can cause the system to downgrade resources gracefully to the original configuration. The system does not deactivate dedicated engines, or the last of in-use shared engines.

Planning: CBU for processors provides a concurrent upgrade. This upgrade can result in more enabled processors, changed capacity settings that are available to a system configuration, or both. You can activate a subset of the CBU features that are ordered for the system. Therefore, more planning and tasks are required for *nondisruptive* logical upgrades. For more information, see "Guidelines to avoid disruptive upgrades" on page 360.

For more information, see the *Capacity on Demand User's Guide*, SC28-6846.

8.8.2 CBU activation and deactivation

The activation and deactivation of the CBU function is your responsibility and does not require the onsite presence of IBM SSRs. The CBU function is activated or deactivated concurrently from the HMC by using the API. On the SE, CBU is activated by using the Perform Model Conversion task or through the API. The API enables task automation.

CBU activation

CBU is activated from the SE by using the HMC and SSO to the SE, by using the Perform Model Conversion task, or through automation by using the API on the SE or the HMC. During a real disaster, use the Activate CBU option to activate the 90-day period.

Image upgrades

After CBU activation, the [IBM z16 A02](#) and [IBM z16 AGZ](#) can have more capacity, more active PUs, or both. The extra resources go into the resource pools and are available to the LPARs. If the LPARs must increase their share of the resources, the LPAR weight can be changed or the number of logical processors can be concurrently increased by configuring reserved processors online. The operating system must concurrently configure more processors online. If necessary, more LPARs can be created to use the newly added capacity.

CBU deactivation

To deactivate the CBU, the extra resources must be released from the LPARs by the operating systems. In some cases, this process involves varying the resources offline. In other cases, it can mean shutting down operating systems or deactivating LPARs. After the resources are released, the same facility on the HMC/SE is used to turn off CBU. To deactivate CBU, select the **Undo temporary upgrade** option from the Perform Model Conversion task.

CBU testing

Test CBUs are provided as part of the CBU contract. CBU is activated from the SE by using the Perform Model Conversion task. Select the test option to start a 10-day test period. A standard contract allows one test per CBU year. However, you can order more tests in increments of one up to a maximum of 15 for each CBU order.

Tip: The CBU test activation is done the same way as the real activation; that is, by using the same Perform a Model Conversion task and selecting the **Temporary upgrades** option. The HMC windows were changed to avoid accidental real CBU activations by setting the test activation as the default option.

The test CBU must be deactivated in the same way as the regular CBU. Failure to deactivate the CBU feature before the expiration date can cause the system to degrade gracefully back to its original configuration. The system does not deactivate dedicated engines or the last in-use shared engine.

CBU example

An example of a CBU operation is shown in Figure 8-12. The permanent configuration is a B02, and a record contains four CP CBU features. During an activation, many target configurations are available. With four CP CBU features, you can add up to 4 CPs within the same MCI, which enables the activation of a B03, B04, B05, or a B06 (the blue path).

Alternatively, two CP CBU features can be used to change the MCI (in the example from a B02 to a E02) and then add the remaining two CP CBU features to upgrade to a E04 (the red path).

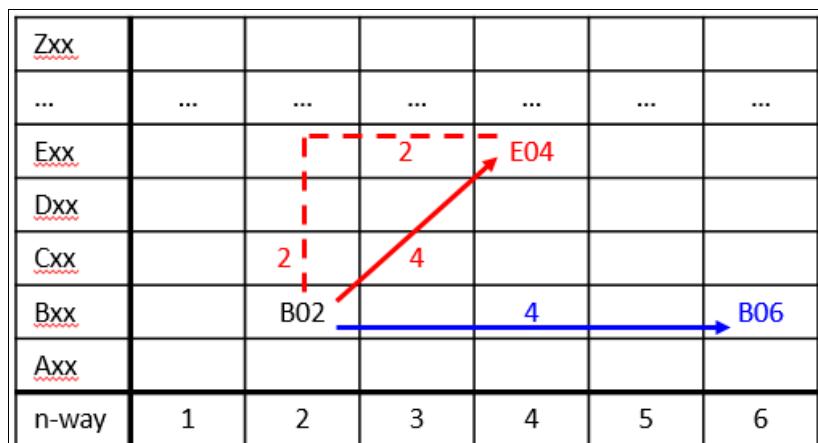


Figure 8-12 CBU example

8.8.3 Automatic CBU enablement for GDPS

The IBM Geographically Dispersed Parallel Sysplex (GDPS) enables automatic management of the PUs that are provided by the CBU feature during a system or site failure. Upon detection of a site failure or planned disaster test, GDPS concurrently adds CPs to the systems in the take-over site to restore processing power for mission-critical production workloads. GDPS automation runs the following tasks:

- ▶ The analysis that is required to determine the scope of the failure. This process minimizes operator intervention and the potential for errors.
- ▶ Automates authentication and activation of the reserved CPs.
- ▶ Automatically restarts the critical applications after reserved CP activation.
- ▶ Reduces the outage time to restart critical workloads from several hours to minutes.

The GDPS service is for z/OS only, or for z/OS in combination with Linux on Z.

8.9 Flexible Capacity Cyber Resiliency

Flexible Capacity for Cyber Resiliency is a new Capacity on Demand (CoD) offering available on IBM z16 A02 and IBM z16 AGZ machines that allows processing capacity flexibility between an organization's primary site and alternate data centers. This section summarizes the main features of Flexible Capacity for Cyber Resiliency. For more information see the Redpaper *IBM Z Flexible Capacity for Cyber Resiliency*, REDP-5702.

Flexible Capacity for Cyber Resiliency can be ordered by contacting your IBM hardware sales representative. The offering requires an order placed against each serial number (SN) involved in capacity transfer with one record per SN.

The following feature codes (FC) are introduced:

- ▶ Flexible Capacity Authorization (#9933),
- ▶ Flexible Capacity Record (#0376),
- ▶ Billing feature codes (#0317 through #0322, and #0378 through #0386).

Installation and setup: The new Flexible Capacity Record is installed and set up on each participating IBM z16 A01, IBM z16 A02 or IBM z16 AGZ.

- ▶ On the IBM z16 A01, IBM z16 A02 or IBM z16 AGZ source system, the permanent capacity is unassigned to the base level
- ▶ The new Flexible Capacity Record is installed and activated on the IBM z16 A01, IBM z16 A02 and IBM z16 AGZ source system to restore capacity back to the purchased level.
- ▶ On the IBM z16 A01, IBM z16 A02 or IBM z16 AGZ target system(s), the new Flexible Capacity Record enables clients to bring the capacity up to the level of the production system when activated. The Flexible Capacity Record remains inactive until capacity is shifted from the base system to the target system(s).
- ▶ After deactivating the Flexible Capacity Record on the base system, the capacity active through Flexible Capacity Transfer records on the target system(s) should not exceed the capacity active on the base system before the swap.
- ▶ If GDPS is used to automate the shift, it must be set up with the correct LIC records to add capacity in the target system(s) and remove capacity in the base system(s) site.

Site swap example

Figure 8-13 on page 354 through Figure 8-15 on page 355 show a sequence of events to move capacity from one site to another using the Flexible Capacity features of the IBM z16 A02 or IBM z16 AGZ.

A Flexible Capacity record is installed on the machines at each site. The machine at Site A has the base capacity configured at the same level as the high water mark of the Site B machine and has flex capacity activated to its HWM.

The flex capacity is then added to the machine at Site B, bringing it up to the Site A HWM and workload is transferred.

Once all workload has been transferred (within 24 hours) the capacity of the Site A machine is reduced to the base level. Workload can continue to run at Site B for up to a year.

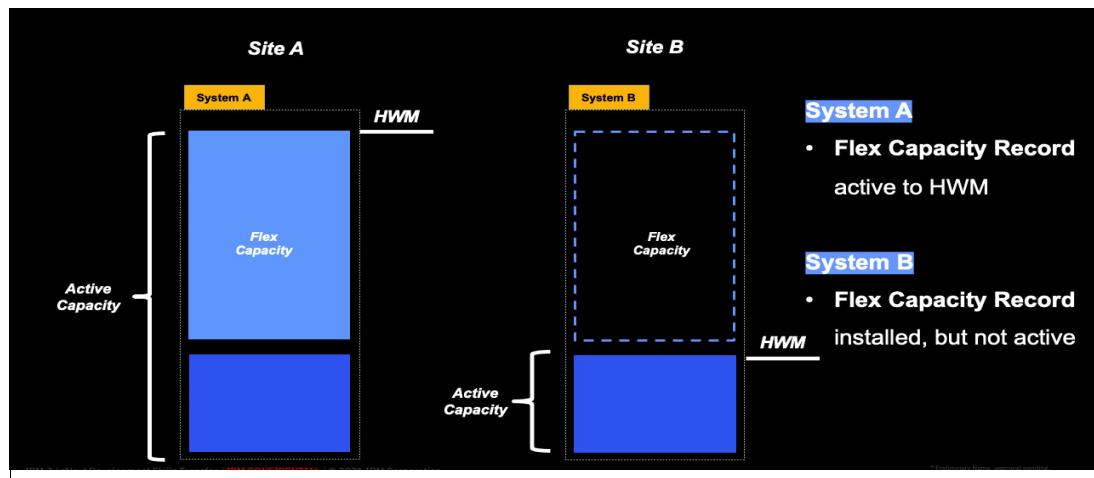


Figure 8-13 Flexible Capacity Normal Operation

Capacity transfer is shown in Figure 8-14. Flexible Capacity Record is ACTIVE in both sites for up to 24 hours.



Figure 8-14 Flexible Capacity record at Site B is activated and workload transferred

After the site swap (capacity active in both sites for up to 24 hours), workload can stay in Site B for up to one year. (see Figure 8-15 on page 355).

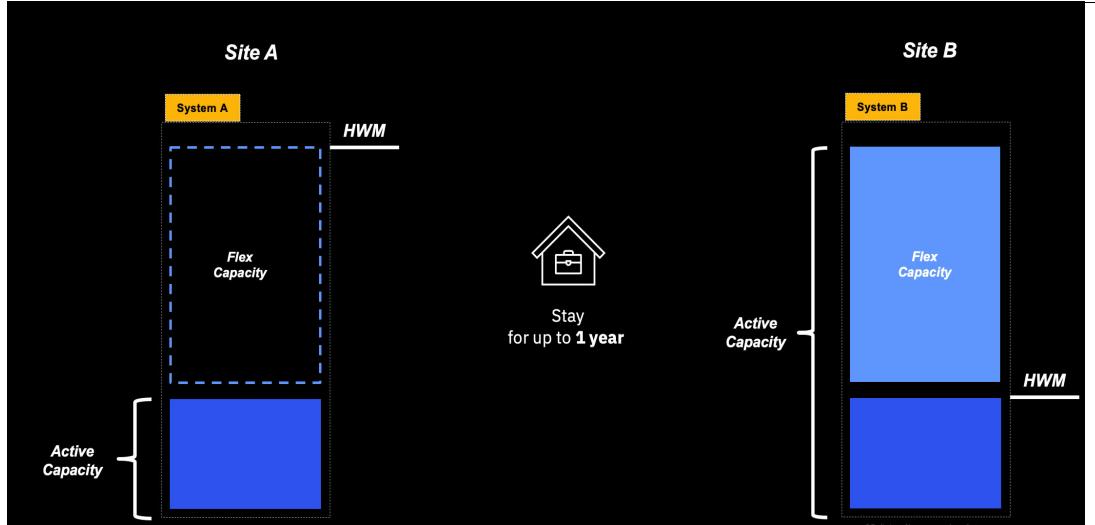


Figure 8-15 Swap and stay at Site B for up to 1 year

For more information please see the Redpaper *IBM Z Flexible Capacity for Cyber Resiliency*, REDP-5702-00.

8.10 Planning for nondisruptive upgrades

Continuous availability is an important requirement for clients, and planned outages are no longer acceptable. Although Parallel Sysplex clustering technology is the best continuous availability solution for z/OS environments, nondisruptive upgrades within a single system can avoid system outages and cover non-z/OS operating systems.

IBM z16 A02 and IBM z16 AGZ allow *concurrent* upgrades, which means that dynamically adding capacity to the system is possible. If the operating system images that run on the upgraded system do not require disruptive tasks to use the new capacity, the upgrade is also *nondisruptive*. This process avoids power-on resets (POR), LPAR deactivation, and IPLs.

If the concurrent upgrade is intended to satisfy an *image* upgrade to an LPAR, the operating system that is running in this partition must concurrently configure more capacity online. z/OS operating systems include this capability. z/VM can concurrently configure new processors and I/O devices online, and memory can be dynamically added to z/VM partitions.

If the concurrent upgrade is intended to satisfy the need for more operating system images, more LPARs can be created *concurrently* on the **IBM z16 A02 and IBM z16 AGZ**. These LPARs include all resources that are needed. These extra LPARs can be activated concurrently.

These enhanced configuration options are available through the HSA, which is an IBM reserved area in system memory.

Linux operating systems, in general, cannot add more resources concurrently. However, Linux, and other types of virtual machines that run under z/VM, can benefit from the z/VM capability to nondisruptively configure more resources online (processors and I/O).

With z/VM, Linux guests can manipulate their logical processors by using the Linux CPU hotplug daemon. The daemon can start and stop logical processors that are based on the

Linux *load average* value. The daemon is available in Linux SLES 10 SP2 and later, and in Red Hat Enterprise Linux (RHEL) V5R4 and up.

8.10.1 Components

The following components can be added, depending on the considerations as described in this section:

- ▶ PUs
- ▶ Memory
- ▶ I/O
- ▶ Cryptographic adapters
- ▶ Special features

PUs

CPs, ICFs, zIIPs, and IFLs can be added concurrently to an [IBM z16 A02](#) and [IBM z16 AGZ](#) if unassigned PUs are available on any installed CPC drawer. The [IBM z16 A02](#) and [IBM z16 AGZ](#) allow the concurrent addition of a second CPC drawer (for IBM z16 AGZ if the CPC reserve FC 2332 is installed).

Tip: The 2:1 zIIP:CP ratio restriction has been removed for IBM z16 A01, IBM z16 A02 and IBM z16 AGZ as of May 2023.

If necessary, more LPARs can be created concurrently to use the newly added processors.

The Coupling Facility Control Code (CFCC) can also configure more processors online to coupling facility LPARs by using the CFCC image operations window.

Memory

Memory can be added concurrently up to the physical installed memory limit. More CPC drawers can be installed concurrently, which allows further memory upgrades by LICCC, and enables memory capacity on the new CPC drawers.

By using the previously defined reserved memory, z/OS operating system images, and z/VM partitions, you can dynamically configure more memory online. This process allows nondisruptive memory upgrades. Linux on IBM Z supports Dynamic Storage Reconfiguration.

I/O

I/O features can be added concurrently if all the required infrastructure (I/O slots and PCIe Fanouts) is present in the configuration. PCIe+ I/O drawers can be added concurrently without planning if free space is available in one of the frames and the configuration permits.

Dynamic I/O configurations are supported by certain operating systems (z/OS and z/VM), which allows nondisruptive I/O upgrades. Dynamic I/O reconfiguration on a stand-alone coupling facility system is also possible using the Dynamic I/O activation for stand-alone CF CPCs features.

Cryptographic adapters

Crypto Express8S features can be added concurrently if all the required infrastructure is in the configuration.

Special features

Special features such as zHyperlink, Coupling Express2 LR, and RoCE features can be added concurrently if all infrastructure is available in the configuration.

8.10.2 Concurrent upgrade considerations

By using an MES upgrade, On/Off CoD, CBU, or CPE, an BM z16 A02 or IBM z16 AGZ can be upgraded concurrently from one model to another (temporarily or permanently).

Enabling and using the extra processor capacity is not apparent to most applications. However, certain programs depend on processor model-related information, such as ISV products. Consider the effect on the software that is running on an BM z16 A02 or IBM z16 AGZ when you perform any of these configuration upgrades.

Processor identification

The following instructions are used to obtain processor information:

- ▶ Store System Information (STSI) instruction

The STSI instruction can be used to obtain information about the current execution environment and any processing level below the current environment. It can be used to obtain processor model and model capacity identifier information from the basic machine configuration form of the system information block (SYSIB). It supports concurrent upgrades and is the recommended way to request processor information.

- ▶ Store CPU ID (STIDP) instruction

STIDP returns information that identifies the execution environment, system serial number, and machine type.

Note: To ensure unique identification of the configuration of the issuing CPU, use the STSI instruction specifying basic machine configuration (SYSIB 1.1.1).

Store System Information instruction

The format of the basic machine configuration SYSIB that is returned by the STSI instruction is shown in Figure 8-16. The STSI instruction returns the model capacity identifier for the permanent configuration and the model capacity identifier for any temporary capacity. This data is key to the functioning of CoD offerings.

0	P	Reserved	M	T	IBM	CCR	CAI
1					Reserved		
8					Manufacturer		
12					Type		
13					Reserved		
16					Model-Capacity Identifier		
20					Sequence Code		
24					Plant of Manufacture		
25					Model		
29					Model-Permanent-Capacity Identifier		
33					Model-Temporary-Capacity Identifier		
37					Model-Capacity Rating		
38					Model-Permanent-Capacity Rating		
39					Model-Temporary-Capacity Rating		
40	Type 1 Pctg.	Type 2 Pctg.	Type 3 Pctg.	Type 4 Pctg.			
41	Type 5 Pctg.				Reserved		
42					Nominal Model-Capacity Rating		
43					Nominal Model-Permanent-Capacity Rating		
44					Nominal Model-Temporary-Capacity Rating		
45							
1023					Reserved		
	0	7	8	16	24	31	

Figure 8-16 Format of system-information block (SYSIB)

The model capacity identifier contains the base capacity, On/Off CoD, and CBU. The Model Permanent Capacity Identifier and the Model Permanent Capacity Rating contain the base capacity of the system. The Model Temporary Capacity Identifier and Model Temporary Capacity Rating contain the base capacity and On/Off CoD.

For more information about the STSI instruction, see *z/Architecture Principles of Operation*, SA22-7832.

Store CPU ID (STIDP) instruction

The STIDP instruction returns information about the processor type, serial number, and LPAR identifier, as shown in Figure 8-17.

Environment	Configuration Identification		
0	8		31
	Machine-Type Number	F	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
32		48 49	63

Figure 8-17 STIDP Information

Consider the following points:

- ▶ Bits 0 - 7:
 - For a program that is run by an IBM machine in a level-1 configuration (basic machine mode), or for a program being run by a level-2 configuration (in a logical partition), the environment field contains 00 hex.
 - For a program that is run natively by the System z Personal-Development Tool, the environment field contains C1 hex or D3 hex.
 - For a program that is run by a level-3 configuration (a virtual machine, such as a z/VM guest), the environment field contains FF hex.
- ▶ Bit positions 8 - 31

Contains six hexadecimal digits. The right-most of these digits can represent the machine's serial number.
- ▶ Bit positions 32 - 47

Contains an unsigned packed-decimal number that identifies the machine type of the CPU.
- ▶ Bit position 48

Specifies the format of the first two hexadecimal digits of the configuration-identification field.
- ▶ Bit positions 49 - 63 are reserved and stored as zeros.

For more information about the STIDP instruction, see *z/Architecture Principles of Operation*, SA22-7832.

Planning for nondisruptive upgrades

Online permanent upgrades, On/Off CoD, CBU, and CPE can be used to upgrade an IBM z16 A02 and IBM z16 AGZ concurrently. However, certain situations require a disruptive task to enable capacity that was recently added to the system. Some of these situations can be avoided if planning is done in advance. Planning ahead is a key factor for nondisruptive upgrades. In a multi-site high-availability configuration, another option is the use of Flexible Capacity for Cyber Resilience to move workload to another site while hardware maintenance is performed.

Disruptive upgrades are performed for the following reasons:

- ▶ LPAR memory upgrades when reserved storage was not previously defined are disruptive to image upgrades. z/OS and z/VM support this function.
- ▶ An I/O upgrade when the operating system cannot use the dynamic I/O configuration function is disruptive to that partition. Linux, z/VSE, and z/TPF do not support dynamic I/O configuration.

You can minimize the need for these outages by carefully planning and reviewing “Guidelines to avoid disruptive upgrades” on page 360.

Guidelines to avoid disruptive upgrades

Based on the reasons for disruptive upgrades (see “Planning for nondisruptive upgrades” on page 360), you can use the following guidelines to avoid or at least minimize these situations, which increases the chances for nondisruptive upgrades:

- ▶ By using an SE function that is called Logical Processor add, which is under Operational Customization tasks, CPs and zIIPs can be added concurrently to a running partition. The CP and zIIP and initial or reserved number of processors can be changed dynamically.
- ▶ The operating system that runs in the targeted LPAR must support the dynamic addition of resources and to configure processors online. The total number of defined and reserved CPs cannot exceed the number of CPs that are supported by the operating system. z/OS V2.R5, V2.R4, and V2.R3 support 200 PUs per z/OS LPAR in non-SMT mode and 128 PUs per z/OS LPAR in SMT mode. For both, the PU total is the sum of CPs and zIIPs. z/VM supports up to 64 processors.
- ▶ Configure reserved storage to LPARs.

Configuring reserved storage for all LPARs before their activation enables them to be nondisruptive upgraded. The operating system that is running in the LPAR must configure memory online. The amount of reserved storage can be greater than the CPC drawer threshold limit, even if no other CPC drawer is installed. With IBM z16 A02 or IBM z16 AGZ, the current partition storage limit is 4TB for z/OS V2.R3 & V2.R4 and 16TB for z/OS V2.R5 and later. z/VM V7.R1 supports 2 TB and z/VM V7.R2 supports 4TB memory partitions.

Considerations when installing second CPC drawer

During an upgrade, a second CPC drawer can be installed concurrently if pre-planned. Depending on the system configuration, a fanout rebalancing might be needed for availability reasons.

8.11 Summary of Capacity on-Demand offerings

The CoD infrastructure and its offerings are based on client requirements for more flexibility, granularity, and better business control over the IBM zSystems infrastructure, operationally, and financially.

After the offerings are installed on the BM z16 A02 or IBM z16 AGZ SE, they can be activated at any time at the client's discretion. No intervention by IBM or IBM personnel is necessary. In addition, the activation of CBU does not require a password.

The [IBM z16 A02](#) and [IBM z16 AGZ](#) can have up to eight offerings installed at the same time, with the limitation that only *one* of them can be an On/Off CoD offering. The others can be any combination. The installed offerings can be activated fully or partially, and in any sequence and any combination. The offerings can be controlled manually through tasks on the HMC, or programmatically through a number of APIs. IBM applications, ISV programs, and client-written applications can control the use of the offerings.

Resource usage (and therefore, financial exposure) can be controlled by using capacity tokens in the On/Off CoD offering records.

The CPM is an example of an application that uses the CoD APIs to provision On/Off CoD capacity that is based on the requirements of the workload. The CPM cannot control other offerings.

For more information about any of the topics in this chapter, see *Capacity on Demand User's Guide*, SC28-6943.



Reliability, availability, and serviceability

From the Quality perspective, the IBM z16 A02 and IBM z16 AGZ reliability, availability, and serviceability (RAS) design is driven by a set of high-level program RAS objectives. The IBM zSystems platform continues to drive toward Continuous Reliable Operation (CRO) at the single footprint level.

The key objectives, in order of priority, are to ensure data integrity, computational integrity, reduce or eliminate unscheduled outages, reduce scheduled outages, reduce planned outages, and reduce the number of Repair Actions.

This chapter includes the following topics:

- ▶ 9.1, “RAS strategy” on page 364
- ▶ 9.2, “Technology” on page 364
- ▶ 9.3, “Structure” on page 369
- ▶ 9.4, “Reducing complexity” on page 370
- ▶ 9.5, “Reducing touches” on page 370
- ▶ 9.6, “IBM z16 A02 and IBM z16 AGZ availability characteristics” on page 370
- ▶ 9.7, “IBM z16 A02 and IBM z16 AGZ RAS functions” on page 374
- ▶ 9.8, “Enhanced drawer availability” on page 378
- ▶ 9.9, “Concurrent Driver Maintenance” on page 385
- ▶ 9.10, “RAS capability for the HMA and SE” on page 387

9.1 RAS strategy

The RAS strategy is to manage change by learning from previous generations and investing in new RAS function to eliminate or minimize all sources of outages. Enhancements introduced with IBM z15 RAS designs are implemented on the IBM z16 A02 and IBM z16 AGZ design through the introduction of new technology, structure, and requirements. Continuous improvements in RAS are associated with new features and functions to ensure that IBM zSystems deliver exceptional value to clients.

RAS can be accomplished with improved concurrent replace, repair, and upgrade functions for processors, memory, drawers, and I/O. RAS also extends to the nondisruptive capability for installing Licensed Internal Code (LIC) updates. In most cases, a capacity upgrade can be concurrent without a system outage. As an extension to the RAS capabilities, environmental controls are implemented in the system to help reduce power consumption and meet cooling requirements.

The following overriding RAS requirements are principles as shown in Figure 9-1:

- ▶ Inclusion of existing (or equivalent) RAS characteristics from previous generations.
- ▶ Learn from current field issues and addressing the deficiencies.
- ▶ Understand the trend in technology reliability (hard and soft) and ensure that the RAS design points are sufficiently robust.
- ▶ Invest in RAS design enhancements (hardware and firmware) that provide IBM Z and Customer valued differentiation.

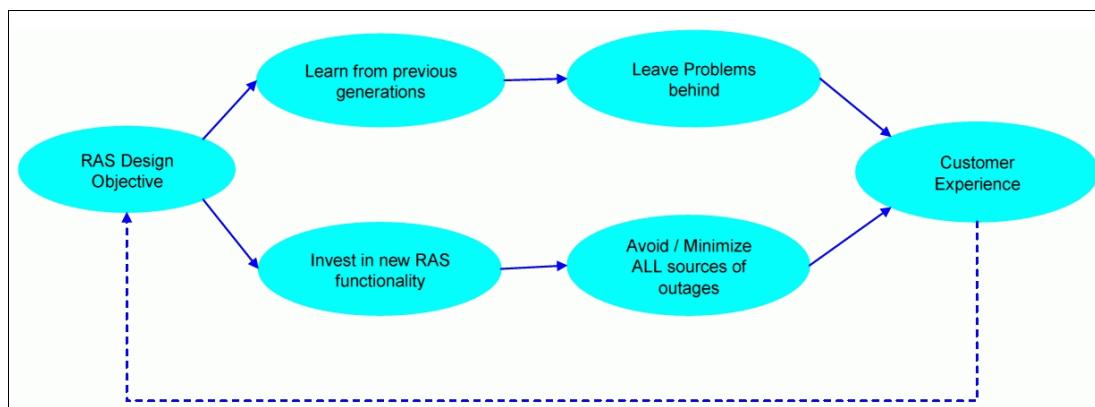


Figure 9-1 RAS design process overview

9.2 Technology

This section introduces some of the RAS features that are incorporated in the IBM z16 A02 and IBM z16 AGZ design.

9.2.1 Processor Unit chip

IBM z16 A02 and IBM z16 AGZ use the Processor Unit (PU) chip with the following technical changes:

- ▶ A processor unit (PU) chip is manufactured using 7nm CMOS FinFET technology featuring EUV lithography, and has eight cores (PUs) per chip (design) running at 4.6 GHz.
- ▶ Each core has private L1 (instruction and data) caches and a semi-private L2 cache, 32 MB in size. The eight cores and L2 caches on the chip communicate through bi-directional high speed on-chip ring and with all SMP, I/O and memory interfaces (see Figure 9-2).

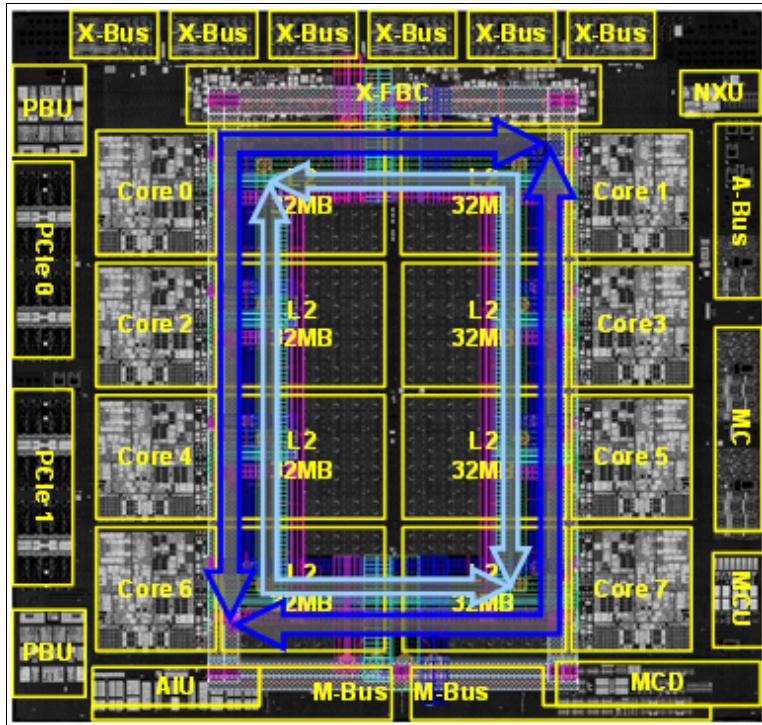


Figure 9-2 PU chip structure (w/ bidirectional ring shown)

- ▶ Two PU chips are packaged on a dual chip module (DCM), shown in Figure 9-3 on page 366. PU chip to PU chip communication is performed through the M-Bus, a high-speed bus and acting as ring-to-ring extension communication at 160 Gbps data rate.
 - ECC on data path and snoop bus
 - Parity on miscellaneous control lines
 - One spare per ~50 data bit bus

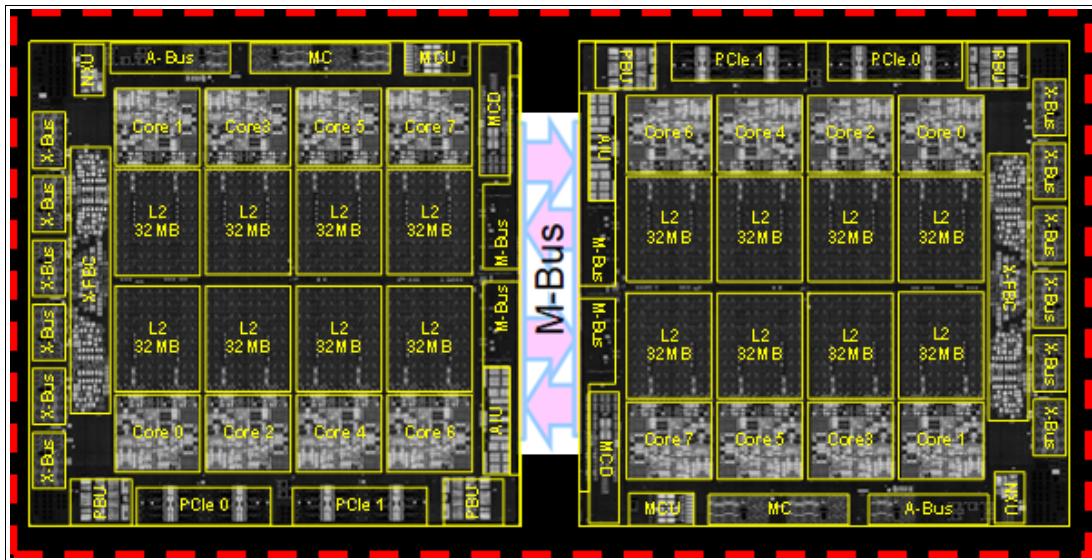


Figure 9-3 IBM z16 A02 and IBM z16 AGZ Dual Chip Module (M-Bus connects the two chips)

- ▶ The processor recovery logic resides on the DCM and has the following RAS characteristics:
 - PU refresh for soft errors
 - Hardened RU (recovery unit) with wordline failure detection
 - Improved latch interleaving (spacing) for soft error resistance
 - Core sparing for hard errors
 - All servers ship with two spare cores
 - Improved long term thresholding
 - Redesigned nest
 - Concurrent repair via CDR (concurrent drawer repair)
 - Concurrent upgrade via LICCC and EDA (enhanced drawer availability)
- ▶ L1 and L1 shadow - L1 shadow is new on IBM z16 A02 and IBM z16 AGZ:
 - Behaves like the L1 (for repairs)
 - Contains changed data
 - All UE¹ checkstop core
 - UE are refetched before acting
 - UE impact dependent on system state
 - Before end OP, IPD w/o Storage validity
 - Before SIE synch, System Damage
- ▶ IBM Z Integrated Accelerator for AI (AIU)
 - AIU is an on-chip AI Accelerator which is new on IBM z16 A02 and IBM z16 AGZ. AIU behaves like a co-processor to process the synchronous instructions but AIU is located in nest. The core control the AIU by issuing instruction (NNPA).
 - The array macro have row and column repair (MD and ABIST)
 - 1MB cache with SECDED² ECC
- ▶ On-chip L2 caches are implemented in dense SRAM. The IBM z16 A02 and IBM z16 AGZ has removed the physical L3 (on-chip for IBM z15) and L4 (on additional storage controller single chip module - SC SCM) and has implemented clustered cache algorithms which

¹ UE - Unrecoverable error

² SECDED - Single error correction double error detection

- provides virtualized L3 (shared victim) and virtual L4 (shared victim) caches (see Figure 9-4).
- ▶ The dedicated L2 cache (dense SRAM) is semi-private to the core, or we can say each core has an associated 32MB slice of the L2 cache (semi-private).
 - Virtual L3 on PU Chip (shared victim 256MB)
 - Virtual L4 on drawer up to 2GB

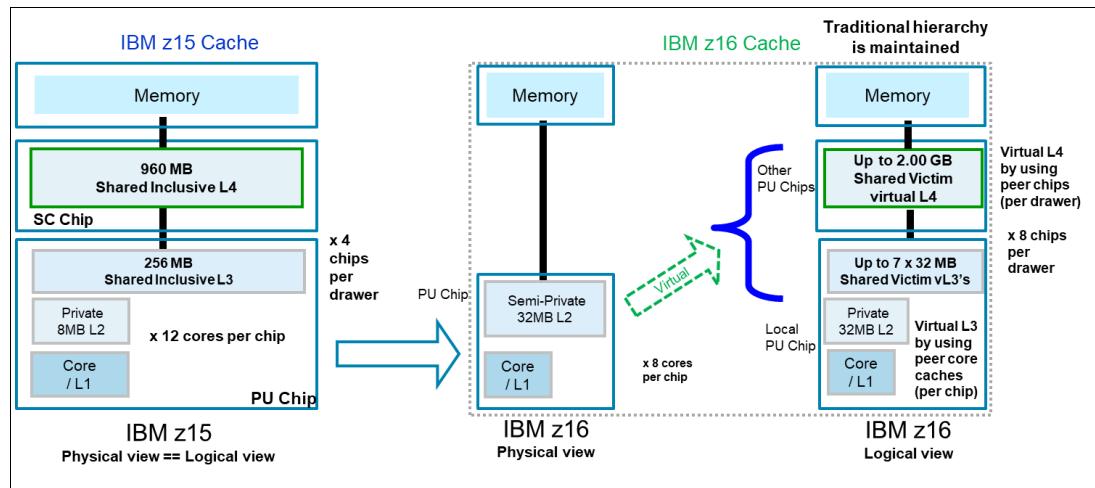


Figure 9-4 Cache hierarchy: IBM z15 vs. IBM z16 A02 and IBM z16 AGZ

- ▶ L2 cache and L2 cache recovery
 - The L2 cache has expected inline Symbol ECC (RAID4) recovery
 - 64:1 interleaving (4:1 physical, 16:1 logical)
 - L2 symbol ECC is inherited by virtual L3 and virtual L4
 - The array macro have row and column repair (Module manufacturing and ABIST)
 - Ring can be fenced from L2 for yield, fences core too (Module manufacturing)
 - If the core checkstops, all 32MB can be used by the system
 - L2 dedicated / shared split is managed in the LRU logic
 - Dynamic macro sparing
 - Four spare macros, one spare for each quarter slice
 - L2 1/8 portion is taken offline when a spare is needed, and none is available (named 7/8 recovery)
 - The core must be spared
 - L2 directory is SECDED ECC protected

IBM z16 A02 and IBM z16 AGZ processor memory and cache structure are shown in Figure 9-5 on page 368. The physical L3 cache (on chip) and System Controller (SC) SCM (physical L4 cache for IBM z15), previously implemented in EDRAM, have been removed and replaced on IBM z16 A02 and IBM z16 AGZ virtual L3 and L4 cache structure.

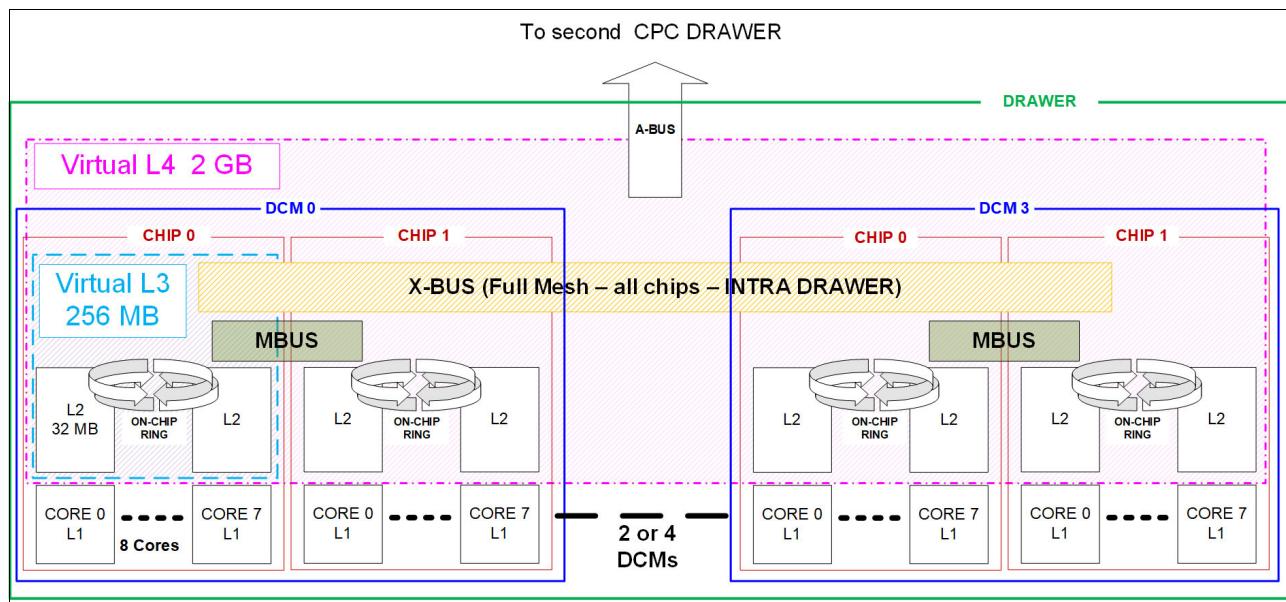


Figure 9-5 IBM z16 A02 and IBM z16 AGZ M/T 3932 fully populated CPC Drawer Cache structure

9.2.2 Main memory

IBM z16 A02 and IBM z16 AGZ memory consist of the following features:

- ▶ Up to 48 DIMM per drawer in two rows (16TB system max):
 - Organized in 8-channel Reed-Solomon RAIM groups
 - 50% reduced RAIM overhead (compared to IBM z15)
 - RAIM protection (similar to IBM z15)
 - Up to 3 chip marks + 1 channel mark
 - DDR4 DRAM with on chip power regulation
 - N+1 voltage regulators
 - Standard Open Memory Interface (OMI); up to 6 OMI per drawer
 - CRC/Retry for soft errors
 - Degrade bus lanes from 4 lanes to 2 lanes on hard error
 - No waiting for all 8 RAIM channels, use first seven to respond
 - RAIM recovery is used to enhance performance without impacting RAS
 - Processor uses data from the first 7 channels to respond
 - Refetch data using all 8 RAIM channels if an error occurs

Note: Flexible Memory feature is not supported on IBM A02 and IBM z16 A02 and IBM z16 AGZ.

- Virtual Flash Memory considerations
 - Each VFM Feature (FC 0644) takes 512 GB of memory. Up to 12 VFM features can be ordered.
 - If VFM is present, it is included in the Flexible memory calculations
- ▶ Concurrent service and upgrade via Concurrent Drawer Repair (CDR) for 2-CPC drawer systems (Max68),

9.2.3 I/O and service

I/O and service consist of the following features:

- ▶ New I/O features on IBM z16 A02 and IBM z16 AGZ:
 - FICON Express 32S
 - Coupling Express2 LR
 - OSA Express7 1.2
 - Crypto Express 8S
 - RoCE Express3
- ▶ The number of PSP, support partitions, for managing native PCIe I/O:
 - Four partitions
 - Reduced effect on MCL updates
 - Better availability
- ▶ Faster Dynamic Memory Relocation engine:
 - Enables faster reallocation of memory that is used for LPAR activations, CDR, and concurrent upgrade
 - Enables also LPAR Optimizations with DSR - Dynamic Storage Reconfiguration
 - Provides faster, more robust service actions

9.3 Structure

The IBM z16 A02 and IBM z16 AGZ have been designed for 19-inch frames format. The IBM z16 A02 and IBM z16 AGZ can be delivered in an IBM rack (IBM z16 A02) or installed in a customer provided rack (IBM z16 AGZ) and fulfills the requirements for ASHRAE A3 class environment.

The IBM z16 A02 and IBM z16 AGZ can have up to three PCIe+ I/O drawers. The structure of the IBM z16 A02 and IBM z16 AGZ are done with the following goals:

- ▶ Enhanced system modularity
- ▶ Standardization to enable rapid integration
- ▶ Platform simplification

Cables are keyed to ensure that correct lengths are plugged. Plug detection ensures correct location, and custom latches ensure retention. Further improvements to the fabric bus include symmetric multiprocessing (SMP9) cables that connect the drawers.

To improve field-replaceable unit (FRU) isolation, advanced continuity checking techniques are applied to the SMP-9 cables (Concurrent Cable Repair), with new SMP9 cables plugging/unplugging tool.

The IBM z16 A02 and IBM z16 AGZ has the following characteristics:

- ▶ Processing infrastructure is designed by using drawer technology.
- ▶ Keyed cables and plugging detection.
- ▶ SMP9 cables that are used for fabric bus connections.
- ▶ Master-master redundant oscillator design in the main memory.
- ▶ Point of load cards are separate FRUs.
- ▶ Improved redundant oscillator design
- ▶ Redundant combined Base Management Card (BMC) and Oscillator Cards (OSC) are provided per CPU drawer.
- ▶ Redundant N+1 Power Supply Units (PSU) to CPC drawer and PCIe+ I/O drawer.

9.4 Reducing complexity

IBM z16 A02 and IBM z16 AGZ continue the IBM z15 enhancements that reduced system RAS complexity.

Independent channel recovery with replay buffers on all interfaces allows recovery of a single DIMM channel, while other channels remain active. Further redundancies are incorporated in I/O pins for clock lines to main memory, which eliminates the loss of memory clocks because of connector (pin) failure. The following RAS enhancements reduce service complexity:

- ▶ Continued use of RAIM ECC.
- ▶ RAIM logic moved on DIMM
- ▶ N+1 on-DIMM voltage regulation
- ▶ Replay buffer for hardware retry on soft errors on the main memory interface.
- ▶ Redundant I/O pins for clock lines to main memory.
- ▶ Staggered refresh for performance enhancement
- ▶ The new RAIM scheme achieves a higher ratio of data to ECC symbols, while also providing an additional chip mark.

9.5 Reducing touches

IBM zSystems RAS efforts focus on the reduction of unscheduled, scheduled, planned, and unplanned outages. IBM zSystems technology has a long history of demonstrated RAS improvements, and this effort continues with changes that reduce service *touches* on the system.

Firmware was updated to improve filtering and resolution of errors that do not require action. Enhanced integrated sparing in processor cores, cache relocates, N+1 SEEPROM and POL³, N+2 redundancies, and DRAM marking also are incorporated to reduce touches. The following RAS enhancements reduce service touches:

- ▶ Improved error resolution to enable filtering
- ▶ Enhanced integrated sparing in processor cores
- ▶ Cache relocates
- ▶ N+1 SEEPROM
- ▶ N+2 POL
- ▶ DRAM marking
- ▶ (Dynamic) Spare BUS lanes for PU-PU, PU-MEM, MEM-MEM fabric
- ▶ N+1 Support Element (SE) (with N+1 SE power supplies)
- ▶ Redundant temperature sensor (one SEEPROM and one temperature sensor per I2C bus)
- ▶ FICON forward error correction
- ▶ A-Bus Lane Sparing
- ▶ OMI Bus Lane Sparing
- ▶ PU Core Sparing

9.6 IBM z16 A02 and IBM z16 AGZ availability characteristics

The following functions include availability characteristics on IBM z16 A02 and IBM z16 AGZ:

- ▶ Enhanced drawer availability (EDA)

³ Point of load

EDA is a *procedure* under which a CPC drawer in a multidrawer system can be removed and reinstalled during an upgrade or repair action with no effect on the workload.

- ▶ Concurrent memory upgrade or replacement (concurrent replacement available on Max68 feature only).

Memory can be upgraded concurrently by using Licensed Internal Code Configuration Control (LICCC) if physical memory is available on the drawers.

The EDA function can be useful if the physical memory cards must be changed in a 2-CPC drawer system (requiring the drawer to be removed).

It requires the availability of more memory resources on the other drawer or reducing the need for memory resources during this action. This option provides more resources to use EDA when repairing a drawer or memory on a drawer. They are also available when upgrading memory when larger memory cards might be required.

- ▶ Concurrent Driver Maintenance (CDM)

One of the greatest contributors to downtime during planned outages is LIC driver updates that are performed in support of new features and functions. IBM z16 A02 and IBM z16 AGZ is designed to support the concurrent activation of a selected new driver level.

- ▶ Concurrent fanout addition or replacement

A PCIe+ fanout card provides the path for data between memory and I/O through PCIe cables. With IBM z16 A02 and IBM z16 AGZ, a hot-pluggable and concurrently upgradeable fanout card is available (feature dependent). Up to 3 PCIe fanout cards for PCIe+ I/O drawers per system are supported for IBM z16 A02 and IBM z16 AGZ. An IBM z16 A02 and IBM z16 AGZ or IBM z16 A02 and IBM z16 AGZ feature Max68 has two CPC drawers and 24 PCIe fanout slots (for PCIe + fanout cards or ICA SR fanout features).

Internal I/O paths from the CPC drawer fanout ports to a PCIe drawer or an I/O drawer are spread across two CPC drawers (for feature Max68) and across different DCMs within a single CPC drawer Feature Max32. During an outage, a fanout card that is used for I/O can be repaired concurrently while redundant I/O interconnect ensures that no I/O connectivity is lost.

- ▶ Redundant I/O interconnect

Redundant I/O interconnect helps maintain critical connections to devices. Max68 system allows a single drawer to be removed and reinstalled concurrently during an upgrade or repair. Connectivity to the system I/O resources is maintained through a second path from a different drawer.

- ▶ Baseboard Management Card (BMC) / Oscillator Cards (OSC).

IBM z16 A02 and IBM z16 AGZ have two combined Baseboard Management Cards (BMC) and Oscillator Cards (OSC) per CPC drawer. The strategy of redundant clock and switchover stays the same as with the IBM z15 generation. One primary and one backup is available. If the primary OSC fails, the backup detects the failure, takes over transparently, and continues to provide the clock signal to the CPC.

- ▶ Processor unit (PU) sparing

IBM z16 A02 and IBM z16 AGZ and IBM z16 A02 and IBM z16 AGZ have two spare PUs per system to maintain performance levels if an active PU, Internal Coupling Facility (ICF), Integrated Facility for Linux (IFL), IBM Z Integrated Information Processor (zIIP), integrated firmware processor (IFP), or system assist processor (SAP) fails. Transparent sparing for failed processors is supported and sparing is supported across the drawers.

- ▶ Application preservation

This function is used when a PU fails and no spares are available. The state of the failing PU is passed to another active PU, where the operating system uses it to successfully resume the task, in most cases without client intervention.

► Cooling change

The IBM z16 A02 and IBM z16 AGZ configurations include front to rear cooling (air flow), which includes blowers, controls, and sensors that are N+1 redundant. The replacement of blowers is concurrent with no affect on performance.

► FICON Express 32S with Forward Error Correction (FEC)

FICON Express32S features continue to provide a new standard for transmitting data over 32 Gbps links by using 256b/257b encoding. The new standard that is defined by T11.org FC-FS-3 is more efficient than the 64b/66b or older 8b/10b encoding.

FICON Express32S channels that are running at 32 Gbps can take advantage of FEC capabilities when connected to devices that support FEC.

FEC allows FICON Express32S channels to operate at higher speeds, over longer distances, with reduced power and higher throughput. They also retain the same reliability and robustness for which FICON channels are traditionally known.

FEC is a technique that is used for controlling errors in data transmission over unreliable or noisy communication channels. When running at 32 Gbps link speeds, clients often see fewer I/O errors, which reduces the potential effect to production workloads from those I/O errors.

Read Diagnostic Parameters (RDP) improve Fault Isolation. After a link error is detected (for example, IFCC, CC3, reset event, or a link incident report), link data that is returned from Read Diagnostic Parameters is used to differentiate between errors that result from failures in the optics versus failures because of dirty or faulty links.

Key metrics can be displayed on the operator console. The results of a display matrix command with the LINKINFO=FIRST parameter, which collects information from each device in the path from the channel to the I/O device (see Figure 9-6 on page 373):

- Transmit (Tx) and Receive (Rx) optic power levels from the PCHID, Switch Input and Output, and I/O device
- Capable and Operating speed between the devices
- Error counts
- Operating System requires new function APAR OA49089

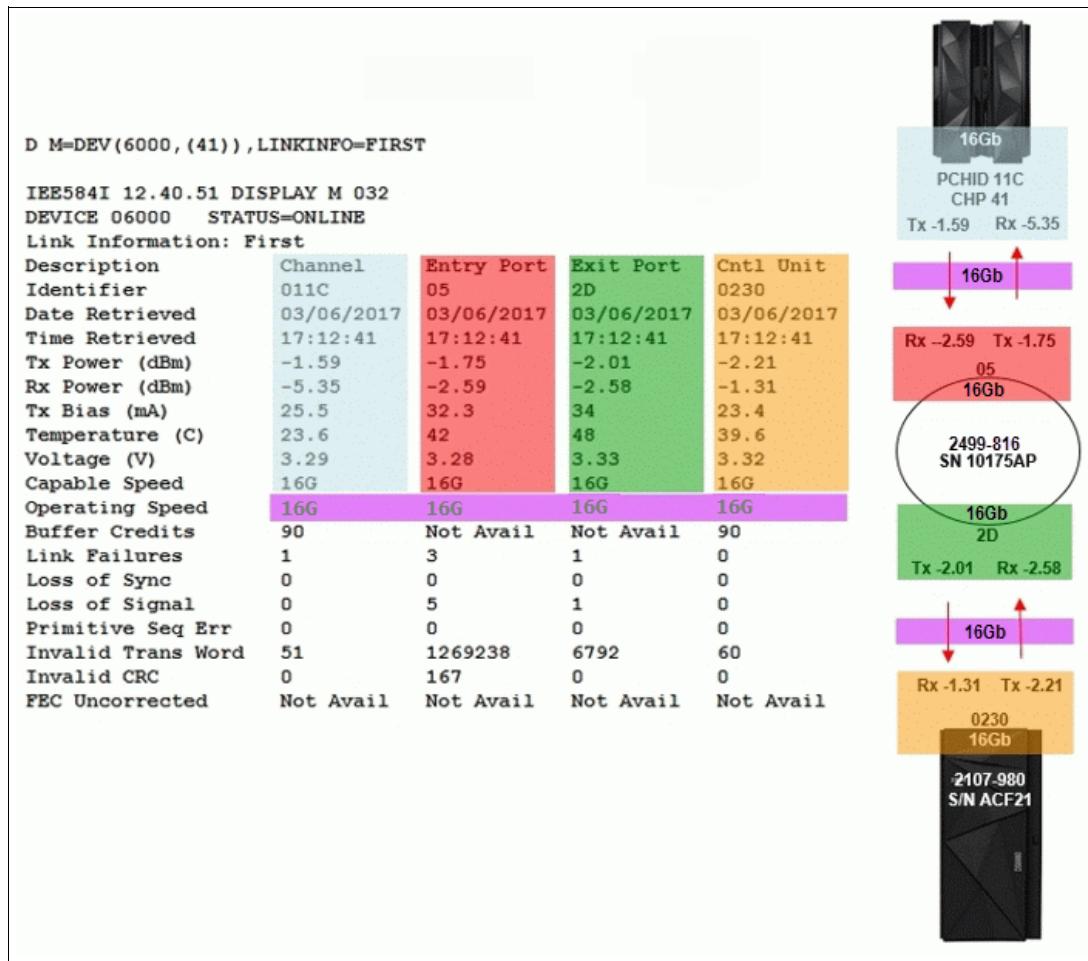


Figure 9-6 Read Diagnostic Parameters function

The new IBM zSystems Channel Subsystem Function performs periodic polling from the channel to the end points for the logical paths that are established and reduces the number of useless Repair Actions (RAs).

The RDP data history is used to validate Predictive Failure Algorithms and identify Fibre Channel Links with degrading signal strength before errors start to occur. The new Fibre Channel Extended Link Service (ELS) retrieves signal strength.

► FICON Dynamic Routing

FICON Dynamic Routing (FIDR) enables the use of storage area network (SAN) dynamic routing policies in the fabric. With the IBM z16 A02 and IBM z16 AGZ, FICON channels are no longer restricted to the use of static routing policies for inter-switch links (ISLs) for cascaded FICON directors.

FICON Dynamic Routing dynamically changes the routing between the channel and control unit based on the Fibre Channel Exchange ID. Each I/O operation has a unique exchange ID. FIDR is designed to support static SAN routing policies and dynamic routing policies.

FICON Dynamic Routing can help clients reduce costs by providing the following features:

- Share SANs between their FICON and FCP traffic.
- Improve performance because of SAN dynamic routing policies that better use all the available ISL bandwidth through higher use of the ISLs,

- Simplify management of their SAN fabrics by using static routing policies that assign different ISL routes with each power-on-reset (POR), which makes the SAN fabric performance difficult to predict.

Clients must ensure that all devices in their FICON SAN support FICON Dynamic Routing before they implement this feature.

9.7 IBM z16 A02 and IBM z16 AGZ RAS functions

Hardware RAS function improvements focus on addressing all sources of outages. Sources of outages feature the following classifications:

- ▶ Unscheduled

This outage occurs because of an unrecoverable malfunction in a hardware component of the system.

- ▶ Scheduled

This outage is caused by changes or updates that must be done to the system in a timely fashion. A scheduled outage can be caused by a disruptive patch that must be installed, or other changes that must be made to the system.

- ▶ Planned

This outage is caused by changes or updates that must be done to the system. A planned outage can be caused by a capacity upgrade or a driver upgrade. A planned outage usually is requested by the client, and often requires pre-planning. The IBM z16 A02 and IBM z16 AGZ design phase focuses on enhancing planning to simplify or eliminate planned outages.

The difference between scheduled outages and planned outages might not be obvious. The general consensus is that scheduled outages occur sometime soon. The time frame is approximately two weeks.

Planned outages are outages that are planned well in advance and go beyond this approximate two-week time frame. This chapter does not distinguish between scheduled and planned outages.

Preventing unscheduled, scheduled, and planned outages was addressed by the IBM Z system design for many years.

IBM z16 A02 and IBM z16 AGZ and IBM z16 A02 and IBM z16 AGZ systems have a fixed size HSA of 160 GB. This size helps eliminate pre-planning requirements for HSA and provides the flexibility to update dynamically the configuration. You can perform the following tasks dynamically:⁴

- ▶ Add a logical partition (LPAR)
- ▶ Add a logical channel subsystem (LCSS)
- ▶ Add a subchannel set
- ▶ Add a logical PU to an LPAR
- ▶ Add a cryptographic coprocessor
- ▶ Remove a cryptographic coprocessor
- ▶ Enable I/O connections
- ▶ Swap processor types
- ▶ Add memory

⁴ Some pre-planning considerations might exist. For more information, see Chapter 8, “System upgrades” on page 317.

- ▶ Add a physical processor

By addressing the elimination of planned outages, the following tasks also are possible:

- ▶ Concurrent driver upgrades
- ▶ Concurrent and flexible customer-initiated upgrades

For more information about the flexible upgrades that are started by clients, see 8.2.2, “Customer Initiated Upgrade facility” on page 326.

- ▶ STP management of concurrent CTN Split and Merge
- ▶ Dynamic I/O for stand-alone CF CPCs

Dynamic I/O configuration changes can be made to a stand-alone CF without requiring a disruptive power on reset. An LPAR with a firmware-based appliance version of an HCD instance is used to apply the new I/O configuration changes. The firmware-based LPAR is driven by updates from an HCD instance that is running in a z/OS LPAR on a different CPC that is connected to the same IBM z16 A02 and IBM z16 AGZ HMA.

- ▶ System Recovery Boost Stage 3

System Recovery Boost enhancements for IBM z16 A02 and IBM z16 AGZ, introduce the possibility of significantly reduce the impact of these disruptions by boosting a set of recovery processes that create significant pain points for our users today.

These recovery processes include:

- SVC Dump boost
- Middleware restart/recycle boost
- Hyperswap configuration load boost

For more information about System Recovery Boost, see *Introducing IBM Z System Recovery Boost*, REDP-5563.

9.7.1 Scheduled outages

Concurrent hardware upgrades, parts replacement, driver upgrades, and firmware fixes that are available with IBM z16 A02 and IBM z16 AGZ all address the elimination of scheduled outages. Also, the following indicators and functions that address scheduled outages are included:

- ▶ Memory data bus lane sparing.
This feature reduces the number of repair actions for memory.
- ▶ Dual tabbed Memory Clock signals
- ▶ Triple DRAM chipkill tolerance.
- ▶ Processor drawer power distribution *N+2* design.

The CPC Drawer uses point of load (POL) cards in a highly redundant *N+2* configuration. POL regulators are daughter cards that contain the voltage regulators for the principle logic voltage boundaries in the IBM z16 A02 and IBM z16 AGZ CPC drawer. They plug onto the CPC drawer system board and are non concurrent FRUs for the affected drawer, similar to the memory DIMMs. If you can use EDA, the replacement of POL cards is concurrent for the IBM zSystems server.

- ▶ Redundant (*N+1*) Ethernet switches.
- ▶ Redundant (*N+2*) ambient temperature sensors.
- ▶ Dual inline memory module (DIMM) field-replaceable unit (FRU) indicators.

These indicators imply that a memory module is not error-free and might fail sometime in the future. This indicator gives IBM a warning and provides time to concurrently repair the storage module for the IBM z16 A02 and IBM z16 AGZ Max68 feature.

The process to repair the storage module is to isolate or “fence off” the drawer, remove the drawer, replace the failing storage module, and then add the drawer. Single processor core checkstop and sparing.

This indicator shows that a processor core malfunctioned and is *spared*. IBM determines what to do based on the system and the history of that system.

- ▶ Point-to-point fabric for symmetric multiprocessing (SMP) RAS design
 - The A-bus is the logical bus and it is split over 2 SMP9 cables
 - 11 bits per A-Bus
 - One SMP9 cable has 6 bits and the other has 5 bits.
 - CRC detection and replay
 - RAS lane Degrade and Callhome Strategy design
 - Concurrent repair of the SMP9 cable. NEW for IBM z16 A02 and IBM z16 AGZ generation.

Having fewer components that can fail is an advantage. In a two drawer system, the drawers are connected by point-to-point connections. A second drawer can be added concurrently if the starting point is IBM z16 A02 and IBM z16 AGZ Max32 feature.

- ▶ The PCIe+ I/O drawer is available for IBM z16 A02 and IBM z16 AGZ. It and all of the PCIe+ I/O drawer-supported features can be installed concurrently.
- ▶ Memory interface logic to maintain channel synchronization when one channel goes into replay. IBM z16 A02 and IBM z16 AGZ can isolate recovery to only the failing channel.
- ▶ Out-of-band access to DIMM (for background maintenance functions).
Out-of-band access (by using an I2C interface) allows maintenance (such as logging) without disrupting customer memory accesses.
- ▶ OMII Memory Bus lane sparing.
- ▶ Improved DIMM exerciser for testing memory during IML.
- ▶ PCIe redrive hub cards plug straight in (no blind mating of connector). Simplified plugging that is more reliable is included.
- ▶ ICA SR (short distance) coupling cards plug straight in (no blind mating of connector). Simplified plugging that is more reliable is included.
- ▶ Coupling Express2 LR (CE2 LR) coupling cards plug into the PCIe+ I/O drawer, which allows more connections while maintaining link compatibility with previous generation Coupling Express features (CE LR).

9.7.2 Unscheduled outages

An *unscheduled outage* occurs because of an unrecoverable malfunction in a hardware component of the system.

The following improvements can minimize unscheduled outages:

- ▶ Continued focus on firmware quality
For LIC and hardware design, failures are eliminated through rigorous design rules; design walk-through; peer reviews; element, subsystem, and system simulation; and extensive engineering and manufacturing testing.
- ▶ Memory subsystem

Redundant Array of Independent Memory (RAIM) on IBM zSystem servers is a concept similar to the concept of Redundant Array of Independent Disks (RAID). The RAIM design detects and recovers from dynamic random access memory (DRAM), socket, memory channel, or DIMM failures. Memory size now includes RAIM protection and recovery.

Memory channels are organized in 8 card RAIM groups providing 50% reduced RAIM overhead (compared to IBM z15).

RAIM protection is similar to IBM z15:

- Up to 3 chip marks + 1 channel mark
- DDR4 DRAM with on chip power regulation
- N+1 voltage regulators

Memory is implemented using Standard Open Memory Interface (OMI) with up to 6 OMI per drawer, has CRC/Retry for soft errors, degrade bus lanes 4->2 on hard error, and no waiting for all eight cards (use first seven to respond)

A precise marking of faulty chips helps ensure timely DIMM replacements. The design of the IBM z16 A02 and IBM z16 AGZ further improved this chip marking technology. Graduated DRAM marking is available, and channel marking and scrubbing calls for replacement on the third DRAM failure is available. For more information about the memory system on IBM z16 A02 and IBM z16 AGZ, see 2.5, "Memory" on page 43.

► Soft-switch firmware

IBM z16 A02 and IBM z16 AGZ is equipped with the capabilities of soft-switching firmware. Enhanced logic in this function ensures that every affected circuit is powered off during the soft-switching of firmware components. For example, when you are upgrading the microcode of a FICON feature, enhancements are implemented to avoid any unwanted side effects that were detected on previous systems.

► Server Time Protocol (STP) recovery enhancement. IBM z16 A02 and IBM z16 AGZ has updated clocking structure:

- System uses a mesosynchronous clocking structure (similar to IBM z15)
- Two redundant oscillator cards in each drawer with dynamic oscillator switchover
- STP now has external clock reference support, separate Ethernet ports for ETS
- PTP and NTP (Ethernet cabling) are now directly connected to the CPC (no SE ETS connection needed)
- Oscillator card shares FRU package with drawer' Base Management Card (BMC)
- Concurrent repair of oscillator/control card

New signal has been implemented for IBM z16 A02 and IBM z16 AGZ - N-mode power signal for STP recovery. In a CTN with both PTS and BTS being IBM z16 A02 and IBM z16 AGZ, CTN recovery can be initiated by the new signal. Systems must have dual power with at least one side of the power capable of holding the power (UPS or similar) for five minutes in case of utility power failure.

When PCIe-based integrated communication adapter (ICA) Short Reach (SR) links are used, an unambiguous "going away signal" is sent when the server on which the coupling link is running is about to enter a failed (check stopped) state.

When the "going away signal" that is sent by the Current Time Server (CTS) in an STP-only Coordinated Timing Network (CTN) is received by the Backup Time Server (BTS), the BTS can safely take over as the CTS without relying on the previous Offline Signal (OLS) in a two-server CTN, or as the Arbiter in a CTN with three or more servers.

Enhanced Console Assisted Recovery (ECAR) contains recovery algorithms during a failing Primary Time Server (PTS) and uses communication over the HMA/SE network to assist with BTS takeover. For more information, see Chapter 10, "Hardware Management Console and Support Element" on page 391.

Coupling Express2 LR does not support the "going away signal"; however, ECAR can be used to assist with recovery in the following configurations:

► Design of pervasive infrastructure controls in processor chips in memory ASICs.

- ▶ Improved error checking in the processor recovery unit (RU) to better protect against word line failures in the RU arrays.

9.8 Enhanced drawer availability

Enhanced drawer availability (EDA) is a procedure in which a drawer in a multidrawer system can be removed and reinstalled during an upgrade or repair action. This procedure has no effect on the running workload⁵.

The EDA procedure and careful planning help ensure that all the resources are still available to run critical applications in an $(n-1)$ drawer configuration. This process allows you to avoid planned outages. Consider the flexible memory option to provide more memory resources when you are replacing a drawer.

To minimize the effect on current workloads, ensure that sufficient inactive physical resources exist on the remaining drawers to complete a drawer removal. Also, consider deactivating non-critical system images, such as test or development LPARs. After you stop these non-critical LPARs and free their resources, you might find sufficient inactive resources to contain critical workloads while completing a drawer replacement.

9.8.1 EDA planning considerations

To use the EDA function, configure enough physical memory and engines so that the loss of a single drawer does not result in any degradation to critical workloads during the following occurrences:

- ▶ A degraded restart in the rare event of a drawer failure
- ▶ A drawer replacement for repair or a physical memory upgrade

The following configurations especially enable the use of the EDA function. These IBM z16 A02 and IBM z16 AGZ features need enough spare capacity so that they can cover the resources of a fenced or isolated drawer. This configuration imposes limits on the following number of the client-owned PUs that can be activated when one CPC drawer is fenced:

- ▶ A maximum of 68 PUs are configured on the Max68.
- ▶ No special feature codes are required for PU and model configuration.
- ▶ Feature Max32 and Max68 have 4 SAPs in each drawer.

The system configuration must have sufficient dormant resources on the remaining drawers in the system for the *evacuation* of the drawer that is to be replaced or upgraded. Dormant resources include the following possibilities:

- ▶ Unused PUs or memory that is not enabled by LICCC
- ▶ Inactive resources that are enabled by LICCC (memory that is not being used by any activated LPARs)
- ▶ Amount of Memory that is available

The I/O connectivity must also support drawer removal. Most of the paths to the I/O feature redundant I/O interconnect support in the I/O infrastructure (drawers) that enable connections through multiple fanout cards.

If sufficient resources are not present on the remaining drawers, certain non-critical LPARs might need to be deactivated. One or more PUs or storage might need to be configured

⁵ With proper planning and depending on your system configuration. A single CPC drawer IBM z16 A02 and IBM z16 AGZ does not support Enhanced Drawer Availability.

offline to reach the required level of available resources. Plan to address these possibilities to help reduce operational errors.

Exception: Single-drawer systems cannot use the EDA procedure.

Include the planning as part of the initial installation and any follow-on upgrade that modifies the operating environment. A client can use the Resource Link machine information report to determine the number of drawers, active PUs, memory configuration, and channel layout.

If the IBM z16 A02 and IBM z16 AGZ is installed, click **Prepare for Enhanced Drawer Availability** in the Perform Model Conversion window of the EDA process on the Hardware Management Appliance (HMA). This task helps you determine the resources that are required to support the removal of a drawer with acceptable degradation to the operating system images.

The EDA process determines which resources, including memory, PUs, and I/O paths, are free to allow for the removal of a drawer. You can run this preparation on each drawer to determine which resource changes are necessary. Use the results as input in the planning stage to help identify critical resources.

With this planning information, you can examine the LPAR configuration and workload priorities to determine how resources might be reduced and still allow the drawer to be concurrently removed.

Include the following tasks in the planning process:

- ▶ Review of the IBM z16 A02 and IBM z16 AGZ configuration to determine the following values:
 - Number of drawers that are installed and the number of PUs enabled. Consider the following points:
 - Use the Resource Link machine information or the HMA to determine the model, number, and types of PUs (CPs, IFLs, ICFs, and zIIPs).
 - Determine the amount of memory (physically installed and LICCC-enabled).
 - Work with your IBM Service Support Representative (IBM SSR) to determine the memory card size in each drawer. The memory card sizes and the number of cards that are installed for each drawer can be viewed from the SE under the CPC configuration task list. Use the View Hardware Configuration option.
 - ICA SR fanout layouts and ICA to ICA connections.
- Use the Resource Link machine information to review the channel configuration. This process is a normal part of the I/O connectivity planning. The alternative paths must be separated as far into the system as possible.
- ▶ Review the system image configurations to determine the resources for each image.
- ▶ Determine the importance and relative priority of each LPAR.
- ▶ Identify the LPAR or workloads and the actions to be taken:
 - Deactivate the entire LPAR.
 - Configure PUs.
 - Reconfigure memory, which might require the use of reconfigurable storage unit (RSU) values.
 - Vary off the channels.

- ▶ Review the channel layout and determine whether any changes are necessary to address single paths.
- ▶ Develop a plan to address the requirements.

When you perform the review, document the resources that can be made available if the EDA is used. The resources on the drawers are allocated during a POR of the system and can change after that process. Perform a review when changes are made to your IBM z16 A02 and IBM z16 AGZ, such as adding a second drawer, PUs, memory, or channels. Also, perform a review when workloads are added or removed, or if the HiperDispatch feature was enabled and disabled since the last time you performed a POR.

9.8.2 Enhanced drawer availability processing

To use the EDA, first ensure that the following conditions are met:

- ▶ Free the used processors (PUs) on the drawer that is removed.
- ▶ Free the used memory on the drawer.
- ▶ For all I/O domains that are connected to the drawer and ensure that alternative paths exist. Otherwise, place the I/O paths offline.

For the EDA process, this phase is the preparation phase. It is started from the SE, directly or on the HMA, by using the Single object operation option on the Perform Model Conversion window from the CPC configuration task list, as shown in Figure 9-7.

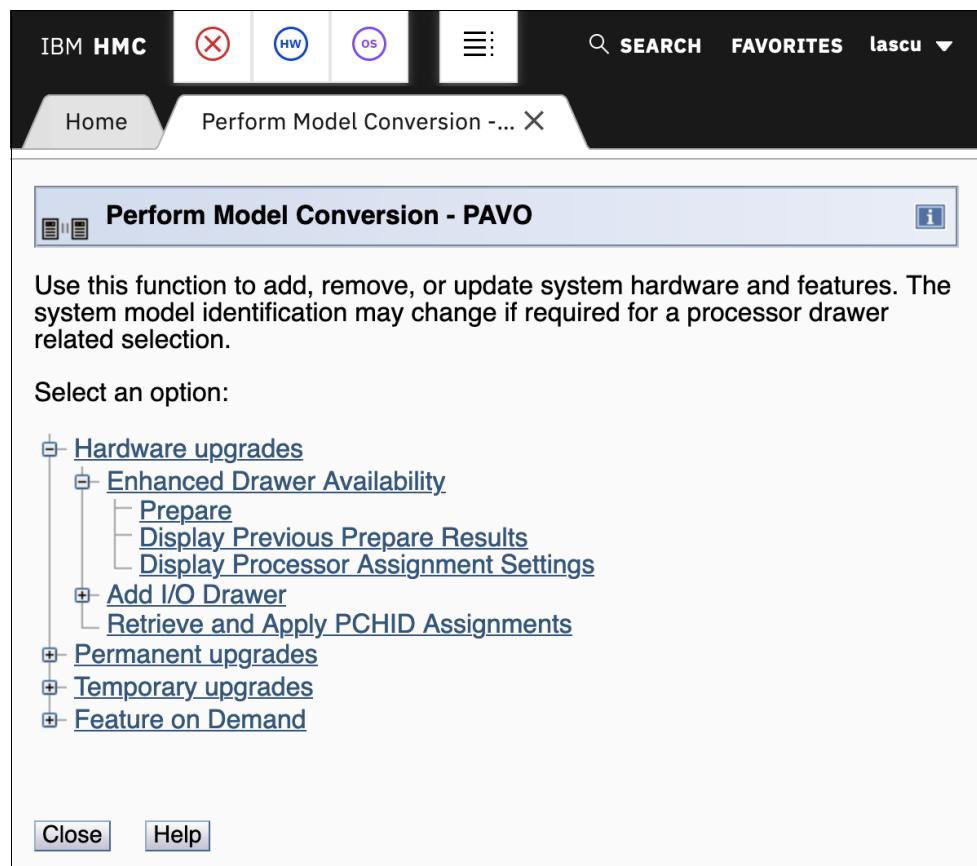


Figure 9-7 Clicking Prepare for Enhanced Drawer Availability option

Processor availability

Processor resource availability for reallocation or deactivation is affected by the type and quantity of the resources in use, such as:

- ▶ Total number of PUs that are enabled through LICCC
- ▶ PU definitions in the profiles that can be dedicated and dedicated reserved or shared
- ▶ Active LPARs with dedicated resources at the time of the drawer repair or replacement

To maximize the PU availability option, ensure that sufficient inactive physical resources are on the remaining drawers to complete a drawer removal.

Memory availability

Memory resource availability for reallocation or deactivation depends on the following factors:

- ▶ Physically installed memory
- ▶ Image profile memory allocations
- ▶ Amount of memory that is enabled through LICCC
- ▶ Virtual Flash Memory if enabled and configured

For more information, see 2.7.2, “Enhanced drawer availability (EDA)” on page 56.

Fan out card to I/O connectivity requirements

The optimum approach is to maintain maximum I/O connectivity during drawer removal. The redundant I/O interconnect (RII) function provides for redundant connectivity to all installed I/O domains in the PCIe+ I/O drawers.

Preparing for enhanced drawer availability

The Prepare Concurrent Drawer replacement option validates that enough dormant resources are available for this operation. If enough resources are not available on the remaining drawers to complete the EDA process, the process identifies those resources. It then guides you through a series of steps to select and free up those resources. The preparation process does not complete until all processors, memory, and I/O conditions are successfully resolved.

Preparation: The preparation step does not reallocate any resources. It is used only to record client choices and produce a configuration file on the SE that is used to run the concurrent drawer replacement operation.

The preparation step can be done in advance. However, if any changes to the configuration occur between the preparation and the physical removal of the drawer, you must rerun the preparation phase.

The process can be run multiple times because it does not move any resources. To view the results of the last preparation operation, click **Display Previous Prepare Enhanced Drawer Availability Results** from the Perform Model Conversion window in the SE.

The preparation step can be run without performing a drawer replacement. You can use it to dynamically adjust the operational configuration for drawer repair or replacement before IBM SSR activity. The Perform Model Conversion window in you click **Prepare for Enhanced Drawer Availability** is shown in Figure 9-7 on page 380.

After you click **Prepare for Enhanced Drawer Availability**, the Enhanced Drawer Availability window opens. Select the drawer that is to be repaired or upgraded; then, select **OK**, as shown in Figure 9-8 on page 382. Only one target drawer can be selected at a time.

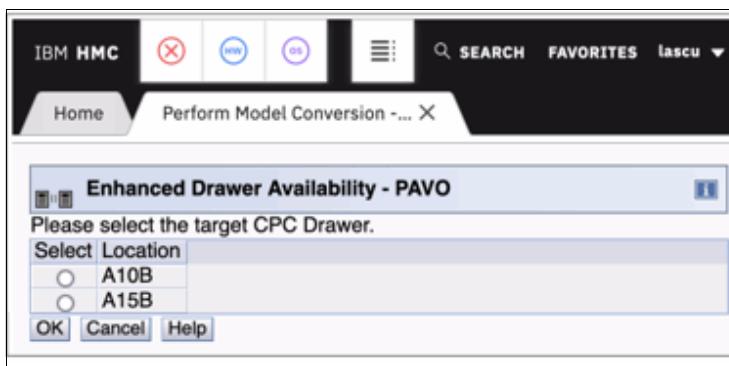


Figure 9-8 Selecting the target drawer

The system verifies the resources that are required for the removal, determines the required actions, and presents the results for review. Depending on the configuration, the task can take from a few seconds to several minutes.

The preparation step determines the readiness of the system for the removal of the targeted drawer. The configured processors and the memory in the selected drawer are evaluated against unused resources that are available across the remaining drawers. The system also analyzes I/O connections that are associated with the removal of the targeted drawer for any single path I/O connectivity.

If insufficient resources are available, the system identifies the conflicts so that you can free other resources.

The following states can result from the preparation step:

- ▶ The system is ready to run the EDA for the targeted drawer with the original configuration.
- ▶ The system is not ready to run the EDA because of conditions that are indicated by the preparation step.
- ▶ The system is ready to run the EDA for the targeted drawer. However, to continue with the process, processors are reassigned from the original configuration.

Review the results of this reassignment relative to your operation and business requirements. The reassessments can be changed on the final window that is presented. However, before making any changes or approving reassessments, ensure that the changes are reviewed and approved by the correct level of support based on your organization's business requirements.

Preparation tabs

The results of the preparation are presented for review in a tabbed format. Each tab indicates conditions that prevent the EDA option from being run. The following tab selections are available:

- ▶ Processors
- ▶ Memory
- ▶ Single I/O
- ▶ Single Domain I/O
- ▶ Single Alternate Path I/O

Only the tabs that feature conditions that prevent the drawer from being removed are displayed. Each tab indicates the specific conditions and possible options to correct them.

For example, the preparation identifies single I/O paths that are associated with the removal of the selected drawer. These paths must be varied offline to perform the drawer removal. After you address the condition, rerun the preparation step to ensure that all the required conditions are met.

Preparing the system to perform enhanced drawer availability

During the preparation, the system determines the PU configuration that is required to remove the drawer. The results and the option to change the assignment on non-dedicated processors are shown in Figure 9-9.

The screenshot shows a software interface titled "Processor Assignments - PAVO". It displays two tabs: "Drawer_A10B" and "Drawer_A15B". The "Drawer_A10B" tab is active, showing the following data:

Processor Type	Dedicated Count	Non-Dedicated Count	Processor Totals	LICCC Count
CPU	0	21	21	24
ICF	0	1	1	2
IFL	0	21	21	24
zIIP	0	14	14	16
SAP	14	0	14	16
Available to use		0	0	
Remaining processor drawer Totals	14	58	72	

At the bottom of the window are "Cancel" and "Help" buttons.

Figure 9-9 Reassign Non-Dedicated Processors results

Important: Consider the results of these changes relative to the operational environment. Understand the potential effect of making such operational changes. Changes to the PU assignment, although technically correct, can result in constraints for critical system images. In certain cases, the solution might be to defer the reassignments to another time that has less effect on the production system images.

After you review the reassignment results and make any necessary adjustments, click **OK** (Figure 9-10).

The screenshot shows a software interface titled "Reassign Non-Dedicated Processors - PAVO". At the top, there is a message: "Accept or reassign non-dedicated processors without exceeding the LICCC count." Below it is a warning message: "Warning: These values should only be reassigned under the direction of the System Programmer. Otherwise accept the provided default values." The table data is identical to Figure 9-9.

Processor Type	Dedicated Count	Non-Dedicated Count	Processor Totals	LICCC Count
CPU	0	21	21	24
ICF	0	1	1	2
IFL	0	21	21	24
zIIP	0	14	14	16
SAP	14	0	14	16
Available to use		0	0	
Remaining processor drawer Totals	14	58	72	

At the bottom of the window are "OK", "Cancel", and "Help" buttons.

Figure 9-10 Reassign Non-Dedicated Processors, message ACT37294

Summary of the drawer removal process steps

To remove a drawer, the following resources must be moved to the remaining active drawers:

- ▶ PUs: Enough PUs must be available on the remaining active drawers, including all types of PUs that can be characterized (CPs, IFLs, ICFs, zIIPs, SAPs, and IFP).
- ▶ Memory: Enough installed memory must be available on the remaining active drawers.
- ▶ I/O connectivity: Alternative paths to other drawers must be available on the remaining active drawers, or the I/O path must be taken offline.

By understanding the system configuration and the LPAR allocation for memory, PUs, and I/O, you can make the best decision about how to free the necessary resources to allow for drawer removal.

Complete the following steps to concurrently replace a drawer:

1. Run the preparation task to determine the necessary resources.
2. Review the results.
3. Determine the actions to perform to meet the required conditions for EDA.
4. When you are ready to remove the drawer, free the resources that are indicated in the preparation steps.
5. Repeat the step that is shown in Figure 9-7 on page 380 to ensure that the required conditions are all satisfied.
6. Upon successful completion, the system is ready for the removal of the drawer.

The preparation process can be run multiple times to ensure that all conditions are met. It does not reallocate any resources; instead, it produces only a report. The resources are not reallocated until the Perform Drawer Removal process is started.

Rules during EDA

During EDA, the following rules are enforced:

- ▶ Processor rules

All processors in any remaining drawers are available to be used during EDA. This requirement includes the two spare PUs or any available PU that is non-LICCC.

The EDA process also allows conversion of one PU type to another PU type. One example is converting a zIIP to a CP during the EDA function. **The preparation for the concurrent drawer replacement task indicates whether any SAPs must be moved to the remaining drawers.**

- ▶ Memory rules

All physical memory that is installed in the system, including flexible memory, is available during the EDA function. **Any physical installed memory, whether purchased or not, is available to be used by the EDA function.**

- ▶ Single I/O rules

Alternative paths to other drawers must be available, or the I/O path must be taken offline.

Review the results. The result of the preparation task is a list of resources that must be made available before the drawer replacement can occur.

Freeing any resources

At this stage, create a plan to free these resources. The following resources and actions are necessary to free them:

- ▶ Freeing any PUs:
 - Vary off the PUs by using the Perform a Model Conversion window, which reduces the number of PUs in the shared PU pool.
 - Deactivate the LPARs.
- ▶ Freeing memory:
 - Deactivate an LPAR.
 - Vary offline a portion of the reserved (online) memory. For example, in z/OS, run the following command:

```
CONFIG_STOR(E=1),<OFFLINE/ONLINE>
```

This command enables a storage element to be taken offline. The size of the storage element depends on the RSU value. In z/OS, the following command configures offline smaller amounts of storage than the amount that was set for the storage element:

```
CONFIG_STOR(nnM),<OFFLINE/ONLINE>
```
 - A combination of both LPAR deactivation and varying memory offline.

Reserved storage: If you plan to use the EDA function with z/OS LPARs, set up reserved storage and an RSU value. Use the RSU value to specify the number of storage units that are to be kept free of long-term fixed storage allocations. This configuration allows for storage elements to be varied offline.

9.9 Concurrent Driver Maintenance

CDM is one more step toward reducing the necessity for and the duration of a scheduled outage. One of the components to planned outages is LIC Driver updates that are run in support of new features and functions.

When correctly configured, IBM z16 A02 and IBM z16 AGZ support concurrently activating a selected new LIC Driver level. Concurrent activation of the selected new LIC Driver level is supported only at specific released sync points. Concurrently activating a selected new LIC Driver level anywhere in the maintenance stream is not possible. Certain LIC updates do not allow a concurrent update or upgrade.

Consider the following key points about CDM:

- ▶ The HMA can query whether a system is ready for a concurrent driver upgrade.
- ▶ Previous firmware updates, which require an initial machine load (IML) of the IBM z16 A02 and IBM z16 AGZ to be activated, can block the ability to run a concurrent driver upgrade.
- ▶ A function on the SE allows you or your IBM SSR to define the concurrent driver upgrade sync point to be used for an CDM.
- ▶ The ability to concurrently install and activate a driver can eliminate or reduce a planned outage.
- ▶ IBM z16 A02 and IBM z16 AGZ introduce Concurrent Driver Upgrade (CDU) cloning support to other CPCs for CDU preinstallation and activation.
- ▶ Concurrent crossover from Driver level N to Driver level $N+1$, then to Driver level $N+2$, must be done serially. No composite moves are allowed.
- ▶ Disruptive upgrades are permitted at any time, and allow for a composite upgrade (Driver N to Driver $N+2$).

- ▶ Concurrently backing up to the previous driver level is not possible. The driver level must move forward to driver level $N+1$ after CDM is started. Unrecoverable errors during an update might require a scheduled outage to recover.

The CDM function does not eliminate the need for planned outages for driver-level upgrades. Upgrades might require a system level or a functional element scheduled outage to activate the new LIC. The following circumstances require a scheduled outage:

- ▶ Specific complex code changes might dictate a disruptive driver upgrade. You are alerted in advance so that you can plan for the following changes:
 - Design data or hardware initialization data fixes
 - CFCC release level change
- ▶ OSA CHPID code changes might require PCHID Vary OFF/ON to activate new code.
- ▶ Crypto code changes might require PCHID Vary OFF/ON to activate new code.

Note: zUDX clients should contact their User Defined Extensions (UDX) provider before installing Microcode Change Levels (MCLs). Any changes to Segments 2 and 3 from a previous MCL level might require a change to the client's UDX. Attempting to install an incompatible UDX at this level results in a Crypto checkstop.

9.9.1 Resource Group and native PCIe features MCLs

Microcode fixes, referred to as *individual MCLs* or *packaged in Bundles*, might be required to update the Resource Group code and the native PCIe features. Although the goal is to minimize changes or make the update process concurrent, the maintenance updates at times can require the Resource Group or the affected native PCIe to be toggled offline and online to implement the updates. The native PCIe features (managed by Resource Group code) are listed Table 9-1.

Table 9-1 Native PCIe cards for IBM z16 A02 and IBM z16 AGZ

Native PCIe adapter type	Feature code	Resource required to be offline
25 GbE RoCE Express3 SR	0452	FIDs/PCHID
25 GbE RoCE Express3 LR	0453	FIDs/PCHID
25 GbE RoCE Express2.1	0450	FIDs/PCHID
25 GbE RoCE Express2	0430	FIDs/PCHID
10 GbE RoCE Express3 SR	0440	FIDs/PCHID
10 GbE RoCE Express3 LR	0441	FIDs/PCHID
10 GbE RoCE Express2.1	0432	FIDs/PCHID
10 GbE RoCE Express2	0412	FIDs/PCHID
zHyperLink Express1.1	0451	FIDs/PCHID
zHyperLink Express	0431	FIDs/PCHID
Coupling Express2 LR	0434	CHPIDs/PCHID

Consider the following points for managing native PCIe adapters microcode levels:

- ▶ Updates to the Resource Group require all native PCIe adapters that are installed in that RG to be offline.

- ▶ Updates to the native PCIe adapter require the adapter to be offline. If the adapter is not defined, the MCL session automatically installs the maintenance that is related to the adapter.

The PCIe native adapters are configured with Function IDs (FIDs) and might need to be configured offline when changes to code are needed. To help alleviate the number of adapters (and FIDs) that are affected by the Resource Group code update, IBM z16 A02 and IBM z16 AGZ have four Resource Groups per system (CPC).

Note: Other adapter types, such as FICON Express, OSA Express, and Crypto Express that are installed in the PCIe+ I/O drawer are not effected because they are not managed by the Resource Groups.

The front, rear, and top view of the PCIe+ I/O drawer and the Resource Group assignment by card slot are shown in Figure 9-11. All PCIe+ I/O drawers that are installed in the system feature the same Resource Group assignment.

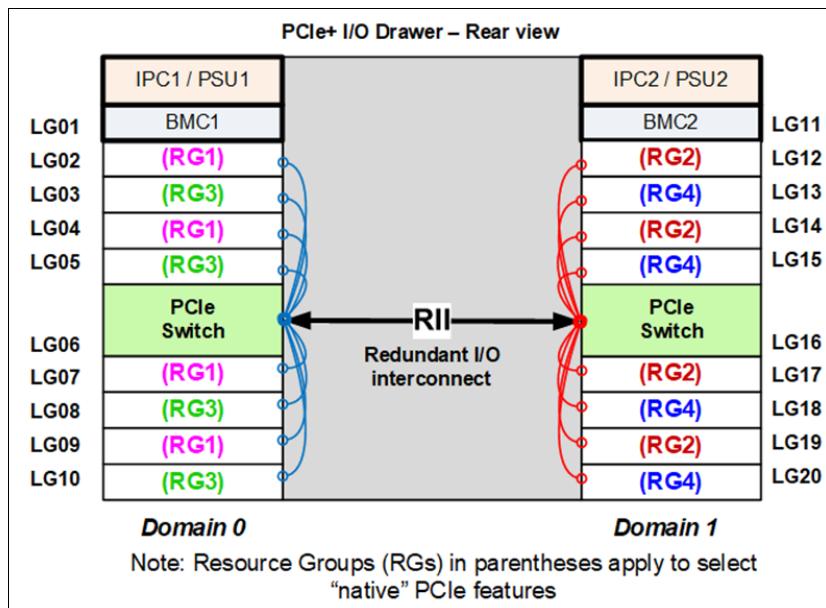


Figure 9-11 Resource Group slot assignment

9.10 RAS capability for the HMA and SE

The HMA and the SE include the following RAS capabilities:

- ▶ Back up from HMA and SE

For the customers who do not have an FTP server that is defined for backups, the HMA can be configured as an FTP server.

On a scheduled basis, the HMA hard disk drive (HDD) is backed up to the USB flash memory drive (UFD), a defined FTP server, or both.

SE HDDs are backed up on to the primary SE HDD and an alternative SE HDD. In addition, you can save the backup to a defined FTP server.

For more information, see 10.2, “HMC and SE new features and changes” on page 393.

- ▶ Remote Support Facility (RSF)

The HMA RSF provides the important communication to a centralized IBM support network for hardware problem reporting and service. For more information, see 10.4, “Remote Support Facility” on page 416.

- ▶ Microcode Change Level (MCL)

Regular installation of MCLs is key for RAS, optimal performance, and new functions. Generally, plan to install MCLs quarterly at a minimum. Review hiper MCLs continuously. You must decide whether to wait for the next scheduled apply session, or schedule one earlier if your risk assessment of the new hiper MCLs warrants.

For more information, see 10.5.5, “HMC and SE Microcode” on page 421.

- ▶ SE

IBM z16 A02 and IBM z16 AGZ are provided with two 1U trusted servers inside the “A” frame: One is always the primary SE and the other is the alternative SE⁶. The primary SE is the active SE. The alternative acts as the backup. Information is mirrored once per day. The SE servers include N+1 redundant power supplies.

For more information, see 10.2.3, “HMC and SE servers” on page 406.

9.11 IBM z16 AGZ specifics

The IBM z16 AGZ offering enables clients to procure and integrate an IBM zSystems air cooled configuration into their own racks and PDU power systems.

- ▶ Client-provided racks and PDUs must meet the requirements specified in the IMPP
- ▶ The IBM z16 AGZ uses the same IBM components as the IBM z16 A02.
- ▶ The components are installed by the IBM SSR in Client racks powered by Client PDUs
- ▶ Planning documents will be provided to assist with frame and power requirements for the Client

The key document for all information regarding the planning is the *IBM z16 and LinuxONE Rockhopper 4 Rack Mount Bundle Installation Manual for Physical Planning (IMPP)*, GC28-7035.

Client provided PDUs

- ▶ PDUs must meet power and port requirements detailed in the IMPP
- ▶ Client must define which IBM z16 AGZ power jumper plugs into exactly which PDU outlet
 - power jumper plug lengths are defined by feature code during the order process

Shipping - POD Concept

- ▶ Systems are shipped in palletized set of PODs for protection during transport and ease of content removal / install into a rack at the client’s data center.
- ▶ Two POD sizes are used (6U and 9U high) with a maximum of three interconnected per pallet.
- ▶ Each POD group is fully enclosed in a wooden and palletized crate with removable side panels.
- ▶ The system configuration will determine the number of Pods

⁶ If HMA feature is installed on the system, special upgrade procedure must be followed to ensure non-disruptive SE upgrade.

- Clients will have to plan to store the system, in its Pod(s) on the pallet(s), until IBM SSR installs the machine
- Clients must provide service clearances for installation

Figure 9-12 displays an example of the wooden shipping container that the pods will ship in. The contents can be any one of the following pods to build the system in the rack.



Figure 9-12 Shipping container and various Pods

Larger configurations may contain multiple shipping containers. The rails used for each pod will also be transferred into the client rack during the installation.

Genie GL8 Lift Tool

The IBM z16 AGZ requires a new Lift Tool, the Genie GL8. This tool is required for installation and service. Recommendation is one Genie GL8 Lift Tool per account. If a client already has one they bought to use with a System P machine, it can be used for the Z system also. A wedge plate add-on is required to install the CPC drawers into the rack, and **two Lift Tool Feature Codes are provided:**

- ▶ EB3Z (Lift Tool), EB4Z (CPC Drawer Shelf)
- ▶ An added (returnable) transfer shelf will be provided for all systems with at least one I/O drawer

Figure 9-13 shows the Genie GL-8 Lift tool, and the CPC and I/O drawer shelves used for transfer into the customer rack.

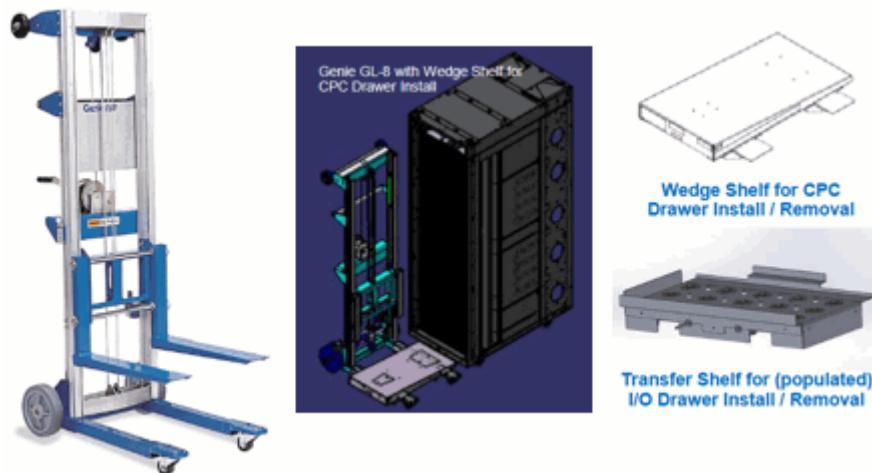


Figure 9-13 Genie GL-8 Lift Tool and Shelves for CPC and I/O drawers

10

Hardware Management Console and Support Element

The Hardware Management Console (HMC) supports the functions and tasks that are required to manage the IBM zSystems. When tasks are performed on the HMC, the commands are sent to the Primary Support Element (SE) of the targeted system, which then issues commands to their respective central processor complex (CPC).

This chapter describes the newest and most important elements for the HMC and SE.

Tip: The Help function is a good starting point to get more information about all of the functions that can be used by the HMC and SE. The Help feature is available by clicking **Help** from the drop-down menu that appears when you click your user ID on the upper right corner.

For more information, see Resource Link (<https://www.ibm.com/servers/resourcelink>), select **Library**, the applicable server, then select either **HMC Version 2.16.0 help file content** or **SE Version 2.16.0 help file content**.

This chapter includes the following topics:

- ▶ 10.1, “Introduction and overview” on page 392
- ▶ 10.2, “HMC and SE new features and changes” on page 393
- ▶ 10.3, “HMC and SE connectivity” on page 409
- ▶ 10.4, “Remote Support Facility” on page 416
- ▶ 10.5, “HMC and SE capabilities” on page 417

10.1 Introduction and overview

The HMC runs a set of management applications. On [IBM z16 A02 and IBM z16 AGZ](#) two HMCs will be delivered with the Hardware Management Appliance (HMA) feature FC 0129. The HMC code runs on the two 1U rack-mounted servers in the frame.

The HMC is a closed system (appliance), which means that no other applications can be installed on it.

The new Driver level for HMC and SE for [IBM z16 A02 and IBM z16 AGZ](#) is Driver 51. Driver 51 is equivalent to Version 2.16.0.

Stand-alone, outside the [IBM z16 A02 and IBM z16 AGZ](#) HMCs (tower or AGZ) can no longer be ordered for new build systems.

The following HMC feature codes can be carried forward from previous orders:

- ▶ FC 0062
- ▶ FC 0063
- ▶ FC 0082
- ▶ FC 0083

On the carry forward HMCs, Driver 51/Version 2.16.0 must be installed to support [IBM z16 A02 and IBM z16 AGZ](#). Also Driver 51/Version 2.16.0 can be installed on the HMCs provided with the HMA feature FC 0100 on IBM z15.

You can use the [IBM z16 A02 and IBM z16 AGZ](#) with carry forward HMCs and no initial order of FC 0129 (HMA for [IBM z16 A02 and IBM z16 AGZ](#)). The HMA, FC0129, can be ordered and installed at a later time as an upgrade.

With [IBM z16 A02 and IBM z16 AGZ](#) the HMA feature shares the 1U server hardware with the SE code. The SE code runs virtualized under the HMC on each of the two 1U rack-mounted servers. One SE is the Primary SE (active) and the other is the Alternate SE (backup). As with the HMCs, the SEs are closed systems (appliances), and no other applications can be installed on the same hardware.

The HMC is used to set up, manage, monitor, and operate one or more IBM zSystems CPCs. It manages IBM zSystems hardware, its logical partitions (LPARs), and provides support applications. At least one HMC is required to operate an IBM zSystems. An HMC can manage multiple IBM zSystems CPCs.

When tasks are performed at the HMC, the commands are routed to the Primary SE of the target IBM zSystems. The SE then issues those commands to the according CPC.

With HMC and SE Driver 41/Version 2.15.0, a number of “traditional” SE-only functions moved to HMC tasks. On the HMC these functions appear as native HMC tasks, but run on the SE. These HMC functions run in parallel with **Single Object Operations** (SOOs), which simplifies and streamlines system management. For more information about SOOs, see “Single Object Operations (SOO)” on page 419.

Check the following introduction videos for the HMC at <https://ibm.biz/IBM-Z-HMC>.

Note: HMC Driver 51/Version 2.16.0 supports managing IBM zSystems N-2 server generations IBM z14, IBM z15 and IBM z16.

10.2 HMC and SE new features and changes

The initial release that is included with [IBM z16 A02 and IBM z16 AGZ](#) is HMC and SE Driver 51/Version 2.16.0. When you log in to the HMC, the Dashboard is displayed. In the dashboard check the “What’s new” widget to examine the new features that are available.

For more information about HMC and SE functions, use the HMC and SE console help system or see Resource Link¹ (<https://www.ibm.com/servers/resourcelink>), select **Library**, the applicable server, then select either **HMC Version 2.16.0 help file content** or **SE Version 2.16.0 help file content**.

10.2.1 Driver 51/Version 2.16.0 HMC and SE new features

The following support was added with Driver 51/Version 2.16.0:

- ▶ Enhanced Multi-Factor Authentication (MFA) functions

With Driver 41/Version 2.15.0, MFA on the HMC is supported via Time-based One-time Password (TOTP) or IBM zSystems Multi-Factor Authentication (z/OS) and RSA Secure ID.

New with Driver 51/Version 2.16.0 the following further MFA possibilities are supported:

- Certificates
 - Personal Identity Verification (PIV)
 - Common Access Card (CAC)
 - Certificates on USB keys
- Generic Remote Authentication Dia-In User Service (RADIUS) allows for support of all various RADIUS factor types. Involves customer provided RADIUS server.
- Support of IBM Z Multi-Factor authentication for RedHat Enterprise Linux Server or SUSE Linux Enterprise Server running on z/VM or native in an LPAR.

For more information see <https://www.ibm.com/docs/en/zma> and 10.3.7, “HMC Multi-factor authentication” on page 415.

- ▶ Login changes to support PCI-DSS

To support Payment Card Industry-Data Security Standard (PCI-DSS) there is now a single GUI login panel for user ID, password, and the authentication code if MFA is used. Before, the entry for the authentication code was on a separate panel. You have to select **Use authentication code** to be able to enter it.

For further security, only in the Audit and Security Log (which needs specific access level) the details about the login failure can be reviewed. For the GUI and Web Services APIs the HMC returns a message indication that the login has failed. No additional information is provided.

- ▶ HMC/SE TLS 1.3 support

With Driver 51/Version 2.16.0, the HMC and SE support TLS 1.3. Before setting the TLS level to 1.3, you have to ensure that all services and servers connecting via TLS to the HMC and SE do support TLS 1.3 as well (for example, the remote browser, LDAP Authentication Servers, WebServices API connections, Fibre Channel End Point Security, FTPS, Single Object Operations).

TLS 1.0 and 1.1 are not supported with Driver 51/Version 2.16.0.

If minimum TLS version is set to 1.2, TLS 1.3 will be attempted first, then fall back to TLS 1.2 if required.

¹ IBM ID is required for authentication and access to Resource Link.

Figure 10-1 shows the HMC panel that allows the selection of the desired TLS version.

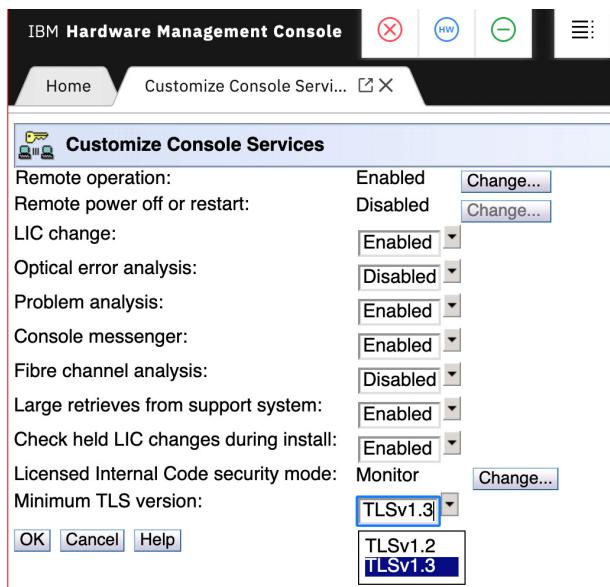


Figure 10-1 Minimum TLS version selection on HMC

- ▶ HMC/SE certification expiration updates

The new default for HMC and SE certification expiration is 398 days and it can be modified. This is driven by industry shorter certificates by the Apple Safari browser and iOS.

A Hardware Message will be posted for every expiration event 90, 30, 7 and 1 day before the expiry date is reached. If the expiration occurs, a HW message will then be posted daily. A call-home to IBM will also be placed 7, and 1 day before the certificate expiration and daily afterwards. These events will also be recorded in the Audit Log and Resource Link.

It is the clients responsibility to manage all certificates.

Figure 10-2 shows an example to change the number of days until the expiration occurs while you create a new certificate. If you like to change the expiration days for an existing certificate go to **Certificate Management** -> select **Valid Until** -> **Selected** -> **Modify**.

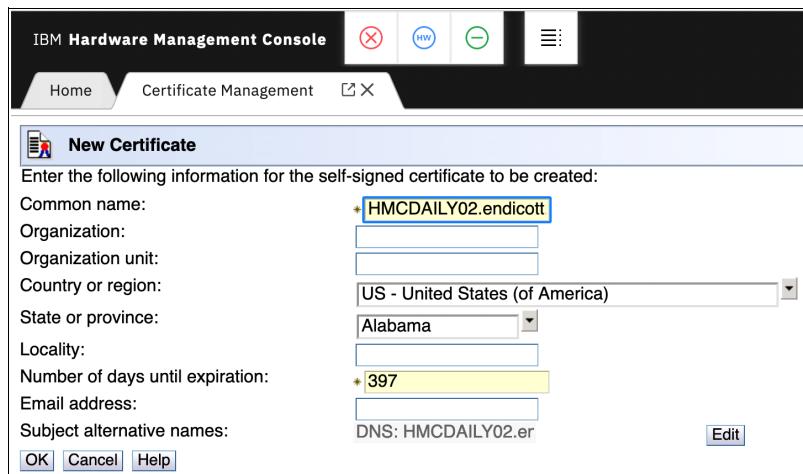


Figure 10-2 Certificate Management on the HMC

- ▶ Task Display Alter

The task **Display or Alter** on the HMC and SE is only possible for an ACSADMIN role. The roles for SYSPROG and SERVICE has no longer access to this task by default.

- ▶ View only HMC tasks

The following tasks can be assigned as **View only** for a specific role with Driver 51/Version 2.16.0:

- Hardware Messages
- Operating System Messages
- Manage Coupling Facility Port Enablement
- OSA Advanced Facilities
- Cryptographic Configuration
- Cryptographic Management
- Change LPAR Controls
- Change LPAR Group Controls
- Configure Channel Path On/Off
- Advanced Facilities
- Configure On/Off
- Manage System Time
- View Activation Profiles

Figure 10-3 shows how you can change the permission from **Edit** to **View only** for a task in **User Management**.

<input checked="" type="checkbox"/> Hardware Messages	<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;"> Edit <div style="float: right;">▼</div> </div> <div style="border: 1px solid #ccc; padding: 2px; display: inline-block; margin-top: 2px;"> Edit </div> <div style="border: 1px solid #ccc; padding: 2px; display: inline-block; background-color: #e0e0ff; color: #333; margin-top: 2px;"> View only </div>	Defined CPC, Fibre Channel Network, HMC Optical Network, Hardware Management Console, LPAR Image	Access Administrator Tasks, Advanced Operator Tasks, Operator Tasks, Service Representative Tasks, System Programmer Tasks
---	--	--	--

Figure 10-3 Selection of View only for a task

- ▶ New HMC Dashboard

With Driver 51/Version 2.16.0 there is a new Dashboard in the Home tab. It is the first page that appears when you login to the HMC. It replaces the Welcome panel used in previous HMC Drivers/Versions. Initially there are 4 widgets available:

- Systems health: Summarizes status of systems managed by the HMC
- Hardware messages: Manage hardware messages for all managed systems
- Frequently used tasks: Display your most frequently launched tasks
- What's new: Display information about the latest console feature

An example of the new Dashboard can be seen in Figure 10-4 on page 396.

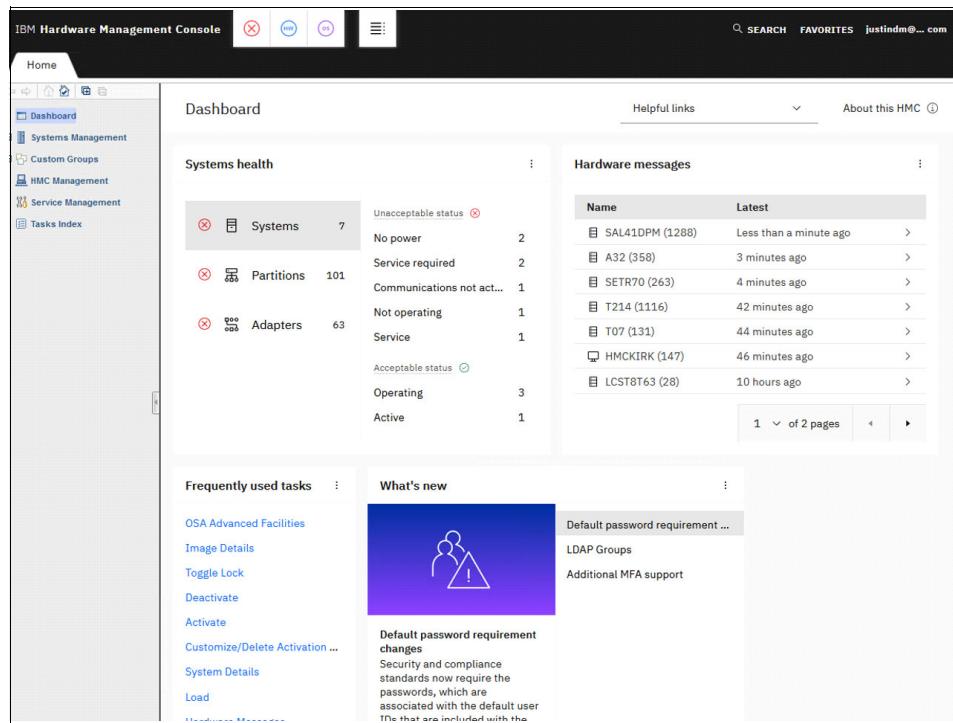


Figure 10-4 The new Dashboard on the HMC

Further under **Helpful links** you will find links to Resource Link, videos, and so on. And under **About this HMC** you can find information like the actual installed Bundle level. If the HMC is part of an HMA, **About this HMC** shows if this HMC has currently the Primary or Alternate SE running. Also the name of the peer HMC, where the 2nd SE is running, is displayed as you can see in Figure 10-5.

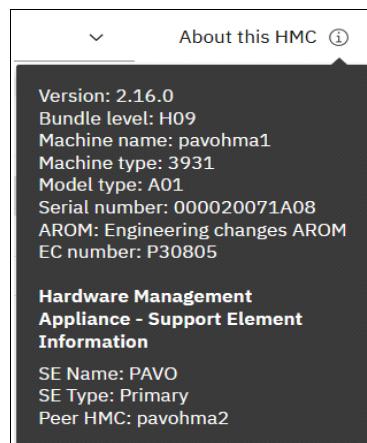


Figure 10-5 Example of About this HMC

- ▶ N-mode Power STP Imminent disruption Signal option

With **IBM z16 A02** and **IBM z16 AGZ** a new STP recovery signal called n-mode power has been implemented. A CTN with IBM z16 for both the Primary Time Server and the Backup Time server can be enabled (configured) to recover from a PTS/CTS failure by using the n-mode power (imminent power failure on the CTS (IBM z16 HW) configured with dual utility power).

For additional information see the *IBM Z Server Time Protocol Guide*, SG24-8480.

- ▶ BCPii enhancements for SE restarts

With [IBM z16 A02](#) and [IBM z16 AGZ](#) the SE and the BCPii for GDPS and System Automation (SA) has an improved communication to not disturb important task on each side. For example if you try to restart an SE and GDPS or SA is doing some important task, the restart panel ask for confirmation to force the restart as shown in Figure 10-6.

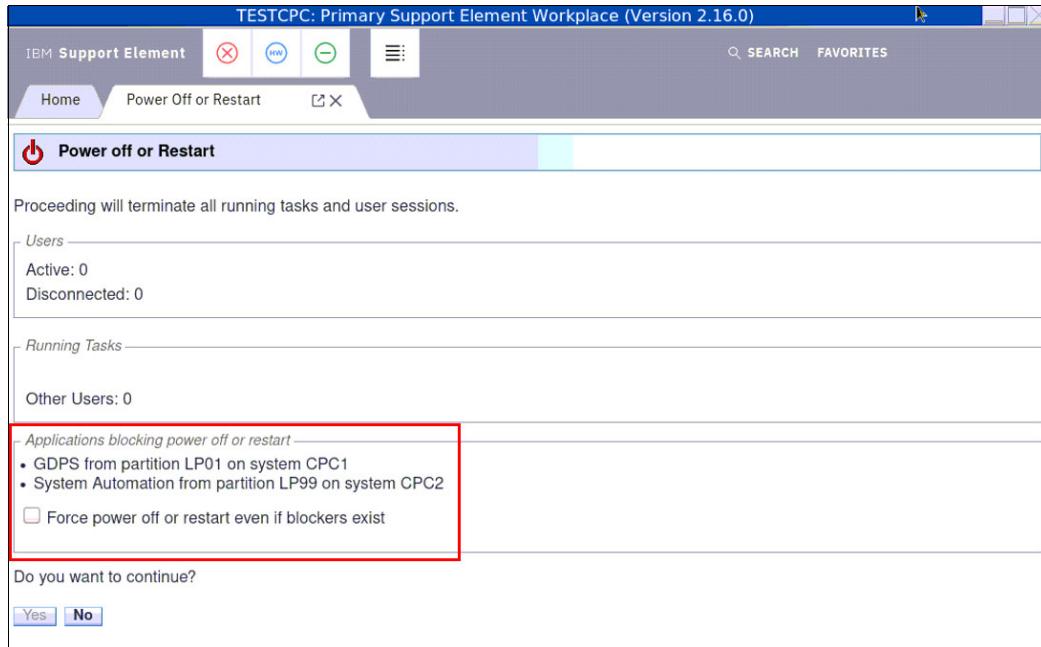


Figure 10-6 New information/question for SE power off or restart regarding GDPS or SA

- ▶ Create and delete profiles on HMC/SE via API/BCPii
 - Support is added for BCPii and Web Services API (WSAPI) on the SE and HMC for the following operations (these operations are not available for CPCs in DPM mode):
 - Create Reset Activation Profile – POST /api/cpcs/{cpc-id}/reset-activation-profiles
 - Create Load Activation Profile - POST /api/cpcs/{cpc-id}/load-activation-profiles
 - Create Image Activation Profile - POST /api/cpcs/{cpc-id}/image-activation-profiles
 - Create Group Profile - POST /api/cpcs/{cpc-id}/group-profiles
 - Delete Reset Activation Profile – DELETE /api/cpcs/{cpc-id}/reset-activation-profiles/{reset-activation-profile-name}
 - Delete Load Activation Profile – DELETE /api/cpcs/{cpc-id}/load-activation-profiles/{load-activation-profile-name}
 - Delete Image Activation Profile – DELETE /api/cpcs/{cpc-id}/image-activation-profiles/{image-activation-profile-name}
 - Delete Group Profile – DELETE /api/cpcs/{cpc-id}/group-profiles/{group-profile-name}

For more information please see *Hardware Management Console Web Services API Version 2.16.0*, SC27-2642.
 - ▶ BCPii v2 HWIREST API enhancements
- BCPii v2 support was added for “List Permitted Adapters” and extended as follows:
- “network-ports” query parameter has been added; returns the physical MAC address and port numbers of the OSA adapters (these operations are not available for CPCs in DPM mode)
 - “additional-properties” query parameter has been added; requests a list of properties (1->n) for retrieval

- Supported “additional-properties” are:
 - > state
 - > crypto-type
 - > physical-channel-status
 - > network-ports
- The “network-info” property has a new inner property known as “alternate-se”. The inner “alternate-se” property is identical in functionality to the “primary-se” property. It’s mapped to an array of “detailed-network-info” objects for each public network interface on the alternate SE. The “detailed-network-info” object contains various info for each interface, including ipv4/ipv6 addresses, domain name, and MAC address among other things.

For more information please see *Hardware Management Console Web Services API Version 2.16.0*, SC27-2642.

► **Web Services API for Secure Execution key management**

The HMC GUI has the functionalities for handling the keys (Host key, Global Hyper Protect key, Host Import Key) for Secure Execution available. Now, the management for these keys can also be done via Web Services API. For example the request URI to import a Secure Execution key is as following: *POST /api/cpcs/{cpc-id}/operations/import-se-key*

Input:

- file-content: KEy bundle file in base64-encoded string form
- type: “host”, “global”, or “host-import”
- force: override certain types of verification failures, if possible

For more information please see *Hardware Management Console Web Services API Version 2.16.0*, SC27-2642.

► **Report a Problem using HMC Web Services API interface**

You can report a problem for the HMC via task **Service Management -> Report a Console Problem**. You can report a problem for the CPC/LPAR via task **Service -> Report a Problem** (RaP).

Now you can also perform these tasks with the Web Services API interface:

- Report a Console Problem can be invoked with the following URI:
 - */api/console/operations/report-problem*
- Report a CPC Problem can be invoked with the following URI:
 - */api/cpcs/{cpc-id}/operations/report-problem*
Where {cpc-id} is the Object ID of the CPC object.
- Report a Logical Partition Problem can be invoked with the following URI:
 - */api/logical-partitions/{logical-partition-id}/operations/report-problem*
Where {logical-partition-id} is the Object ID of the Logical Partition object.
- Report a Partition Problem can be invoked with the following URI:
 - */api/partitions/{partition-id}/operations/report-problem*
Where {partition-id} is object ID of the Partition object.
- **Pattern Match Group child permission**

The **Custom Groups** task provides a mechanism for you to group system resources together in a single view. If you create a new group you can define a Resource Pattern (for example if you specified *abc.**, all the resources that begin with abc will be included in that group). With **IBM z16 A02** and **IBM z16 AGZ** also the for the child management permission for Pattern Match Groups is implemented.

- ▶ Summary of API version updates

In general there are lot of changes / updates in the HMC Web Services API. Not all are mentioned here in this Redbook. Please see the *Hardware Management Console Web Services API Version 2.16.0*, SC27-2642, “Summary of API version updates” for all new functions.

- ▶ Manage System Time Enhancements

Starting with IBM z14 there was a major change how STP management is handled on the HMC and SE. All STP management is now done via the **Manage System Time** on the HMC. With [IBM z16 A02](#) and [IBM z16 AGZ](#) further improvements are made on the wizards to guide you through the different configuration possibilities. The most of the changes are visual and should help you to have an even more clear guidance.

- ▶ System Availability Data changes

Transmit System Availability Data (TSAD) contains important data for the client and IBM. For example this task sends data to IBM Resource Link to have MCL and configuration information available. Also Capacity on Demand (CoD) information is exchanged. IBM proactively monitors RAS (Reliability, Availability, Serviceability) information via TSAD, for other potential actions or system improvements.

Users sometimes wanted to modify the execution time of this TSAD and accidentally deleted this TSAD Scheduled Operation.

With previous IBM zSystems generations than [IBM z16 A02](#) and [IBM z16 AGZ](#), Transmit System Availability Data (TSAD) was scheduled in the **Customize Scheduled Operation** task for the HMC and the SE/CPC. With [IBM z16 A02](#) and [IBM z16 AGZ](#) this TSAD functionality moved to the **Customize Console Services** task as you can see in Figure 10-7.

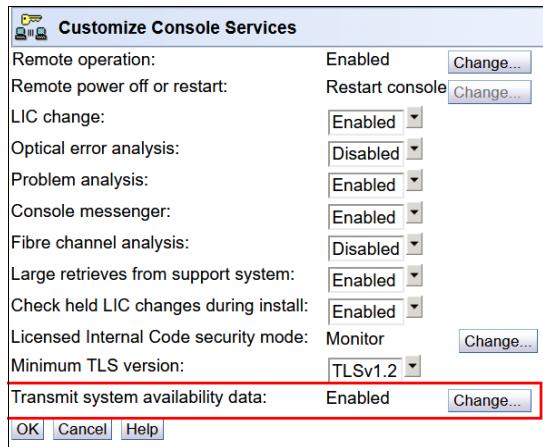


Figure 10-7 New option in Customize Console Services

In the new panel there is a weekly schedule for health and diagnostic data (which is also sent to IBM Resource Link). The active health checking feeds data on daily basis to IBM for automatically analyze your IBM zSystem for potential problems. The new panel is showed in Figure 10-8 on page 400.

Manage system availability collection

Enable sending system availability data to support teams for analysis.

Enable system availability analysis

Health and diagnostic data

Choose the date and time each week to send health and diagnostic data. Send now to immediately send data for analysis when requested by support.

Set weekly schedule

Monday	01:30	AM	GMT-1
--------	-------	----	-------

Send data now

GUIDANCE

This data is collected weekly which you can examine within Resource Link.

Active health checking

Choose the time each day to send active health checking data. Send now to immediately send data for analysis when requested by support.

Set daily schedule

15:30	AM	GMT-1
-------	----	-------

Send data now

GUIDANCE

This data is collected daily and automatically analyzed for potential problems.

Figure 10-8 Panel for Manage system availability collections

We highly recommend to enable System availability collection on your IBM zSystem, to exploit the full value of RAS support.

► HMC Monitor Systems Events e-mail domain setting enhancement

So far the e-mail for HMC System Events was sent from CONSOLENAME_EventMonitor@CONSOLENAME.DOMAINNAME. With [IBM z16 A02](#) and [IBM z16 AGZ](#) you can change the sender name as shown in Figure 10-9.

Event Monitor Summary

Provide the Simple Mail Transfer Protocol (SMTP) server information to enable monitoring. Create monitors by editing the samples or adding new ones and then enabling them.

SMTP Settings

Server:

Port: * 25

Send Mail As:

Notification delay (seconds): * 300

Figure 10-9 New sender option for HMC Event e-mail notifications

► HMC Data Replication enhancements

The task **Configure Data Replication** provides you the ability to exchange configuration data between linked HMCs. For example the User Profile Data can be replicated. An HMC can have 3 different roles for replication:

- Primary

An HMC in the role of primary is a data replication source to other HMCs that are in the

role of replica. A primary HMC can only be a source for data replication, it can not receive replicated data from other HMCs.

- Peer

An HMC in the role of peer is data replication source to other HMCs that are in the role of peer or replica. An HMC in this role receives replicated data from other HMCs that are in the role of peer or primary.

- Replica

An HMC in the role of replica receives replicated data from other HMCs that are in the role of primary or peer. Modifications to replicated data types on a replica HMC are disallowed.

Further a HMC can also be disabled for data replication. An example of selection of the roles can be seen on Figure 10-10.

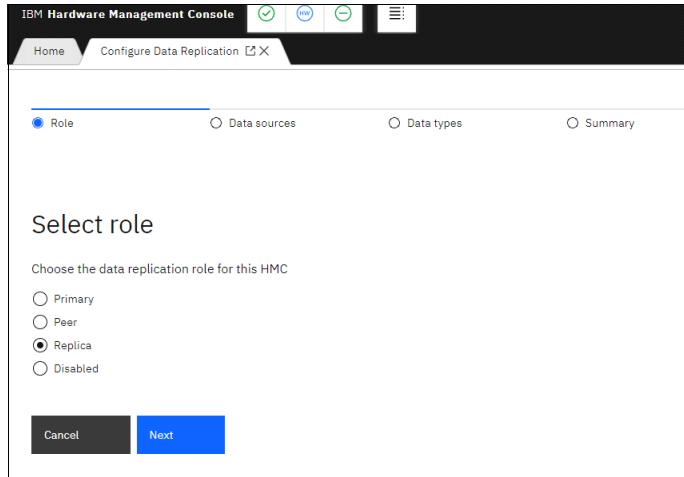


Figure 10-10 Selection of different roles for Data Replication on the HMC

With Driver 51/Version 2.16.0 we improved the notification if you are on a HMC with role Replica and open a panel where data replication is involved. On Figure 10-11 you can see an example of an notification on a HMC with role Replica.

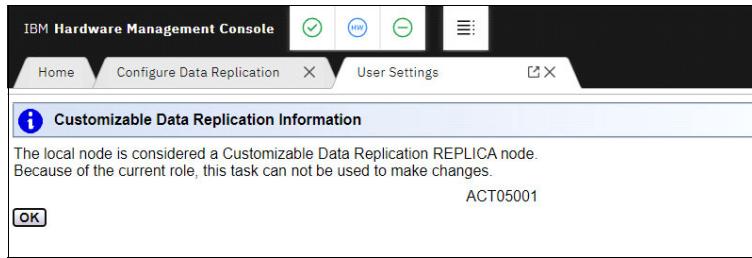


Figure 10-11 Notification for a task with has relation to Data Replication on a HMC with Replica role

Also the wizard to do the settings for Data Replication has changed for better and common visibility.

There is a new category for “Data types” with the name “Certificates”. Trusted server certificates (not the secure boot certificates) can be replicated. Trusted server certificates can be for:

- LDAP server
- MFA server
- 3270 server

- Syslog server
 - FTP server
- Remote Code Load (RCL)

The Remote Code Load for IBM zSystems allows IBM to upgrade a machine remotely by working with the client to schedule a date and time for the code load and monitor the process to make sure it completes successfully.

This feature allows you to schedule one or multiple Single Step Code Load for an HMC or SE.

On an [IBM z16 A02](#) and [IBM z16 AGZ](#) with the HMA feature, RCL can be scheduled to update both HMCs and the firmware will automatically manage Alternate SE switches to ensure each HMC can be updated without the Primary SE being present.

In general you first have to generate a token on the HMC task **Manage Remote Firmware Updates**. Afterwards you can schedule a RCL on Resource Link -> **Fixes -> Licensed internal code -> Remote Code Load request**.

Important to understand is, that for IBM zSystems, IBM support will not connect to your IBM zSystems from outside to do the RCL. The RCL is managed on your HMC.

More info about RCL can be found here:

<https://www-01.ibm.com/servers/resourcelink/lib03010.nsf/pages/remoteCodeLoadForIbmZFirmware?OpenDocument>

Statement of direction:

Firmware update process: IBM z16 is planned to be the last server family to support IBM service support representatives (SSRs) onsite performing firmware updates without an additional premium service contract. The IBM Z Remote Code Load (RCL) option, which was introduced on IBM z15, is available without an additional premium service contract. With IBM z15, and now IBM z16, clients can request a remote code load or they can choose the SSR onsite method for their firmware update. IBM recommends that clients try the RCL option on IBM z15 or IBM z16 to see for themselves that IBM provides the same quality service through RCL.

- BCPii Query of Crypto info

Prior to Driver 51/Version 2.16.0 automation support for querying Crypto information was only provided with SNMP API interfaces. With [IBM z16 A02](#) and [IBM z16 AGZ](#) support is added for BPCII v2/HWIREST APIs.

- Crypto automatic toggling

In rare cases, after an microcode update, the Crypto adapter has to be configured off/on to get the loaded firmware activated. With [IBM z16 A02](#) and [IBM z16 AGZ](#) there is a new function that gives you the possibility to select one or more Crypto adapters and do an automated serialized update of each card.

Important: Configure off/on operation will be concurrent to the Operating Systems activities only if the Crypto adapters are configured redundantly. The automated process will not do the redundancy check for you. If you are unsure, we recommend to toggle each adapter off/on manually.

The Figure 10-12 on page 403 shows two examples of the panel for automatic firmware activation on Crypto adapters.

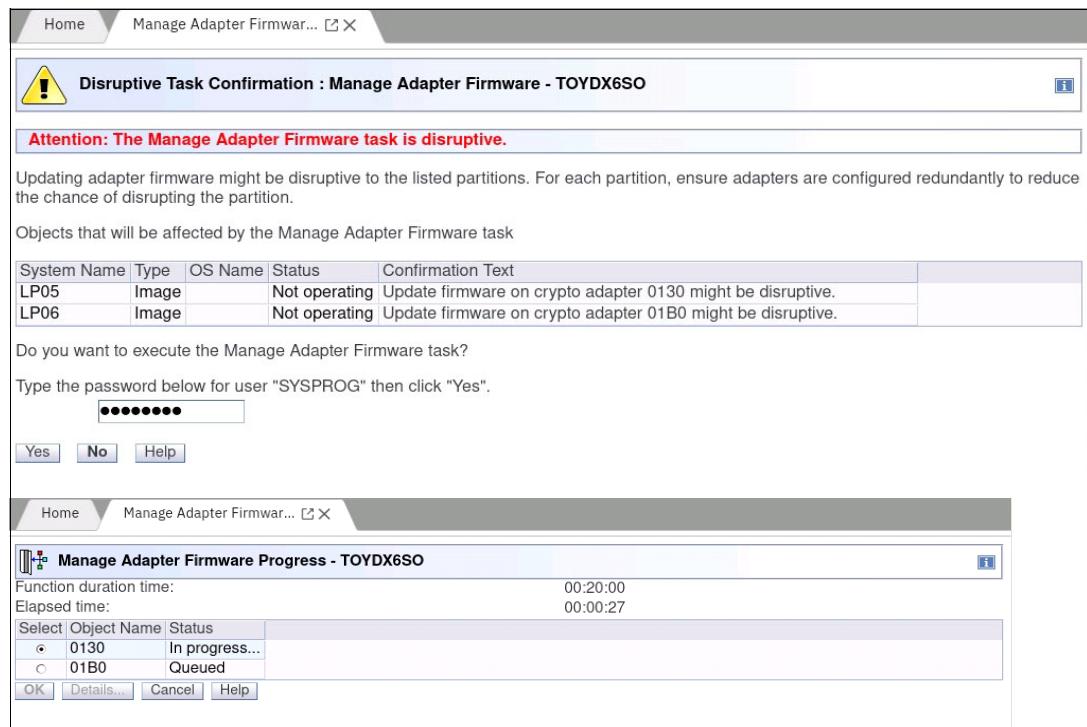


Figure 10-12 Crypto adapters automatic firmware activation

- ▶ Default HMC and SE users security changes

With [IBM z16 A02](#) and [IBM z16 AGZ](#) default *users* ADVANCED, OPERATOR, STORAGEADMIN, and SYSPROG will no longer be pre installed on the new shipped HMCs and SEs.

Default *user roles* for ADVANCED, OPERATOR, STORAGEADMIN, and SYSPROG will be shipped and the user IDs can be created using the provided default roles.

Any default user ID which was part of a previous HMC can be carried forward to the new HMCs as part of a MES upgrade or via the tasks, **User Profile Data** for the **Save/Restore Customizable Console Data** or **Configure Data Replication**.

At first logon, you have to change the default password for the ACSADMIN and SERVICE user.

Note:

- ▶ It is client responsibility to maintain the passwords and user IDs.
- ▶ For the SERVICE user the client must be able to provide the password at any time to the IBM System Service Representative (SSR). The client need an established process to avoid service delays, not having the password available for the IBM technician.

- ▶ Report a Problem update

If your IBM zSystems has an hardware or microcode problem and no automatic call went out to IBM, you can open a case with the task **Report a Problem** (RaP). Before [IBM z16 A02](#) and [IBM z16 AGZ](#) you could do this task for the HMC (**Service Management -> Report a Problem**) or an CPC (**Service -> Report a Problem**). Further its now possible to select an LPAR for RaP.

For better clarification the RaP for the HMC is now called **Report a Console Problem**.

- ▶ CoD automation scripts

Statement of Direction: IBM z16 A02 and IBM z16 AGZ is planned to be the last server to support legacy CoD unique record type automation interfaces. For example, legacy command HWMCA_ACTIVATE_CBU_COMMAND has to change to HWMCA_ADD_CAPACITY_COMMAND. We suggest you to start now to change your automation scripts accordingly. For more information see *Capacity on Demand User's Guide*, SC28-7025.

- ▶ Sustainability / Environmental monitoring

As sustainability gets more important for many clients, there was a request to have more monitoring possibilities. The following features at HMC level are improved:

- The previous task **Environmental Efficiency Statistics** is replaced by a new task called **Environmental Dashboard**. With the new **Environmental Dashboard** the following new observations are now possible:
 - View of power utilized by components assigned to individual partitions
 - View of power used by the infrastructure components (including top of rack switches, SE/HMAs, and PDUs) which are not included in the power view for partitions
 - View of power of unused I/O adapters / components that are not assigned to any partition (including standby components)
 - Broader selectable time ranges for metrics view for historical trending data
 - Display selected system and partition metrics in line chart and tabular views
 - Filters for different views
 - Export metric data
- The task **Monitors Dashboard** is enhanced with the following new information and possibilities:
 - Total Partition Power Consumption (kW)
 - Total Infrastructure Power Consumption (kW)
 - Total Unassigned Power Consumption (kW)
 - Partition power consumption per Partition (kW)
- HMC Web Services API has new parameters to get information for common Data Center Infrastructure Management (DCIM) tools. To get the new Metric field names for the values mentioned in the last bullet, please check the *Hardware Management Console Web Services API Version 2.16.0*, SC27-2642.

For more information about this topic see 10.5.6, “Monitoring” on page 424 and use the HMC console help system or see Resource Link² (<https://www.ibm.com/servers/resourcelink>), select **Library**, the applicable server, then select the **HMC Version 2.16.0 help file content**.

10.2.2 Hardware Management Appliance (HMA)

Started with IBM z15, the two 1U rack-mounted servers provide increased hardware capacity, which allows instances of both HMC and SE to run collocated on the same physical server. The SE code runs as a virtual guest of the Hardware Management Console code.

Figure 10-13 on page 405 illustrates the HMA and relation to HMCs and SEs.

² IBM ID is required for authentication and access to Resource Link.

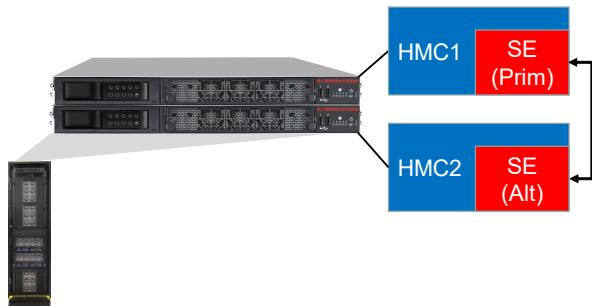


Figure 10-13 HMA: HMCs and SEs

The SE interface can be accessed from the HMC by using the task **Single Object Operation** on the HMC.

The optional HMA feature FC 0129 consists of the HMC code installed on the two 1U servers collocated with the SE code. The servers are configured with processor, memory, storage and networking resources to support all processing and security requirements for running both HMC and SE code. The two HMCs, named HMC1 and HMC2 from manufacturing (you can change the names) are configured as independent HMCs. They are not a Primary or Alternate HMCs. HMC Data Replication can be established if so desired. For example Data Replication can be used to replicate user credentials among different HMCs.

Important: With the IBM Hardware Management Appliance, shutdown or restart of the HMC which has the Primary SE code as guest, restarts also the Primary SE code. Just an application restart of the HMC is not disruptive to the guest SE code.

The two SE code instances are clustered for high availability. One SE code runs the Primary SE the other the Alternate SE. These two SEs perform data mirroring and their role can be switched for maintenance purposes.

Switching the Primary and Alternate SE roles is important, as HMC microcode installation can only be performed on the HMC which runs the Alternate SE as a guest.

Recommendation: It is recommended to limit the number of HMA instances per location or datacenter to two (2). Cabling, management and microcode updates can become overwhelmingly complicated.

You can use the [IBM z16 A02](#) and [IBM z16 AGZ](#) with carry forward HMCs and no initial order of FC 0129 (HMA for [IBM z16 A02](#) and [IBM z16 AGZ](#)). Order and install of the HMA FC 0129 is possible at a later time as an upgrade.

10.2.3 HMC and SE servers

The two 1U rack-mounted hardware servers that are installed in the rack are used for HMC (optional HMA feature) and SE functionality shown in Figure 10-14.



Figure 10-14 HMC and SE servers (front view)

HMC and SE Keyboard Mouse Monitor

With IBM z16 A02 and IBM z16 AGZ, a Keyboard Mouse Monitor (KMM) device located in a cubby at the front of the frame can be used to work with the HMCs and SEs.

The KMM device information for the IBM z16 A02 are shown in Figure 10-15 on page 407.

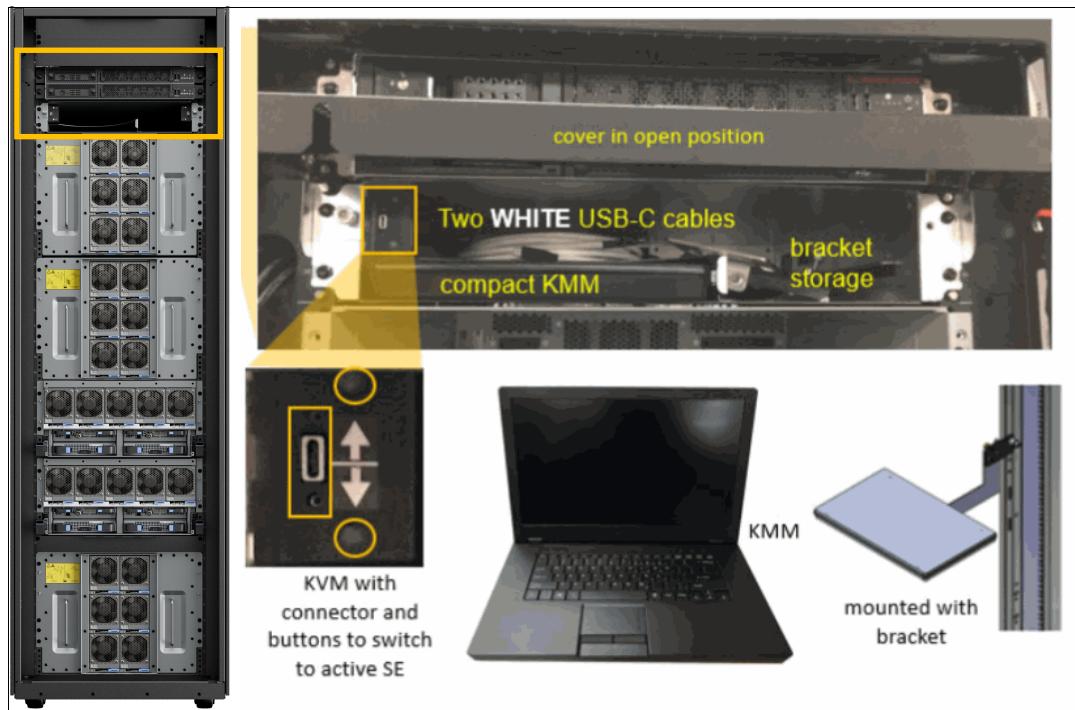


Figure 10-15 HMC / SE KMM device

For the IBM z16 AGZ, no mounting bracket is shipped. The KMM is expected to be supported by a locally provided (non-IBM supplied) cart.

Consider the following points:

- ▶ The KMM device is intended to be used by the IBM System Service Representative (SSR) only. In case the remote access to the HMC and SE is not possible, the client can use it as an emergency option. The SEs can be local / on-site managed by the Virtual Support Element Management task, as shown in Figure 10-16.

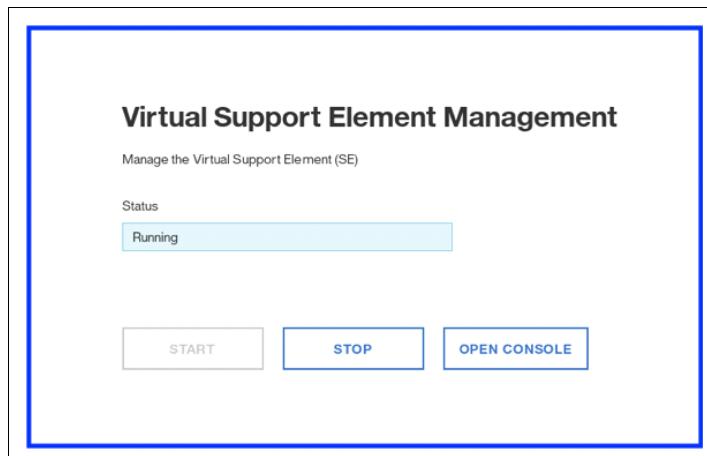


Figure 10-16 Local Virtual Support Element Management

- ▶ One KMM is provided
- ▶ The USB-C cable can be used to plug the device into a KVM switch at the front or the rear of the rack when servicing the system.

- ▶ Switching between servers is done by using buttons next to the USB-C port (see Figure 10-15 on page 407). The KMM screen also indicates which server is selected
- ▶ The KMM mounting bracket (only provided with z16 AGZ) can be used to mount the device to any frame in the system (either in front or rear)
- ▶ The KMM can be used on any IBM z16 system (no affinity to system with which it is shipped)

For more information about the KMM and how to attach it to the frame, see *3932 Installation Manual*, GC28-7041.

10.2.4 USB support for HMC and SE

Because a DVD drive is not available on the HMC or SE, this section describes two according service and functional operations for HMC Driver 51/Version 2.16.0.

Microcode load

Microcode can be loaded by using the following options:

- ▶ USB
 - If the HMC and SE code is shipped on a USB drive when a new system is ordered, the load procedure is similar to that used with a DVD.
- ▶ Electronic
 - If USB load is not allowed, or if FC 0846 (no physical media option) is ordered, an ISO image is used for a firmware load over a local area network (LAN). The ISO image can be downloaded through zRSF or an FTP (/FTPS/SFTP) server accessible in the LAN.

Important: The ISO image server *must* be in the same IP subnet with the target system to load the HMC or SE ISO.

Operating system load from removable media or server

z/OS, z/VM, z/VSE, and Linux on IBM Z are available via USB or network distribution. z/TPF does not use the HMC for code load.

10.2.5 SE Driver/Version support with the HMC Driver 51/Version 2.16.0

The Driver of the HMC and SE is equivalent to a specific HMC and SE Version:

- ▶ Driver 36 is equivalent to Version 2.14.1
- ▶ Driver 41 is equivalent to Version 2.15.0
- ▶ Driver 51 is equivalent to version 2.16.0

An HMC with Driver 51/Version 2.16.0 supports N-2 IBM zSystems server generations. Some functions that are available on Driver 51/ Version 2.16.0 and later are supported only when the HMC is connected to an IBM zSystems with Driver 51/Version 2.16.0.

The following SE Drivers/Versions are supported by the HMC Driver 51/Version 2.16.0 as listed in Table 10-1.

Table 10-1 SE supported with HMC Driver 51/Version 2.16.0

IBM zSystems product name	Machine type	SE Driver	SE Version
IBM z16 A02 IBM z16 AGZ	3932	51	2.16.0
IBM z16 A01	3931	51	2.16.0
IBM z15 T02	8562	41	2.15.0
IBM z15 T01	8561	41	2.15.0
IBM z14 ZR1	3907	36	2.14.1
IBM z14 M0x	3906	36	2.14.1

10.3 HMC and SE connectivity

The connectivity for multiple CPC generations and mixing standalone HMCs and HMAs environments (IBM z16 N-2 only) is shown in Figure 10-17.

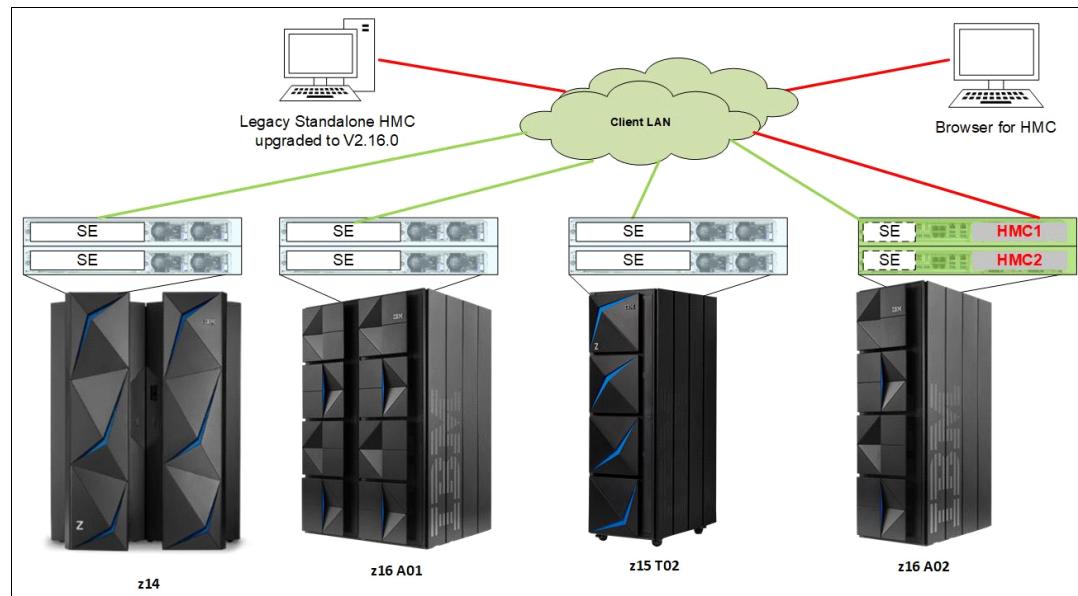


Figure 10-17 IBM z16 A02 and IBM z16 AGZ HMC/SE connectivity with multiple CPCs

Various methods are available for setting up the network. Designing and planning the HMC and SE connectivity is the clients' responsibility, based on the environment's connectivity and security requirements. For further references to setting up the HMC/SE communication please, see (per your configuration):

- ▶ *3932 Single Frame Installation Manual for Physical Planning (A02/LA2), GC28-7040*
- ▶ *IBM 3932 Installation Manual for Physical Planning (AGZ/AGL), GC28-7035*
- ▶ *IBM 3932 Installation Manual (A02/LA2), GC28-7041*
- ▶ *IBM 3932 Installation Manual (AGZ/AGL), GC28-7036.*

Security: The configuration of network components, such as routers or firewalls, is beyond the scope of this document. Whenever the networks are interconnected, security exposures can exist. For more information about HMC security, see *Hardware Management Console Security*, SC28-7027 and *Integrating the Hardware Management Console's Broadband Remote Support Facility into your Enterprise*, SC28-7026.

For more information about the HMC settings that are related to access and security, see Resource Link (<https://www.ibm.com/servers/resourcelink>), select **Library**, the applicable server, then select either **HMC Version 2.16.0 help file content** or **SE Version 2.16.0 help file content**.

10.3.1 Hardware Management Appliance (HMA) HMC/SE connectivity

On **IBM z16 A02** and **IBM z16 AGZ**, two HMCs will be delivered with the Hardware Management Appliance (HMA) feature FC 0129. The HMC code runs on the two integrated 1U rack-mounted servers in the frame.

The SE code runs virtualized on the integrated two HMCs on the two integrated 1U rack-mounted servers. One SE is the Primary SE (active) and the other is the Alternate SE (backup). With the HMA, each HMC and each SE has its own two physical Ethernet RJ45 ports as shown in Figure 10-18. The red lines illustrates the RJ45 Ethernet connectivity for the SEs and the green lines for the HMCs.

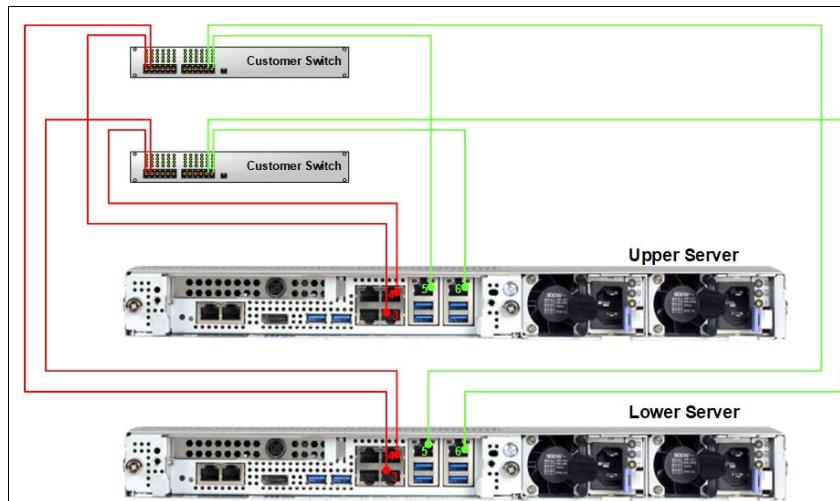


Figure 10-18 Hardware Management Appliance connectivity

Note: The HMC *must* be connected to the SEs by using a customer provided switch. Otherwise you will never get a communication from the HMCs to the SEs. Direct Ethernet connections between HMC and SE is *not* supported.

10.3.2 Legacy standalone HMC connectivity

You can use the IBM z16 A02 and IBM z16 AGZ with previous ordered standalone HMCs (Tower or Rack) and no HMA. It is possible to order the HMA FC 0129 at a later point in time as an MES.

The HMC communicates with the SE through a customer-supplied Ethernet Switch (two Switches are highly recommended for redundancy). Each SE and each HMC has two Ethernet RJ45 ports. An example of how the connections are without having an HMA is shown in Figure 10-19.

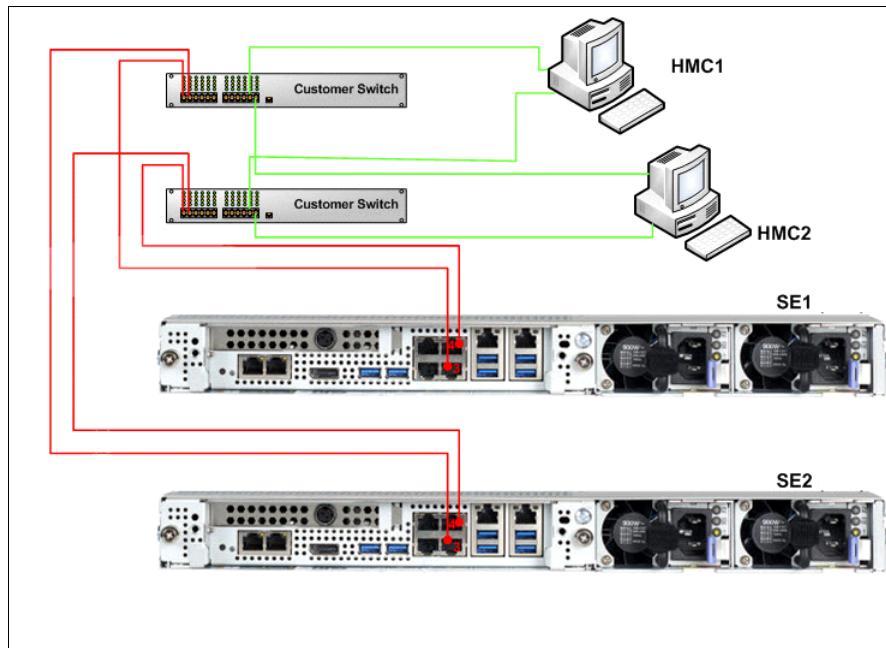


Figure 10-19 HMC / SE connectivity without HMA

Note: The HMC *must* be connected to the SEs by using a customer provided switch. Otherwise you will never get a communication from the HMCs to the SEs. Direct connection between HMC and SEs is *not* supported.

10.3.3 Network planning for the HMC, SE, and ETS

Plan the HMC and SE network connectivity carefully to allow for current and future use. Many of the IBM zSystems capabilities benefit from the various network connectivity options. The following functions are examples which depend on the HMC connectivity:

- ▶ Lightweight Directory Access Protocol (LDAP) support, which can be used for HMC user authentication
- ▶ Network Time Protocol (NTP)
- ▶ RSF through broadband
- ▶ HMC remote access and HMC Mobile
- ▶ RSA SecurID support
- ▶ Multi-Factor Authentication with TOTP³

External Time Source (ETS) for STP

Other than for previous IBM zSystems generations, [IBM z16 A02](#) and [IBM z16 AGZ](#) is connected to the ETS via two separated RJ45 Ethernet cables which are connected in the CPC drawer(s) front BMC/OSC cards. These two cables and according IP configuration have

³ Time-based One Time Password

to be planned if ETS is used. For further information see <<Chapter 2, Oscillator cards / 10.3.6, “Assigning TCP/IP addresses to the HMC, SE, and ETS” on page 414.>>

HMC File Transfer support

FTP, FTPS, and SFTP protocols are supported on the HMC and SE. All three file transfer protocols require login ID and password (credentials).

FTPS is based on Secure Sockets Layer cryptographic protocol (SSL) and requires certificates to authenticate the servers. SFTP is based on Secure Shell protocol (SSH) and requires SSH keys to authenticate the servers. Certificates and key pairs are hosted in the [IBM z16 A02](#) and [IBM z16 AGZ](#) HMC.

The following FTP server requirements must be met:

- ▶ Support passive data connections
- ▶ A server configuration that allows the client to connect on an ephemeral (temporary or non-registered) port

The following FTPS server requirements must be met:

- ▶ Operate in “explicit” mode
- ▶ Allows a server to offer secure and unsecured connections
- ▶ Must support “passive” data connections
- ▶ Must support secure data connections

The SFTP server must support password-based authentication.

The file transfer server choices for HMC are shown in Figure 10-20.

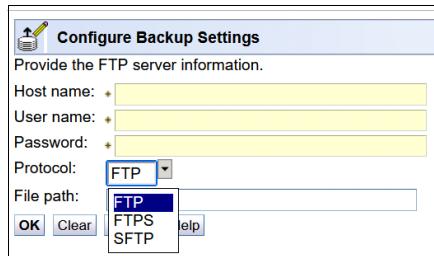


Figure 10-20 FTP protocols drop-down list

FTP through HMC

It is recommended to keep the HMCs and SEs on an isolated network. This approach prevents SEs initiating FTP connections with outside networks and applies to all supported file transfer protocols (FTP, FTPS, and SFTP).

All FTP connections that originate from the SE are taken to the HMC. Secure FTP server credentials must be imported to one or more managing HMC consoles.

After the HMC console completes all FTP operations, the HMC console performs the FTP operation on SE’s behalf and returns the results.

Secure console-to-console communications

The [IBM z16 A02](#) and [IBM z16 AGZ](#) HMC has an industry standard-based, password-driven cryptography system. The Domain Security Settings are used to provide authentication and high-quality encryption. We recommend that clients use unique Domain Security settings to provide maximum security. This provides greater security than anonymous cipher suites, even if the default settings are used.

To allow greater flexibility in password selection, the password limit was increased to 64 characters and special characters are allowed.

For more information about HMC networks, see the following resources:

- ▶ The HMC and SE Driver 51/Version 2.16.0 console help system
- ▶ Resource Link (<https://www.ibm.com/servers/resourcelink>), select **Library**, the applicable server, then select either **HMC Version 2.16.0 help file content** or **SE Version 2.16.0 help file content**.

10.3.4 Hardware considerations

The following Hardware considerations are important for IBM z16 A02 and IBM z16 AGZ:

- ▶ IBM does not provide Ethernet RJ45 cables with the system for connections between HMC, SE, Switches and ETS
- ▶ IBM does not provide Ethernet switches with the system for HMC and SE communication

Ethernet switches

Ethernet switches for HMC and SE connectivity must be provided by the client. Existing supported switches can still be used.

Ethernet switches often include the following characteristics:

- ▶ A total of 16 auto-negotiation ports
- ▶ 100/1000 Mbps data rate
- ▶ Full or half duplex operation
- ▶ Auto medium-dependent interface crossover (MDIX) on all ports
- ▶ Port status LEDs
- ▶ Copper RJ45 connections

Note: The recommendation is to use an Ethernet switch with 1000 Mbps/Full duplex support.

10.3.5 TCP/IP Version 6 on the HMC and SE

The HMC and SE can communicate by using IPv4, IPv6, or both.

IPv6 link-local addresses feature the following characteristics:

- ▶ Every IPv6 network interface is assigned a link-local IP address.
- ▶ A link-local address is used on a single link (subnet) only and is never routed.
- ▶ Two IPv6-capable hosts on a subnet can communicate by using link-local addresses, without having any other IP addresses assigned. This is the reason, if HMC to SE IPv6 link-local is working, that the SE/CPC appears in **System Management -> Unmanaged Systems** on the HMC.

10.3.6 Assigning TCP/IP addresses to the HMC, SE, and ETS

Use the worksheet in the Installation Manual or GC28-7041 (A02), GC28-7036 (AGZ) *Hardware Management Appliance and Support Element customer configuration requirements*, to plan your HMC, SE, and ETS IP configuration.

An **HMC** can have the following IP configurations:

- ▶ Statically assigned IPv4 or statically assigned IPv6 addresses
- ▶ Dynamic Host Configuration Protocol (DHCP)-assigned IPv4 or DHCP-assigned IPv6 addressees
- ▶ Auto-configured IPv6:
 - Link-local is assigned to every network interface.
 - Router-advertised, which is broadcast from the router, can be combined with a Media Access Control (MAC) address to create a unique address.
 - Privacy extensions can be enabled for these addresses as a way to avoid the use of the MAC address as part of the address to ensure uniqueness.

An **SE** can have the following IP addresses:

- ▶ Statically assigned IPv4 or statically assigned IPv6
- ▶ Auto-configured IPv6 as link-local or router-advertised

IP addresses on the SE cannot be dynamically assigned through DHCP to ensure repeatable address assignments. DHCP privacy extensions are not used on the SE.

The HMC uses IPv4 and IPv6 multicasting⁴ to automatically discover the SEs. The HMC Network Diagnostic Information task can be used to identify the IP addresses (IPv4 and IPv6) which are used by the HMC to communicate to the SEs (of a CPC).

IPv6 addresses are easily identified. A fully qualified IPV6 address features 16 bytes. It is written as eight 16-bit hexadecimal blocks that are separated by colons, as shown in the following example:

2001:0db8:0000:0000:0202:b3ff:fe1e:8329

Because many IPv6 addresses are not fully qualified, shorthand notation can be used. In shorthand notation, the leading zeros can be omitted, and a series of consecutive zeros can be replaced with a double colon. The address in the previous example also can be written in the following manner:

2001:db8::202:b3ff:fe1e:8329

If an IPv6 address is assigned to the HMC for remote operations that use a web browser, browse to it by specifying that address. The address must be surrounded with square brackets in the browser's address field, as shown in the following example:

`https://[fdab:1b89:fc07:1:201:6cff:fe72:ba7c]`

The use of link-local addresses must be supported by your browser.

An **STP/ETS Network** can have the following IP addresses:

- ▶ Statically assigned IPv4 or statically assigned IPv6
- ▶ Auto-configured IPv6 as link-local or router-advertised

⁴ For the customer-supplied switch, multicast must be enabled at the switch level.

The configuration for the STP/ETS Networks is done on the SE in task **Customize Network Settings** as you can see in Figure 10-21.

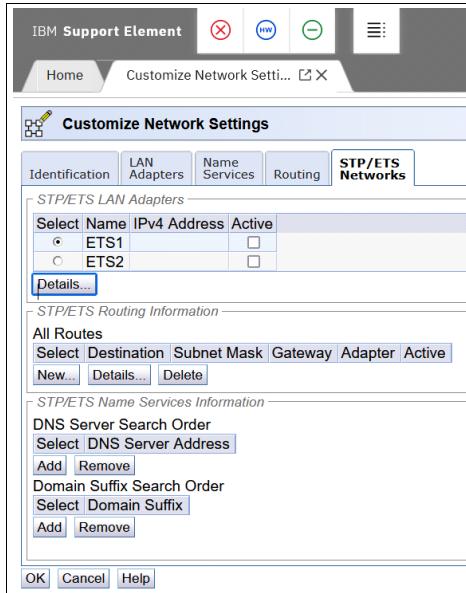


Figure 10-21 STP/ETS Networks setting on the Support Element

10.3.7 HMC Multi-factor authentication

Multi-factor authentication is an optional and configurable feature on per-user, per-template basis. It enhances security by requiring not only what you know (which is first factor) but also what you have available, which means that only a person who owns a specific phone number can log in.

Multi-factor authentication first factor is login and password; the second factor is TOTP (Time-based One-Time Password) that is sent to your smartphone, desktop, or app (for example, Google Authenticator or IBM Verify). This TOTP is defined in RFC 6238 standard and uses a cryptographic hash function that combines a secret key with the current time to generate a one-time password.

The secret key is generated by HMC/SE/TKE while the user is performing first factor logon. The secret key is known only to HMC/SE/TKE and to the user's smartphone. For that reason, it must be protected as much as your first factor password.

Multi-factor authentication code (MFA code) that was generated as a second factor is time-sensitive. Therefore, it is important to remember that it should be used soon after it is generated.

The algorithm within the HMC that is responsible for MFA code generation changes the code every 30 seconds. However, to make things easier, the HMC and SE console accepts current, previous, and next MFA codes. It is also important to have HMC, SE, and smartphone clocks synchronized. If the clocks are not synchronized, the MFA logon attempt fails. Time zone differences are irrelevant because the MFA code algorithm uses UTC.

On IBM z15, HMC Driver 41/Version 2.15.0 provided integration of HMC authentication and z/OS MFA support, which means RSA SecurID authentication is achieved by way of centralized support from IBM MFA for z/OS, with the MFA policy defined in RACF and the HMC IDs assigned to RACF user IDs. The RSA SecurID passcode (from an RSA SecurID

Token) is verified by the RSA authentication server. This authentication is supported on HMC only, *not* on the SE.

The following support was added with IBM z16 Driver 51/Version 2.16.0:

- ▶ Enhanced Multi-Factor Authentication (MFA) functions
 - Certificates
 - Personal Identity Verification (PIV)
 - Common Access Card (CAC)
 - Certificates on USB keys
 - Generic Remote Authentication Dia-in User Service (RADIUS) allows for support of all various RADIUS factor types. Involves customer provided RADIUS server.

Also, Driver 51/Version 2.16.0 provides support for IBM zSystems Multi-Factor authentication for RedHat Enterprise Linux Server or SUSE Linux Enterprise Server running on z/VM or native in an LPAR.

10.4 Remote Support Facility

The HMC Remote Support Facility (RSF) provides important communication to a centralized IBM support network for hardware problem reporting and service. The following types of communication are provided:

- ▶ Problem reporting and repair data
- ▶ Microcode Change Level (MCL) delivery
- ▶ Hardware inventory data, which is also known as vital product data (VPD)
- ▶ Health and diagnostic data
- ▶ Capacity on Demand (CoD) enablement

10.4.1 Security characteristics

The following security characteristics are in effect:

- ▶ RSF requests are started always from the HMC to IBM. An inbound connection is never started from the IBM Service Support System to the HMC.
- ▶ All data that is transferred between the HMC and the IBM Service Support System is encrypted with Transport Layer Security (TLS) encryption.
- ▶ When starting the TLS-encrypted connection, the HMC validates the trusted host with the digital signature that is issued for the IBM Service Support System.
- ▶ Data that is sent to the IBM Service Support System consists of hardware problem and configuration data.

More information: For more information about the benefits of Broadband RSF and the TLS-secured protocol, and a sample configuration for the Broadband RSF connection, see *Integrating the HMC Broadband Remote Support Facility into your Enterprise*, SC28-7026.

10.4.2 RSF connections to IBM and Enhanced IBM Service Support System

To have the best availability and redundancy and to be prepared for the future, the HMC connects to IBM by using the internet to the IBM Remote Support Facility (RSF) in the following manner: transmission using a domain name server (DNS). The DNS has to be

configured on the HMC if a proxy for RSF is not used. If a proxy for RSF is used, the proxy can provide the DNS.

The following host names and IP addresses are used and your network infrastructure must allow the HMC to access RSF:

- ▶ esupport.ibm.com on port 443
- ▶ Using IPv4 requires outbound connectivity to the following IP addresses with port 443:
 - 129.42.21.70
 - 129.42.18.70
 - 129.42.19.70
 - 129.42.54.189
 - 129.42.56.189
 - 129.42.60.189
- ▶ Using IPv6 requires outbound connectivity to the following IP addresses with port 443:
 - 2607:f0d0:3901:33:129:42:21:70
 - 2607:f0d0:1f01:9f:129:42:18:70
 - 2607:f0d0:2601:13:129:42:19:70
 - 2620:0:6c0:200:129:42:54:189
 - 2620:0:6c2:200:129:42:56:189
 - 2620:0:6c4:200:129:42:60:189

Note: All other previous IP addresses are no longer supported.

10.5 HMC and SE capabilities

The HMC and SE has many capabilities. This section describes some key areas. For more information about these capabilities, see the HMC and SE Driver 51/Version 2.16.0 console help system or see Resource Link (<https://www.ibm.com/servers/resourcelink>), select **Library**, the applicable server, then select either **HMC Version 2.16.0 help file content** or **SE Version 2.16.0 help file content**.

With the introduction of the DPM mode for mainly LinuxONE management, the user interface and user interaction with the HMC changed dramatically; the capabilities underneath are still the same. The figures and descriptions in this section only covers the traditional Processor Resource/Systems Manager (PR/SM) mode.

10.5.1 Central processor complex management

The HMC is the primary place for CPC control. For example, the input/output configuration data set (IOCDs) includes definitions of LPARs, channel subsystems, control units, and devices, and their accessibility from LPARs. IOCDs can be created and put into production from the HMC.

The HMC is used to start the power-on reset (POR) of the system. During the POR, processor units (PUs) are characterized and placed into their respective pools, memory is put into a single storage pool, and the IOCDs is loaded and started into the hardware system area (HSA).

The hardware messages task displays hardware-related messages at the CPC, LPAR, or SE level. It also displays hardware messages that relate to the HMC.

10.5.2 LPAR management

Use the HMC to define LPAR properties, such as the number of processors of each type, how many are reserved, and how much memory is assigned to it. These parameters are defined in LPAR profiles and stored on the SE.

Because Processor Resource/Systems Manager (PR/SM) must manage LPAR access to processors and the initial weights of each partition, weights are used to prioritize partition access to processors.

You can use the Load task on the HMC to perform an IPL of an operating system. This task causes a program to be read from a designated device, and starts that program. You can perform the IPL of the operating system from storage, the USB flash memory drive (UFD), or an FTP server.

When an LPAR is active and an operating system is running in it, you can use the HMC to dynamically change certain LPAR parameters. The HMC provides an interface to change partition weights, add logical processors to partitions, and add memory.

Channel paths can be dynamically configured on and off (as needed for each partition) from an HMC.

10.5.3 HMC and SE remote operations

As stand-alone, outside the [IBM z16 A02](#) and [IBM z16 AGZ](#) HMCs (tower or rack mount) can no longer be ordered, most of you will connect to the HMC (and SE) via a browser. Direct browser access to the SE is not possible, you have to use the **Single Object Operations** (SOO) task on the HMC.

Note: Remote web browser access is the default for the Hardware Management Appliance HMCs.

Access to the USB Device on HMC and SE requires physical access to the HMC/SE.

Logon security for a web browser is provided by the local HMC user logon procedures. Certificates for secure communications are provided, and can be changed by the user. You can limit remote web browser access by specifying an IP address from the Customize Console Services task. To enable or disable the Remote operation service, click **Change...** in the **Customize Console Services** window, as shown in Figure 10-22 on page 419.

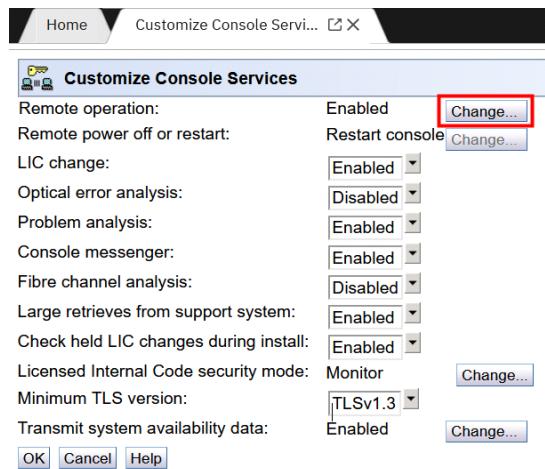


Figure 10-22 Customizing HMC remote operation

Note: If the setting in **Change Remote Access Setting -> IP Access Control** is set to **Allow specific IP addresses**, but if no or the wrong IP addresses are configured, you are not able to have a remote HMC connection via a web browser.

Microsoft Edge, Mozilla Firefox, Safari and Google Chrome were tested as remote web browsers. For more information about web browser requirements, see the HMC and SE console help system or [Resource Link](#), select **Library**, the applicable server, then select either Hardware Management Console Operations Guide or Support Element Operations Guide.

Single Object Operations (SOO)

It is not necessary to be physically at the location of an SE to use it. The HMC can be used to access the SE remotely by using the SOO task. The interface is the same as the interface that is used on the SE. For more information, see the HMC and SE console help system or [Resource Link](#), select **Library**, the applicable server, then select either Hardware Management Console Operations Guide or Support Element Operations Guide.

Note: With HMC Driver 41/Version 2.15.0 and Driver 51/Version 2.16.0, certain tasks that required in the past to access to the SE in SOO mode were implemented as HMC tasks. With this enhancement, the HMC runs the tasks on the SE directly, without the need to logon the SE in SOO mode.

IBM HMC Mobile

IBM HMC Mobile is an iOS and Android app that allows you to monitor all of your IBM zSystems and partitions and to receive alerts when messages or status changes come up.

You can also start, stop, or change the activation profile for a partition.

A full set of granular security controls are provided from the HMC including multi-factor authentication. This mobile interface is optional and is disabled by default. More and more functionalities from the HMC will also be available on IBM HMC Mobile.

On <https://ibm.biz/IBM-Z-HMC> you can find a short introduction video for **HMC Mobile**. Further information can be found on this page: <http://ibm.biz/hmc-mobile>.

10.5.4 Operating system communication

The **Operating System Messages** task displays messages from an LPAR. You can also enter operating system commands and interact with the system. This task is especially valuable for entering Coupling Facility Control Code (CFCC) commands.

The HMC also provides integrated 3270 and ASCII consoles. These consoles allow an operating system to be accessed without requiring other network or network devices, such as TCP/IP or control units.

Updates to x3270 support

The Configure 3270 Emulators task on the HMC and TKE consoles was enhanced to verify the authenticity of the certificate that is returned by the 3270 server when a secure and encrypted SSL connection is established to an IBM host. This 3270 Emulator with encrypted connection is also known as *Secure 3270*.

Use the Certificate Management task if the certificates that are returned by the 3270 server are not signed by a well-known trusted certificate authority (CA) certificate, such as VeriSign or Geotrust. An advanced action within the Certificate Management task, Manage Trusted Signing Certificates, is used to add trusted signing certificates.

For example, if the certificate that is associated with the 3270 server on the IBM host is signed and issued by a corporate certificate, it must be imported, as shown in Figure 10-23.

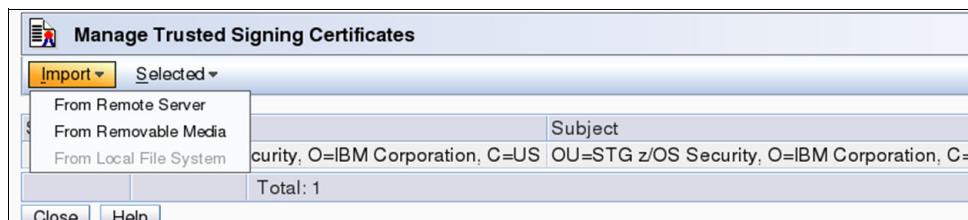


Figure 10-23 Manage Trusted Signing Certificates

The import from the remote server option can be used if the connection between the console and the IBM host can be trusted when the certificate is imported, as shown in Figure 10-24. Otherwise, import the certificate by using removable media.

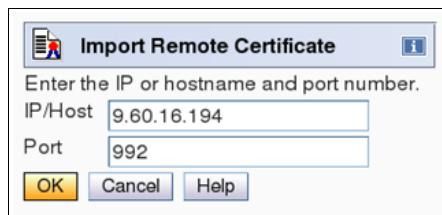


Figure 10-24 Import Remote Certificate example

A secure Telnet connection is established by adding the prefix L: to the IP address:port of the IBM host, as shown in Figure 10-25 on page 421.

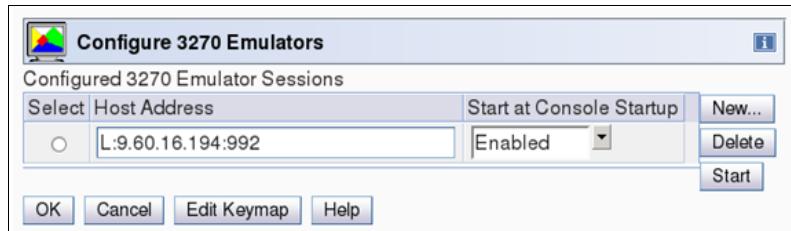


Figure 10-25 Configure 3270 Emulators

10.5.5 HMC and SE Microcode

The Microcode for the HMC, SE, and CPC is included in the driver/version. At the time of this writing, Driver 51/Level 2.16.0 is current for IBM z16 A02 and IBM z16 AGZ.

If in the future a new Driver/Level is available, the HMC provides the management of the driver upgrade through Enhanced Driver Maintenance (EDM). EDM provides the installation of the latest functions and patches (MCLs) of the new driver. When you perform a driver upgrade, always check the Driver/Level Customer Exception Letter option in the Fixes section at IBM Resource Link.

Microcode Change Level

Regular installation of Microcode Change Levels (MCLs) is key for reliability, availability, and serviceability (RAS), optimal performance, and new functions. We recommend to:

- ▶ Install MCLs on a quarterly basis at a minimum.
- ▶ Review hiper MCLs continuously to decide whether to wait for the next scheduled fix application session or to schedule one earlier if the risk assessment warrants.
- ▶ Sign On the “IBM zSystems Security Portal” (<https://www.ibm.com/it-infrastructure/z/capabilities/system-integrity>) web site and review for security alerts and related MCL fixes.

Tip: The IBM Resource Link provides access to the system information for your IBM zSystems according to the system availability data that is sent on a scheduled basis. It provides more information about the MCL status of your IBM zSystems.

At the Resource Link webbiest (<https://www.ibm.com/servers/resourcelink>), click **Tools** → **Machine Information**, choose your IBM zSystem, and then click on **EC/MCL**.

Microcode terms

The Microcode features the following characteristics:

- ▶ The Driver contains engineering change (EC) streams.
- ▶ Each EC stream covers the code for a specific component of the IBM zSystems. It includes a specific name and an ascending number.
- ▶ The EC stream name and a specific number are one MCL.
- ▶ MCLs from the same EC stream must be installed in sequence.
- ▶ MCLs can include installation dependencies on other MCLs.
- ▶ Combined MCLs from one or more EC streams are in one Bundle.
- ▶ An MCL contains one or more Microcode Fixes (MCFs).

How the Driver, Bundle, EC stream, MCL, and MCFs interact with each other is shown in Figure 10-26.

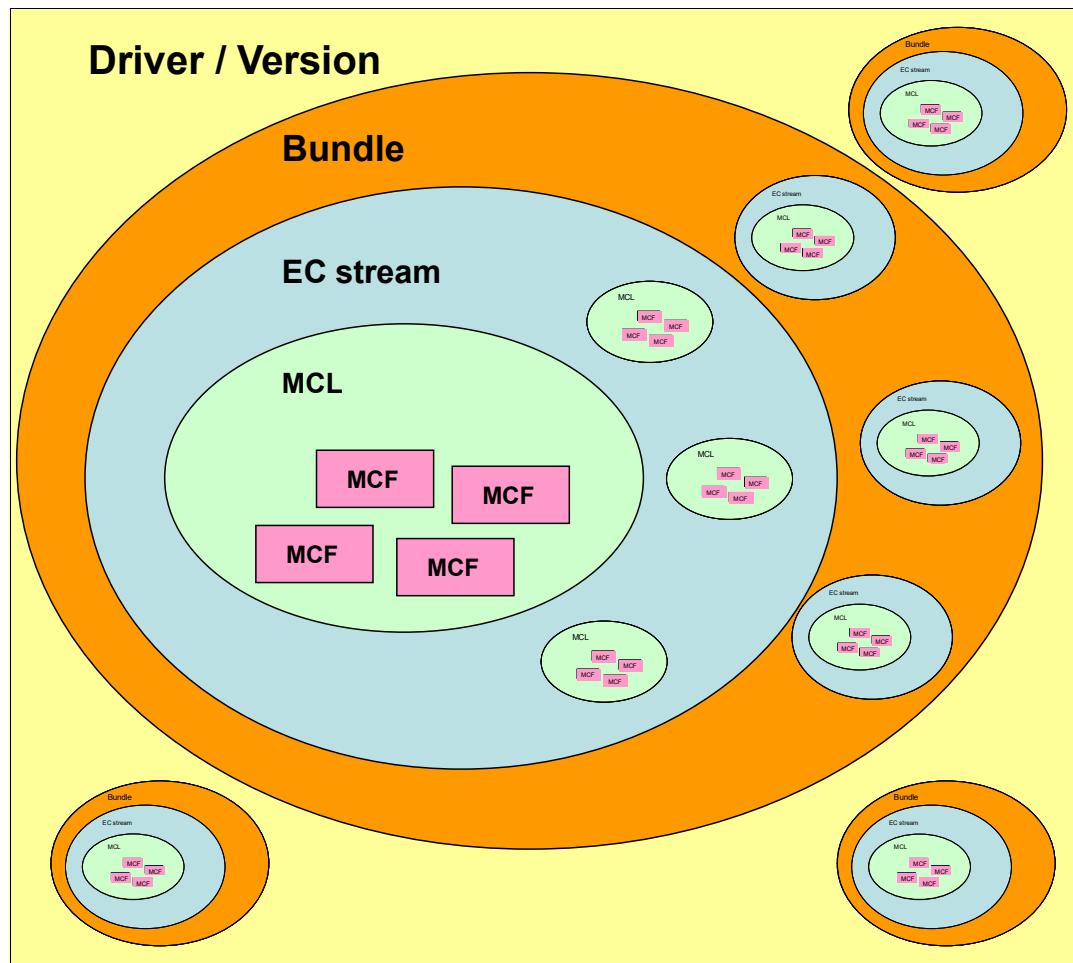


Figure 10-26 Microcode terms and interaction

MCL Activation

By design and with planning, MCLs can be activated concurrently. Consider the following points:

- ▶ Most MCLs activate concurrently when applied.
- ▶ A few MCLs need a disruptive config off/on to activate the newly loaded microcode.
- ▶ Activate traditional I/O Feature Pended MCL – LIC on the hardware feature:
 - Display Pending MCLs using HMC function or Resource Link Machine Information Reports
 - Activate using HMC function on a feature basis by PCHID one at a time – disruptive: CONFIG the CHPID OFF to all sharing LPARs, activate, and then CONFIG ON to all
- ▶ Activate Native PCIe Pended MCL – LIC on a hardware feature OR Resource Group (RG) LIC:
 - Display Pending MCLs using HMC function or Resource Link Machine Information Reports

- Feature LIC: Activate using HMC function on a one feature (PCHID) at a time basis - disruptive: CONFIG FUNCTIONS mapped to the feature OFF to all LPARs, activate, and then CONFIG ON
- RG LIC: Activate using HMC function to each RG in turn – disruptive to all PCHIDs in the RG: CONFIG all FUNCTIONS mapped to all PCHIDs in RG1 OFF, activate, then CONFIG ON. Repeat for all PCHIDs in RG2, RG3, RG4

Note: For hardware that does not need CHPID or a FUNCTION definition (for example, Crypto Express), a different method that is specific to the feature is used.

- ▶ Alternative: Apply and activate all Pended MCLs disruptively with a scheduled Power On Reset (POR)

To discover this “Pended” situation, the following actions are done whenever an MCL is applied:

- ▶ Logon the HMC and select your CPC under “System Management”
- ▶ Change Management
- ▶ System Information
- ▶ Query Additional Actions...

Or:

- ▶ Logon the SE and select CPC under “System Management”
- ▶ Change Management
- ▶ Query Channel/Crypto Configure Off/On Pending

Microcode installation by MCL Bundle target

A *Bundle* is a set of MCLs that are grouped during testing and released as a group on the same date. You can install an MCL to a specific target Bundle level. The System Information window is enhanced to show a summary Bundle level for the activated level, as shown in Figure 10-27.

System Information - A214						
Machine Information						
EC number:	P30713	LIC control level:	0001	Engineering Changes AROM		
Type:	3932	Model number:	A02	Serial number:	000020000214	
Version:	2.16.0	Driver level:	51	Bundle level:	S13+	
Internal Code Change Information						
Select	EC Number	Retrieved Level	Installable Concurrent	Activated Level	Accepted Level	Description
<input type="radio"/>	P30713 009	009	009	000	000	SE Framework
<input type="radio"/>	P30714 010	010	010	000	000	Firmware Management
<input type="radio"/>	P30715 024	024	024	000	000	SE Licensed Internal Code Alerts
<input type="radio"/>	P30716 013	013	013	000	000	I390/PU-ML LIC
<input type="radio"/>	P30717 004	004	004	000	000	LPAR HV LIC
<input type="radio"/>	P30718 004	004	004	000	000	CFCC (COUPLING) LIC
<input type="radio"/>	P30719 001	001	001	000	000	PCX LIC
<input type="radio"/>	P30720 009	009	009	000	000	PSCN Microcode SE and Cage Controller
<input type="radio"/>	P30721 010	010	010	000	000	CEC Microcode
<input type="radio"/>	P30722 000	000	000	000	000	MISR DATA LIC
<input type="radio"/>	P30724 004	004	004	000	000	Power FRU
<input type="radio"/>	P30725 002	002	002	000	000	Feature enablement stream for CEC
EC Details...						
Pending Actions						
Some actions might be pending. Click Query Additional Actions... for more information.						
Query Additional Actions...						
OK Help						

Figure 10-27 System Information: Firmware bundle level

Remote Code Load (RCL)

The Remote Code Load for IBM zSystems allows IBM to upgrade a machine remotely by working with the client to schedule a date and time for the code load and monitor the process to make sure it completes successfully.

This feature allows you to schedule one or multiple Single Step Code Load for an HMC or SE.

On an [IBM z16 A02 and IBM z16 AGZ](#) system with the HMA feature, RCL can be scheduled to update both HMCs and the firmware will automatically manage Alternate SE switches to ensure each HMC can be updated without the Primary SE being present.

In general you first have to generate a token on the HMC task **Manage Remote Firmware Updates**. Afterwards you can schedule a RCL on Resource Link -> **Fixes** -> **Licensed internal code** -> **Remote Code Load request**.

Important to understand is, that for IBM zSystems, IBM support will not connect to your IBM zSystems from outside to do the RCL. The RCL is managed on your HMC.

More info about RCL can be found here:

<https://www-01.ibm.com/servers/resourcelink/lib03010.nsf/pages/remoteCodeLoadForIBMZFirmware?OpenDocument>

Firmware update process^a: IBM z16 is planned to be the last server family to support IBM service support representatives (SSRs) onsite performing firmware updates without an additional premium service contract. The IBM Z Remote Code Load (RCL) option, which was introduced on IBM z15, is available without an additional premium service contract. With IBM z15, and now IBM z16, clients can request a remote code load or they can choose the SSR onsite method for their firmware update. IBM recommends that clients try the RCL option on IBM z15 or IBM z16 to see for themselves that IBM provides the same quality service through RCL.

- a. IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality.

10.5.6 Monitoring

This section describes systems monitoring considerations.

Monitor task group

The Monitor task group on the HMC and SE includes monitoring-related tasks for IBM zSystems CPCs, as shown in Figure 10-28.

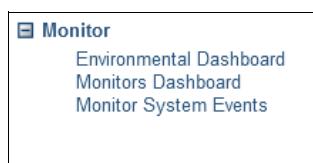


Figure 10-28 HMC Monitor Task Group

Environmental Dashboard task

The **Environmental Dashboard** task is part of the Monitor task group. It provides historical power consumption and thermal information for the IBM zSystems CPC. Figure 10-29 show an example of the Environmental Dashboard.

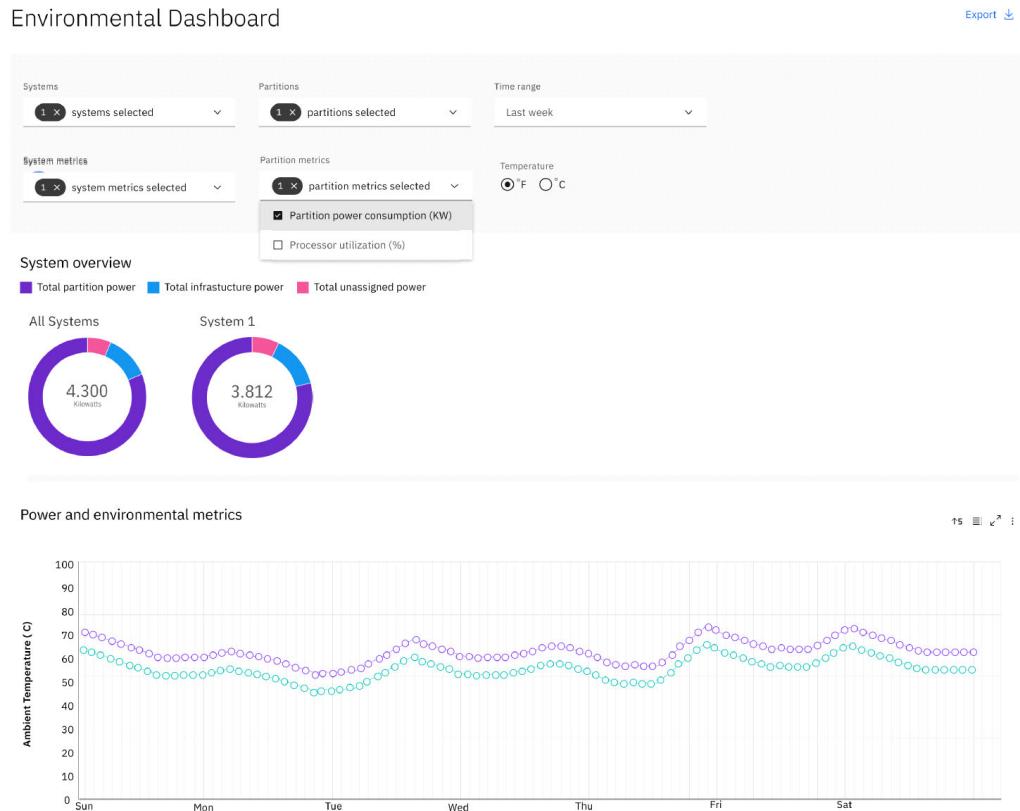


Figure 10-29 Example of the Environmental Dashboard

The data is presented in table format and graphical “histogram” format. The data also can be exported to a .csv-formatted file so that the data can be imported into a spreadsheet. For this task, you must use a web browser to connect to an HMC. The following can be displayed:

- ▶ View of power utilized by components assigned to individual partitions
- ▶ View of power of infrastructure components (including top of rack switches, SE/HMAs, and PDUs) which are not included in the power view for partitions
- ▶ View of power of unused I/O adapters / components that are not assigned to any partition (including standby components)
- ▶ Broader selectable time ranges for metrics view for historical trending data
- ▶ Display selected system and partition metrics in line chart and tabular views
- ▶ Filters for different views

Monitors Dashboard task

The Monitors Dashboard task in the Monitor task group provides a tree-based view of resources.

Multiple graphical views are available for displaying data, including history charts. The Monitors Dashboard monitors processor and channel usage. It produces data that includes power monitoring information, power consumption, and the air input temperature for the system.

You can display information for the following components:

- ▶ Power consumption
- ▶ Aggregated Processors
- ▶ Processors (with SMT information)
- ▶ System Assist Processors
- ▶ Logical Partitions
- ▶ Channels
- ▶ Adapters: Crypto use percentage is displayed according to the physical channel ID (PCHID number)
- ▶ Environmentals - New with Driver 51/Version 2.16.0:
 - Total Partition Power Consumption (kW)
 - Total Infrastructure Power Consumption (kW)
 - Total Unassigned Power Consumption (kW)
 - Partition power consumption per Partition (kW)

Monitor System Events

The Monitor System Events task allows you to create and manage event monitors. An event monitor listens for events from managed objects. When an event is received, the monitor tests it with user-defined criteria. If the event passes the tests, the monitor enables an e-mail to be sent to interested users.

An example of the settings for Event Monitor Summary can be see on Figure 10-30.

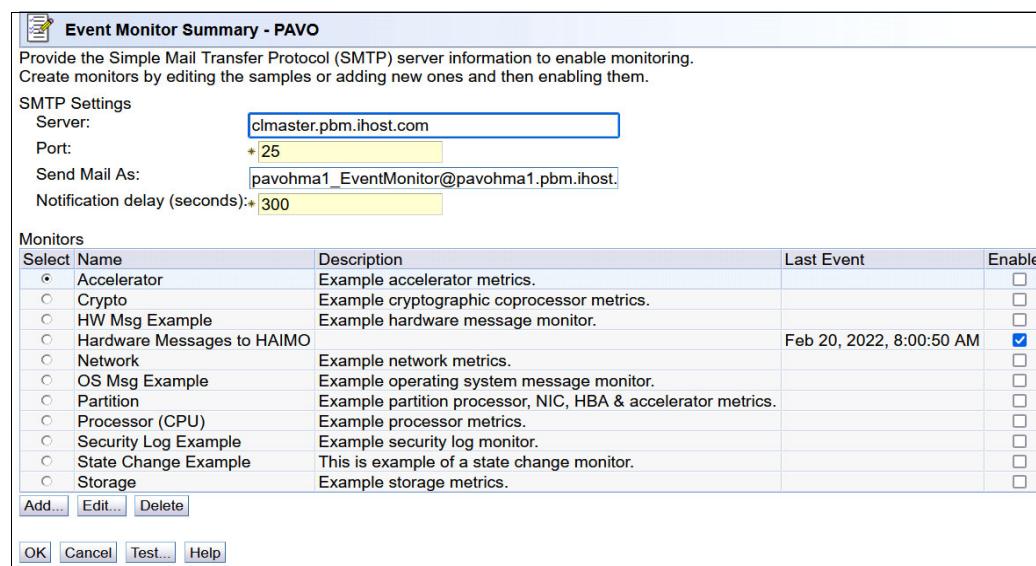


Figure 10-30 Event Monitor Summary

HMC Web Services API

With the Web Services API you can call information and use it with common Data Center Infrastructure Management (DCIM) tools.

10.5.7 Capacity on-demand support

All capacity on demand (CoD) upgrades are performed by using the **Perform a Model Conversion** task on the HMC or SE. Use the task to retrieve and activate a permanent upgrade, and to retrieve, install, activate, and deactivate a temporary upgrade. The task shows a list of all installed or staged LICCC records to help you manage them. It also shows a history of recorded activities.

The HMC for IBM z16 A02 and IBM z16 AGZ features the following CoD capabilities:

- ▶ SNMP API support:
 - API interfaces for granular activation and deactivation
 - API interfaces for enhanced CoD query information
 - API event notification for any CoD change activity on the system
 - CoD API interfaces, such as On/Off CoD and Capacity Back Up (CBU)
- ▶ SE window features (accessed through HMC Single Object Operations):
 - Window controls for granular activation and deactivation
 - History window for all CoD actions
 - Description editing of CoD records
- ▶ HMC/SE provides the following CoD information:
 - Millions of service units (MSU) and processor tokens
 - Last activation time
 - Pending resources that are shown by processor type instead of only a total count
 - Option to show more information about installed and staged permanent records
 - More information for the Attention state by providing seven more flags

HMC and SE are a part of the z/OS Capacity Provisioning environment. The Capacity Provisioning Manager (CPM) communicates with the HMC through IBM zSystems APIs, and enters CoD requests. For this reason, SNMP must be configured and enabled by using the **Customize API Settings** task on the HMC.

Note: Statement of Direction: IBM z16 is planned to be the last server to support legacy CoD unique record type automation interfaces. For example, legacy command HWMCA_ACTIVATE_CBU_COMMAND has to change to HWMCA_ADD_CAPACITY_COMMAND. We suggest you to start now to change your automation scripts accordingly. For more information see *Capacity on Demand User's Guide*, SC28-7025.

For more information about using and setting up CPM, see the following publications:

- ▶ *z/OS MVS Capacity Provisioning User's Guide*, SC34-2661
- ▶ *Capacity on-Demand User's Guide*, SC28-7025

10.5.8 Server Time Protocol (STP) support

In the task **Manage System Time** on the HMC (for the related CPC/SE), STP functions can be managed.

Detailed instructions and guidelines are provided within task workflow. A preview of the configuration is shown in a display. An example of the topology view is shown in Figure 10-31 on page 428.

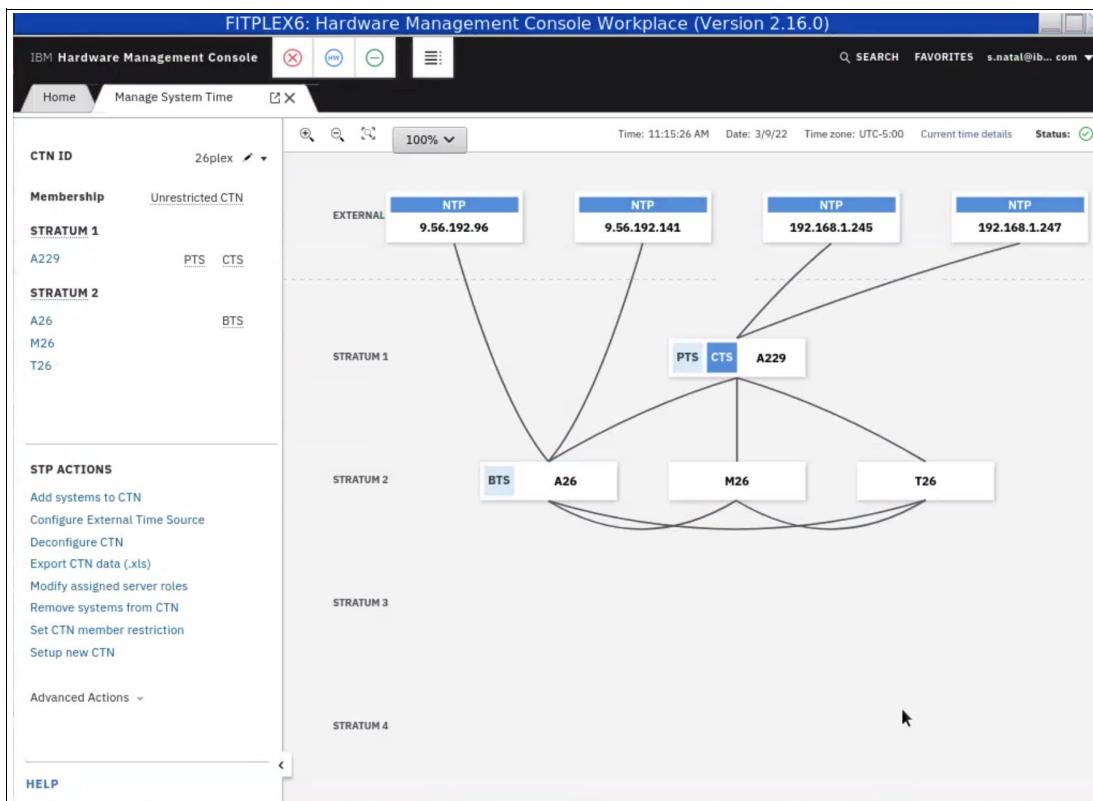


Figure 10-31 STP topology visible on HMC Manage System Time window

Important: The Sysplex Time task on the SE was discontinued with IBM z15. Therefore, with Driver 51/Version 2.16.0 the task **Manage System Time** on the HMC has to be used to managing STP and the according timing functions.

For more information about planning and setup, see the following publication:

- ▶ *IBM zSystems Server Time Protocol Guide*, SG24-8480.

STP changes and enhancements

IBM z16 A02 and IBM z16 AGZ implements the following major enhancements in support of the Server Time Protocol functionality:

- ▶ n-mode Power STP Imminent Disruption Signal option
On IBM zSystems, losing a Preferred Time Server (PTS) has significant consequences to the timing network and the overall workload execution environment of the IBM zSystems sysplex. The IBM zSystems and the HMC have had longtime automated failover protection for various cases that can arise.
New for IBM z16 A02 and IBM z16 AGZ, since there is no longer an integrated battery facility, support was added by the HMC to allow the client to configure an option to monitor for n-mode power conditions (wall power or line cord loss), and if detected, an automated failover will occur to the Backup Time Server (BTS). Note that you should provide some backup power method to hold power for 60 seconds on the PTS to allow failover to successfully complete.
There are also **Manage System Time** user interface controls to manage to failback to the PTS when the full power state is restored.
- ▶ CPC drawer direct Ethernet connectivity for the external time source (ETS).
In previous generation IBM zSystems, the ETS for the STP was provided by connecting

the Support Element to the client network.

With [IBM z16 A02](#) and [IBM z16 AGZ](#) the ETS, either PTP (IEEE 1588) or NTP network connectivity is provided using the [IBM z16 A02](#) and [IBM z16 AGZ](#) CPC drawer oscillator (OSC) cards dedicated network ports (RJ45) to client LAN for accessing the time synchronization information.

Pulse-per-second connectivity is also provided for higher timing information accuracy. Connection of the ETS direct to the IBM zSystems CPC provides less delay in accessing the time source than connection through the Support Element. For more information see [**<<chapter 2 , Oscillator card>>**](#)

Enhanced Console Assisted Recovery

Enhanced Console Assisted Recovery (ECAR) speeds up the process of BTS takeover by performing the following steps:

1. When the Preferred Time Server (PTS/CTS) detects a checkstop condition, the CEC informs its SE and HMC.
2. The PTS SE recognizes the checkstop pending condition, and calls the PTS SE STP code.
3. The PTS SE sends an ECAR request thorough HMC to the Backup Time Server (BTS) SE.
4. The BTS SE communicates with the BTS to start the takeover.

ECAR support is faster than the original CAR support because the console path changes from a 2-way path to a 1-way path. Also, almost no lag time is incurred between the system checkstop and the start of CAR processing. Because the request is generated from the PTS before system logging, it avoids the potential of recovery being held up.

For more information about planning and understanding STP server roles, see the following publications:

- ▶ *IBM Z Server Time Protocol Guide*, SG24-8480

10.5.9 NTP client and server on the HMC

For time synchronisation of the HMC as an NTP client, NTP servers from the client/Internet can be configured (with authentication support) in the task **Customize Console Date/Time**.

In the same task, the HMC can be configured to act as an NTP server for the CPCs (*Enable as time Server*). With this support, the [IBM z16 A02](#) and [IBM z16 AGZ](#) can receive the time from the HMC without accessing a LAN other than the HMC and SE network.

An example of this task can be see Figure 10-32 on page 430.

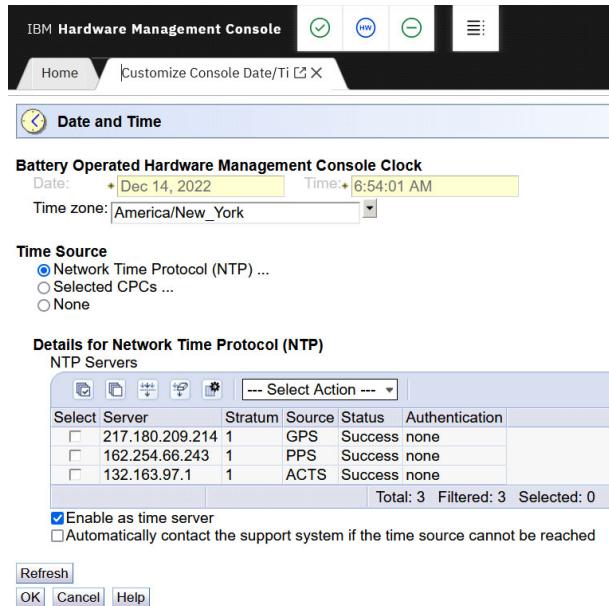


Figure 10-32 HMC Customize Date/Time task

For more information about planning and setup for STP and NTP, see the following publications:

- ▶ *IBM zSystems Server Time Protocol Guide*, SG24-8480

10.5.10 Security and user ID management

This section addresses security and user ID management considerations.

On [IBM z16 A02 and IBM z16 AGZ](#):

- ▶ Password is never stored in clear (one-way hash)
- ▶ The HMC and SE are closed appliances
- ▶ All network traffic is TLS encrypted
- ▶ HMC/SE features embedded firewall
- ▶ Firmware is digitally signed and validated for delivery
- ▶ Firmware Integrity Monitoring is used for any attempted tempering post delivery
- ▶ HMC/SE HDD encryption uses Trusted Platform Module (TPM) and Linux Unified Key Setup (LUKS) technology.

HMC and SE security audit

With the Audit and Log Management task, audit reports can be generated, viewed, saved, and offloaded. The **Customize Scheduled Operations** task allows you to schedule audit report generation, saving, and offloading. The **Monitor System Events** task allows Security Logs to send email notifications by using the same type of filters and rules that are used for hardware and operating system messages.

You can off load the following HMC and SE log files for customer audit:

- ▶ Console event log
- ▶ Console service history
- ▶ Tasks performed log
- ▶ Security logs

- ▶ System log

Full log offload and delta log offload (since the last offload request) are provided. Offloading to removable media and to remote locations by FTP is available. The offloading can be manually started by the new Audit and Log Management task or scheduled by the Customize Scheduled Operations task. The data can be offloaded in the HTML and XML formats.

HMC user ID templates and LDAP user authentication

Lightweight Directory Access Protocol (LDAP) user authentication and HMC user ID templates enable the addition and removal of HMC users according to your own corporate security environment. These processes use an LDAP server as the central authority.

Each HMC user ID template defines the specific authorization levels for the tasks and objects for the user who is mapped to that template. The HMC user is mapped to a specific user ID template by user ID pattern matching. The system then obtains the name of the user ID template from content in the LDAP server schema data.

Default HMC user IDs

For HMC Driver 51/Version 2.16.0 the default user ids are limited to ACSADMIN and SERVICE.

ADVANCED, OPERATOR, STORAGEADMIN, SYSPROG default users are no longer shipped. Default user roles for ADVANCED, OPERATOR, STORAGEADMIN, and SYSPROG are provided, and user IDs can be created from those.

Any Default User IDs which are part of a previous HMC level can be carried forward to new HMC levels as part of a MES Upgrade or via the selection of User Profile Data for the Save/Restore Customizable Console Data or Configure Data Replication tasks.

Multi-Factor Authentication (MFA)

MFA can be implemented to increase the security for the login procedure at the HMC and SE. For more information see 10.3.7, “HMC Multi-factor authentication” on page 415.

HMC and SE secure FTP support

You can use a secure FTP connection from a HMC/SE FTP client to a customer FTP server location. This configuration is implemented by using the Secure Shell (SSH) File Transfer Protocol, which is an extension of SSH. You can use the **Manage SSH Keys** task, which is available to the HMC and SE, to import public keys that are associated with a host address. For this task you need admin or service role.

The Secure FTP infrastructure allows HMC and SE applications to query whether a public key is associated with a host address and to use the Secure FTP interface with the appropriate public key for a host. Tasks that use FTP now provide a selection for the secure host connection.

When selected, the task verifies that a public key is associated with the specified host name. If a public key is not provided, a message window opens that points to the Manage SSH Keys task to enter a public key. The following tasks provide this support:

- ▶ Import/Export IOCDs
- ▶ Advanced Facilities FTP IBM Content Collector Load
- ▶ Audit and Log Management (Scheduled Operations only)
- ▶ FCP Configuration Import/Export
- ▶ OSA view Port Parameter Export
- ▶ OSA-Integrated Console Configuration Import/Export

10.5.11 System Input/Output Configuration Analyzer on the SE and HMC

The System Input/Output Configuration Analyzer task supports the system I/O configuration function.

The information that is needed to manage a system's I/O configuration must be obtained from many separate sources. The System Input/Output Configuration Analyzer task enables the system hardware administrator to access, from one location, the information from those sources. Managing I/O configurations then becomes easier, particularly across multiple systems.

The System Input/Output Configuration Analyzer task runs the following functions:

- ▶ Analyzes the current active IOCDS on the SE.
- ▶ Extracts information about the defined channel, partitions, link addresses, and control units.
- ▶ Requests the channels' node ID information. The Fibre Channel connection (FICON) channels support remote node ID information, using the pull-down menu **View -> Node ID**.

The System Input/Output Configuration Analyzer is a view-only tool. It does not offer any options other than viewing. By using the tool, data is formatted and displayed in five different views. The tool provides various sort options, and data can be exported to a USB or FTP.

The following views are available:

- ▶ PCHID Control Unit View shows PCHIDs, channel subsystems (CSS), CHPIDs, and their control units.
- ▶ PCHID Partition View shows PCHIDs, CSS, CHPIDs, and the partitions in which they exist.
- ▶ Control Unit View shows the control units, their PCHIDs, and their link addresses in each CSS.
- ▶ Link Load View shows the Link address and the PCHIDs that use it.
- ▶ Node ID View shows the Node ID data under the PCHIDs.

10.5.12 Automated operations

As an alternative to manual operations at the HMC and SE GUI, an application can interact with the HMC and SE through an API.

There are different API options to interact with the HMC:

- ▶ **HMC Web Services API**
The Web Services API is a web-oriented programming interface that makes the underlying zManager capabilities available for use by higher level management applications, system automation functions, or custom scripting. The functions that are exposed through the API support several important usage scenarios in virtualization management, including resource inventory, provisioning, monitoring, automation and workload-based optimization among others. For more details see *Hardware Management Console Web Services API Version 2.16.0*, SC27-2642.
- ▶ **SNMP Application Programming Interfaces**
The SNMP API provide monitoring and control functions through SNMP. The API can get and set a managed object's attributes, issue commands, receive asynchronous notifications, and generate SNMP traps. For more information see *SNMP Application Programming Interfaces*, SB10-7179.

10.5.13 Cryptographic support

This section describes the cryptographic management and control functions that are available in the HMC and SE.

Cryptographic hardware

IBM z16 A02 and IBM z16 AGZ include standard cryptographic hardware and optional cryptographic features for flexibility and growth capability.

The HMC/SE interface provides the following capabilities:

- ▶ Defining the cryptographic controls
- ▶ Dynamically adding a Crypto feature to a partition for the first time
- ▶ Dynamically adding a Crypto feature to a partition that already uses Crypto
- ▶ Dynamically removing a Crypto feature from a partition

The Crypto Express8S, which is a new Peripheral Component Interconnect Express (PCIe) cryptographic coprocessor, is an optional **IBM z16 A02 and IBM z16 AGZ** exclusive feature. Crypto Express8S provides a secure programming and hardware environment on which crypto processes are run. Each Crypto Express8S adapter can be configured by the installation as a Secure IBM CCA coprocessor, a Secure IBM Enterprise Public Key Cryptography Standards (PKCS) #11 (EP11) coprocessor, or an accelerator.

When EP11 mode is selected, a unique Enterprise PKCS #11 firmware is loaded into the cryptographic coprocessor. It is separate from the Common Cryptographic Architecture (CCA) firmware that is loaded when a CCA coprocessor is selected. CCA firmware and PKCS #11 firmware cannot coexist in a card.

The Trusted Key Entry (TKE) Workstation with smart card reader feature is required to support the administration of the Crypto Express8S when configured as an Enterprise PKCS #11 coprocessor.

To support the new Crypto Express8S card, the TKE 10.0 is needed. An example of the Cryptographic Configuration window is shown in Figure 10-33.

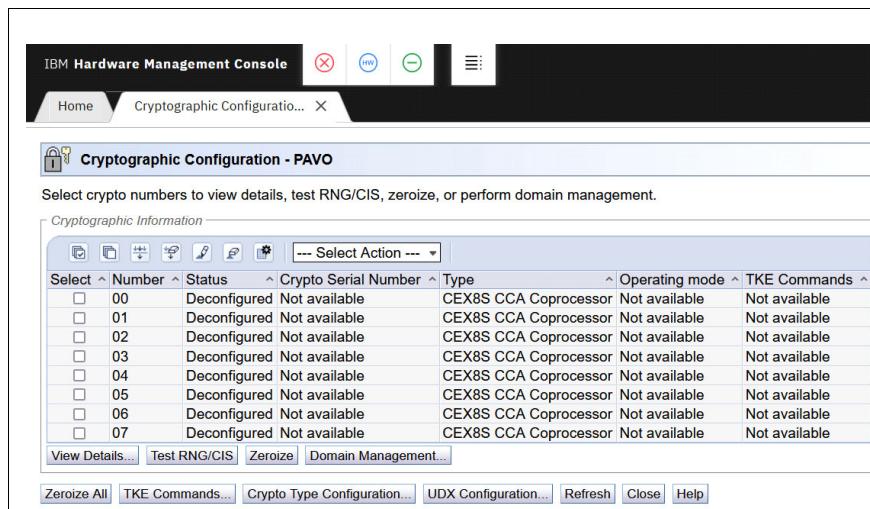


Figure 10-33 Cryptographic Configuration window

The Usage Domain Zeroize task is provided to clear the appropriate partition crypto keys for a usage domain when you remove a crypto card from a partition. Crypto Express8/7/6S in EP11 mode is configured to the standby state after the zeroize process.

For more information, see *IBM z16 Configuration Setup*, SG24-8960.

Digitally signed firmware

Security and data integrity are critical issues with firmware upgrades. Procedures are in place to use a process to digitally sign the firmware update files that are sent to the HMC, SE, and TKE. By using a hash algorithm, a message digest is generated that is then encrypted with a private key to produce a digital signature.

This operation ensures that any changes that are made to the data are detected during the upgrade process by verifying the digital signature. It helps ensure that no malware can be installed on IBM zSystems products during firmware updates. It also enables the [IBM z16 A02](#) and [IBM z16 AGZ](#) Central Processor Assist for Cryptographic Function (CPACF) functions to comply with Federal Information Processing Standard (FIPS) 140-3 Level 1 planned for Cryptographic LIC changes. The enhancement follows the IBM zSystems focus of security for the HMC and the SE.

The Crypto Express8S (CEX8S) is compliant with CCA PCI HSM. TKE workstation is optional when used to manage a Crypto Express8S feature that is defined as a CCA coprocessor in normal mode. However, it is mandatory when it is used to manage a Crypto Express8S feature that is defined as a CCA coprocessor in PCI-HSM mode or is defined as an EP11 coprocessor (CCA in PCI-HSM mode and EP11 also require a smart card reader plus smart cards with FIPS certification). <>For more information, link to Crypto>>

10.5.14 Installation support for z/VM that uses the HMC

Starting with z/VM V5R4 and z10, Linux on IBM Z can be installed in a z/VM virtual machine from HMC workstation media. This Linux on IBM Z installation can use the communication path between the HMC and the SE. No external network or extra network setup is necessary for the installation.

10.5.15 Dynamic Partition Manager (DPM)

DPM is an administrative mode (GUI to PR/SM) that was introduced for Linux only systems for IBM z13 and following systems. With DPM, you can use your Linux and virtualization skills while taking advantage of the full value of IBM zSystems hardware, robustness, and security in a workload optimized environment.

A system can be configured in DPM mode or in PR/SM mode (POR is required to switch modes). DPM supports in general the following functions:

- ▶ Create, provision, and manage partitions (processor, memory, and adapters) and storage
- ▶ Monitor and troubleshoot the environment

The following LPAR modes are available for DPM:

- ▶ z/VM
- ▶ Linux on IBM Z (also used for KVM deployments)
- ▶ Secure Service Container

If DPM is enabled, the [IBM z16 A02](#) and [IBM z16 AGZ](#) cannot run z/OS, IBM z/VSE®, and z/TPF LPARs.

The [IBM z16 A02](#) and [IBM z16 AGZ](#) can be initialized in PR/SM mode or in Dynamic Partition Manager (DPM) mode. No mix.

DPM provides a GUI for PR/SM to manage resources. Tools like HCD are in DPM mode not necessary.

This book does not cover scenarios that use DPM. For more information about the use of DPM, see *IBM Dynamic Partition Manager (DPM) Guide*, SB10-7182.

Important:

- ▶ The Enabling Dynamic Partition Manager task is run on the SE and is performed by your IBM system service representative (SSR) at installation time.
- ▶ If DPM is enabled, the [IBM z16 A02](#) and [IBM z16 AGZ](#) cannot run z/OS, IBM z/VSE[®], and z/TPF images in an LPAR or as a second level guest of z/VM.



Environmentals

This chapter describes the environmental requirements for IBM z16 A02 and IBM z16 AGZ. It also lists the power and cooling requirements that are needed to plan for the installation of these servers.

Important: The purpose of this chapter is to provide general information about the environmental characteristics for *IBM z16 A02* and *IBM z16 AGZ*, available configuration options, and how the configuration options affect the environmental requirements of these machines.

For detailed information regarding the specifications and requirements, always refer to the Installation Manual for Physical Planning (IMPP) for the specific configurations (available at the [IBM Resource Link website](#) (login required)):

- ▶ *3932 Single Frame Installation Manual for Physical Planning (A02/LA2)*, GC28-7040
- ▶ *IBM z16 and LinuxONE Rockhopper 4 Rack Mount Bundle Installation Manual for Physical Planning (IMPP)*, GC28-7035

This chapter includes the following topics:

- ▶ Introduction
- ▶ IBM z16 A02 environmental considerations
- ▶ IBM z16 AGZ environmental considerations
- ▶ Energy Management

11.1 Introduction

The IBM z16 A02 and IBM z16 AGZ is the first generation of IBM zSystems to deliver two different air-cooled configurations, which can be installed based on clients' requirements; The configurations are:

- ▶ The IBM z16 A02: single frame, factory assembled in an IBM 19-inch rack
- ▶ The IBM z16 AGZ: a component bundle which enables the core compute, I/O and networking components to be installed into a client provided standard 19-inch rack and powered by client supplied power distribution units (PDUs).

The following options are available for physically installing the server:

- ▶ Single phase and three phase AC utility power, configuration dependant.
- ▶ Installation on a raised floor or non-raised floor
- ▶ I/O and power cables can exit under the raised floor or off the top of the server frames

In the following two sections the different environmental specifications and requirements of both the IBM z16 A02 and IBM z16 AGZ are described separately.

11.2 IBM z16 A02 environmental considerations

The IBM z16 A02 is build in a 19-inch format industry standard frame. It supports installation on a raised floor or non-raised floor. The servers are air cooled, with air flow direction from front to back.

11.2.1 Power infrastructure

The IBM z16 A02 is available in the following power options:

- ▶ Single phase Intelligent Power Distribution Unit-based power (iPDU) - or PDU
Single phase is only available single CPC drawer configurations
- ▶ Three-phase Intelligent Power Distribution Unit-based power (iPDU) - or PDU

Note: All configurations can be powered from three-phase utility power.

Intelligent Power Distribution Unit

The IBM z16 A02 supports two or four iPDU's (two or four line cords) and depending on the machine configuration, these iPDU's can have the following specifications:

- ▶ Single phase:
 - 32A, 200-240 V AC
- ▶ Three phase:
 - 30A, 200-240 V AC ("Δ" - Delta wiring)
 - 32A, 380-415 V AC ("Y" - Wye wiring)

The iPDU design also offers some standardization and ease of data center installation planning, which allows the IBM z16 A02 to easily coexist with other platforms within the data center.

Power requirements

The IBM z16 A02 is designed with a fully redundant power system, using redundant iPDU's. iPDU's are installed in pairs. To make full use of the redundancy that is built into the server, the iPDU's within one pair must be powered from different power distribution panels. In that case, if one iPDU in a pair fails, the second iPDU ensures continued operation of the server without interruption.

Depending on system configuration (number of CPC drawers and/or PCIe+ I/O drawers), the IBM z16 A02 can have two or four PDUs installed. PDUs cannot be ordered separately; instead, they are always determined by the system configuration and are factory installed.

Note: IBM z16 A02 configurations with one CPC drawer (Max5, Max16, and Max32) can be ordered with single-phase power cords.

Compared to the previous generation, the CPC1 reserve feature is not available on the IBM z16 A02 since that rack space is reserved by default. However, for a future nondisruptive add of the second CPC drawer (CPC1) the initial order of the system must include three-phase power.

Configurations with two CPC drawers (Max68) are automatically shipped with three-phase power cords.

Power consumption

The utility power consumption for the IBM z16 A02 depends on a number of things:

- ▶ Number of installed PU DCM's / CPC drawers and installed amount of memory
- ▶ Number of installed PCIe+ I/O drawers and the number of cards in those drawers
- ▶ Environmental conditions in the data center

The *Installation Guide for Physical Planning* on the Resource Link website contains several tables that give an indication of the power consumption for several generic configurations.

For specific configurations the *Power & weight estimation tool*, also on Resource Link, can be used to estimate a more precise power consumption based on the configuration and data center conditions.

On Resource Link click **Tools → Power and weight estimation**.

11.2.2 Cooling requirements

The IBM z16 A02 is an air cooled system. All internal components are air cooled (heatsink cooled by forced air). PCIe+ I/O drawers, power enclosures, and CPC drawers are also cooled by chilled air with internal blowers.

The IBM z16 A02 support ASHRAE¹ Class A3 guidelines, which recommends a (long-term) ambient temperature range of 18°C (64.4°F) - 27°C (80.6°F). The minimum allowed ambient temperature is 5°C (41°F) and the maximum allowed temperature is 40°C (104°F).

The system requires chilled air to fulfill the air-cooling requirements. The airflow is from the front of the frame (intake, chilled air) to the rear (exhausts, warm air) of the frame. The chilled air is provided through perforated floor panels in front of the system.

¹ American Society of Heating Refrigeration and Air Conditioning Engineers

The hot and cold airflow and the arrangement of server aisles are shown in Figure 11-1 on page 440.

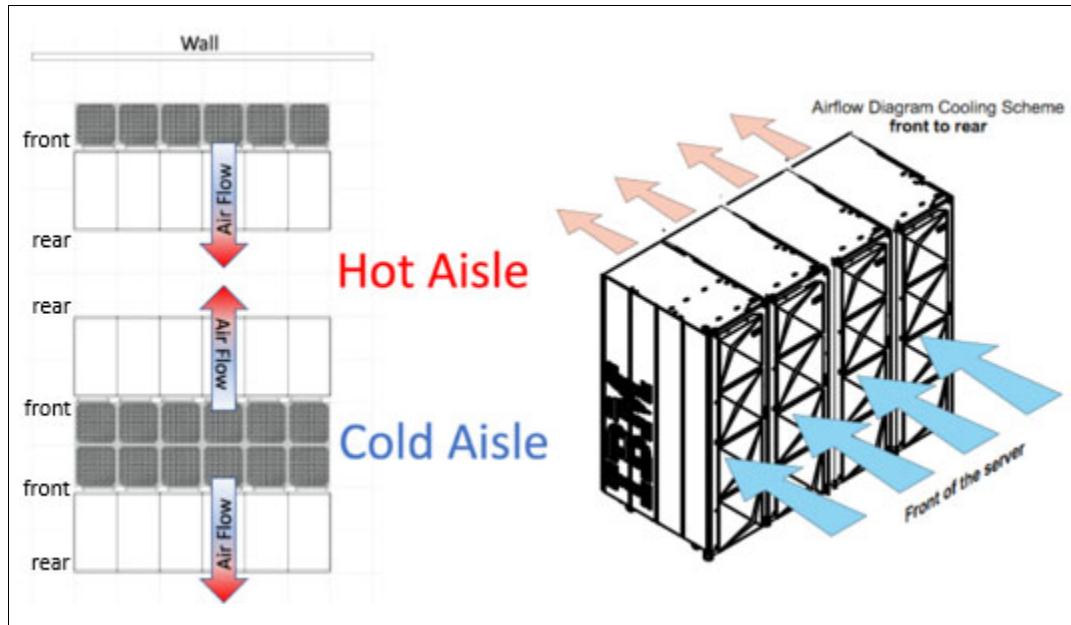


Figure 11-1 Hot and cold aisles

As shown in Figure 11-1, rows of servers must be placed front-to-front. Chilled air is provided through perforated floor panels that are placed in rows between the fronts of servers (the cold aisles).

For more information about cooling requirements, see *3932 Single Frame Installation Manual for Physical Planning* (Models A02/LA2), GC28-7040 on Resource Link.

11.2.3 Physical specifications

The IBM z16 A02 is built in an industry standard 19-inch frame and can be installed on a raised or non-raised floor.

For more information about dimensions, weight and floor loading tables, *3932 Single Frame Installation Manual for Physical Planning* (Models A02/LA2), GC28-7040 on Resource Link.

The *Power and weight estimation tool* for IBM zSystem servers, also available on Resource Link, can provide the estimated weight for any specific configuration.

On the Resource Link website, click **Tools** → **Power and weight estimation**.

11.2.4 Physical planning

This section describes the floor mounting, power, and I/O cabling options. For more information, see *3932 Single Frame Installation Manual for Physical Planning (Models A02/LA2)*, GC28-7040 on Resource Link.

Raised floor or non-raised floor

IBM z16 A02 can be installed on a raised or non-raised floor. The following options and features are available for I/O cabling and line cords:

- ▶ Top Exit cabling without Tophat feature
- ▶ Top Exit cabling with Tophat feature
- ▶ Bottom Exit cabling (*not available on a non-raised floor*)

Note: On the IBM z16 A02, all I/O cabling and line cords come from the rear of the machine; therefore, all related features for Bottom and Top Exit cabling are in the rear of the frame.

Top Exit cabling feature

When Top Exit cabling is required, the *optional* Top Exit Tophat cabling feature adds cable management options, such as trunking and retainer brackets.

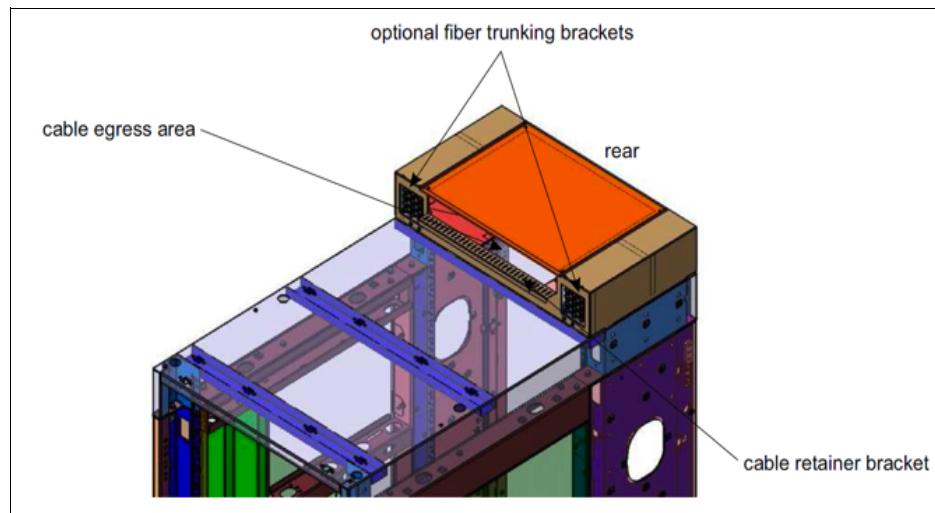


Figure 11-2 Top Exit Tophat cabling feature

The Top Exit cabling Tophat feature can be placed as shown in Figure 11-2, with the exit area toward the front of the frame, or with the exit area toward the rear of the frame. Additional height and weight will be added to the frame. For more information see *3932 Single Frame Installation Manual for Physical Planning (Models A02/LA2)*, GC28-7040 on Resource Link.

If the Top Exit Tophat cabling feature is *not* ordered, two sliding plates are standard available on the top of the frame (one on each side of the rear of the frame) that can be partially opened. By opening these plates, I/O cabling and power cords can exit the frame.

Bottom Exit cabling feature

The Bottom Exit cabling feature is required for raised floor environments, where I/O cabling and/or power cords must exit from the bottom of the frame. This feature includes the hardware to allow bottom exit, and other components for cable management and filler plates to preserve the recommended air circulation.

Frame cable management

A vertical cable management guide (“spine”) can assist with proper cable management for fiber, copper, and coupling cables (see Figure 11-3 on page 442). The spine is shipped with configurations that contain either three PCIe+ I/O drawers and/or two CPC drawers. All external cabling to the system (from top or bottom) can use the spine to minimize interference with the PDUs mounted on the sides of the frame.

If necessary, the spine easily can be relocated for service procedures.

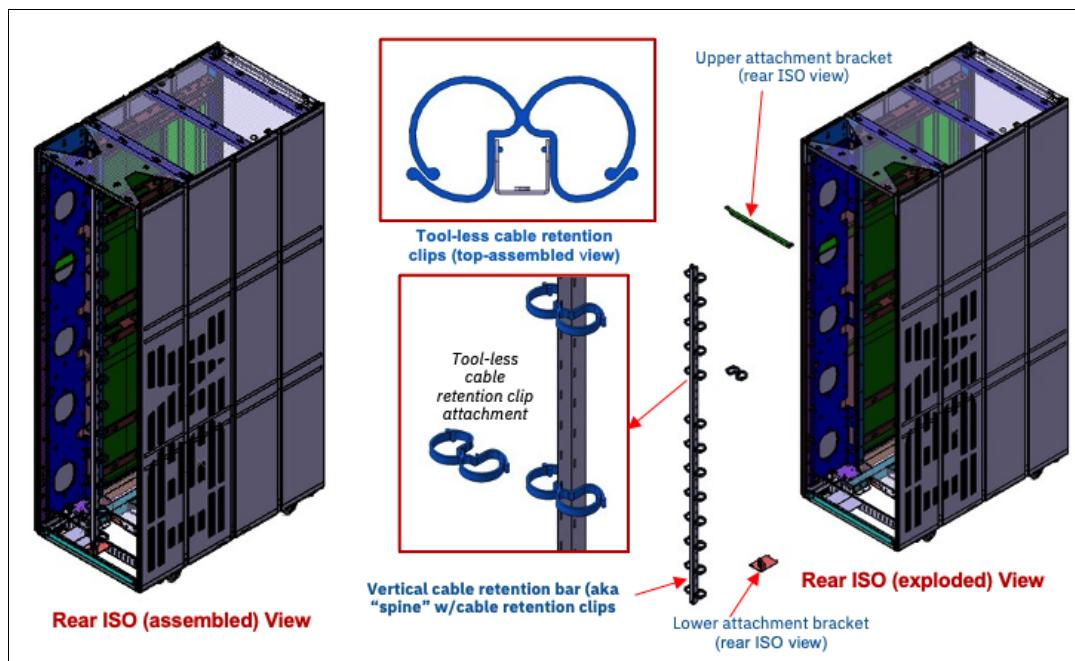


Figure 11-3 I/O cable management spine (rear view)

Frame Bolt-down kit

An Earthquake Kit is available for the IBM z16 A02. The kit provides hardware to enhance the ruggedness of the frame, the frame stiffener, and to tie down the frame to a concrete floor.

The frame tie-down kit can be used on a non-raised floor where the frame is secured directly to a concrete floor, or on a raised floor where the frame is secured to the concrete floor underneath the raised floor.

The kits help secure the frames and their contents from damage when they are exposed to shocks and vibrations, such as in a seismic event.

Service clearance areas

IBM z16 A02 requires specific service clearance to ensure the fastest possible repair in the unlikely event that a part must be replaced. Failure to provide enough clearance to open the front and rear covers results in extended service times or outages.

For more information, see *IBM 3932 Factory Frame A02/LA2 Installation Manual for Physical Planning*, GC28-7040 on Resource Link.

11.3 IBM z16 AGZ environmental considerations

The IBM z16 AGZ enables the core compute, I/O and networking components to be installed into a client provided rack. The components of IBM z16 AGZ will be installed into the client rack by IBM System Service Representative (SSR).

11.3.1 Power infrastructure

For the IBM z16 AGZ, customer is required to provide their own PDU's to power the components in their down rack. The following power options are available:

- ▶ Single phase Intelligent Power Distribution Unit-based power (PDU)
Single phase is only available single CPC drawer configurations
- ▶ Three-phase Intelligent Power Distribution Unit-based power (PDU). All configurations can be powered from three-phase facility power.

Intelligent Power Distribution Unit

The IBM z16 AGZ supports two or four PDUs (two or four line cords) and depending on the system configuration, these PDUs can have the following specification:

- ▶ Single phase:
 - 32A, 200-240 V AC
- ▶ Three phase:
 - 30A, 200-240 V AC (" Δ " - Delta wiring)
 - 32A, 380-415 V AC (" Y " - Wye wiring)

The PDU design also offers some standardization and ease of data center installation planning, which allows the IBM z16 AGZ to easily coexist with other platforms within the data center.

Note: For the IBM z16 AGZ it is the clients' responsibility to provide PDUs that meet the required specifications. Please refer to the *IBM z16 and LinuxONE Rockhopper 4 Rack Mount Bundle Installation Manual for Physical Planning (IMPP)* GC28-7035 on Resource Link.

Power cables

To provide power from the customer supplied PDUs to the different components, the right power cables must be ordered as part of the IBM z16 AGZ ordering process. Several options are available depending on the receptacles of the customer provided PDU, the component it powers and system configuration.

▶ *Receptacles options*

- C13: a power cable with a C13 type receptacle is available to provide power to the following IBM z16 AGZ components:
 - Service Element / Hardware Management Appliance
 - Network switches
 - PCIe+ I/O drawers
- C19: a power cable with a C19 type receptacle is available to provide power to the following IBM z16 AGZ components:
 - Service Element / Hardware Management Appliance
 - Network switches
 - PCIe+ I/O drawers
 - *CPC drawers*

Important: CPC drawers can only be connected through a power cable with C19 receptacle, but when deciding on the other components make sure customer provided has corresponding receptacles.

► **Power cables length**

Power cables can be ordered in three different lengths: 1, 2 and 3 meter. A power cable length planning matrix will be added to Resource Link which will help order the correct (minimum) length for each specific cable.

The required length of the power cable can be influenced by factors such as location of the IBM z16 A02 and IBM z16 AGZ components in the rack, placement of the PDUs in the rack and any cabling best practices.

Power requirements

The IBM z16 AGZ is designed to have a fully redundant power system, using redundant PDUs. PDUs must be installed in pairs. To make full use of the redundancy that is built into the server, the PDUs within one pair must be powered from different power distribution panels. In that case, if one PDU in a pair fails, the second PDU ensures continued operation of the server without interruption.

Depending on system configuration (number of CPC drawers and/or PCIe+ I/O drawers), the IBM z16 AGZ can have two or four PDUs installed which the customer must provide.

Note: IBM z16 AGZ with one CPC drawer (Max5, Max16, and Max32) can have the PDUs powered with single-phase power cords. Systems with two CPC drawers (Max68) can only use three-phase PDUs and power cords.

For a future nondisruptive upgrade from single CPC drawer system to one with two CPC drawers, plan the initial machine with three-phase power and FC 2332 (CPC1 Reserve space).

Power consumption

The utility power consumption for the IBM z16 AGZ depends on a number of things:

- the number of installed PU DCMs / CPC drawers and installed amount of memory
- the number of installed PCIe+ I/O drawers and the number of cards in those drawers
- the environmental conditions in the data center

The *Implementation Guide for Physical Planning* on the Resource Link website contains several tables that give an indication of the power consumption for several generic configurations.

For specific configurations the *Power & weight estimation tool*, also on Resource Link, can be used to estimate a more precise power consumption based on the configuration and data center conditions.

On Resource Link click **Tools → Power and weight estimation**.

11.3.2 Cooling requirements

The IBM z16 AGZ is an air cooled system. All internal components are air cooled (heatsink cooled by forced air). PCIe+ I/O drawers, power enclosures, and CPC drawers are also cooled by chilled air with blowers.

The IBM z16 AGZ support ASHRAE² Class A3 guidelines, which recommends a (long-term) ambient temperature range of 18°C (64.4°F) - 27°C (80.6°F). The minimum allowed ambient temperature is 5°C (41°F) and the maximum allowed temperature is 40°C (104°F).

² American Society of Heating Refrigeration and Air Conditioning Engineers

The system requires chilled air to fulfill the air-cooling requirements. The airflow is from the front of the frame (intake, chilled air) to the rear (exhausts, warm air) of the frame. The chilled air is provided through perforated floor panels in front of the system.

Note: Since the rack for the IBM z16 AGZ is provided by the client, some requirements for this rack need to be considered. For example: front and rear covers/doors must be perforated with a minimum of 45% open area and any open spaces in the rack must be sealed with filler plates, also provided by the customer.

For more information on these cooling requirements, see *IBM z16 and LinuxONE Rockhopper 4 Rack Mount Bundle Installation Manual for Physical Planning (IMPP)* GC28-7035 on Resource Link.

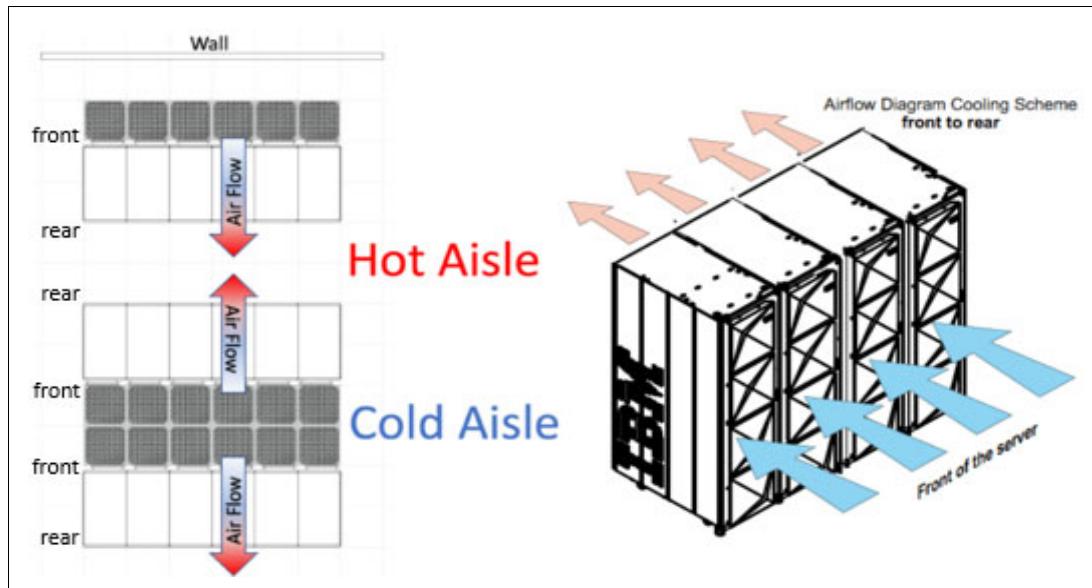


Figure 11-4 Hot and cold aisles

Figure 11-4 shows an example on how rows of racks must be placed front-to-front. Chilled air is provided through perforated floor panels that are placed in rows between the front of servers (the cold aisles).

11.3.3 Physical specifications

Note: For the installation and operation of the IBM z16 AGZ, it is the clients' responsibility to provide a 19-inch rack that meets the specifications for the server components.

Please refer to the *IBM z16 and LinuxONE Rockhopper 4 Rack Mount Bundle Installation Manual for Physical Planning (IMPP)*, GC28-7035 on Resource Link

The *Power and weight estimation tool* for IBM zSystems servers, also available on Resource Link, can provide the estimated weight for any specific configuration.

On the Resource Link website, click **Tools** → **Power and weight estimation**.

11.3.4 Physical planning

This section describes planning considerations when installing the IBM z16 AGZ into a customer provided rack. A majority of these options are inherited from or related to this customer provided rack. Therefore some features that you would need to configure for the IBM z16 A02 are now provided by the rack and not part of the IBM z16 AGZ configuration.

For more information, see *IBM z16 and LinuxONE Rockhopper 4 Rack Mount Bundle Installation Manual for Physical Planning (IMPP)*, GC28-7035 on Resource Link.

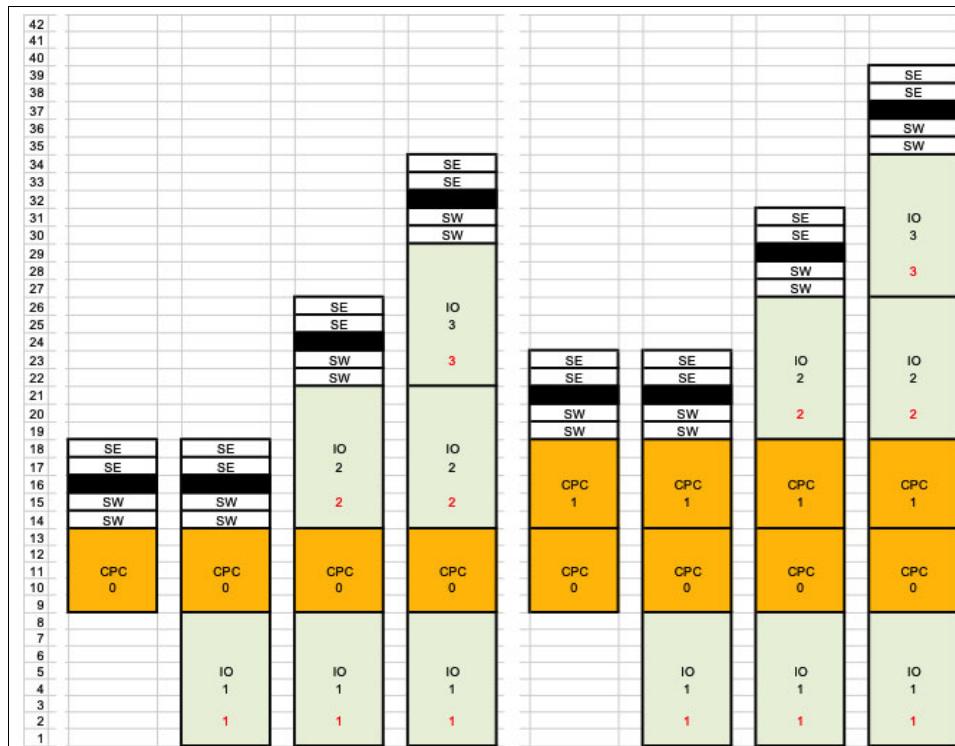
System component placement

The IBM z16 AGZ can consist of the following components:

- ▶ The Hardware Management Appliance (HMA)
- ▶ System ethernet switches
- ▶ CPC drawer(s)
- ▶ PCIe+ I/O drawer(s)

These components must be installed as a contiguous portion in the client rack, and in a specific order. In case CPC drawer or PCIe+ I/O drawer plan-ahead features are ordered for future non-disruptive upgrades, required free space will be left open but filler plates are required to be installed to minimize front-back airflow circulation. These specific filler plates are part of the plan-ahead features, in contrary to the filler plates for the rest of the customer provided rack above and below the IBM z16 AGZ.

Figure 11-5 shows the possible systems configurations. One or more systems are allowed in one rack, as long as within defined height limits.



Raised floor or non-raised floor

The IBM z16 AGZ can be installed on a raised or non-raised floor, but it's the customer provided rack that must provide the infrastructure to guide any cabling for top or bottom exit.

Frame cable management

The predominance of system cabling resides in the rear of the system and make use of the cabling management hardware of the customer provided rack, but some custom brackets will come with certain components which needs to be used during installation:

- ▶ Two vertical brackets per CPC drawer
- ▶ One horizontal bracket per PCIe+ I/O drawer
- ▶ Cable retention clips for all power cable connections

Service clearance areas

IBM z16 AGZ requires specific service clearance to ensure the fastest possible repair in the unlikely event that a part must be replaced. Failure to provide enough clearance to open the front and rear covers results in extended service times or outages.

For more information, see *IBM z16 and LinuxONE Rockhopper 4 Rack Mount Bundle Installation Manual for Physical Planning (IMPP)*, GC28-7035 on Resource Link.

11.4 Energy Management

This section describes the elements of energy management to help you understand the requirements for power and cooling, monitoring and trending, and reducing power consumption.

The energy management structure for the server is shown in Figure 11-6.

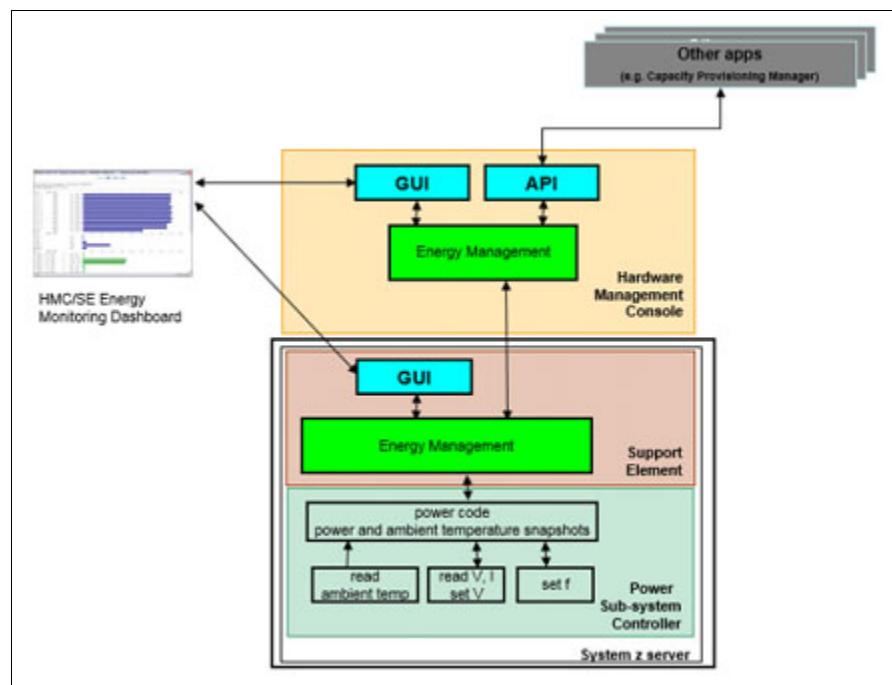


Figure 11-6 IBM z16 Energy Management

The hardware components in the IBMz16 A02 and IBM z16 AGZ are monitored and managed by the energy management component in the Support Element (SE) and HMC. The graphical user interfaces (GUI) of the SE and HMC provide views, such as the Monitors Dashboard, Environmental Efficiency Statistics, Environmental Dashboard and Energy Optimization Advisor.

The following tools are available to plan and monitor the energy consumption of the IBM z16 servers:

- ▶ Power estimation tool on Resource Link
- ▶ Energy Optimization Advisor task for maximum potential power on HMC and SE
- ▶ Monitors Dashboard, Environmental Efficiency Statistics and Environmental Dashboard tasks on HMC and SE

11.4.1 Environmental monitoring

This section describes energy monitoring HMC and SE tasks.

Energy Optimization Advisor

This window is started from the HMC targeting the system and task under Energy Management. The window displays the following recommendations:

- ▶ Thermal Advice
- ▶ Processor Utilization Advice

Select the advice hyperlinks to provide specific recommendations for your system, as shown in Figure 11-7.

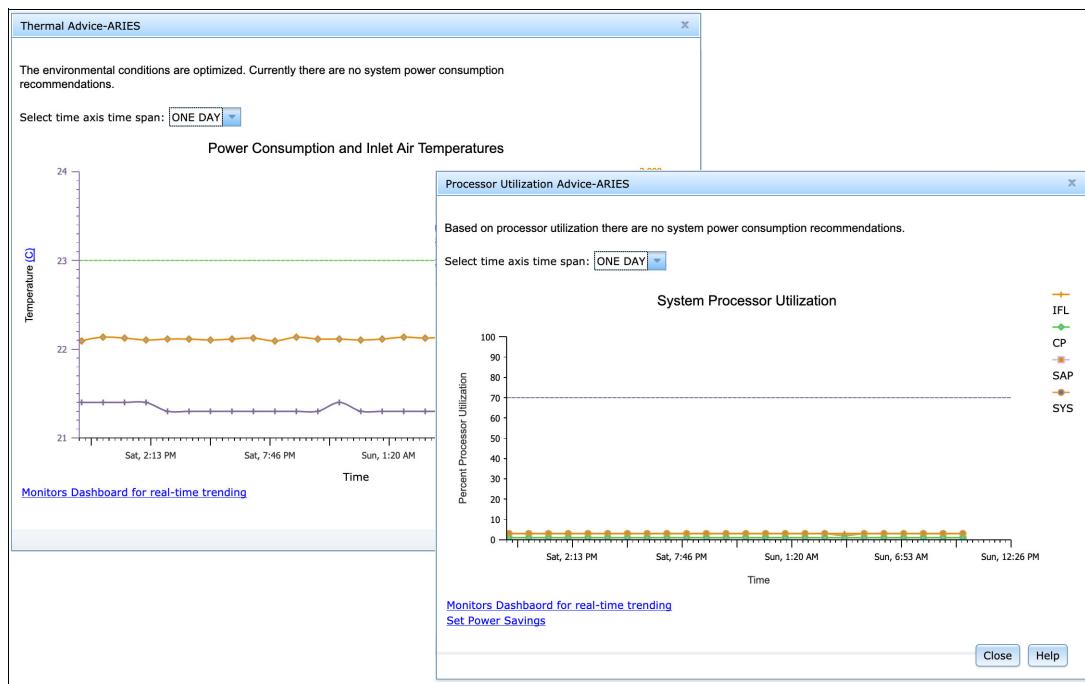


Figure 11-7 Energy Optimization Advisor

Monitors Dashboard task

In IBM z16 A02 and IBM z16 AGZ, the Monitors Dashboard task in the Monitor task group provides a tree-based view of resources. Multiple graphical views display data, including history charts. This task monitors processor and channel usage. It produces data that includes power monitoring information, power consumption, and the air input temperature for the server.

An example of the Monitors Dashboard task is shown in Figure 11-8.

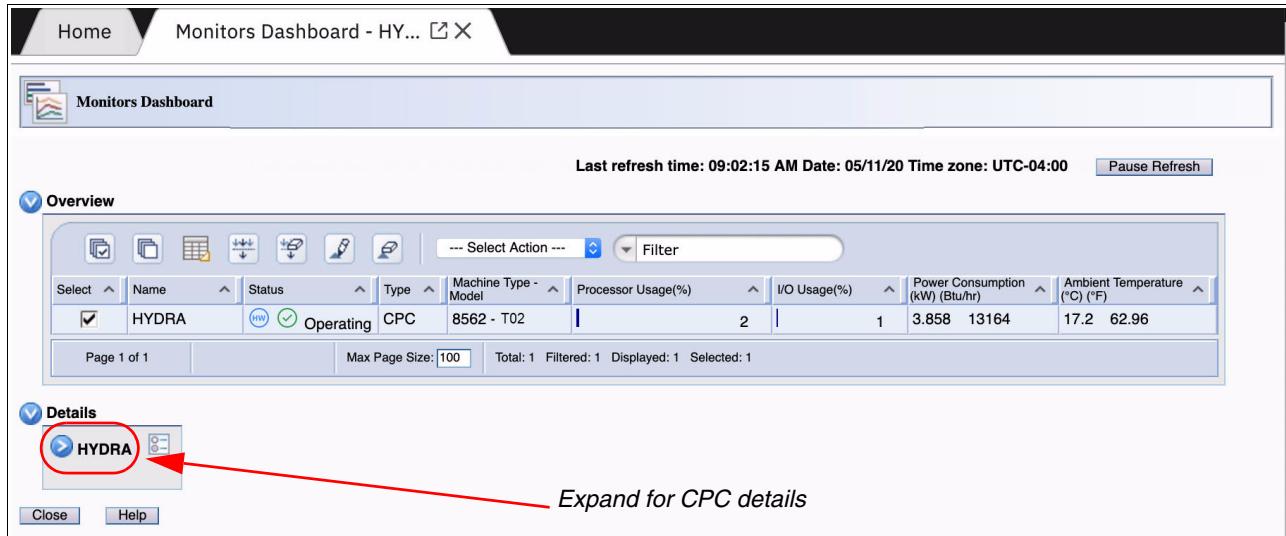


Figure 11-8 Monitors Dashboard task

Environmental Efficiency Statistics task

The Environmental Efficiency Statistics task (see Figure 11-9) has been redesigned and renamed to Environmental Dashboard. See next topic.

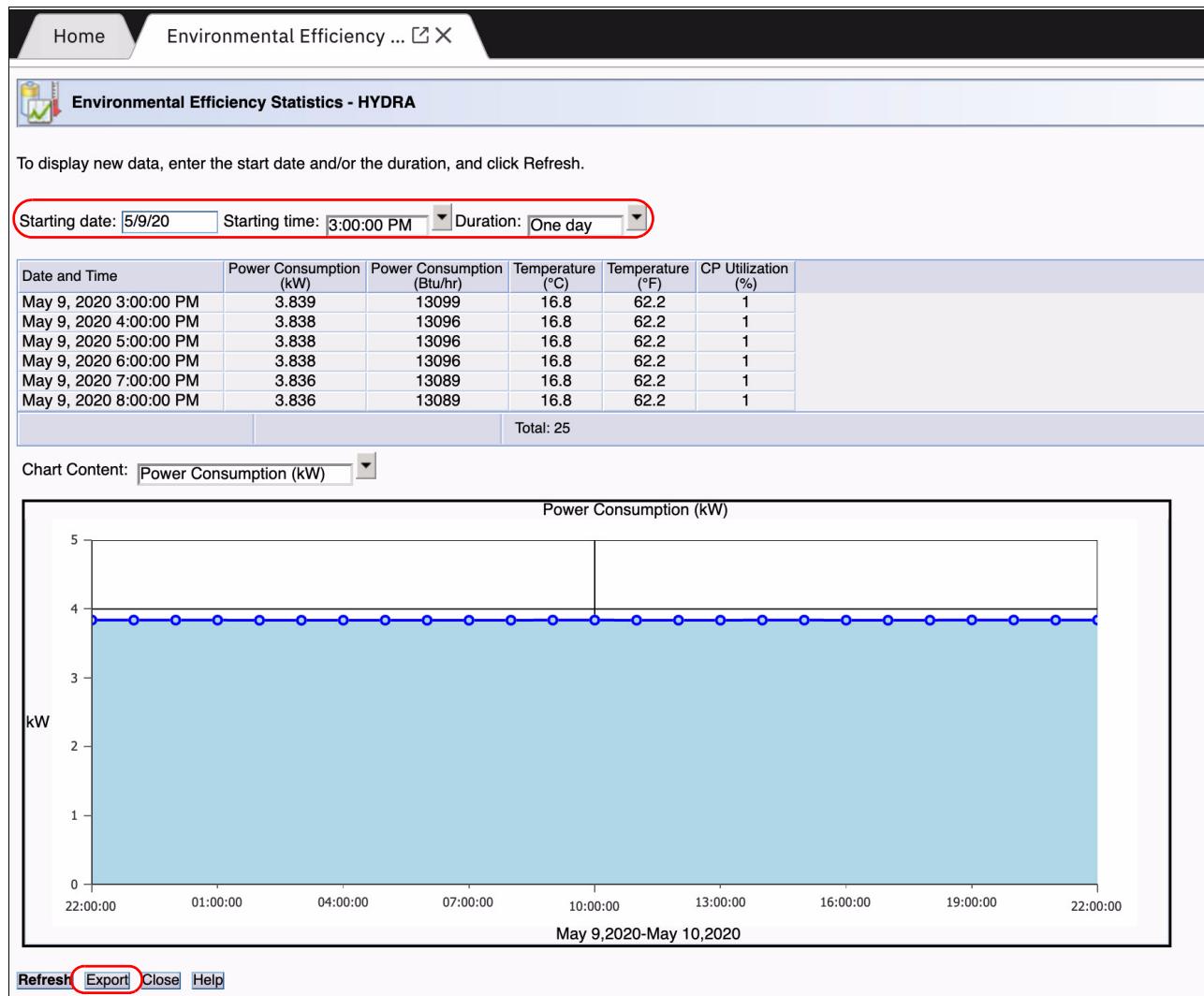


Figure 11-9 Environmental Efficiency Statistics task

The data is presented in table format and graphical “histogram” format. The data can also be exported to a .csv-formatted file so that the data can be imported into a spreadsheet. For this task, you must use a web browser to connect to an HMC.

Environmental Dashboard

The Environmental Dashboard is new with the IBM z16 A02 and IBM z16 AGZ. It enhances the HMC Monitors Dashboard with system and partition level power consumption. The Environmental Dashboard displays:

- ▶ System and LPAR power consumption
- ▶ Real Time data access
- ▶ Historical Trending Data
- ▶ Exported data and reports

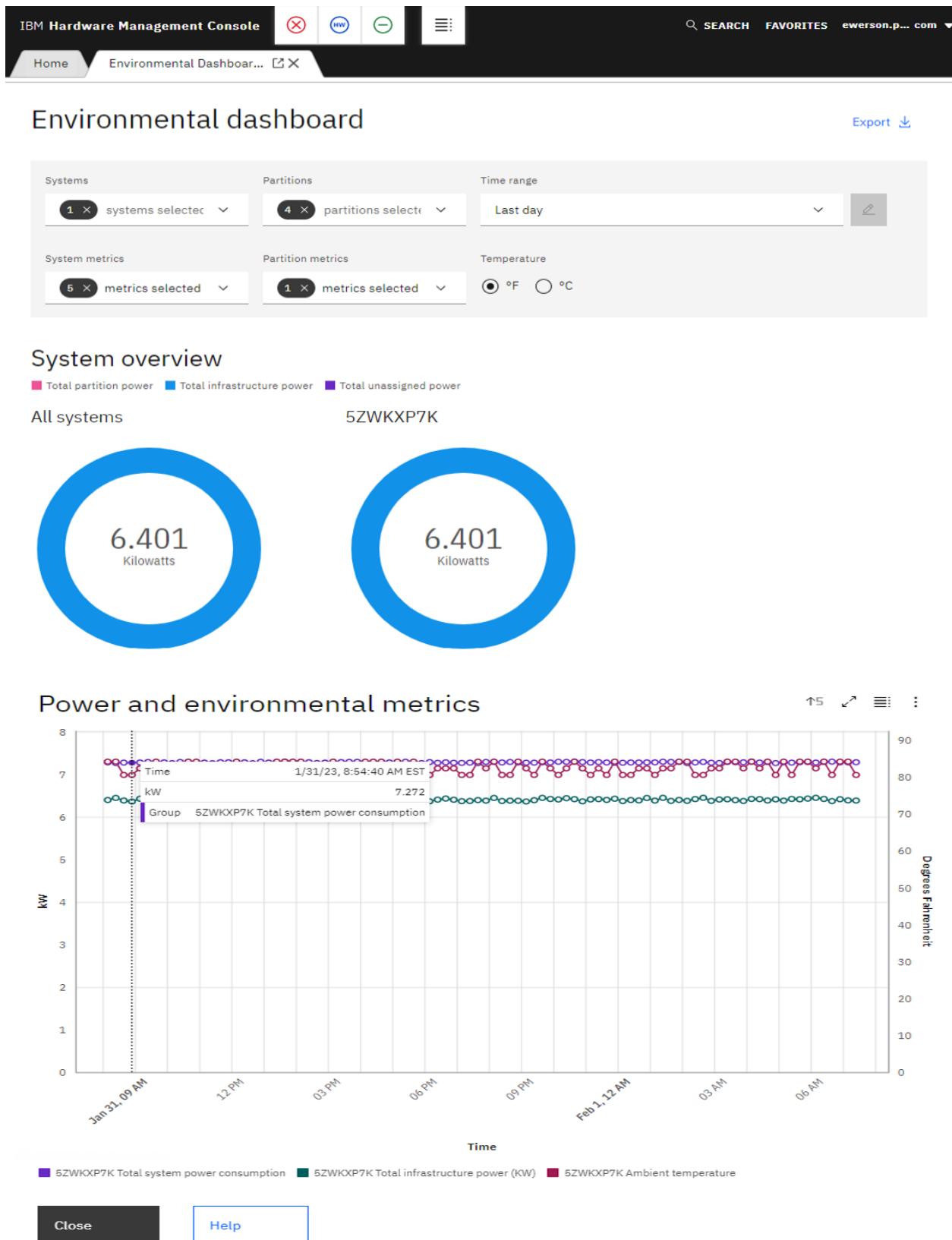


Figure 11-10 Environmental Dashboard display example

**12**

Performance

This chapter describes the performance and capacity planning of IBM z16 A02 and IBM z16 AGZ .

This chapter includes the following topics:

- ▶ 12.1, “IBM z16 A02 and IBM z16 AGZ performance overview” on page 454
- ▶ 12.2, “IBM z16 Large System Performance Reference ratio” on page 455
- ▶ 12.3, “Fundamental components of workload performance” on page 457
- ▶ 12.4, “Relative Nest Intensity” on page 459
- ▶ 12.5, “LSPR workload categories based on RNI” on page 460
- ▶ 12.6, “Relating production workloads to LSPR workloads” on page 461
- ▶ 12.7, “CPU MF counter data and LSPR workload type” on page 461
- ▶ 12.8, “Workload performance variation” on page 462
- ▶ 12.9, “Capacity planning for z16 A02 and AGZ” on page 463

12.1 IBM z16 A02 and IBM z16 AGZ performance overview

IBM z16 A02 and IBM z16 AGZ model Z06¹ is designed to offer up to 12-13% more capacity and twice the amount of memory than the IBM z15 T02 model Z06.

Uniprocessor performance also increased. On average, a z15 T02 model Z01 offers average performance improvements of up to 12% over the IBM z15 T02 model Z01. Figure 12-1 shows a uniprocessor performance comparison of successive IBM zSystems servers.

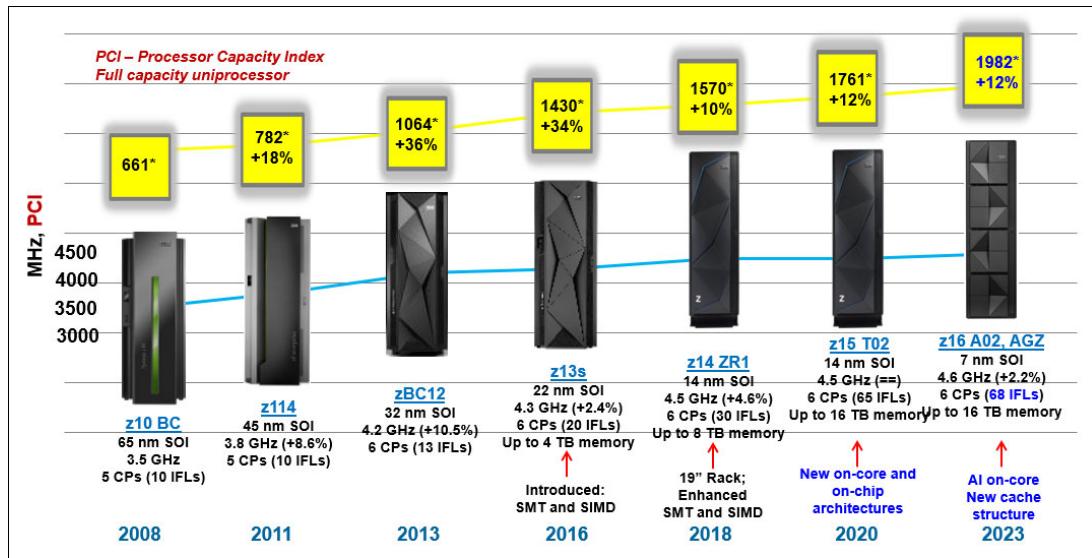


Figure 12-1 Uniprocessor performance evolution

Note: PCI = Processor Capacity Index.

Operating system support for the number of “engines” varies.

12.1.1 IBM z16 A02 and IBM z16 AGZ single-thread capacity

The IBM z16 A02 and IBM z16 AGZ processor chip runs at 4.6 GHz clock speed, which is slightly faster than the IBM z15 T02 processor chip, which runs at 4.5 GHz. For N-way processors model, it increases 1.13x on average at equal N-way configuration. These numbers differ depending on the workload type and LPAR configuration. For Linux on IBM Z capacity, the IBM z16 A02 and IBM z16 AGZ and IBM z16 A02 and IBM z16 AGZ Feature Max68 is designed to offer up to 25% more Linux capacity compared to an IBM z15 T02 Max65.

12.1.2 IBM z16 A02 and IBM z16 AGZ SMT capacity

From IBM z13 to IBM z16 A02 and IBM z16 AGZ , customers can choose to run two threads on IFL and zIIP cores by using SMT mode. SMT increases throughput by 10 - 40% (average 25%), depending on workload.

The SMT performance is increased by up to 13% on IBM z16 A02 and IBM z16 AGZ ,

¹ Model Capacity Identifier Z06. “Z” is the full CP capacity engine. IBM z16 A02 and IBM z16 AGZ support up to six CP engines and up to 26 capacity levels for CPs.

compared to IBM z15 T02. It means more workload can be executed by IBM z16 A02 and IBM z16 AGZ in the same time window than IBM z15 T02 when using SMT.

12.1.3 IBM Integrated Accelerator for zEnterprise Data Compression

Starting with z13, IBM introduced the zEnterprise Data Compression (zEDC) Express PCIe feature, bringing efficiency and economies for data storing and data transfers.

The zEDC Express feature was adopted by enterprises because it helps with software costs for compression/decompression operations (by offloading these operations), and increases data encryption (compression before encryption) efficiency.

With z15, the zEDC Express functionality was moved off from the PCIe infrastructure into the processor nest. By moving the compression and decompression into the processor nest (on-chip), IBM z15 processor provides a new level of performance for these tasks and eliminates the need for the zEDC Express feature virtualization. It also brings new use cases to the platform.

The IBM z16 A02 and IBM z16 AGZ continue to support IBM Integrated Accelerator for zEnterprise Data Compression. For more information, see Appendix C, “IBM Integrated Accelerator for zEnterprise Data Compression” on page 475.

12.1.4 Primary performance improvement drivers with z16

The attributes and design points of IBM z16 A02 and IBM z16 AGZ contribute to overall performance and throughput improvements as compared to the IBM z15. The following major items contribute to IBM z16 A02 and IBM z16 AGZ performance improvements:

- ▶ IBM z16 A02 and IBM z16 AGZ microprocessor architecture:
 - 7nm lithography FinFET silicon technology
 - New Core-Nest Interface
 - Brand new branch prediction design using SRAM
 - Significant architecture changes - COBOL compiler & more
 - On chip Artificial Intelligence (AI) - deep learning focus for inference
 - Fourth-generation SMT processing for zIIPs, IFLs and SAPs
- ▶ Cache:
 - L2 Private cache (unified) increased to 32 MB
 - Virtual L3 cache up to 8x32 = 256 MB/ CP chip
 - Virtual L4 cached up to 8x32x8 = 2048² MB/ drawer
- ▶ Software and hardware:
 - z/OS HiperDispatch Optimizations
 - PR/SM Algorithm Improvements (, including LPAR Resource Placement)
 - Quantum safe Encryption
 - New System Recovery Boost functionality

12.2 IBM z16 Large System Performance Reference ratio

The Large System Performance Reference (LSPR) provides capacity ratios among various processor families that are based on various measured workloads. It is a common practice to

² IBM z16 A02 and IBM z16 AGZ Max5 and Max16 Virtual L4 capacity is 1024MB.

assign a capacity scaling value to processors as a high-level approximation of their capacities.

For z/OS V2R5 studies, the capacity scaling factor that is commonly associated with the reference processor is set to a 2094-701 with a Processor Capacity Index (PCI) value of 593. This value is unchanged since z/OS V1R11 LSPR. The use of the same scaling factor across LSPR releases minimizes the changes in capacity results for an older study and provides more accurate capacity view for a new study.

Performance data for IBM z16 A02 and IBM z16 AGZ was obtained with z/OS V2R4 (running ***Db2 for z/OS V12, CICS TS V5R3, IMS V14, Enterprise COBOL V6R2, and WebSphere Application Server for z/OS V9.0.0.8***). All IBM zSystems server generations are measured in the same environment with the same workloads at high usage.

Note: If your software configuration is different from what is described here, the performance results might vary.

Consult the LSPR when you consider performance on the IBM z116 A02 or IBM z16 AGZ. The range of performance ratings across the individual LSPR workloads is likely to include a large spread. Performance of the individual logical partitions (LPARs) varies depending on the fluctuating resource requirements of other partitions and the availability of processor units (PUs). Therefore, it is important to know which LSPR workload type suite your production environment.

For more information, see 12.8, “Workload performance variation” on page 462.

For more information about performance, see the [Large Systems Performance Reference for IBM zSystems page](#) of the Resource Link website.

For more information about millions of service units (MSU) ratings, see the [IBM zSystems hardware and software consumptions solutions](#) of the IBM IT infrastructure website.

12.2.1 LSPR workload suite

Historically, LSPR capacity tables, including pure workloads and mixes, were identified with application names or a *software* characteristic; for example, CICS, IMS, OLTP-T,³ CB-L,⁴ LoIO-mix,⁵ and TI-mix.⁶ However, capacity performance is more closely associated with how a workload uses and interacts with a particular processor *hardware* design.

The CPU Measurement Facility (CPU MF) data that was introduced on the z10 provides insight into the interaction of workload and *hardware design* in production workloads. 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 processor access to caches and memory is called 12.4, “Relative Nest Intensity” on page 459. By using this data, LSPR introduces three workload capacity categories that replace all older primitives and mixes.

LSPR contains the internal throughput rate ratios (ITRRs) for the z16 and the previous generation processor families. These ratios are based on measurements and projections that use standard IBM benchmarks in a controlled environment.

³ Traditional online transaction processing workload (formerly known as IMS).

⁴ Commercial batch with long-running jobs.

⁵ Low I/O Content Mix Workload.

⁶ Transaction Intensive Mix Workload.

The throughput that any user experiences can vary depending on the amount of multiprogramming in the user's job stream, the I/O configuration, and the workload processed. Therefore, no assurance can be given that an individual user can achieve throughput improvements that are equivalent to the performance ratios that are stated.

12.3 Fundamental components of workload performance

Workload performance is sensitive to the following major factors:

- ▶ Instruction path length
- ▶ Instruction complexity
- ▶ Memory hierarchy and memory nest

These factors are described next.

12.3.1 Instruction path length

A transaction or job runs a set of instructions to complete its task. These instructions are composed of various paths through the operating system, subsystems, and application. The total count of instructions that are run across these software components is referred to as the *transaction or job path length*.

The path length varies for each transaction or job, and depends on the complexity of the tasks that must be run. For a particular transaction or job, the application path length tends to stay the same, assuming that the transaction or job is asked to run the same task each time.

However, the path length that is associated with the operating system or subsystem can vary based on the following factors:

- ▶ Competition with other tasks in the system for shared resources. As the total number of tasks grows, more instructions are needed to manage the resources.
- ▶ The number of logical processors (*n-way*) of the image or LPAR. As the number of logical processors grows, more instructions are needed to manage resources that are serialized by latches and locks.

12.3.2 Instruction complexity

The type of instructions and the sequence in which they are run interacts with the design of a microprocessor to affect a performance component. This factor is defined as *instruction complexity*. The following design alternatives affect this component:

- ▶ Cycle time (GHz)
- ▶ Instruction architecture
- ▶ Pipeline
- ▶ Superscalar
- ▶ Out-of-order execution
- ▶ Branch prediction
- ▶ Transaction Lookaside Buffer (TLB)
- ▶ Transactional Execution (TX)
- ▶ Single instruction multiple data instruction set (SIMD)
- ▶ Simultaneous multithreading (SMT)⁷

⁷ Only available for IFL, zIIP, and SAP processors

As workloads are moved between microprocessors with various designs, performance varies. However, when on a processor, this component tends to be similar across all models of that processor.

12.3.3 Memory hierarchy and memory nest

The *memory hierarchy* of a processor generally refers to the caches, data buses, and memory arrays that stage the instructions and data that must be run on the microprocessor to complete a transaction or job.

The following design choices affect this component:

- ▶ Cache size
- ▶ Latencies (sensitive to distance from the microprocessor)
- ▶ Number of levels, the Modified, Exclusive, Shared, Invalid (MESI) protocol, controllers, switches, the number and bandwidth of data buses, and so on.

Certain caches are *private* to the microprocessor core, which means that only that microprocessor core can access them. Other caches are shared by multiple microprocessor cores. The term *memory nest* for an IBM zSystems processor refers to the shared caches and memory along with the data buses that interconnect them.

A memory nest in a z16 A02 and IBM z16 AGZ CPC drawer is shown in Figure 12-2.

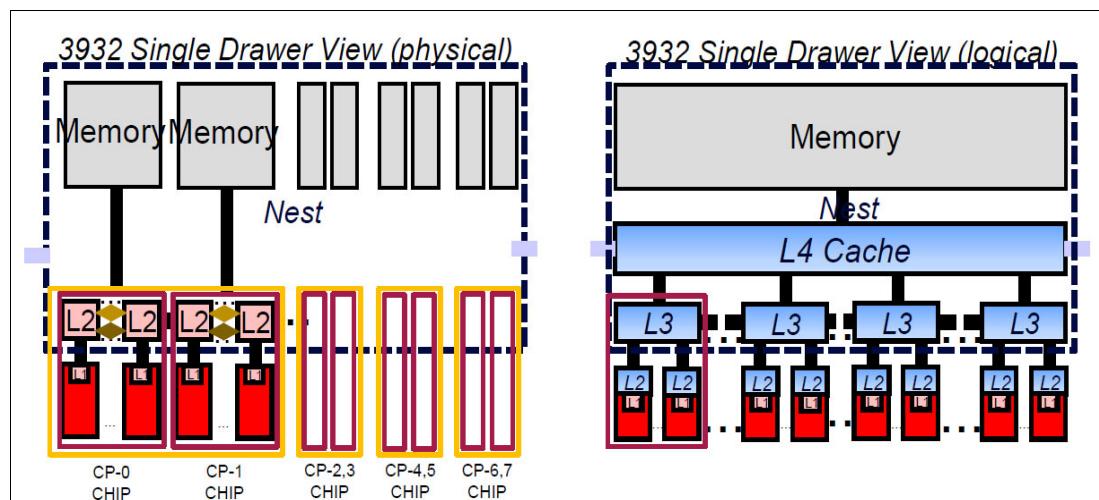


Figure 12-2 IBM z16 A02 and IBM z16 AGZ physical and virtual single drawer memory hierarchy

Workload performance is sensitive to how deep into the memory hierarchy the processor must go to retrieve the workload instructions and data for running. The best performance occurs when the instructions and data are in the caches nearest the processor because little time is spent waiting before running. If the instructions and data must be retrieved from farther out in the hierarchy, the processor spends more time waiting for their arrival.

As workloads are moved between processors with various memory hierarchy designs, performance varies because the average time to retrieve instructions and data from within the memory hierarchy varies. Also, when on a processor, this component continues to vary because the location of a workload's instructions and data within the memory hierarchy is affected by several factors that include, but are not limited to, the following factors:

- ▶ Locality of reference

- ▶ I/O rate
- ▶ Competition from other applications and LPARs

12.4 Relative Nest Intensity

The most performance-sensitive area of the memory hierarchy is the activity to the memory nest. This area is the distribution of activity to the shared caches and memory.

The term *Relative Nest Intensity* (RNI) indicates the level of activity to this part of the memory hierarchy. By using data from CPU MF, the RNI of the workload that is running in an LPAR can be calculated. The higher the RNI, the deeper into the memory hierarchy the processor must go to retrieve the instructions and data for that workload.

RNI reflects the distribution and latency of sourcing data from shared caches and memory, as shown in Figure 12-3.

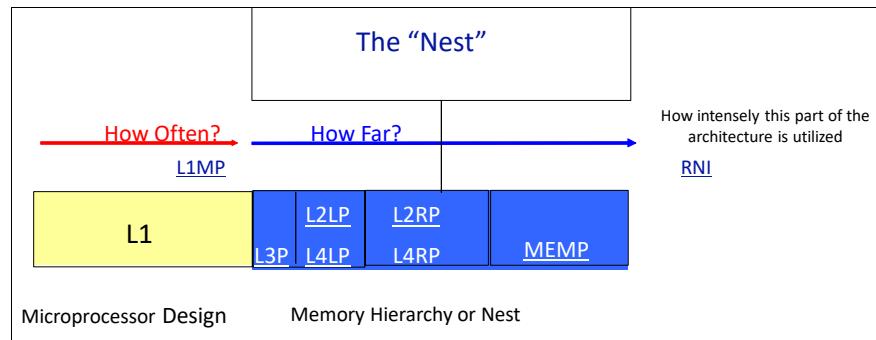


Figure 12-3 Relative Nest Intensity

Many factors influence the performance of a workload. However, these factors often are influencing the RNI of the workload. The interaction of all these factors results in a net RNI for the workload, which in turn directly relates to the performance of the workload.

These factors are tendencies, not absolutes. For example, a workload might have a low I/O rate, intensive processor use, and a high locality of reference, which all suggest a low RNI. However, it might be competing with many other applications within the same LPAR and many other LPARs on the processor, which tends to create a higher RNI. It is the net effect of the interaction of all these factors that determines the RNI.

The traditional factors that were used to categorize workloads in the past are shown with their RNI tendency in Figure 12-4 on page 460.

Relative Nest Intensity		
Low		High
Batch	Application Type	Transactional
Low	IO Rate	High
Single	Application Mix	Many
Intensive	CPU Usage	Light
Low	Dispatch Rate	High
High locality	Data Reference Pattern	Diverse
Simple	LPAR Configuration	Complex
Extensive	Software Configuration Tuning	Limited

Figure 12-4 Traditional factors that were used to categorize workloads

Little can be done to affect most of these factors. An application type is whatever is necessary to do the job. The data reference pattern and processor usage tend to be inherent to the nature of the application. The LPAR configuration and application mix are mostly a function of what must be supported on a system. The I/O rate can be influenced somewhat through buffer pool tuning.

However, one factor, *software configuration tuning*, is often overlooked but can have a direct effect on RNI. This term refers to the number of address spaces (such as CICS application-owning regions (AORs) or batch initiators) that are needed to support a workload. This factor always existed, but its sensitivity is higher with the current high frequency microprocessors. Spreading the same workload over more address spaces than necessary can raise a workload's RNI. This increase occurs because the working set of instructions and data from each address space increases the competition for the processor caches.

Tuning to reduce the number of simultaneously active address spaces to the optimum number that is needed to support a workload can reduce RNI and improve performance. In the LSPR, the number of address spaces for each processor type and *n-way* configuration is tuned to be consistent with what is needed to support the workload. Therefore, the LSPR workload capacity ratios reflect a presumed level of software configuration tuning. Retuning the software configuration of a production workload as it moves to a larger or faster processor might be needed to achieve the published LSPR ratios.

12.5 LSPR workload categories based on RNI

A workload's RNI is the most influential factor in determining workload performance. Other more traditional factors, such as application type or I/O rate, have RNI tendencies. However, it is the net RNI of the workload that is the underlying factor in determining the workload's performance. The LSPR now runs various combinations of former workload primitives, such as CICS, Db2, IMS, OSAM, VSAM, WebSphere, COBOL, and utilities, to produce capacity curves that span the typical range of RNI.

The following workload categories are represented in the LSPR tables:

- ▶ LOW (relative nest intensity)
 - A workload category that represents light use of the memory hierarchy.
- ▶ AVERAGE (relative nest intensity)

A workload category that represents average use of the memory hierarchy. This category is expected to represent most production workloads.

- ▶ HIGH (relative nest intensity)

A workload category that represents a heavy use of the memory hierarchy.

These categories are based on the RNI. The RNI is influenced by many variables, such as application type, I/O rate, application mix, processor usage, data reference patterns, LPAR configuration, and the software configuration that is running. CPU MF data can be collected by z/OS System Measurement Facility on SMF 113 records or z/VM Monitor starting with z/VM V5R4.

12.6 Relating production workloads to LSPR workloads

Historically, the following techniques were used to match production workloads to LSPR workloads:

- ▶ Application name (a client that is running CICS can use the CICS LSPR workload)
- ▶ Application type (create a mix of the LSPR online and batch workloads)
- ▶ I/O rate (the low I/O rates that are used a mix of low I/O rate LSPR workloads)

The IBM Processor Capacity Reference for IBM zSystems (zPCR) tool supports the following workload categories:

- ▶ Low
- ▶ Low-Average
- ▶ Average
- ▶ Average-high
- ▶ High

For more information about the no-charge IBM zPCR tool (which reflects the latest IBM LSPR measurements), see the [Getting Started with zPCR \(IBM's Processor Capacity Reference\) page](#) of the IBM Techdoc Library website.

As described in 12.5, “LSPR workload categories based on RNI” on page 460, the underlying performance sensitive factor is how a workload interacts with the processor hardware.

12.7 CPU MF counter data and LSPR workload type

Beginning with the z10 processor, the hardware characteristics can be measured by using CPU MF (SMF 113) counters data. A production workload can be matched to an LSPR workload category through these hardware characteristics.

For more information about RNI, see 12.5, “LSPR workload categories based on RNI” on page 460.

The AVERAGE RNI LSPR workload is intended to match most client workloads. When no other data is available, use the AVERAGE RNI LSPR workload for capacity analysis.

Low-Average and Average-High categories allow better granularity for workload characterization but these categories can apply on zPCR only.

The CPU MF data can be used determine workload type. When available, this data allows the RNI for a production workload to be calculated.

By using the RNI and another factor from CPU MF, the L1MP (percentage of data and instruction references that miss the L1 cache), a workload can be classified as LOW, AVERAGE, or HIGH RNI. This classification and resulting hit are automated in the zPCR tool. It is preferable to use zPCR for capacity sizing.

Starting with z/OS V2R1 with APAR OA43366, zFS file is not required any more for CPU MF and Hardware Instrumentation Services (HIS). HIS is a z/OS function that collects hardware event data for processors in SMF records type 113, and a z/OS UNIX System Services output files.

Only SMF 113 record is required to know proper workload type by using CPU MF counter data. CPU overhead of CPUMF is minimal. Also, the amount of SMF 113 record is 1% of typical SMF 70 and 72 which RMF writes.

CPU MF and HIS can use not only for deciding workload type but also use another purpose. For example, starting with z/OS V2R1, you can record Instruction Counts in SMF type 30 record when you activate CPU MF. Therefore, we strongly recommend that you *always* activate CPU MF.

For more information about getting CPUMF counter data, see the CPU MF - 2022 Update and WSC Experiences of the IBM Techdoc Library website.

12.8 Workload performance variation

As the size of transistors approaches the size of atoms that stand as a fundamental physical barrier, a processor chip's performance can no longer double every two years (Moore's Law⁸ does not apply).

A holistic performance approach is required when the performance gains are reduced because of frequency. Therefore, hardware and software synergy becomes an absolute requirement.

Starting with z13, Instructions Per Cycle (IPC) improvements in core and cache became the driving factor for performance gains. As these microarchitectural features increase (which contributes to instruction parallelism), overall workload performance variability also increases because not all workloads react the same way to these enhancements.

The workload variability for moving from z15 T02 to IBM z16 A02 and IBM z16 AGZ is expected to be stable. Workloads that are migrating from z14 ZR1s and previous generations to z16 A02 and IBM z16 AGZ can expect to see similar results with slightly less variability than the typical z14 ZR1 experience.

The effect of this variability is increased deviations of workloads from single-number metric-based factors, such as millions of instructions per second (MIPS), MSUs, and CPU time charge-back algorithms.

Experience demonstrates that IBM zSystems servers can be run at up to 100% utilization levels, sustained. However, most clients prefer to leave some room and run at 90% or slightly under.

⁸ For more information, see the [Moore's Law website](#).

For any capacity comparison exercise that uses a single metric, such as MIPS or MSU, is not a valid method. When deciding the number of processors and the uniprocessor capacity, consider the workload characteristics and LPAR configuration. For these reasons, the use of zPCR and involving IBM technical support are recommended when you plan capacity.

12.9 Capacity planning for z16 A02 and AGZ

In this section, we describe recommended ways conduct capacity planning for z16 A02 and AGZ.

Do not use MIPS or MSUs for capacity planning: Do *not* use “one number” capacity comparisons, such as MIPS or MSUs. IBM does not officially announce the processor performance as “MIPS”. MSU is only a number for software license charge and it does *not* represent for performance for the processor.

12.9.1 Collect CPU MF counter data

It is important to recognize the LSPR workload type of your production system. As described in 12.7, “CPU MF counter data and LSPR workload type” on page 461, the capacity of the processor is different from the LSPR workload type. By collecting the CPU MF SMF 113 record, you can recognize the workload type in a specific IBM-provided capacity planning tool. Therefore, collecting CPU MF counter data is a first step to begin the capacity planning.

12.9.2 Creating EDF file with CP3KEXTR

EDF file is an input file of the IBM zSystems capacity planning tool. You can create this file by using the CP3KEXTR program. The CP3KEXTR program reads SMF records and extracts needed data as input to IBM’s Processor Capacity Reference (zPCR) and z Systems Batch Network Analyzer (zBNA) tools.

CP3KEXTR is offered as a “no-charge” application. It can also create the EDF file for ZCP3000. ZCP3000 is an IBM internal tool, but you can create the EDF file for it on your system. For more information about CP3KEXTR, see the [IBM Techdoc z/OS Data Extraction Program \(CP3KEXTR\) for zPCR and zBNA](#).

12.9.3 Loading EDF file to the capacity planning tool

By loading EDF file to IBM capacity planning tool, you can see the LSPR workload type based on CPU MF counter data. Figure 12-5 on page 464 shows a sample zPCR window of a workload type. In this example, the workload type displays in the “Assigned Workload” column. When you load the EDF file to zPCR, it automatically sets your LPAR configuration. It also makes easy to define the LPAR configuration to the zPCR.

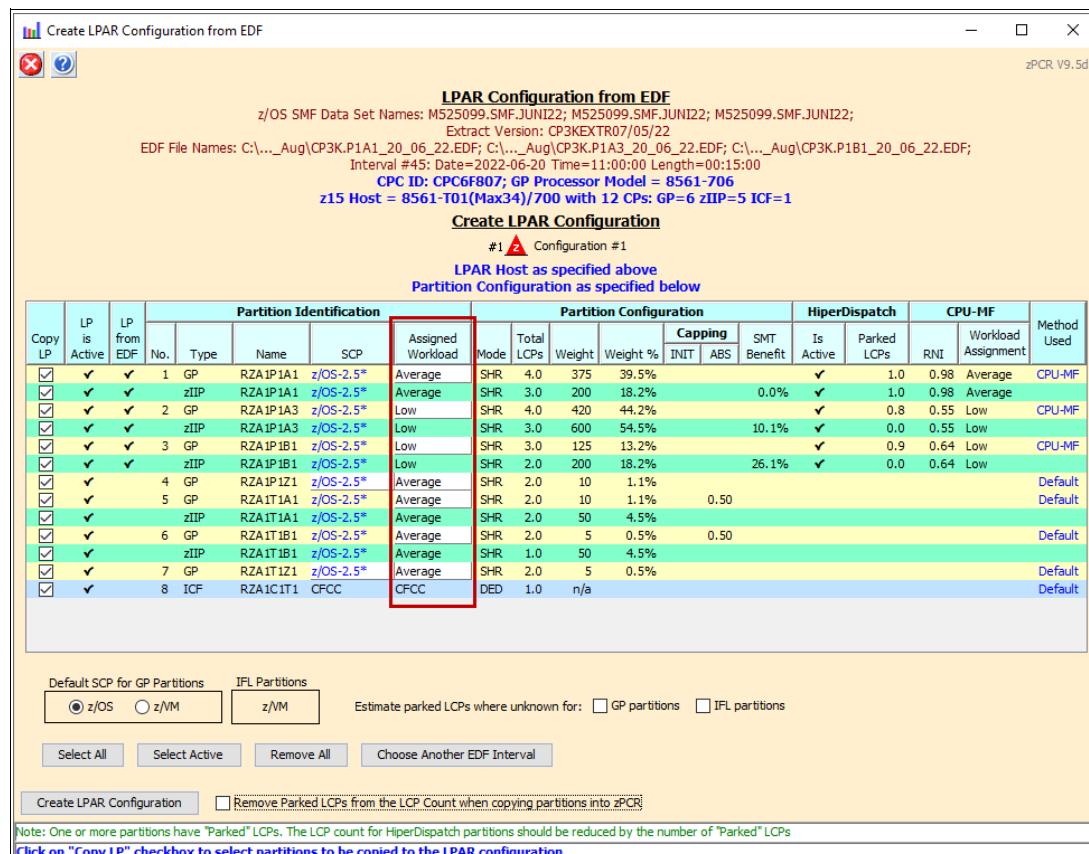


Figure 12-5 zPCR LPAR Configuration from EDF window

12.9.4 Tips to maximize IBM z16 A02 and IBM z16 AGZ capacity

The capacity of the IBM z16 A02 and IBM z16 AGZ can be maximized by using the following tips:

- Turn on HiperDispatch in every LPARs. Hiperdispatch optimizes processor cache usage by creating an affinity between a PU and the workload.
- Assign an appropriate number of logical CPUs. If you assign too many logical CPUs to the LPAR, unnecessary LPAR management cost is exhausted. This issue reduces the efficiency of the cache.

The capacity declines relative to the LCP:RCP ratio (sum of logical CPUs defined in all LPARs: the number of physical CPUs on your configuration). Therefore, assigning the correct number of CPU to LPAR is important.

If your LPARs configuration LCP:RCP ratio reaches its limit, zPCR warns your configuration. Figure 12-6 on page 465 shows a sample zPCR error message window when the practical LCP:RCP ratio is exceeded.

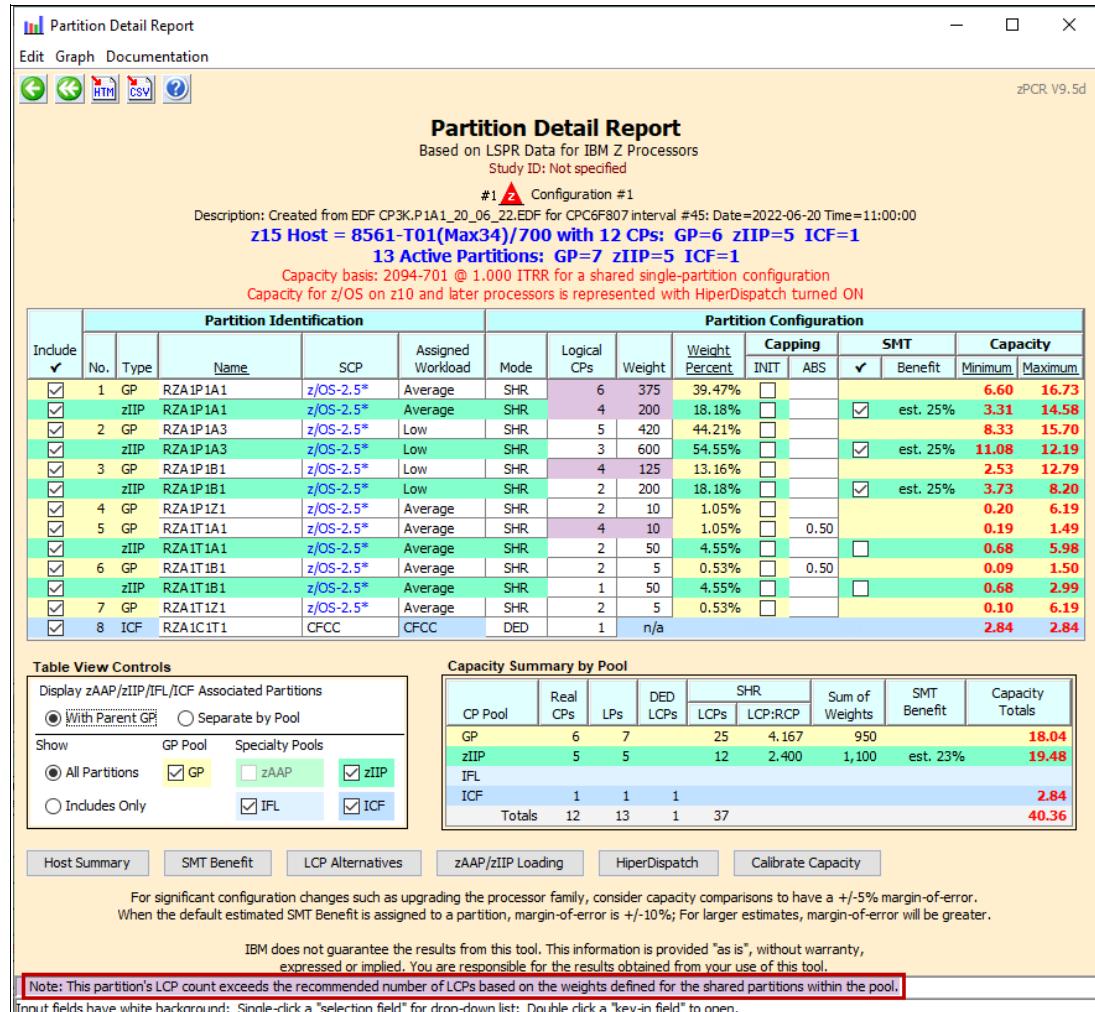
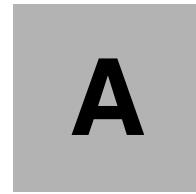


Figure 12-6 zPCR message window



Channel options

This appendix describes all channel attributes, the required cable types, the maximum unrepeated distance, and the bit rate for IBM z16 A02 and IBM z16 AGZ . The features that are hosted in the PCIe drawer for Cryptography are also listed.

For all optical links, the connector type is LC Duplex (except for the zHyperLink) and the ICA SR connections, which are established with multifiber push-on (MPO) connectors.

The MPO connector of the zHyperLink, and the ICA connection features two rows of 12 fibers and are interchangeable.

The electrical Ethernet cable for the Open Systems Adapter (OSA) connectivity is connected through an RJ45 jack.

The attributes of the channel options that are supported on IBM z16 A02 and IBM z16 AGZ are listed in Table A-1.

Table A-1 IBM z16 A02 and IBM z16 AGZ channel feature support

Channel feature	Feature codes	Bit rate ^a in Gbps (or stated)	Cable type	Maximum unrepeated distance ^b	Ordering information
zHyperLink and Fiber Connection (FICON)					
zHyperlink Express1.1	0451	8 Gbps	OM4, OM3	See Table A-2 on page 469	New build, Carry forward
zHyperlink Express	0431				Carry forward
FICON Express32S LX	0461	8, 16, or 32	SM 9 μm	5 km @ 32 Gbps ^c (3.1 miles) 10 km @ 16 or 8 Gbps (6.2 miles)	New build
FICON Express32S SX	0462	8, 16, or 32	OM1, OM2, OM3, or OM4	See Table A-3 on page 469.	New build
FICON Express16SA LX	0436	8, or 16	SM 9 μm	10 km (6.2 miles)	Carry forward

Channel feature	Feature codes	Bit rate ^a in Gbps (or stated)	Cable type	Maximum unrepeated distance ^b	Ordering information
FICON Express16SA SX	0437	8, or 16	OM1, OM2, OM3, OM4	See Table A-3 on page 469.	Carry forward
FICON Express16S+ LX	0427	4, 8, or 16	SM 9 µm	10 km (6.2 miles)	Carry forward
FICON Express16S+ SX	0428	4, 8, or 16	OM1, OM2, OM3, OM4	See Table A-3 on page 469.	Carry forward
FICON Express16S LX	0418	4, 8, or 16	SM 9 µm	10 km (6.2 miles)	Carry forward
Open Systems Adapter (OSA) and Remote Direct Memory over Converged Ethernet (RoCE)					
OSA-Express7S 1.2 25GbE LR	0460	25	SM 9 µm	10 km (6.2 miles)	New build
OSA-Express7S 1.2 25GbE SR	0459	25	MM 50 µm	70 m (2000) 100 m (4700)	New build
OSA-Express7S 1.2 10GbE LR	0456	10	SM 9 µm	10 km (6.2 miles)	New build
OSA-Express7S 1.2 10GbE SR	0457	10	MM 62.5 µm MM 50 µm	33 m (200) 82 m (500) 300 m (2000)	New build
OSA-Express7S 1.2 GbE LX	0454	1.25	SM 9 µm	10 km (6.2 miles)	New build
OSA-Express7S 1.2 GbE SX	0455	1.25	MM 62.5 µm MM 50 µm	275 m (200) 550 m (500)	New build
OSA-Express7S 1.2 1000BASE-T	0458	1000Mbps	Cat 5, Cat 6 unshielded twisted pair (UTP)	100 m	New build
OSA-Express7S 25GbE SR1.1	0449	25	MM 50 µm	70 m (2000) 100 m (4700)	Carry forward
OSA-Express7S 25GbE SR	0429				
OSA-Express7S 10GbE LR	0444	10	SM 9 µm	10 km (6.2 miles)	Carry forward
OSA-Express6S 10GbE LR	0424				
OSA-Express7S 10GbE SR	0445	10	MM 62.5 µm MM 50 µm	33 m (200) 82 m (500) 300 m (2000)	Carry forward
OSA-Express6S 10GbE SR	0425				
OSA-Express7S GbE LX	0442	1.25	SM 9 µm	5 km (3.1 miles)	Carry forward
OSA-Express6S GbE LX	0422				
OSA-Express7S GbE SX	0443	1.25	MM 62.5 µm MM 50 µm	275 m (200) 550 m (500)	Carry forward
OSA-Express6S GbE SX	0423				
OSA-Express7S 1000BASE-T	0446	1000 Mbps	Cat 5, Cat 6 unshielded twisted pair (UTP)	100 m	Carry forward
OSA-Express6S 1000BASE-T	0426	100 or 1000 Mbps			
25GbE RoCE Express3 LR	0453	25	SM 9 µm		New build
25GbE RoCE Express3 SR	0452	25	OM3, OM4		New build

Channel feature	Feature codes	Bit rate ^a in Gbps (or stated)	Cable type	Maximum unrepeated distance ^b	Ordering information
10GbE RoCE Express3 LR	0441	10	SM 9 μm		New build
10GbE RoCE Express3 SR	0440	10	OM2, OM3 OM4	33 m (200) 82 m (500) 300m (2000) ^d	New build
25GbE RoCE Express2.1	0450	25	OM3, OM4	70 m (2000) 100 m (4700)	Carry forward
25GbE RoCE Express2	0430				
10GbE RoCE Express2.1	0432	10	OM2, OM3 OM4	33 m (200) 82 m (500) 300m (2000) ^d	Carry forward
10GbE RoCE Express2	0412				
Parallel Sysplex					
CE2 LR	0434	10 Gbps	SM 9 μm	10 km (6.2 miles)	New build
ICA SR1.1	0176	8 GBps	OM4 OM3	150 m 100 m	New build, Carry forward
ICA SR	0172				

- a. The link data rate does not represent the performance of the link. The performance depends on many factors, including latency through the adapters, cable lengths, and the type of workload.
- b. Where applicable, the minimum fiber bandwidth distance in MHz-km for multi-mode fiber optic links is included in parentheses.
- c. For 32 Gbps links, point to point (to another switch, director, DWDM equipment or another FICON Express3sS) distance is limited to 5 km.
- d. A 600 meters maximum when sharing the switch across two RoCE Express features.

The unrepeated distances for different multimode (MM) fiber optic types for zHyperLink Express are listed in Table A-2.

Table A-2 Unrepeated distances

Cable type (modal bandwidth)	8 Gbps
OM3 (50 μm at 2000 MHz-km)	100 meters
	328 feet
OM4 (50 μm at 4700 MHz-km)	150 meters
	492 feet

The maximum unrepeated distances for FICON SX features are listed in Table A-3.

Table A-3 Maximum unrepeated distance for FICON SX features

Cable type/bit rate	8 Gbps	16 Gbps	32 Gbps
OM1 (62.5 μm at 200 MHz-km)	21 meters	15 meters	N/A
	69 feet	49 feet	N/A
OM2 (50 μm at 500 MHz-km)	50 meters	35 meters	20 meters
	164 feet	115 feet	65 feet

Cable type/bit rate	8 Gbps	16 Gbps	32 Gbps
OM3 (50 µm at 2000 MHz·km)	150 meters	100 meters	70 meters
	492 feet	328 feet	229 meters
OM4 ^a (50 µm at 4700 MHz·km)	190 meters	125 meters	100 meters
	693 feet	410 feet	328 feet

a. Fibre Channel Standard (not certified for Ethernet)



IBM Z Integrated Accelerator for AI

This appendix provides a short overview of the new IBM Z Integrated Accelerator for Artificial Intelligence implemented in the IBM z16 A02 and IBM z16 AGZ processor (Telum).

I

Overview

Each generation of the IBM zSystems processing chip has been enhanced with new on-chip functions such as compression, sort, cryptography and vector processing. The purpose-built accelerators that provide these functions mean lower latency and higher throughput for specialized operations. These work together with advanced chip design features such as data pre-fetch, high capacity L1 and L2 caches, branch prediction, and other innovations.

This also supports and enables compliance with security policies, as the data is not leaving the platform to be processed. The hardware, firmware and software are vertically integrated to deliver this functionality seamless to the applications.

In August 2021 IBM announced new generation of [IBM zSystems processor, Telum](#) with new Artificial Intelligence (AI) accelerator (Figure 12-7), an innovation that will bring incredible value to the applications and workloads running on IBM Z.

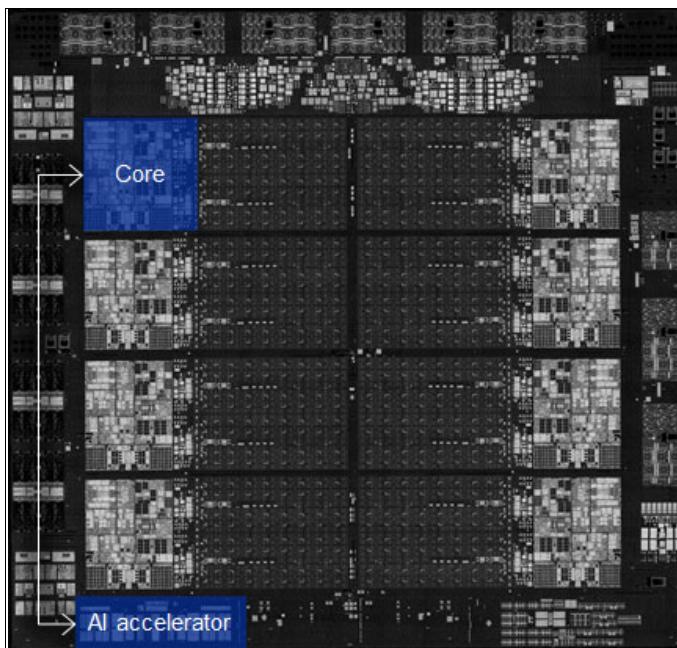


Figure 12-7 IBM z16 A02 and IBM z16 AGZ Processor chip - AIU location

With the new IBM Z Integrated Accelerator for AI clients can benefit from the acceleration of AI operations such as fraud detection, customer behavior predictions and streamlining of supply chain operations - all in real time, clients are able to derive the valuable insights from their data instantly. AI accelerator is designed to deliver AI inference in real time, at large scale and rate, with no transaction left behind without the need to offload data off the IBM zSystems for performing AI inference.

The AI capability is applied directly into the running transaction - shifting the traditional paradigm of applying AI to the transactions that have already completed. This innovative technology can be used for intelligent IT workloads placement algorithms, contributing to the better overall system performance. The co-processor is driven by the new NNPA (Neural Networks Processing Assist) instruction.

NNPA and IBM z16 A02 and IBM z16 AGZ Hardware

NNPA is a new non-privileged CISC (Complex Instruction Set Computer) memory-to-memory instruction that operates on tensor objects that are in user programs' memory. AI functions and macros are abstracted by NNPA.

The logical diagram in Figure 12-8 shows the AI accelerator and its components: the data movers surrounding the compute arrays composed by the Processor Tiles (PT), the Processing Elements (PE) and the Special Function Processors (SFP).

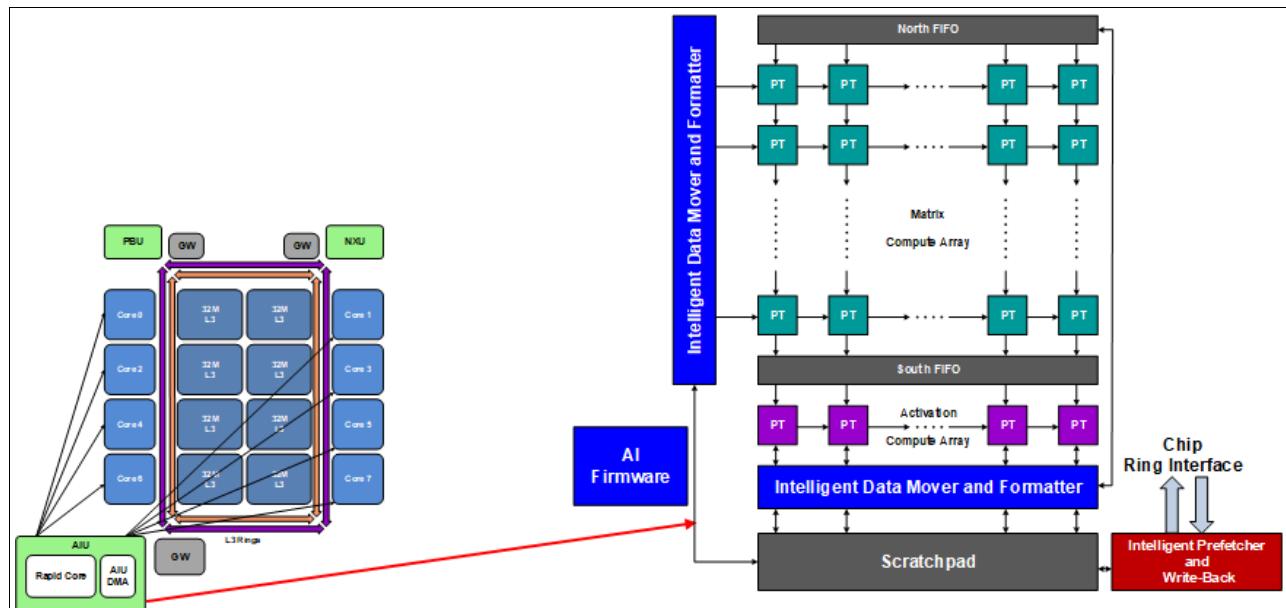


Figure 12-8 AIU logical diagram

Intelligent data movers and prefetchers are connected to the chip via ring interface for high speed low latency read-write cache operations: 200+GB/s read/store bandwidth; and 600+GB/s bandwidth between engines.

Compute Arrays consist of 128 processor tiles with 8-way FP-16 FMA SIMD, optimized for matrix multiplication and convolution, and 32 processor tiles with 8-way FP-16/FP-32 SIMD, optimized for activation functions and complex functions.

The integrated AI accelerator delivers more than 6 TFLOPs per chip and over 200 TFLOPs in the 32 chip system (a fully configured IBM z16 A02 and IBM z16 AGZ with four CPC drawers).

The AI accelerator is shared by all cores on the chip. The firmware, running on the cores and accelerator, orchestrates and synchronizes the execution on the accelerator.

How to leverage IBM Z Integrated AI Accelerator in your enterprise

This chart shows the high-level of seamless integration of AI accelerator into enterprise AI/ML solution stack. There's a great flexibility and interoperability for training and building models.

Acknowledging the very diverse AI training frameworks, clients can train their models on platforms of their choice, including IBM zSystems (on-prem and in hybrid cloud) and then deploy it efficiently on IBM Z, in collocation with the transactional workloads. There's no additional development effort needed to enable this strategy.

To allow this flexible “Train anywhere, Deploy on IBM Z” approach, IBM invests into ONNX (Open Neural Network Exchange) [<https://onnx.ai>] technology (Figure 12-9).

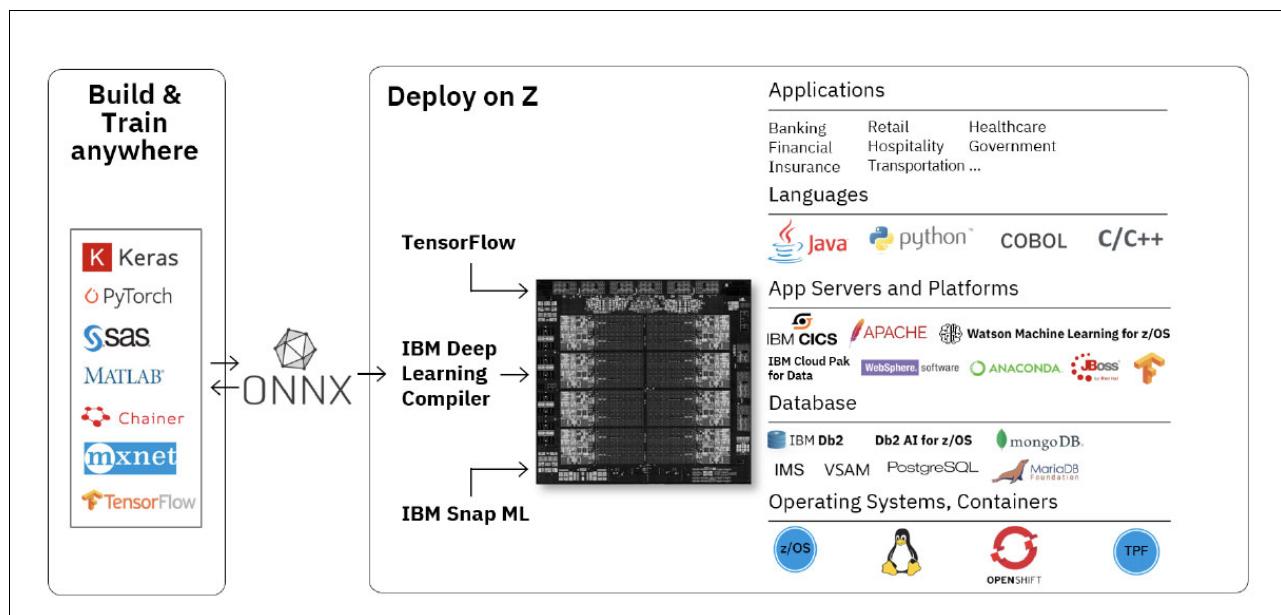


Figure 12-9 ONNX ecosystem

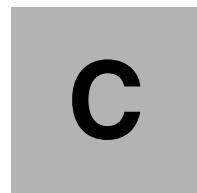
This is a standard format for representing AI models allowing a data scientist to build and train a model in the framework of choice without worrying about the downstream inference implications. To enable deployment of ONNX models, IBM provides an ONNX model compiler that is optimized for IBM Z. In addition to this, IBM is optimizing key open source frameworks such as TensorFlow (and TensorFlow Serving) for use on IBM Z.

IBM open sourced [zDNN library](#) provides common APIs for the functions allowing conversion from tensor format to the accelerator required format. Clients can run zDNN both under z/OS¹ and Linux on IBM Z. There's also a DLC (Deep Learning Compiler) for z/OS and Linux, providing the AI functionality to the applications.

References

- ▶ IBM Telum Processor: the next-gen microprocessor for IBM zSystems and IBM LinuxONE:
<https://www.ibm.com/blogs/systems/ibm-telum-processor-the-next-gen-microprocessor-for-ibm-z-and-ibm-linuxone/>
- ▶ Leveraging ONNX Models on IBM zSystems and LinuxONE
<https://community.ibm.com/community/user/ibmz-and-linuxone/blogs/andrew-sica/2021/10/29/leveraging-onnx-models-on-ibm-z-and-linuxone>
- ▶ Jump starting your experience with AI on IBM Z
<https://blog.share.org/Article/jump-starting-your-experience-with-ai-on-ibm-z>

¹ zDNN is zCX eligible which it runs on zIIPs under z/OS



IBM Integrated Accelerator for zEnterprise Data Compression

This appendix describes the new IBM Integrated Accelerator for zEnterprise Data Compression (zEDC) that is implemented in IBM zSystems hardware.

The appendix includes the following topics:

- ▶ “Client value of IBM zSystems compression” on page 476
- ▶ “IBM z16 A02 and IBM z16 AGZ IBM Integrated Accelerator for zEDC” on page 476
- ▶ “IBM z16 A02 and IBM z16 AGZ migration considerations” on page 478
- ▶ “Software support” on page 478
- ▶ “Compression acceleration and Linux on IBM Z” on page 479

Client value of IBM zSystems compression

The amount of data that is captured, transferred, and stored continues to grow. Software-based compression algorithms can be costly in terms of processor resources, storage costs, and network bandwidth.

An optional PCIe feature that was available for IBM z14, zEDC Express, addressed customer requirements by providing hardware-based acceleration for data compression and decompression. zEDC provided data compression with lower CPU consumption than compression technology that was available on the IBM zSystems server.

Clients deployed zEDC compression to deliver the following types of compression:

- ▶ Storage
- ▶ Data transfer
- ▶ Database
- ▶ In-application

Data compression delivers the following benefits:

- ▶ Disk space savings
- ▶ Improved elapse times
- ▶ Reduced CPU consumption
- ▶ Reduced network bandwidth requirements and transfer times

Many clients increased their zEDC footprint to 8GBps with up to 16 features per IBM z14 system at 1 GBps throughput per feature (redundancy reduces total throughput to 8 GBps).

While the zEDC PCIe feature provided CPU savings by offloading the select compression and decompression operations, it had also the drawback of limited virtualization capabilities (one zEDC PCIe feature could be shared across a maximum of 15 LPARs) as well as limited bandwidth.

IBM z15 introduced for the first time an on-chip accelerator (implemented in the PU chip) for compression and decompression operations, which was tied directly into processor's L3 cache, thus providing much higher bandwidth as well as removing the virtualization limitations of a PCIe feature.

The IBM z16 A02 and IBM z16 AGZ further address the growth of data compression requirements with the integrated on-chip compression unit (implemented in processor Nest, one per PU chip) that significantly increases compression throughput and speed compared to previous zEDC deployments.

IBM z16 A02 and IBM z16 AGZ IBM Integrated Accelerator for zEDC

IBM z16 A02 and IBM z16 AGZ on-chip compression provides value for existing and new compression users by bringing the compression facility into the PU chip, which is tied in L3¹ cache.

¹ Virtual L3 (shared victim) cache for IBM z16 A02 and IBM z16 AGZ - see Chapter 2, "Central processor complex hardware components" on page 21

The IBM z16 A02 and IBM z16 AGZ Integrated Accelerator for zEDC delivers industry-leading throughput and replaces the zEDC Express PCIe adapter that is available on the IBM z14 and earlier servers.

IBM z16 A02 and IBM z16 AGZ compression/decompression is implemented in the Nest Accelerator Unit (NXU, see Figure 3-15 on page 92) on each processor chip and replaces the existing zEDC Express adapter in the PCIe+ I/O drawer.

One Nest Accelerator Unit is available per processor chip, which is shared by all cores on the chip and features the following benefits:

- ▶ New concept of sharing and operating an accelerator function in the nest
- ▶ Supports DEFLATE compliant compression/decompression and GZIP CRC/ZLIB Adler
- ▶ Low latency
- ▶ High bandwidth
- ▶ Problem state execution
- ▶ Hardware and firmware interlocks to ensure system responsiveness
- ▶ Designed instruction
- ▶ Run in millicode

On-Chip Compression provides an up to 5% improvement in compression ratios for BSAM/VSAM datasets over zEDC while maintaining full compatibility.

Eliminating adapter sharing by using Nest Compression Accelerator

Sharing of zEDC cards is limited to 15 LPAR guests per adaptor. The Nest Compression Accelerator removes this virtualization constraint because it is shared by all PUs on the processor chip and therefore is available to all LPARs and guests.

Moving the compression function from the (PCIe) I/O drawer to the processor chip means that compression can operate directly in L3 cache and data does not need to be passed by using I/O operations.

Compression modes

Compression is run in one of the following modes:

- ▶ Synchronous

Execution occurs in problem state where the user application starts the instruction in its virtual address space.

- ▶ Asynchronous

Execution is optimized for Large Operations under z/OS for authorized applications (for example, BSAM) and issues I/O by using EADMF for asynchronous execution.

This type of execution maintains the current user experience and provides a transparent implementation for authorized users of zEDC.

Note: The zEDC Express feature does *not* carry forward to IBM z16 A02 and IBM z16 AGZ.

IBM z16 A02 and IBM z16 AGZ migration considerations

The IBM Integrated Accelerator for zEDC is fully compatible with zEDC. Data compressed by zEDC can be read by IBM z16 A02 and IBM z16 AGZ (the on-chip) nest accelerator unit and vice versa.

All z/OS configuration stay the same

No changes are required when moving from earlier systems using zEDC to IBM z16 A02 and IBM z16 AGZ.

The IFAPRDxx feature is still required for authorized services. For problem state services, such as zlib usage of Java, it is not required.

Consider fail-over and DR sizing

The order of magnitude throughput increase on IBM z16 A02 and IBM z16 AGZ means that the throughput requirements need to be considered whether failing over to earlier systems with zEDC.

Performance metrics

On-chip compression introduces the following system reporting changes:

- ▶ No RMF PCIE reporting for zEDC
- ▶ Synchronous executions are not recorded (just an instruction invocation)
- ▶ Asynchronous executions are recorded:
 - SMF30 information is captured for asynchronous usage
 - RMF EADM reporting is enhanced (RMF 74.10)
 - SAP utilization is updated to include time spent compressing and decompressing

zEDC to IBM z16 A02 and IBM z16 AGZ zlib Program Flow for z/OS

The z/OS provided zlib library is statically linked into many IBM and ISV products and remains functional. However, to get the best optimization for IBM z16 A02 and IBM z16 AGZ, some minor changes are made to zlib.

The current zlib and the new zlib function is available for IBM z14 and earlier servers and IBM z16 A02 and IBM z16 AGZ hardware. It functions with or without the IBM z16 A02 and IBM z16 AGZ z/OS PTFs on IBM z14 and earlier servers.

Software support

Support of the On-Chip Compression function is compatible with zEDC support and is available in z/OS V2R2 and later for data compression and decompression. Support for data recovery (decompression) in the case that zEDC or On-Chip Compression not available; however, it is provided through software in z/OS V2R2 with the appropriate program temporary fixes (PTFs).

Software decompression is slow and can involve considerable processor resources. Therefore, it is not recommended for production environments.

A specific fix category that is named IBM.Function.zEDC identifies the fixes that enable or use the zEDC and On-Chip Compression function.

z/OS guests that run under z/VM V7.R1 with PTFs and later can use the zEDC Express feature and IBM z16 A02 and IBM z16 AGZ On-Chip Compression.

For more information, look for the Enhancements to z/VM 7.1 presentation on the z/VM Presentations page on the IBM website.

IBM 31-bit and 64-bit SDK for z/OS Java Technology Edition, Version 7 Release 1 (5655-W43 and 5655-W44) (IBM SDK 7 for z/OS Java) provides use of the zEDC Express feature and Shared Memory Communications-Remote Direct Memory Access (SMC-R), which is used by the 10GbE RoCE Express feature.

For more information about how to implement and use the IBM zSystems compression features, see *Reduce Storage Occupancy and Increase Operations Efficiency with IBM zEnterprise Data Compression*, SG24-8259.

C.0.1 IBM Z Batch Network Analyzer

IBM Z Batch Network Analyzer (zBNA) is a no-charge, “as is” tool. It is available to clients, IBM Business Partners, and IBM employees.

zBNA is based on Microsoft Windows, and provides graphical and text reports, including Gantt charts, and support for alternative processors.

zBNA can be used to analyze client-provided System Management Facilities (SMF) records to identify jobs and data sets that are candidates for zEDC and IBM z16 A02 and IBM z16 AGZ On-Chip Compression across a specified time window (often a batch window).

zBNA can generate lists of data sets by the following jobs:

- ▶ Jobs that perform hardware compression and might be candidates for On-Chip Compression.
- ▶ Jobs that might be On-Chip Compression candidates, but are not in extended format.

Therefore, zBNA can help you estimate the use of On-Chip Compression features and help identify savings. The following resources are available:

- ▶ IBM Employees can obtain zBNA and other CPS tools at the [IBM Z Batch Network Analyzer \(zBNA\) Tool page](#) of the IBM Techdoc website.
- ▶ IBM Business Partners can obtain zBNA and other CPS tools at the [IBM PartnerWorld website](#) (log in required).
- ▶ IBM clients can obtain zBNA and other CPS tools at the [IBM Z Batch Network Analyzer \(zBNA\) Tool page](#) of the IBM Techdoc Library website.

Compression acceleration and Linux on IBM Z

The zEDC I/O adapter use is limited in many Linux on IBM zSystems environments because SR-IOV does not provide a high degree of virtualization; therefore, the user must pick and choose which guests are granted access to the accelerator.

the IBM z16 A02 and IBM z16 AGZ On-Chip Compression accelerator solves these virtualization limitations because the function is no longer an I/O device and is available as a problem state instruction to all Linux on IBM Z guests without constraints.

This feature enables pervasive usage in highly virtualized environments.

IBM z16 A02 and IBM z16 AGZ On-Chip Compression is available to open source applications by way of zlib.

**D**

Rack configurations

This appendix lists the various CPC drawer and PCIe+ I/O drawer configurations for both IBM z16 A02 and IBM z16 AGZ . All the diagrams are views from the rear of the system.

The common building blocks are displayed and range from 1 - 2 CPC drawers and 1 - 3 PCIe+ I/O drawers.

Topics:

- ▶ IBM z16 A02 (Factory) Frame configurations
- ▶ IBM z16 AGZ configurations

IBM z16 A02 (Factory Frame) configurations

Factory installed configurations come with fixed rack positions. There are no space reserve features for IBM z16 A02.

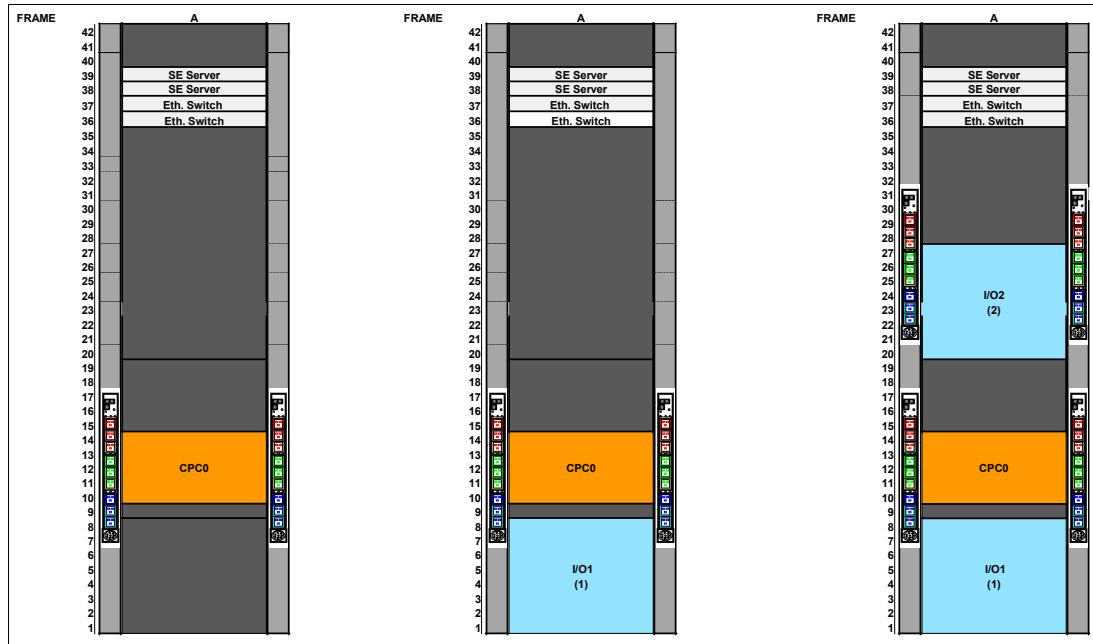


Figure 12-10 Single CPC drawer, 0, 1, and 2 PCIe+ I/O drawers

Figure 12-11 shows IBM z16 A02 configurations w/ two CPC drawers.

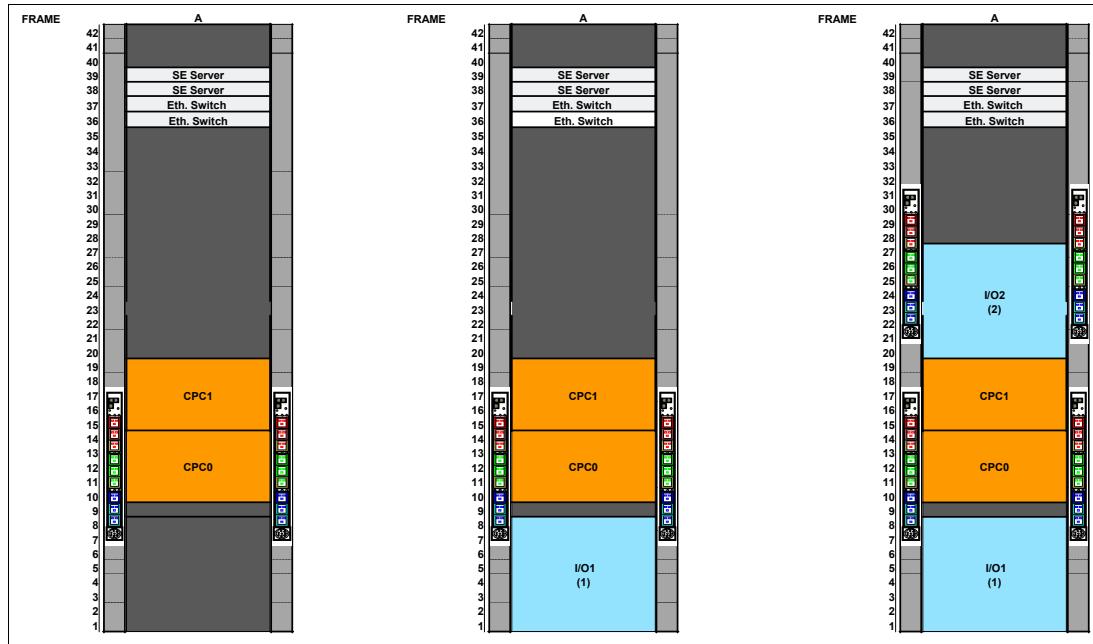


Figure 12-11 Dual CPC drawer, 0, 1, and 2 PCIe+ I/O drawers

Important: Configurations with two CPC drawers are delivered with three phase iPDU's. However, for configurations with single CPC drawer, if addition of the second CPC drawer is required, the power infrastructure must be changed to three-phase iPDU's. This is disruptive.

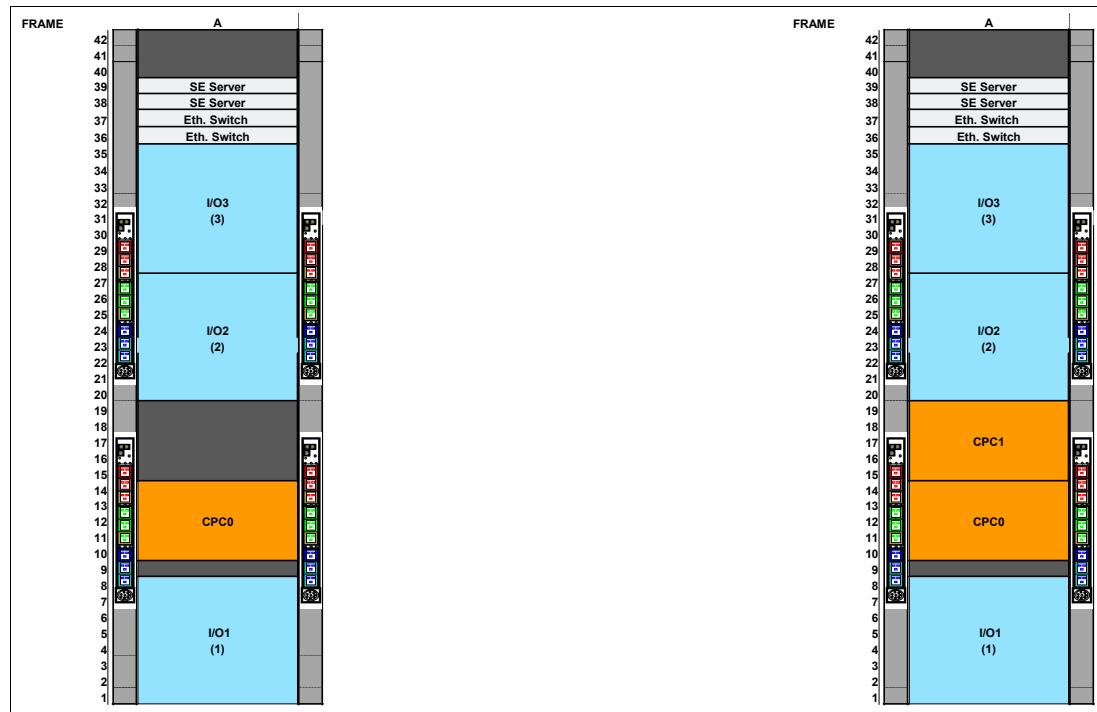


Figure 12-12 One or two CPC drawers and three PCIe+I/O drawers

IBM z16 AGZ configurations

Considerations:

- ▶ It is Clients' responsibility to plan ahead and reserve rack space for future I/O addition
- ▶ IBM Offers FC 2332 (CPC1 Reserved) only for IBM z16 AGZ
- ▶ System components install order is fixed;
- ▶ Single rack install required; no split between racks; PDUs must be in the same rack

PDUs are provided and installed by the Client, or iPDU's can be ordered from IBM, appropriate power rating and redundancy - per IBM z16 AGZ bundle Installation Manual, GC28-7036

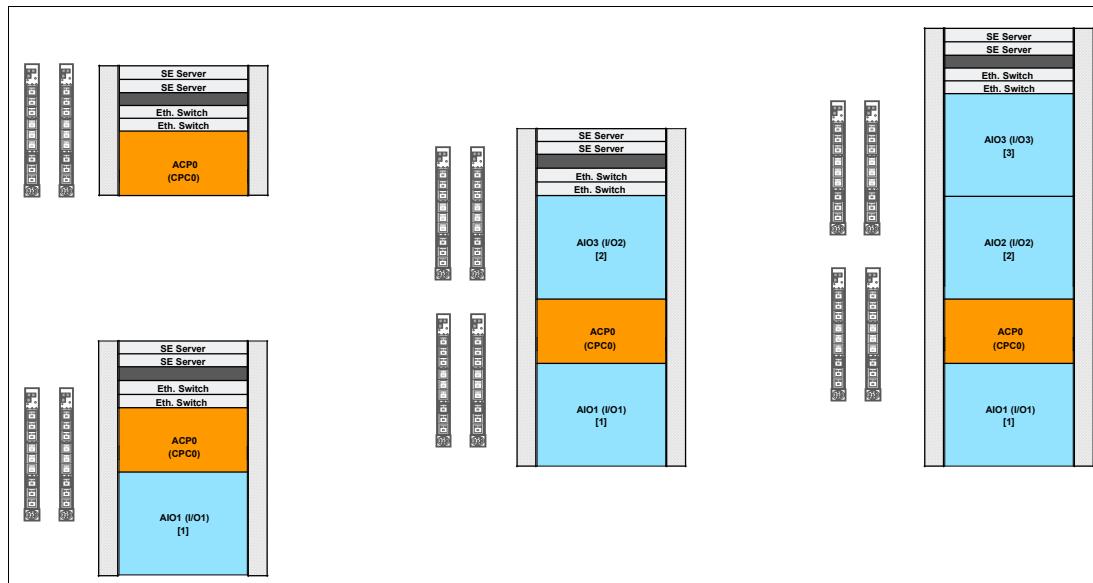


Figure 12-13 IBM z16 AGZ - one CPC drawer, 0, 1, 2, and 3 PCIe+ I/O drawers

Important: Two CPC Drawer configurations (Figure 12-14) require three-phase PDUs. Also, CPC1 Reserve feature (FC 2332) requires three-phase PDUs for non-disruptive add of the second CPC drawer.

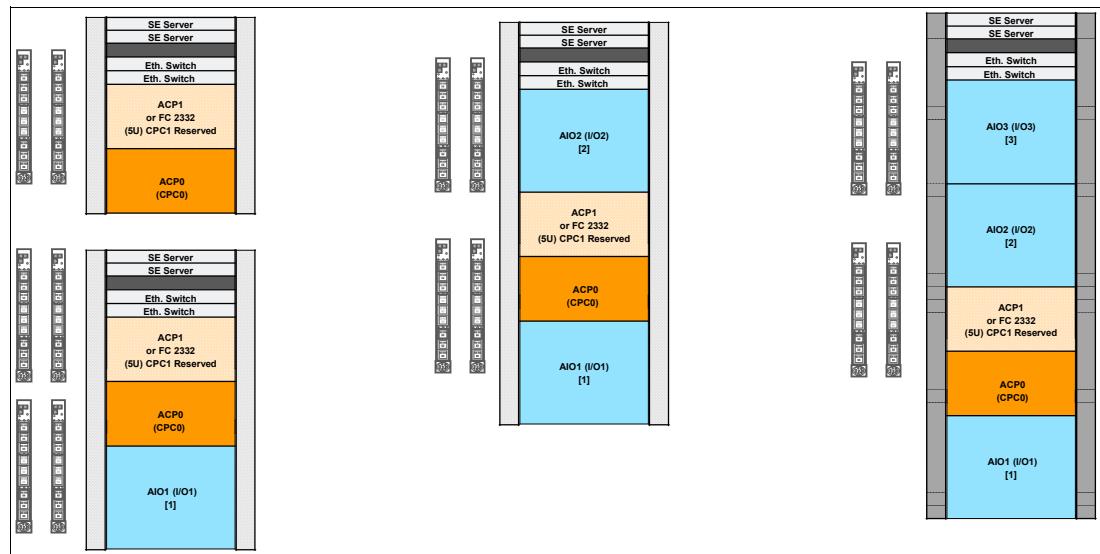


Figure 12-14 Two CPC drawers and 0, 1, 2, and 3 PCIe+ I/O drawers



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

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.

- ▶ *????full title???????, xxxx-xxxx*
- ▶ *????full title???????, SG24-xxxx*
- ▶ *????full title???????, REDP-xxxx*
- ▶ *????full title???????, TIPS-xxxx*

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:

- ▶ *????full title???????, xxxx-xxxx*
- ▶ *????full title???????, xxxx-xxxx*
- ▶ *????full title???????, xxxx-xxxx*

Online resources

These websites are also relevant as further information sources:

- ▶ Description1
[http://?????????.???.??/?](http://?????????.???.??/)
- ▶ Description2
[http://?????????.???.??/?](http://?????????.???.??/)
- ▶ Description3
[http://?????????.???.??/?](http://?????????.???.??/)

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services



Additional material

This book refers to additional material that can be downloaded from the Internet as described in the following sections.

Locating the web material

The web material associated with this book is available at:

<https://www.redbooks.ibm.com/abstracts/SG24????>

Using the web material

The additional web material that accompanies this book includes the following files:

<i>File name</i>	<i>Description</i>
?????????.zip	????Zipped Code Samples????
?????????.zip	????Zipped HTML Documents????
?????????.zip	????Zipped Presentations????

System requirements for downloading the web material

The web material requires the following system configuration:

Hard disk space:	????MB minimum????
Operating System:	????Windows/Linux????
Processor:	???? or higher????
Memory:	????MB????

Downloading and extracting the web material

Create a subdirectory (folder) on your workstation, and extract the contents of the web material .zip file into this folder.

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize-->Hide:>Set**. Move the changed Conditional text settings to all files in your book by opening the book file with the spine.fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.

Draft Document for Review May 7, 2023 1:21 pm

8952spine.fm 1



IBM z16 A02 and IBM z16 AGZ

Technical Guide

SG24-8952-00
ISBN DocISBN



(1.5" spine)
1.5" <-> 1.998"
789 <-> 1051 pages



IBM z16 A02 and IBM z16 AGZ

Technical Guide

SG24-8952-00
ISBN DocISBN



(1.0" spine)
0.875" <-> 1.498"
460 <-> 788 pages



IBM z16 A02 and IBM z16 AGZ Technical Guide

SG24-8952-00
ISBN DocISBN



(0.5" spine)
0.475" <-> 0.873"
250 <-> 459 pages



IBM z16 A02 and IBM z16 AGZ Technical Guide

(0.2" spine)
0.17" <-> 0.473"
90 <-> 249 pages

(0.1" spine)
0.1" <-> 0.169"
53 <-> 89 pages

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 326. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize-->Hide:>Set**. Move the changed Conditional text settings to all files in your book by opening the book file with the spine.fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.

Draft Document for Review May 7, 2023 1:21 pm

8952spine.fm 2



IBM z16 A02 and IBM z16 AGZ Technical Guide

SG24-8952-00
ISBN DocISBN



(2.5" spine)
2.5"->nnnn.n"
1315-> nnnn pages

IBM z16 A02 and IBM z16 AGZ Technical Guide

SG24-8952-00
ISBN DocISBN



(2.0" spine)
2.0"-> 2.498"
1052 <-> 1314 pages





SG24-8952-00

ISBN DocISBN

Printed in U.S.A.

Get connected

