

Technical Introduction to the IBM z14 ZR1 and LinuxONE Rockhopper II – Part 2

Parwez Hamid

6 November 2018

Session **BG**



Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

BladeCenter*	FICON*	IBM*	LinuxONE	Storwize*	z13s	z/OS*
Db2*	Flash Systems	IBM (logo)*	LinuxONE Emperor II	System Storage*	z14	z/VM*
DFSMSdss	GDPS*	ibm.com	LinuxONE Rockhopper II	WebSphere*	zEnterprise*	z/VSE*
DFSMSshm	HiperSockets	IBMZ*	Power Systems	z13*	zHyperLink	
ECKD	HyperSwap*	InfiniBand*	PR/SM			

* Registered trademarks of IBM Corporation

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

IT Infrastructure Library is a Registered Trade Mark of AXELOS Limited.

ITIL is a Registered Trade Mark of AXELOS Limited.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

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

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

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

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

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

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

VMware, the VMware logo, VMware Cloud Foundation, VMware Cloud Foundation Service, VMware vCenter Server, and VMware vSphere are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.

Other product and service names might be trademarks of IBM or other companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

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

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the 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.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

This information provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g. zIIPs, zAAPs, and IFLs) ("SEs"). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the "Authorized Use Table for IBM Machines" provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html ("AUT"). No other workload processing is authorized for execution on an SE. IBM offers SE at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

I/O Infrastructure

Designed for data

I/O options that protect, access, share



Pervasive Encryption

- New** CF Encryption
- New** Crypto Express6S
- Update** Speed of CPACF Payment Card Industry (PCI) HSM
- New** TKE 9.0



Getting to Data

- New** zHyperLink Express
- New** IBM Virtual Flash Memory
- New** FICON Express16S+
zHPF – Extended Distance II
zEDC Express



Accessing the Web

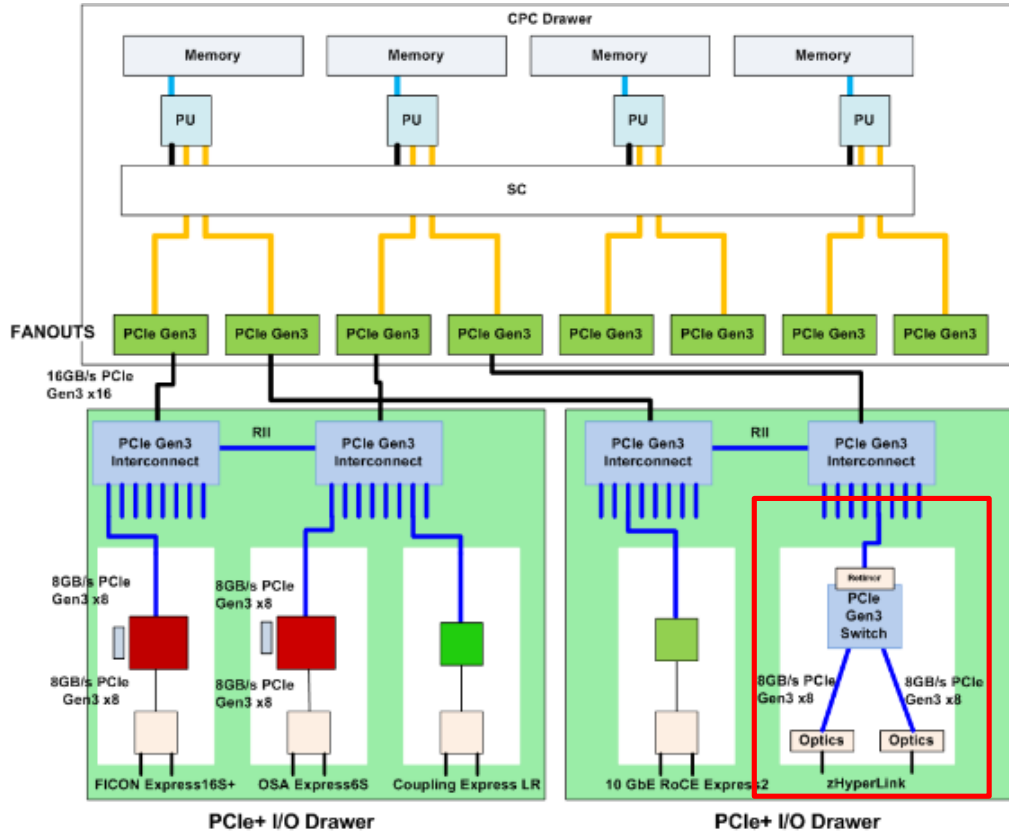
- New** 10GbE RoCE Express2
- New** OSA-Express6S
- HiperSocket
- SMC-D



Clustering to Protect

- New** Coupling Express LR ICA SR
- Plus** Improved CF scalability, constraint relief and diagnostic enhancements

z14 ZR1/LR1 I/O Infrastructure



Notes:

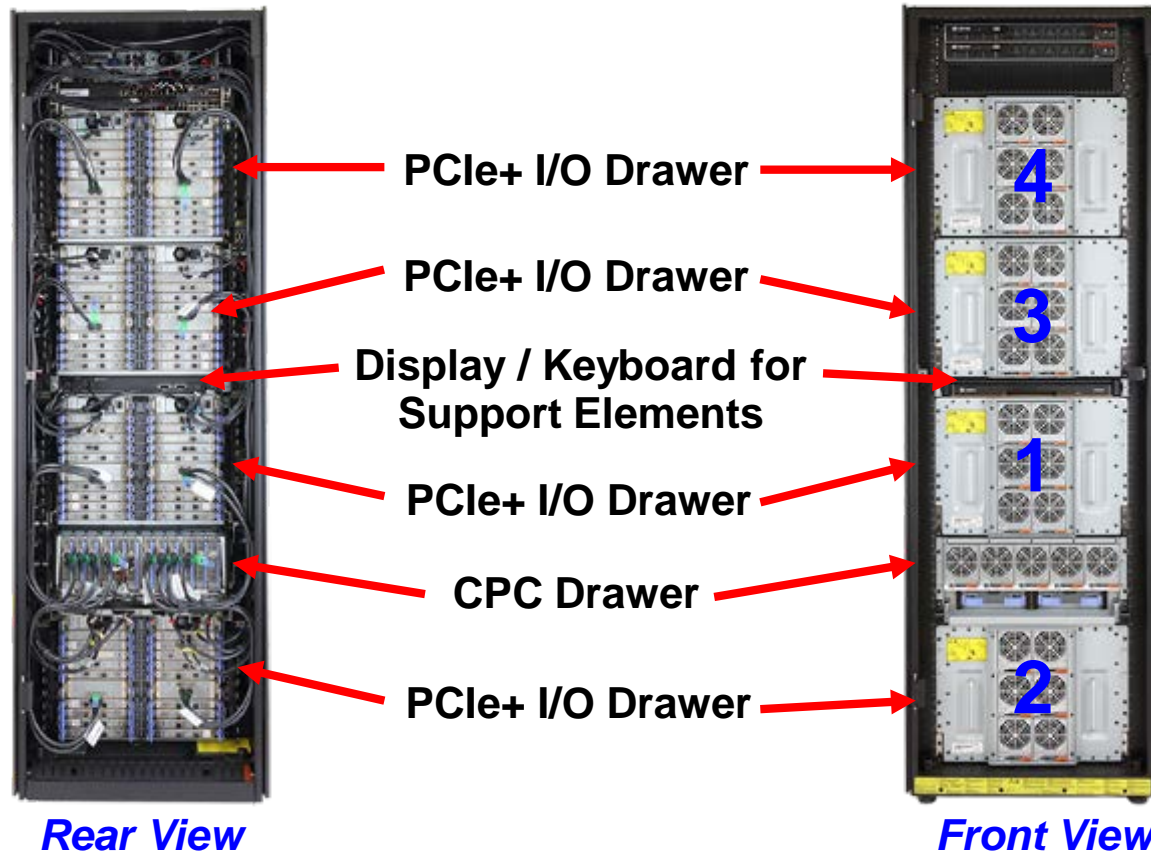
ICA-SR features use the same slot as the PCIe Gen3 Fanouts in the CPC Drawer

Not shown:

Crypto Express5S/6S, OSA Express7S, 25GbE RoCE Express2, IBM FCP Express32S and IBM Adapter for NVMe use a I/O slot in the PCIe+ I/O Drawer



IBM z14 ZR1/LR1 CPC drawer and I/O drawers



16-slot PCIe+ I/O Drawer

z14 ZR1 “New Build” I/O and MES Features Supported

▪ Features – in the PCIe+ I/O drawer

- FICON Express16S+ LX, SX (FC 0427, #0428)
- OSA-Express7S 25GbE SR (FC 0429)
- OSA-Express6S: 1 GbE (LX, SX), 10 GbE (LR, SR), and 1000BASE-T (FC 0422, 0423, 0424, 0425, 0426)
- 25GbE RoCE Express2 (FC 0430)
- 10GbE RoCE Express2 (FC 0412)
- zEDC Express (FC 0420)
- Crypto Express6S (FC 0893)
- Regional Crypto Enablement (RCE) (FC 0901)
- zHyperLink Express (FC 0431)
- Coupling Express LR (FC 0433)
- IBM FCP Express32S
- IBM Adapter for NVMe



PCIe+ I/O drawer (FC 4001)



16 I/O slots

CPC drawer



8 PCIe Fanout I/O slots

▪ PCIe Coupling Link Feature (CPC Drawer PCIe Fanout)

- ICA SR - two 8GBps PCIe Gen3 Coupling Links (FC 0172)

*Note: z14 ZR1 DOES NOT support InfiniBand features

**Note: The link data rates do not represent the performance of the links. The actual performance is dependent upon many factors including latency through the adapters, cable lengths, and the type of workload.

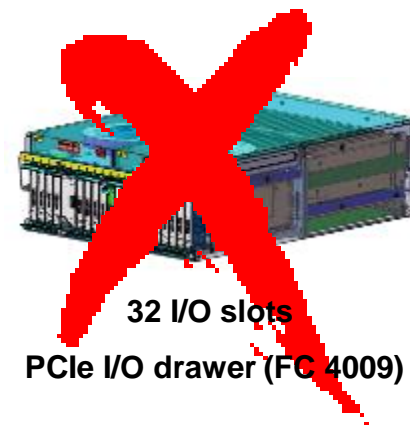
z14 ZR1 Carry Forward I/O Features Supported

▪ Features – which can be carried forward to the NEW PCIe+ I/O drawer

- FICON Express16S
- FICON Express8S
- OSA-Express5S
- OSA-Express4S – all EXCEPT 1000Base-T
- 10GbE RoCE Express (FC 0411)
- zEDC Express
- Crypto Express5S
- Regional Crypto Enablement (RCE)
- Coupling Express LR

▪ PCIe Coupling Link Feature (Fanout)

- ICA SR two 8GBps** (PCIe Gen3) Coupling Links

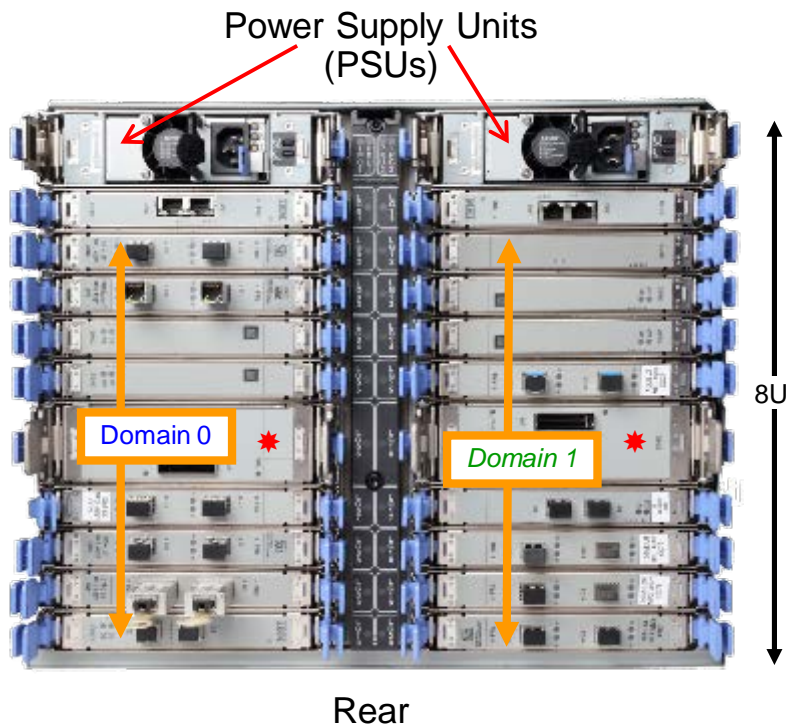


NOTE: PCIe I/O Drawer (FC 4009) cannot be carried forward during an upgrade to z14 ZR1. Supported PCIe features will be carried forward into the new PCIe+ I/O Drawer (FC 4001)

***Note:** z14 ZR1 DOES NOT support InfiniBand features

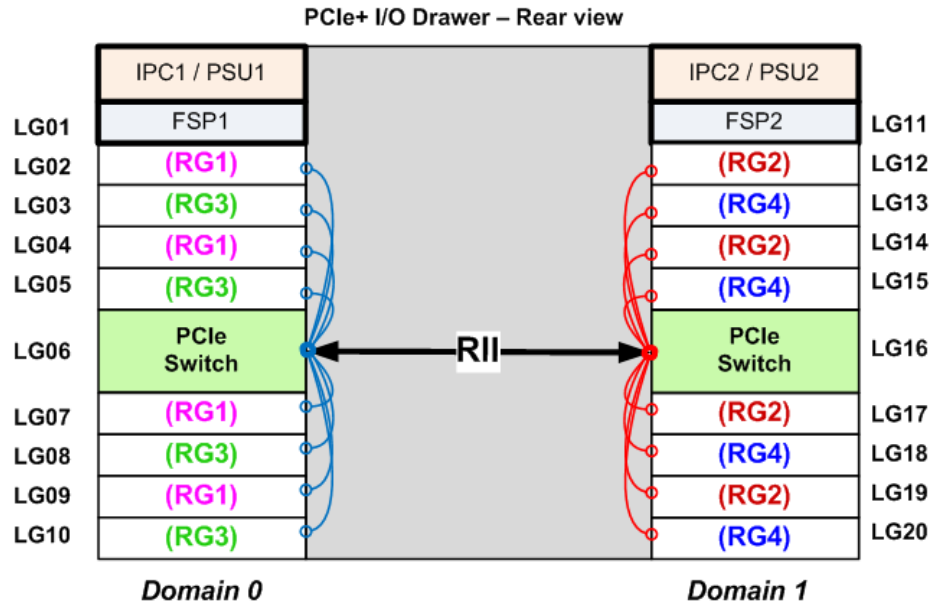
****Note:** The link data rates do not represent the performance of the links. The actual performance is dependent upon many factors including latency through the adapters, cable lengths, and the type of workload.

PCIe+ I/O drawer – 16 slots



- Supports only PCIe I/O cards
 - z14 ZR1: Up to four drawers
- Supports 16 PCIe I/O cards, horizontal orientation, in two 8-card domains (shown as 0 and 1)
- Requires two 16 GBps PCIe switch cards (*), each connected to a 16 GBps PCIe I/O interconnect to activate the two domains.
- Internal Redundant I/O Interconnect (RII) connects the two domains.
- Two redundant PSUs installed on top of the drawer provide power for the cards and are connected to the rack PDUs.
- All PCIe features and cabling are accessible from the rear of the rack.
- Concurrent field install and repair.
- Requires 8 EIA Units of rack space (14 inches ≈ 355 mm)

z14 ZR1 “Native” PCIe feature Plugging and Resource Groups (RGs)



Note: Resource Groups (RGs) in parentheses apply to select “native” PCIe features

- Configurator (eConfig) card placement:
 - Places features of the same type in I/O slots to balance the number in I/O Domains and Resource Groups
 - Reports RG assigned in “AO Data”
- Each slot in a drawer is assigned to an RG*
 - Note: These rules are NOT the same as z14 M0x.
 - Domain Pair 0 and 1
 - RG1: slots 2, 4, 7 and 9
 - RG2: slots 12, 14, 17 and 19
 - RG3: slots 3, 5, 8 and 10
 - RG4: slots 13, 15, 18 and 20
- Each feature’s physical channel ID (PCHID) is assigned based on its slot and the PCIe I/O drawer location.
- For high availability: Note the RG assignment and PCHID of each native PCIe feature*. Balance assignment of each PCIe feature type to every LPAR between the four RGs.

16U Reserved Feature

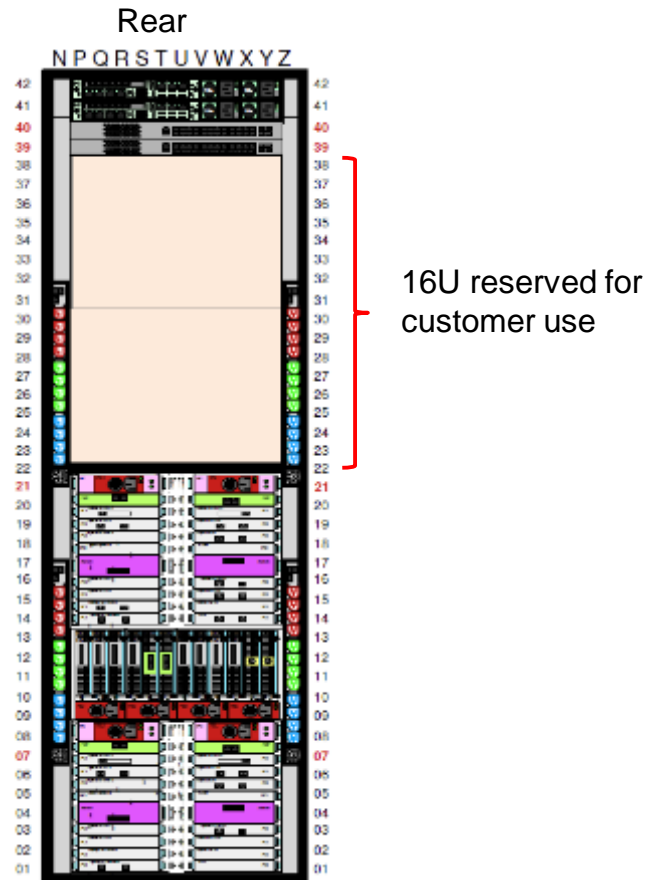
FC 0617

IBM DS8882F (Rack Mounted)

16U Reserved Feature Detail

- Feature code is based on the ability to allocate (and use) contiguous 16U space within system configuration in lieu of 3rd and 4th PCIe+ I/O Drawers.
- The 16U Reserved Rack Space feature is documented in the M/T 3907 IMPP document.
- Some examples* of equipment that can be hosted in the space allocated through the 16U Reserved feature:
 - SAN switches
 - Network Switches
 - Rack mountable HMC and TKE
 - Storage devices
 - etc.,

* Note: All equipment installed in the space allocated for the 16U Reserved feature must comply with requirements described in the IBM 3907 Installation Manual for Physical Planning, GC28-6974



z14 ZR1/LR1 -16U Reserved Feature Code – FC #0617

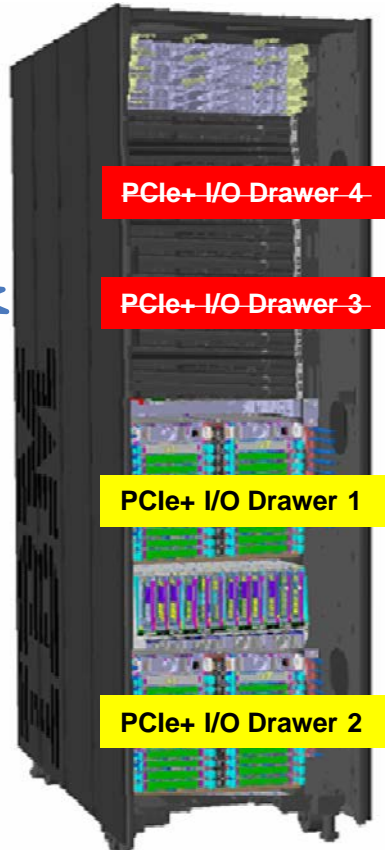
Service Element 2x (1U)
Switch 2x (1U)

Delete of FC 0617 permitted if there is a need for the 3rd PCIe+ I/O Drawer (or more).

GA2

Monitor (1U)
IO Drawer (8U)
CPC Drawer (5U)
IO Drawer (8U)

Must be merged/installed at the client site



16 rack units (16U) of open space tagged for “Flex Usage” via Feature Code in customer configuration

The 16U space utilizes the PDU’s that would’ve otherwise been used by PCIe+ drawers 3 & 4 (10 outlets).

Fit within 19” rail-to-rail width, and 28 1/4” front-to-rear depth.

Front to rear airflow.

No more than 20.4 kg weight per EIA location. (e.g. a 4U drawer can weight no more than 180 lbs ~81.6 kg)

Total power capacity available for the 16U’s of “flex frame space” is 3200W.

16U Reserved – Feature Code #0617

For the first time on IBM Z, clients can populate the rack with their choice of non-IBM elements.

Feature Code	FC0617 Yes or No?	# of PCIe+ I/O Drawers / Fanouts	Max CPs	Max Characterizable Cores	Maximum Memory
0636	Yes	1 (16 slots) / 2	4	4	2 TB
0637	Yes	2 (32 slots) / 4	6	12	4 TB
0638	No	3 or 4 (48-64 slots) / 8	6	24	8 TB
0639	No	3 or 4 (48-64 slots) / 8	6	30	8 TB
0638	Yes	2 (32 slots) / 8	6	24	8 TB
0639	Yes	2 (32 slots) / 8	6	30	8 TB

The 3rd and 4th PCIe+ I/O Drawers prevents FC0617 16U Reserved. If FC0617 is ordered, there is a maximum of 2 PCIe+ drawers (32 I/O slots).

IBM Z and Storage synergy

Storage Networking



SAN256B-6



SAN512B-6



SAN64B-6



SAN42B-R



Flash and Hybrid Storage Systems

z/VM®, Linux on Z (FCP only)

z/OS, z/VM, z/TPF, zVSE®,
Linux on Z



FlashSystem™
A9000



Storwize®
V7000 / V7000F



FlashSystem
V9000



FlashSystem
FS900

**NEW!!!
DS8880F
Storage**

DS8880



Other examples of uses for 16u Reserved include IBM 1u HMC, TKE, Power Systems™, NVMe

IBM DS8882F Rack Mounted

IBM introduced a new member of the DS8880F family of all-flash data systems that span a wide range of business-critical application workloads

Can be integrated into **IBM z14 ZR1** or **IBM LinuxONE Rockhopper II** systems

Feature code (FC 0617) based ability to use 16U of contiguous space in the **IBM z14 ZR1** or **LinuxONE Rockhopper II** system frames

Provides a midrange product with the same advanced functions as the larger DS8880F systems

From 6.4 TB to 368.64 TB of all-flash capacity

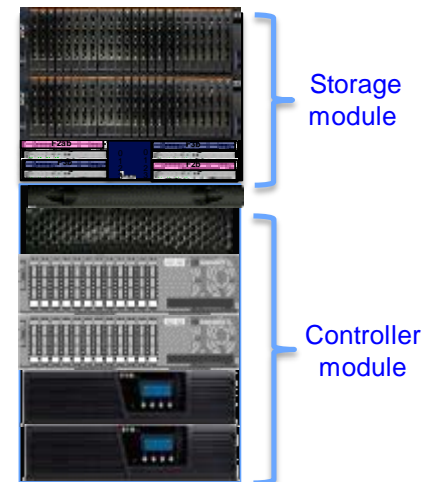
Guidelines for physical structures as well as restriction of interaction with the 'mainframe server' provided



DS8880F in a flexible rack mounted solution

1. Provides a market replacement path for
 - DS6000/DS8100/DS8300/DS8700/DS8800 EoS Systems
 - Small capacity DS8870 Systems
2. Reduces the datacenter footprint and power infrastructure requirements
3. Can be integrated into existing client 19-inch FF standard racks
4. Can be integrated into IBM z14 ZR1 or IBM LinuxOne Rockhopper II systems
5. Provides a midrange product with the same capabilities as the larger DS8880F systems:
 - Same advanced function capabilities
 - Same availability
 - Same application response time

Optional
keyboard
and monitor



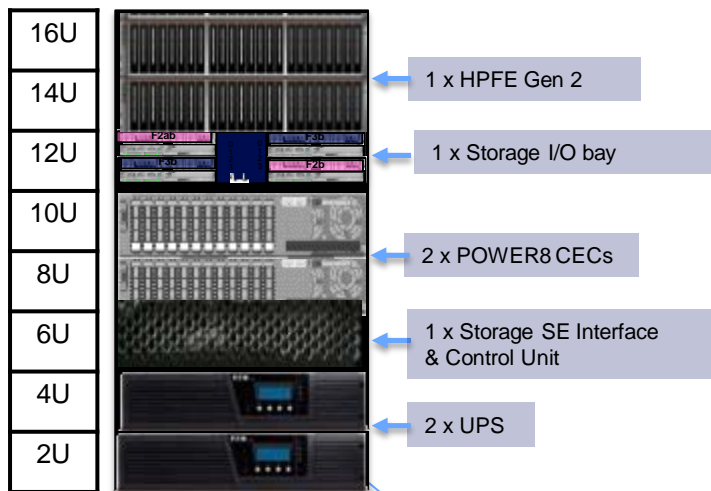
16U to be mounted into:

- IBM Z ZR1
- IBM LinuxOne Rockhopper II

- ✓ From 6.4 TB to 368.64 TB of all-flash capacity
- ✓ From 64 GB to 256 GB of memory cache (DRAM)
- ✓ From 8 to 16 FCP/FICON ports
- ✓ 2 flexible options:

DS8882F fixed 16U configuration details

16U fixed configuration



The redundant UPS units safeguard the information in case of a power outage

Processor complex (CEC)	2 x IBM Power Systems S822
POWER 8 cores per CEC	6
System memory (min / max)	64 GB / 256 GB
16Gb FICON/FCP Ports (min / max)	8 / 16
Storage I/O bay	1
Storage SE Interface & Control Unit - contains two HMC's, RPC's, Ethernet switches	1
High-Performance Flash Enclosures Gen2	1
Flash cards (min / max)	16 / 48
Flash card size options	Either High-Performance flash (HPF) 400 GB, 800 GB, 1.6 TB, 3.2 TB or High-Capacity flash (HCF) 3.84 TB and 7.68 TB
Intermix	Intermix up to 3 HPF capacities No intermix of HPF and HCF drives in the same enclosure
Capacity (min / max)	6.4TB / 368.64 TB
Keyboard and Monitor	Required to share with IBM Z ZR1 or IBM LinuxONE Rockhopper II via KVM switch
Advanced copy services functions	All supported

Uninterruptible Power Supply - UPS

Input	208V default (200/208/220/230/240V)
Power rating (VA/Watts)	3000/2700
Power phase	Single
Dimensions (W x D x H, mm)	440 x 610 x 89
Weight (kg)	28
Batteries recharge time	3 hours to 90%
Hold up time	7.5 minutes load of 75%
Fire hose dump	Supported for write in-flight on power loss (NVS)



Delta RT single phase 3kVA UPS is used to provide battery backup

For more information visit:

<http://www.deltapowersolutions.com/en/mcis/1kva-3kva-single-phase-ups-rt-series-introduction.php>

Common Keyboard / Monitor Tray (Optional)

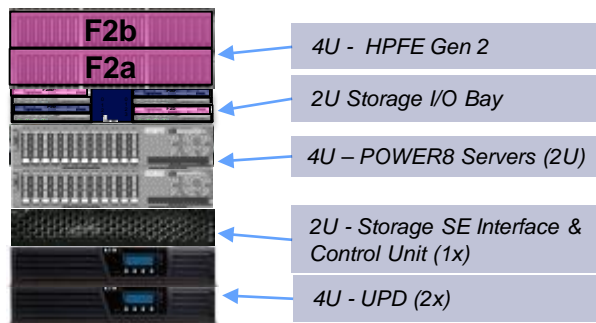
- Form Factor occupies 1U rack space
- Monitor/Keyboard will be shared between ZR1/Rockhopper II Flex Frame and DS8882F and use KVM to select target HMC
 - DS8882F ship group for this configuration includes a cable set to connect the HMCs to the shared KVM switch
- Monitor/Keyboard Tray will be optional for DS8882F
 - Swapping from one HMC to the other requires swapping the cable connection if there is no KVM switch



DS8882F Installation and Service

- Installation of the DS8882F will be performed by IBM
 - Customer installation is not supported
 - Order of component placement is fixed for core components (UPS, CEC, IO Bay, HPFE)

DS8882F Component Placement



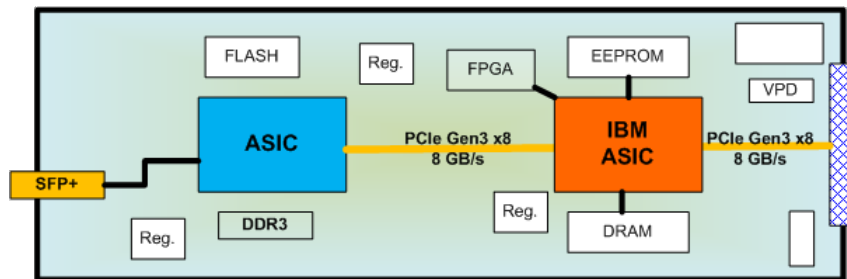
*Keyboard is optional in client provided rack
Shared keyboard in ZR1 / LinuxONE*

Open Systems Adapter (OSA) Connectivity

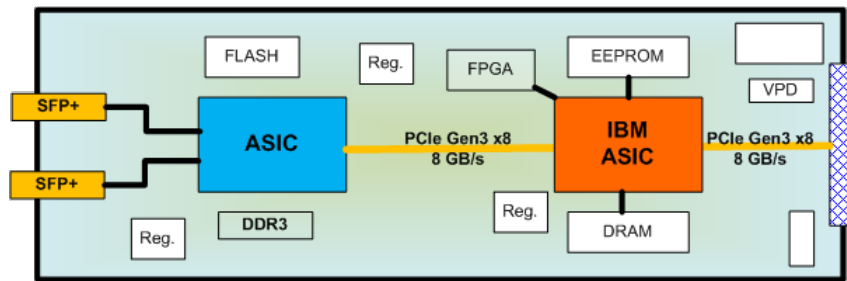
OSA-Express6S Fiber Optic Features

- 10 Gigabit Ethernet (10 GbE)
 - CHPID types: OSD, OSX
 - Single mode (LR) or multimode (SR) fiber
 - One port of LR or one port of SR
 - 1 PCHID/CHPID
 - Small form factor pluggable (SFP+) optics
 - LC duplex

- Gigabit Ethernet (1 GbE)
 - CHPID types: OSD (OSN not supported)
 - Single mode (LX) or multimode (SX) fiber
 - Two ports of LX or two ports of SX
 - 1 PCHID/CHPID
 - Small form factor pluggable (SFP+) optics
 - Concurrent repair/replace action for each SFP
 - LC Duplex



FC 0424 – 10 GbE LR, FC 0425 – 10 GbE SR

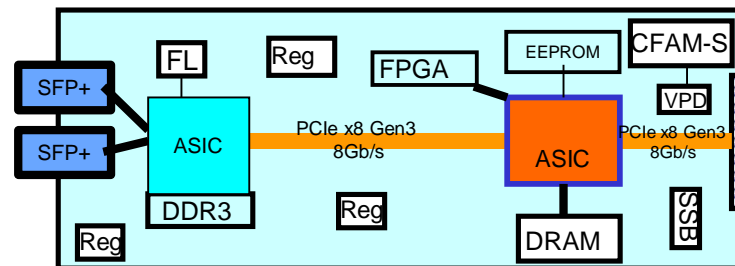


FC 0422 – GbE LX, FC 0423 – GbE SX



OSA-Express6S 1000BASE-T Ethernet feature

- PCIe form factor feature supported by PCIe I/O drawer
 - One two-port CHPID per feature
 - Half the density of the OSA-Express3 version
- Small form factor pluggable (SFP+) transceivers
 - Concurrent repair/replace action for each SFP
- Exclusively Supports: Auto-negotiation to 100* or 1000 Mbps and full duplex only on Category 5 or better copper
 - No 10Mbps
 - RJ-45 connector
 - Operates at “line speed”



FC 0426

Connector = RJ-45

CHPID TYPE Support:

Operation Mode	TYPE	Description
OSA-ICC	OSC	TN3270E, non-SNA DFT, OS system console operations
QDIO	OSD	TCP/IP traffic when Layer 3, Protocol-independent when Layer 2
Non-QDIO	OSE	TCP/IP and/or SNA/APPN/HPR traffic
Unified Resource Manager	OSM	Connectivity to intranode management network (INMN)
OSA for NCP (LP-to-LP)	OSN	NCPs running under IBM Communication Controller for Linux (CCL)

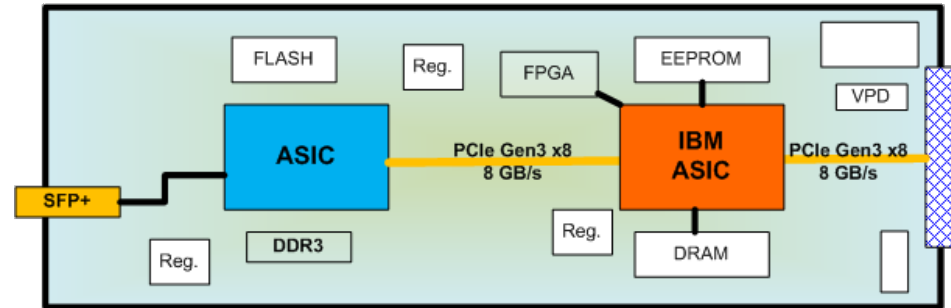
* OSA-Express6S 1000BASE-T adapters (#0426) will be the last generation of OSA 1000BASE-T adapters to support connections operating at 100 Mb/second link speed. Future OSA-Express 1000BASE-T adapter generations will support operation only at 1000 Mb/second (1Gb/s) link speed.

- CHPID TYPE OSC (1000Base-T features only)
 1. OSA-ICC Transport Layer Security (TLS) **Certificate renewal by port**, versus by System.
 - Useful for those who are hosting workloads across multiple business units and/or datacenters (where coordination across all of them would be required).
 - Also, added the ability to edit certain fields of the certificate: 1) Domain Name, 2) Server Name, 3) Location, 4) Expiration Date.
 2. The Server Configuration panel in Advanced Facilities (SE) and OSA Advanced Facilities (HMC) is updated with a new control that allows the customer to **specify the minimum supported version of the TLS protocol** which can be used by the OSA-ICC when a connection is created.
 - Supported on z14 OSA-Express6S at **TLS 1.0, 1.1 or 1.2**.
 - PCI standards now require nothing lower than TLS version 1.2.
 3. OSA-ICC – **Adding support for IPV6**

OSA-Express7S 25GbE SR

A new generation of OSA - OSA-Express7S 25GbE:

- Provides one 25GbE physical port (one 25GbE port per feature)
- CHPID Type = OSD
- Up to 48 features on z14 M0x/ZR1 or LinuxONE Emperor II / Rockhopper II
- Uses 50 micron multimode fiber
- **Auto negotiate not supported**
- Requires 25GbE optics and Ethernet switch 25GbE support
- Ethernet Switch:
 - Switch port **must** support 25GbE (negotiation down to 10GbE is not supported).
 - There are no other new or unique switch requirements (i.e. other than 25GbE)



FC 0429 – OSA-Express7S 25GbE SR

Available for ordering April 9, 2019

OSA Express7S 25 GbE SR – Software requirements

- z/OS 2.2 and 2.3 APARs
 - OA55256 (VTAM)
 - PI95703 (TCP/IP)
- z/VM V6.4 APAR
 - PI99085
- Linux on Z
 - Update required for all supported levels of Linux (Ubuntu, SLES12 & SLES15, RHEL 7... details are TBD).
- z/TPF: APAR to be created (number not available yet)
- z/VSE: N/A
 - No APAR required. z/VSE does not display OSA speed.
- There are required software updates
- There could be minor migration actions: evaluate your QDIOSTG needs and your available fixed (real) storage
- There will be minor documentation updates: updates to Netstat displays, Display OSAINFO, SMF, NMI and SNMP reflecting 25GbE

No impact to the system configuration support (CommServer, HCD, IOCDs)

Remote Direct Memory Access over Converged Ethernet (RoCE) Connectivity

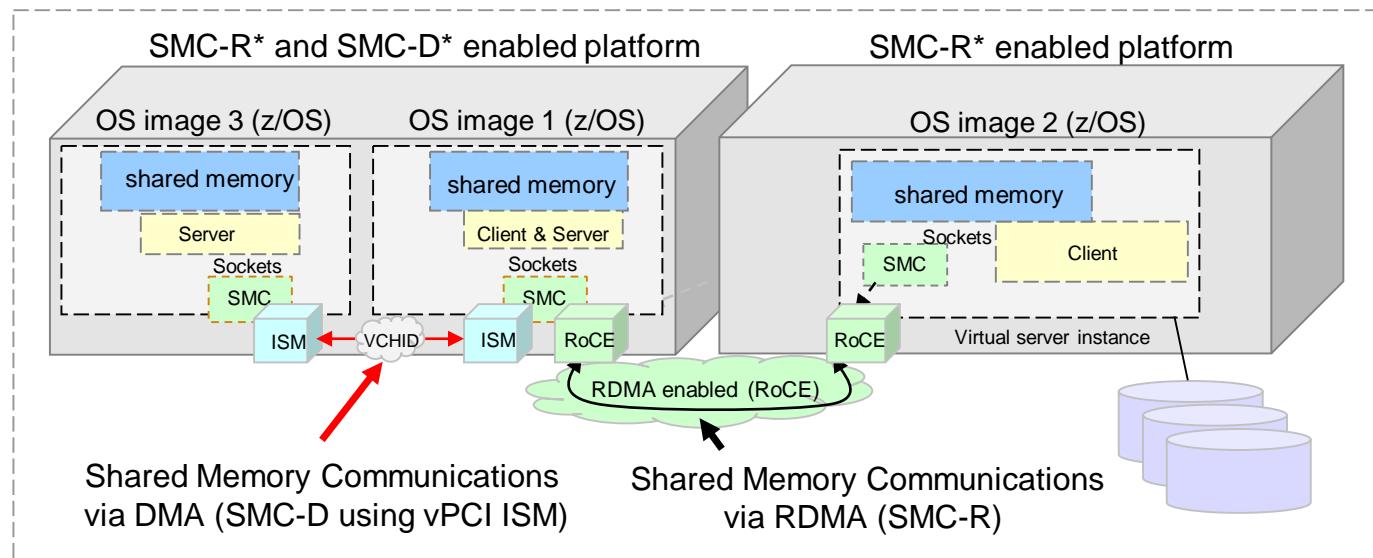
What is SMC (Shared Memory Communications)?

- **Two forms of implementations:**

- **SMC-R** – Shared Memory Communication – Remote Direct Memory Access using 10GbE RoCE Express adapter – Intra- and inter- CPC communications.
SMC-R is an *open* sockets over RDMA protocol that provides transparent exploitation of RDMA for TCP based applications, while preserving key functions and qualities of service from the TCP/IP ecosystem that enterprise level servers/network depend on
 - **SMC-D** – Shared Memory Communications – Direct Memory Access over Internal Shared Memory (ISM) – Intra CPC communications for TCP based applications.
When ISM is exploited by z/OS using SMC-D, the combined solutions provide improved transaction rates for interactive (request/response) workloads due to reduced network latency and lower CPU cost for workloads with larger payloads (i.e. analytics, streaming, big data, or web services).
- **Current implementation supports z/OS only**
 - Any z/OS TCP sockets-based workload can seamlessly use SMC-R or SMC-D without application changes

Shared Memory Communications - Remote and Direct

Clustered Systems: Multi-Tier Application Solution



- Transparent to applications
- Resiliency / High availability
- Secure internal communications
- Response time improvements

- Network latency time reduction (up to 90%)
- Potential CPU savings (up to 60%)
- High scalability & throughput (up to 20%)
- VLAN isolation

**SMC-R requires z/OS 2.1 + PTFs or higher; SMC-D requires z/OS 2.2 + PTFs or higher*

SMC-D Performance Overview

- **Shared Memory Communication – Direct Memory Access (SMC-D) Streaming workload test case results*:**
 - Up to 89% reduction in latency, 9 times the throughput, and 87% reduction in CPU cost compared to **HiperSockets***
 - Up to 95% reduction in latency, 20 times the throughput, and 83% reduction in CPU cost compared to **OSA***
 - Up to 94% reduction in latency, 16 times the throughput, and 58% reduction in CPU cost compared to **SMC-R (RoCE)***
- See the z/OS Communications Server web page for a link to information on detailed SMC-D performance test results

*This performance data was measured in a controlled environment under z/OS. The actual latency, throughput, and CPU cost that any client will experience will vary depending upon considerations such the I/O configuration, the storage configuration, and the characteristics of the communications workload.

SMC Applicability Tool – SMC-AT

- A tool called SMC Applicability Tool (SMCAT) has been created that will help customers determine the value of SMC-R and SMC-D in their environment with minimal effort and minimal impact.
- SMCAT is integrated within the TCP/IP stack:
 - Gathers new statistics that are used to project SMC-R and SMC-D applicability and benefits for the current system
 - Minimal system overhead, no changes in TCP/IP network flows
 - Produces reports on potential benefits of enabling SMC-R / SMC-D
 - Does not require RoCE or ISM hardware or SMC-R/L function. No IP configuration changes are required (measures your existing TCP/IP traffic)
- Available via the service stream on existing z/OS releases:
 - z/OS V2R1 - APAR PI39612, PTF UI28867 or higher

Exploitation Considerations

- Linux can exploit RoCE Express2 as a standard NIC (Network Interface Card) for Ethernet. A specific Linux distribution level is required (reference PSP bucket for additional details).
 - SLES 11 SP4, SLES 12 SP2
 - RHEL 6.8, RHEL 7.3
 - Ubuntu 16.04 (+ additional patches)
 - **Note. Linux does not currently support SMC-R.**
- Configuration or deployment issues that should be considered by field personnel or the customer. In addition to the existing RoCE Express hardware installation procedures, when the FID is configured in HCD the RoCE Express2 port number is also required. Port number must be specified. There is no default.
- RoCE Express2 supports a greater number of Virtual Functions per physical port (63). This aspect will benefit the Linux shared RoCE environment.
- z/VM guest support for both SMC-R and SMC-D

10 GbE RoCE Express2 and 10 GbE RoCE Express

Description	Feature Code	Ports	Available
10GbE RoCE Express2	0412	2	New

- 10GbE RoCE Express2 → FC0412 (2-Ports)
- Enhancements
 - Improved performance with 10GbE
 - Virtualization capabilities for z14 M/T 3906 – 63 Virtual Functions per port (126 VFs per feature)
 - Virtualization capabilities for z14 M/T 3907 – 31 Virtual Functions per port (62 VFs per feature)
 - Improved RAS - ECC double bit correction
- Old 10GbE RoCE Express → FC 0411 (2-Ports on z14 M0x/z14 ZR1//z13/z13s and up to 31 VFs per feature, 1-Port on zEC12)

Description	Feature Code	Ports	Available
10GbE RoCE Express	0411	2	Carry Forward

Notes:

- **z14 M0x** Max. Number of RoCE Express and RoCE Express2 features (combined) per system is 8 (16 ports).
- **z14 ZR1** Max. Number of RoCE Express and RoCE Express2 features (combined) per system is 4 (8 ports).

IBM 25GbE RoCE Express2 – FC 0430



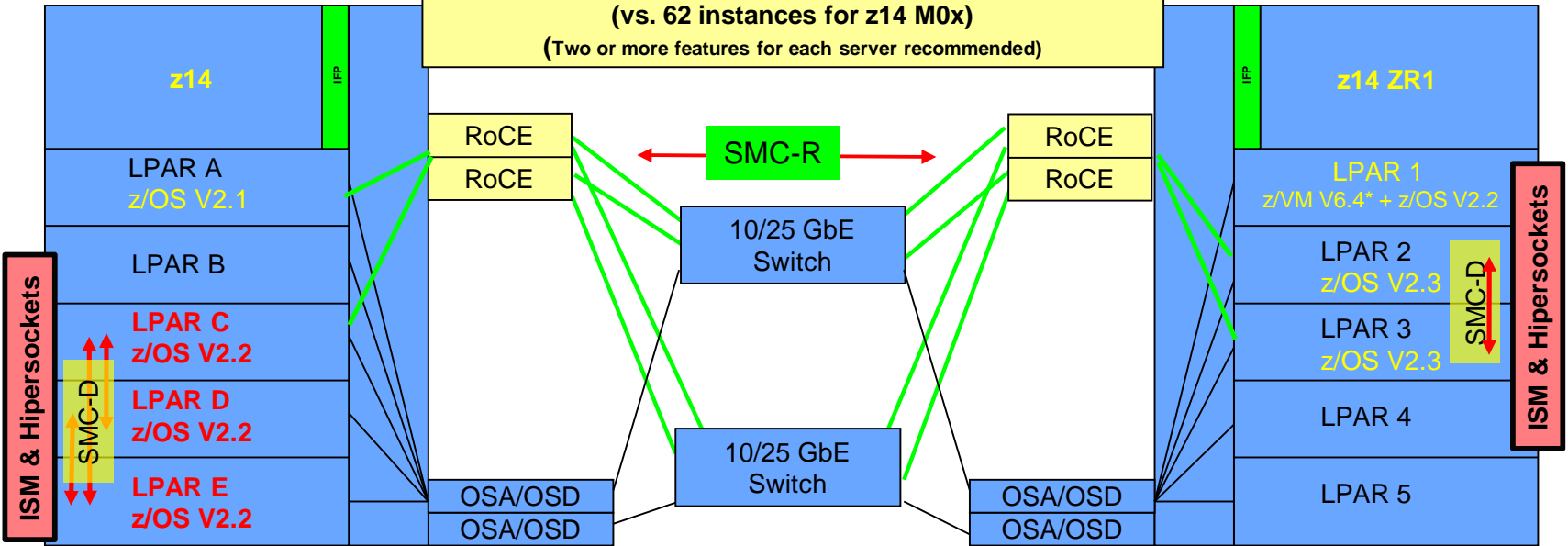
- Based on the existing RoCE Express2 generation hardware
- Requires 25GbE optics and Ethernet switch 25GbE support
- Two physical Ports (D1; D2)
- One PCHID
- There are **no z/OS Communication Server software changes** required for 25GbE RoCE Express2
- There is **no impact to the system configuration support** for RoCE 25GbE (this includes Communication Server, HCD, IOCDs etc.)
- **Note 1:** Communicate with another 25GbE RoCE Express2 through 25Gb switch ports.
Although, this is possible.....
25GbE RoCE <=> 25gb-switch-port/10gb-switch-port <=> 10GbE RoCE
- **Note 2:** The same SMC-R Link Group can have both 10GbE RoCE and 25GbE RoCE assigned, but not recommended due to load balancing conflicts



IBM 25GbE RoCE Express2

z14 Shared Memory Communications (SMC-R and SMC-D)

On z14 ZR1, each 25 GbE FC 0430 and 10 GbE RoCE FC 0412 supports up to 31 instances per port (vs. 62 instances for z14 M0x)
(Two or more features for each server recommended)



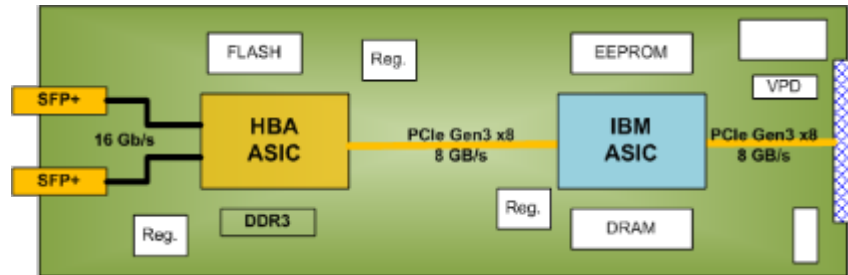
- This configuration allows SMC-D connectivity among LPAR C, LPAR D, and LPAR E.
- SMC-D within one machine is better than using HyperSockets alone.
- For LPAR to LPAR, HyperSockets or OSD connections are required to establish the SMC-D communication.
- ISM = Internal Shared Memory
- No additional hardware purchase required.
- z/VM Guest support

25GbE RoCE Express2 should not be mixed with 10GbE RoCE Express2 or RoCE Express in the same SMC-R Link Group

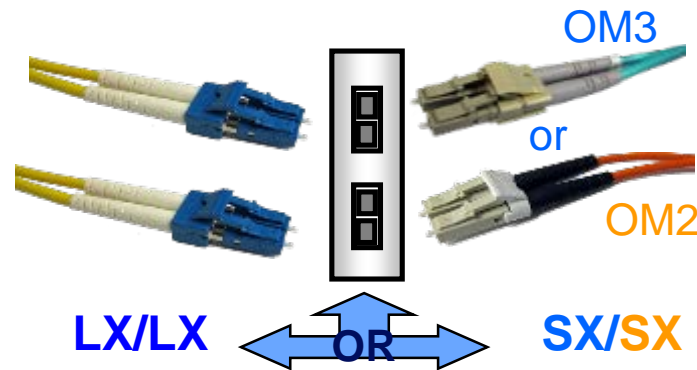
FICON

FICON Express16S+

- For FICON, zHPF, and FCP
 - CHPID types: FC and FCP
 - *Both ports must be same CHPID type*
 - 2 PCHIDs / CHPIDs
- Auto-negotiates to 4, 8, or 16 Gbps
 - 2 Gbps connectivity not supported
 - **FICON Express8S will be available for 2Gbps (carry forward only)**
- Increased performance compared to FICON Express16S
- Small form factor pluggable (SFP) optics
 - Concurrent repair/replace action for each SFP
 - 10KM LX - 9 micron single mode fiber
 - Unrepeated distance - 10 kilometers (6.2 miles)
 - SX - 50 or 62.5 micron multimode fiber
 - Distance variable with link data rate and fiber type
- Two channels of LX or SX (no mix)



FC 0427 – 10KM LX, FC 0428 – SX



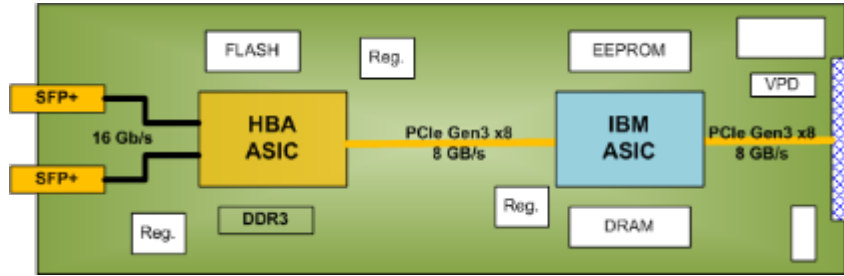
FICON Express16S+ and IOCP Rules

- Both ports must be same CHPID type (either FICON or FCP).
 - z14 (M/T 3906 and 3907) only.
 - z14 M0x and ZR1 can also have older FICON adapters (FICON Express16S/FICON Express8S) from an upgrade which do not have this restriction.
- To permit the mix of CHPID types for an older card, the user will need to specify a new keyword on at least one channel for an adapter where a mix is desired.
 - The new keyword is: **MIXTYPE**.
The keyword only needs to be on one of the PCHIDs for the card.
- The IOCP PCHID Summary and Channel Path Reports will be updated to indicate this. The current design is to add a suffix of '(M)' after the PCHID number.
- IOCP will ignore MIXTYPE keyword for processors prior to z14 – warning message

FICON Express16S+ and IOCP Rules (cont.)

Both: FC or FCP

Valid syntax examples
Invalid syntax examples

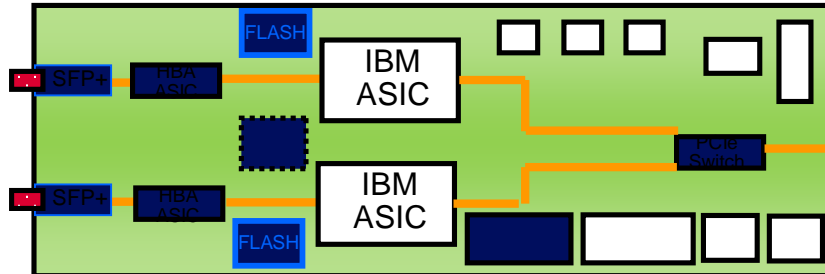


FICON Express16S+

FICON EXPRESS16S	CHPID PCHID=100,PATH=(CSS(0),80),SHARED,TYPE=FC,MIXTYPE	NOT VALID
FICON EXPRESS16S+	CHPID PCHID=104,PATH=(CSS(0),84),SHARED,TYPE=FCP	
FICON EXPRESS16S+	CHPID PCHID=105,PATH=(CSS(0),82),SHARED,TYPE=FC	
FICON EXPRESS8S	CHPID PCHID=108,PATH=(CSS(0),83),SHARED,TYPE=FCP,MIXTYPE	
FICON EXPRESS8S	CHPID PCHID=109,PATH=(CSS(0),84),SHARED,TYPE=FCP	
FICON EXPRESS8S	CHPID PCHID=10C,PATH=(CSS(0),85),SHARED,TYPE=FC,MIXTYPE	
FICON EXPRESS8S	CHPID PCHID=10D,PATH=(CSS(0),86),SHARED,TYPE=FCP,MIXTYPE	
FICON EXPRESS16S	CHPID PCHID=110,PATH=(CSS(0),87),SHARED,TYPE=FC,MIXTYPE	
FICON EXPRESS16S+	CHPID PCHID=114,PATH=(CSS(0),88),SHARED,TYPE=FC	
FICON EXPRESS16S+	CHPID PCHID=115,PATH=(CSS(0),89),SHARED,TYPE=FC	

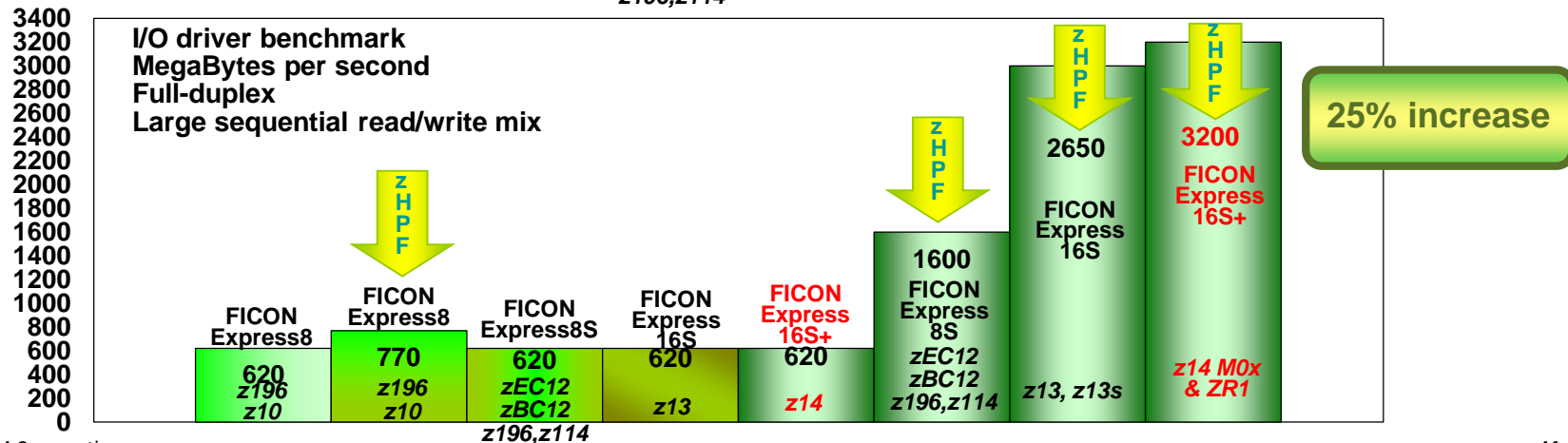
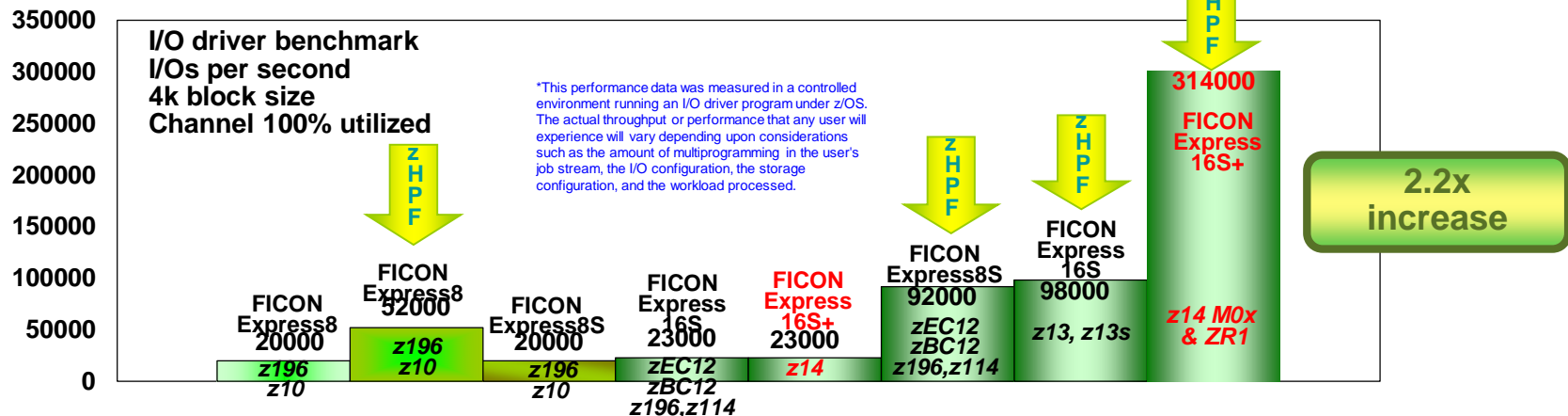
10C -- FC

10D -- FCP

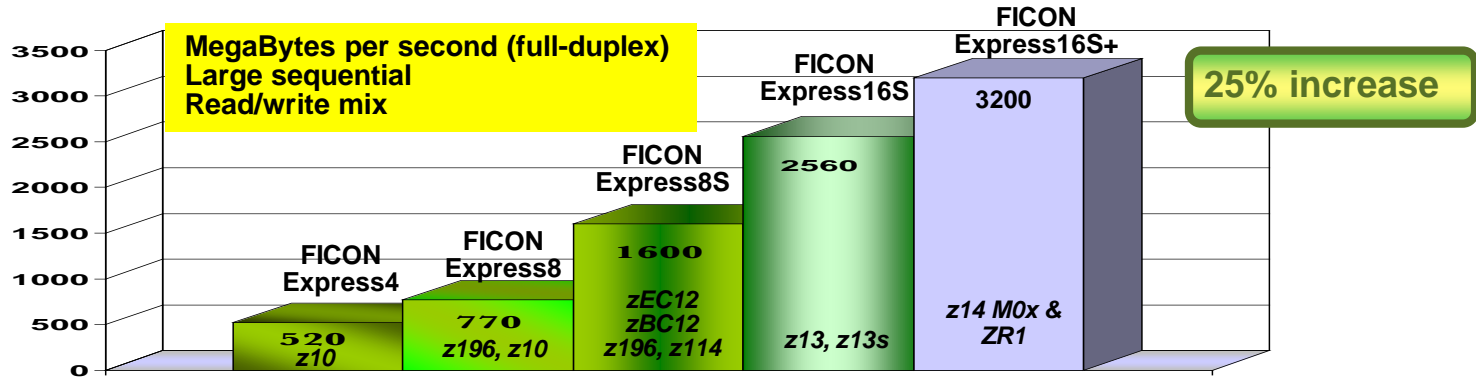
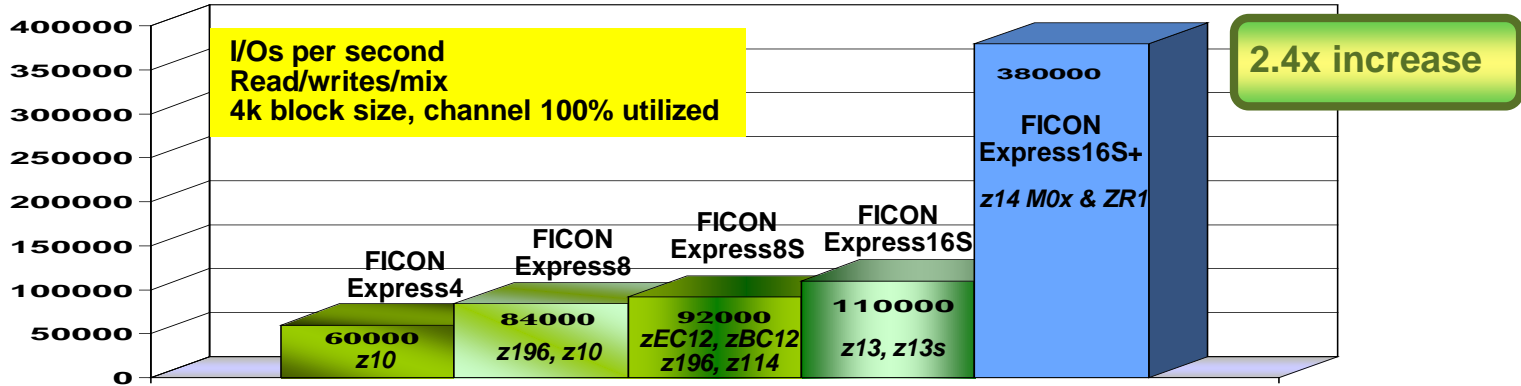


FICON Express16S or FICON Express8S

zHPF and FICON Performance z14



FCP Performance for z14



*This performance data was measured in a controlled environment running an I/O driver program under z/OS. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed.

zHyperlink

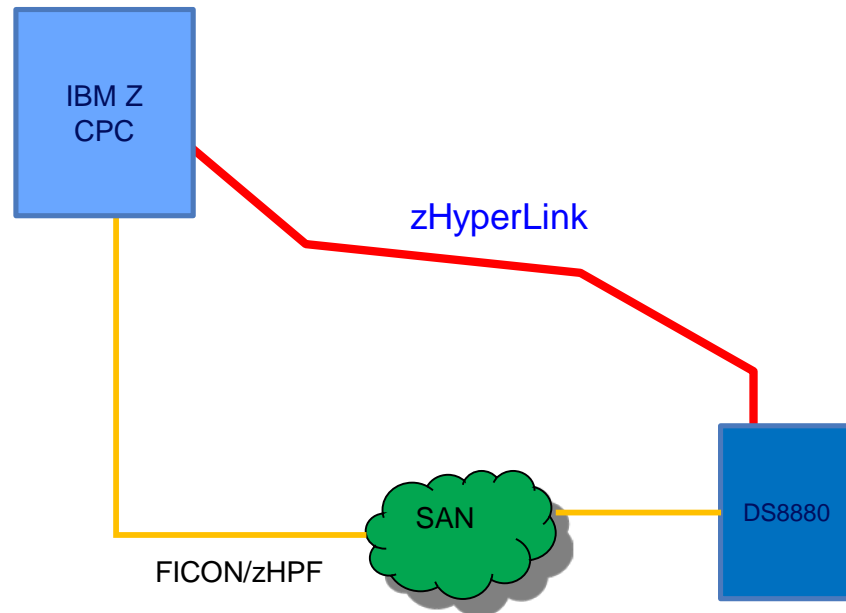
What is IBM zHyperLink™?

- zHyperLink Express is a direct connect short distance IBM Z I/O feature designed to work in conjunction with a FICON or High Performance FICON SAN infrastructure
- IBM zHyperLink™ dramatically reduces latency by interconnecting the z14 CPC directly to the I/O Bay of the DS8880
- zHyperLink improves application response time, cutting I/O sensitive workload response time in half without significant application changes.



How does IBM zHyperLink™ change the game?

- zHyperLink™ is FAST enough the CPU can just wait for the data
 - No Un-dispatch of the running task
 - No CPU Queueing Delays to resume it
 - No host CPU cache disruption
 - Very small I/O service time
- Operating System and Middleware (e.g. Db2) are changed to keep running over an I/O
- Transparently gives Db2 apps fundamentally better latency than applications on platforms without zHyperLink
 - Excluding 100% in memory databases



New with GA2

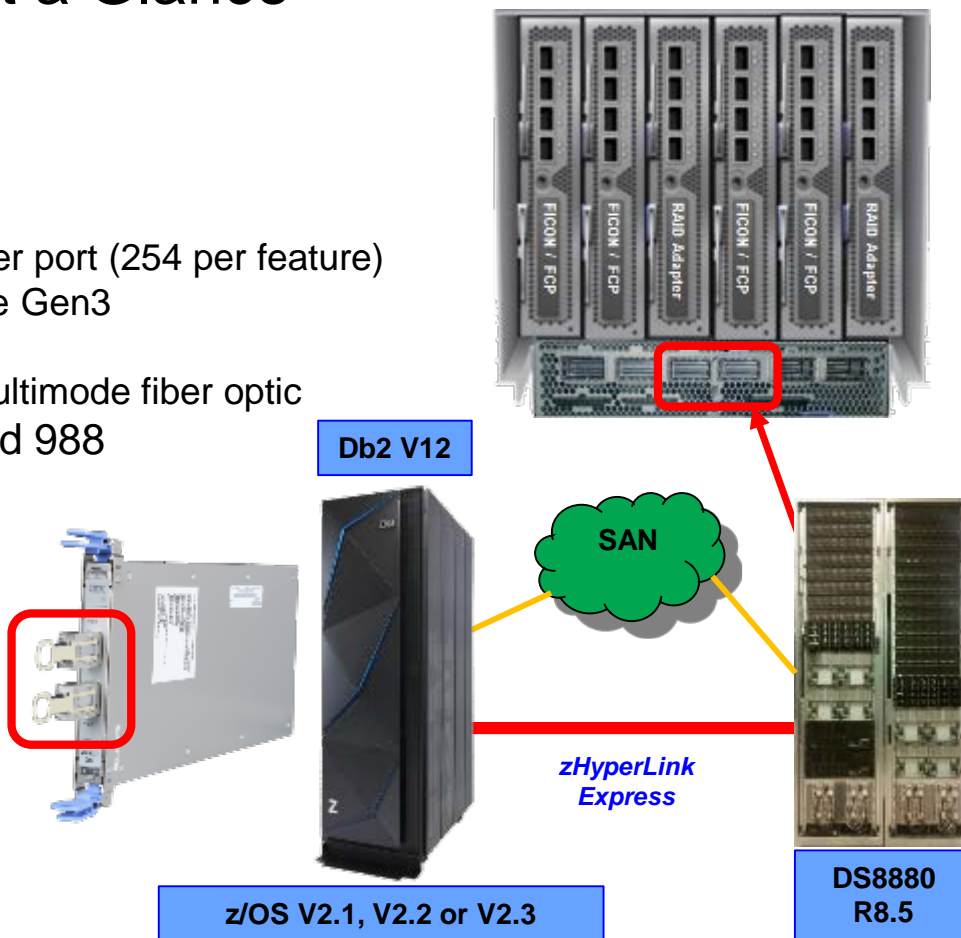
Support for zHyperLink Writes can accelerate Db2 log writes to help meet clients' most stringent requirements and deliver superior service levels by processing high-volume Db2 transactions at speed. IBM zHyperLink Express (FC 0431) requires compatible levels of DS8K hardware and firmware R8.5.1, and Db2 V12 with PTFs



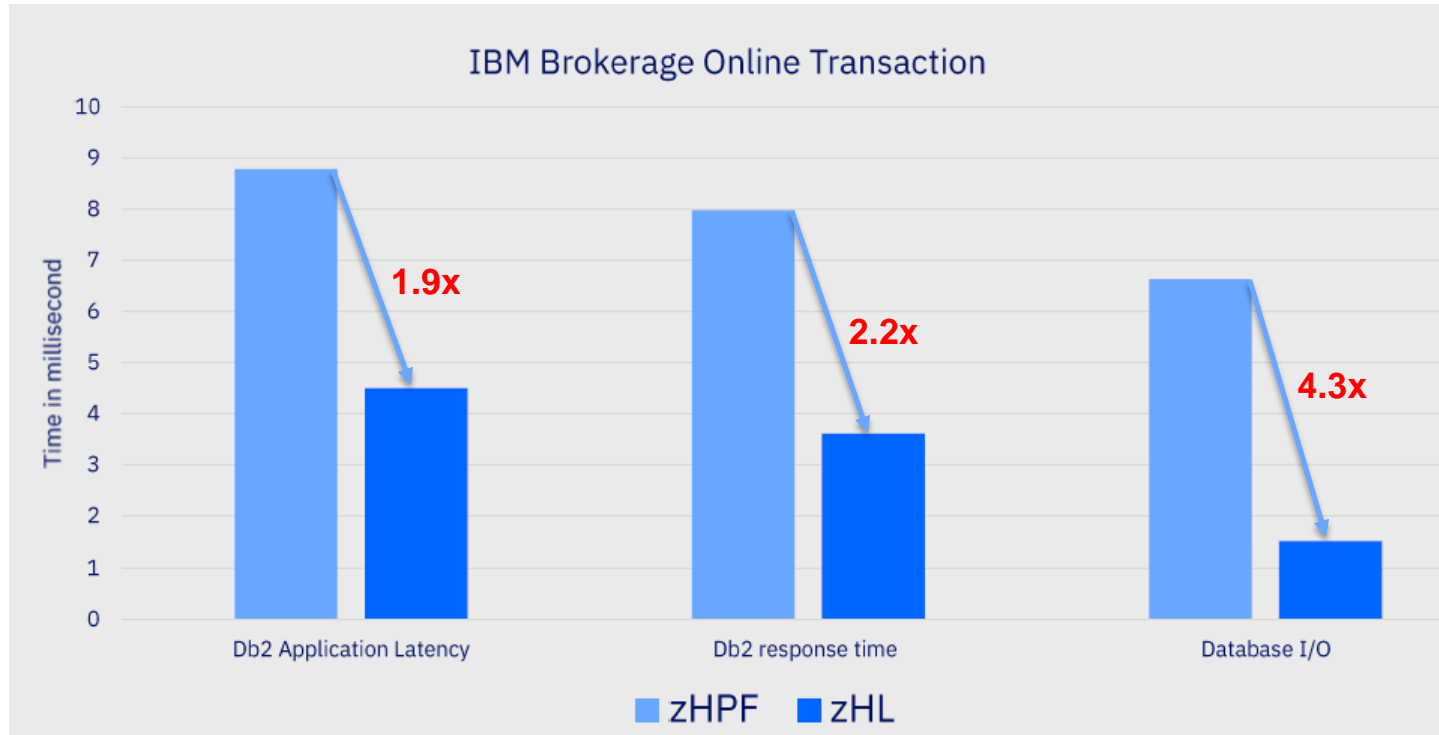
zHyperLink Express® at a Glance

- Feature Code #0431
 - Two ports per feature
 - Maximum of 16 features (32 ports)
 - Function ID Type = HYL
 - Up to 127 Virtual Functions (VFs) per port (254 per feature)
 - Point to point connection using PCIe Gen3
 - Maximum distance: 150 meters
 - Cables: OM4 or OM3 50/125 μm multimode fiber optic
- DS8880 models 984, 985, 986 and 988

Note: A standard FICON channel (CHPID type FC) is required for exploiting the zHyperLink Express feature



Db2 Online Transaction with zHyperlinks (10GB Buffer Pool)



IBM FCP Express32S for LinuxOne

What is FCP (Fibre Channel Protocol)?



What is FCP (Fibre Channel Protocol)

- High-speed network technology used to connect servers to data storage area networks
- Handles high-performance disk storage for applications on many corporate networks, supporting data backups, clustering, and replication

Why is Fibre Channel important?

- Fibre Channel is the Industry Standard for I/O
- >80-90% of all Flash Arrays were connected with Fibre Channel in 2017
- >90% of those Flash Arrays were connected with 16 or 32G Fibre Channel

Fibre Channel is data center storage protocol of choice for the next decade

- Orders of magnitude performance improvement low latency requires higher-throughput protocols
- Bottlenecks exist: 10GbE, 8 Gbps Fibre Channel
- 16 Gbps Fibre Channel will be too slow for the next generation of storage arrays
- Plan for higher throughput, e.g. 32 Gbps Fibre Channel



Gartner Research: The Future of Storage Protocols, G00307902 June 2016



IBM FCP Express32S – FC 0438 & FC 0439



- Initially, these features will only be available on the **LinuxONE** machines. Emperor II (3906-LM1-LM5) Rockhopper II (3907-LR1)
- FC 0438 LX, FC 0439 SX
- Two Ports (D1; D2)
- Maximum of 32 features on LinuxONE
- LC Duplex Connector
- Auto-negotiate to 8/16/32 Gbps



IBM FCP Express32S

IBM FCP Express32S – Software pre-requisites

z/VM support:

- z/VM V7.1
- z/VM V6.4 with PTFs

Linux on Z support:

- SuSE SLES 12 SP2 with service
- SuSE 11 SP4 with service
- Redhat RHEL 7.3 with service
- RedHat RHEL 6.9 with service
- Ubuntu 18.04 LTS with service
- Ubuntu 16.04 LTS with service



IBM FCP Express32S



IBM LinuxONE™

IBM Adapter for NVMe for LinuxOne

NVMe is an open source protocol for accessing storage

Wikipedia

NVM Express (NVMe) or Non-Volatile Memory Host Controller Interface Specification (NVMHCIS) is an open logical device interface specification for accessing non-volatile storage media attached via a PCI Express (PCIe) bus.

By its design, NVM Express allows host hardware and software to fully exploit the levels of parallelism possible in modern SSDs. As a result, NVM Express reduces I/O overhead and brings various performance improvements relative to previous logical-device interfaces, including multiple, long command queues, and reduced latency.

www.nvmexpress.org

NVM Express® is an open collection of standards and information to fully expose the benefits of non-volatile memory in all types of computing environments from mobile to data center. NVMe™ is designed from the ground up to deliver high bandwidth and low latency storage access for current and future NVMe technologies.

Why is NVMe important?

- Why is NVMe important?
- Serial Advanced Technology Attachment (SATA) & Serial Attached SCSI (SAS) protocols were designed for standard hard drives (1 queue → 32K commands)
- NVMe protocol designed for today's Solid-State Drive (SSD) (64K queues → each capable of 64K commands)
- Technology Benefits over SATA and SAS:
 - Reduced Latency
 - Higher IOPS
 - Lower PWR Consumption
 - End-to-End Protection with rated life expectancy (Drive Writes Per Day or Terabytes Written)
 - prevents errors from being written when device reaches maximum endurance
 - will warn when it is reaching end of life
 - can be queried to determine how much life each device has left (planned vs. unplanned)



Market Trends

▪ **Disk Density:**

- Largest HDDs today are in the **12TB range** whereas SSDs are currently available **up to 30TB**. Vendors already talking about 32TB and 64TB single flash devices. March 2018, Nimbus Data announced a **100TB SSD**.

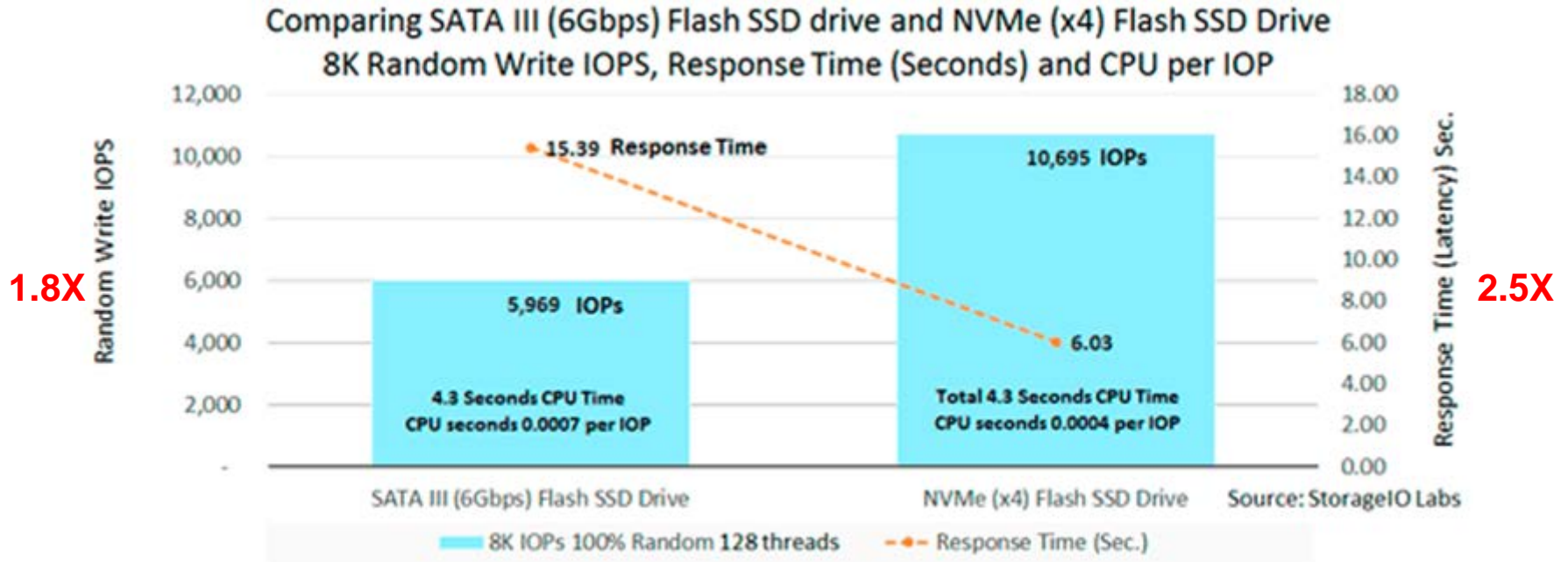
▪ **Price:**

- Past: NVMe flash drives and storage arrays were always 30–50% more expensive than equivalent all-flash arrays
- Today: NVMe volumes are higher, costs have come down, and multiple NVMe SSD suppliers are mainstream

▪ **Cognitive/AI Workloads:**

- Will be valuable for the cognitive/AI workloads that operate in 10–20 microsecond realm, instead of the 100–200 microsecond range for flash. This 10x performance improvement will manifest as both storage cache and tier to deliver better, faster storage.

NVMe I/O Throughput and Latency Study

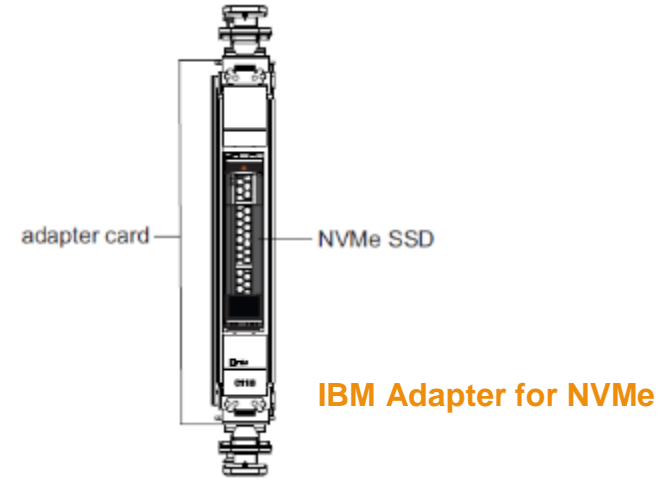


<http://www.datacenterjournal.com/answer-nvme-questions>

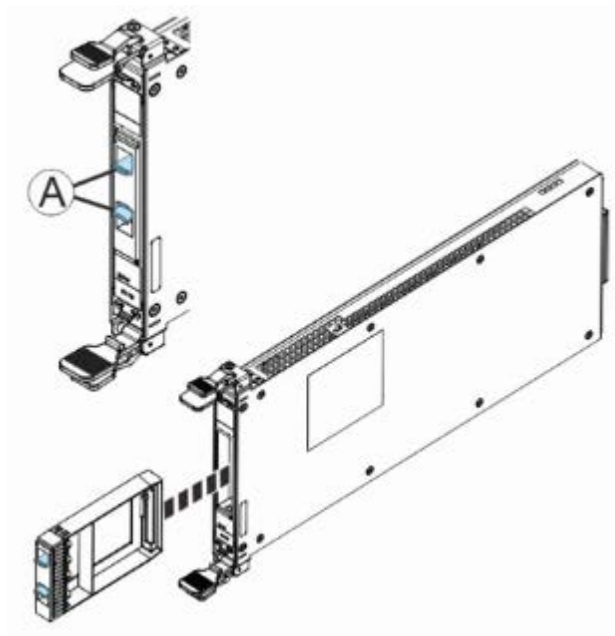
IBM Adapter for NVMe – FC 0435



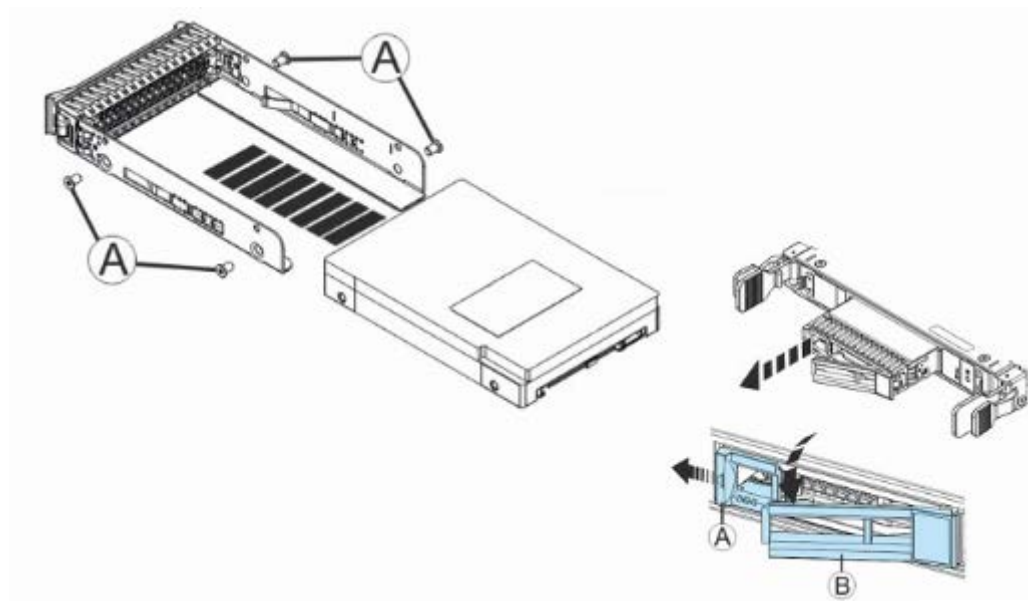
- LinuxONE Emperor II or Rockhopper II only
- “Built in” storage. No boot support initially.
- Uses the normal z14 PCIe EC Stream.
- Carrier Card
 - Zero ports
 - IBM provides a carrier card into which NVMe SSDs can be plugged.
 - IBM service will install the vendor SSD concurrently into the carrier card on-site. Hot/cold plug.
- Up to 16 features in increments of one.



IBM Adapters for NVMe Components



Based card with Filler (A)

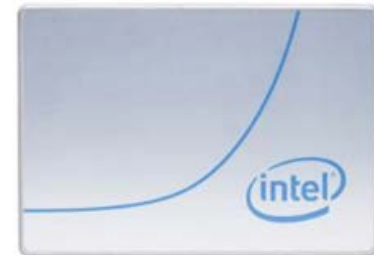


Sled/Carrier with SSD

IBM Adapter for NVMe – FC 0435



- The vendor SSD card will be purchased by the client from a reseller or directly from the vendor.
- Tested in IBM Z
 - Intel PN SSDPE2KX010T701 (1TB) – Up to 16 TB
 - Intel PN SSDPE2KX040T701 (4TB) - Up to 64 TB
 - Performance testing (ongoing).
 - Both can coexist on the same system and same I/O Drawer.
- SAMSUNG NVMe SSDs are also being tested
- IBM will likely maintain a list of tested vendors – possibly in ResourceLink and/or the z14 IMPP.
 - However, it is up to the client to own the SSD end of life and time to market vendor announcements.



NVMe Definitions

- No virtualization initially a single Function Identifier (FID)
 - VF keyword not allowed
 - Single Root I/O Virtualization and Sharing (SR-IOV) in the future
 - Multiple LPARs can be specified as reconfigurable between LPARs
- FID Name = NVME
- Each feature uses a single PCHID
- z/VM Guest Support for Linux
 - z/VM will not support guest IPL from NVMe initially.
 - z/VM will not provide virtualization of an NVMe device.
 - z/VM must give the entire device to one Linux guest.
 - KVM can give smaller chunks of the device to multiple Linux guests (virtio-block).



NVMe Use Cases and Benefits

- Exploitation of built-in storage as another storage option in addition to existing Fibre Channel based options. Time to value.
- Memory-intensive workloads, real-time analytics, fast storage workloads (ex. Kafka streaming, time-sensitive DBs, Apache Spark, and NoSQL databases all share a common architectural feature – the use of in-memory data stores)
- High-frequency financial trading, high-performance analytics, paging/sorting, and all latency-sensitive applications
- Traditional applications like relational databases where individual I/O response time is critical (ex. Banking)
- Reduced Latency and higher I/O's per second
- Lower power consumption
- Fast Sort



IBM Adapter for NVMe – Software pre-requisites

- z/VM Support:
 - z/VM V7.1 with PTFs for guest exploitation.
 - z/VM V6.4 with PTFs for guest exploitation.
- Linux on IBM Z Support:
 - Ubuntu 18.04 LTS with service and Ubuntu 18.04 LTS with service.
 - IBM is working with its Linux distribution partners to include support in future distribution releases.



Sysplex Connectivity

Integrated Coupling Adapter (ICA SR) – FC #0172

- IBM Integrated Coupling Adapter (ICA SR) – FC 0172
 - Coupling Connectivity (Short Distance)
 - ICA SR is Recommended for Short Distance Coupling z13/z13s to z13/z13s and beyond
 - Coupling channel type: CS5
 - Performance similar to Coupling over Infiniband 12X IFB3 protocol
 - PCIe Gen3, Fanout in the CPC drawer, 2-ports per fanout, up to 150m;
 - 8 GigaBytes per second (GBps)**.
 - z13/z13s/z14(M0x and ZR1) to z13/z13s/z14(M0x and ZR1) and up connectivity
 - ICA requires new cabling for single MTP connector; cables - 150m: OM4; 100m OM3
 - Differs from 12X Infiniband split Transmit/Receive connector;
 - Available as of z13 GA1

****Note: The link data rates do not represent the performance of the links. The actual performance is dependent upon many factors including latency through the adapters, cable lengths, and the type of workload.**

z14/z14 ZR1/z13/z13s Coupling Express LR – FC #0433

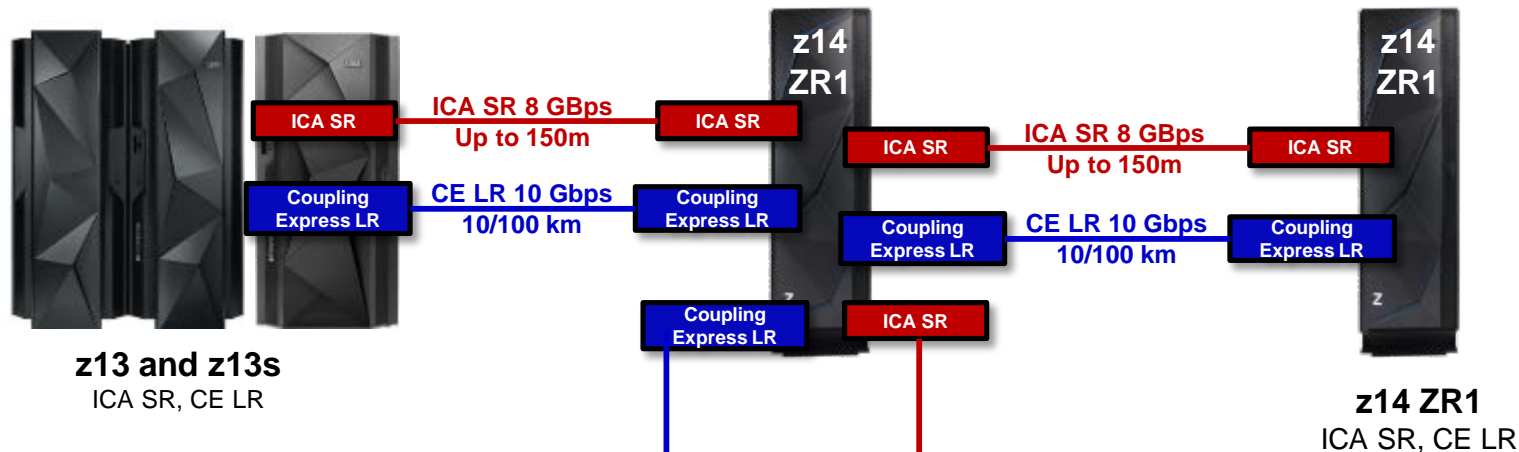
- **Coupling Connectivity:** ONLY from z14/z14 ZR1/z13/z13s to z14/z14 ZR1/z13/z13s
- Up to 16 features on the z14 ZR1 – CHPID type = **CL5**
- PCIe+ I/O drawer required for Stand Alone Coupling
- CFCC Level 22
- Adapter (2-port card): same adapter as 10GbE RoCE Express but with Coupling Optics and FW 10 Gbps**,
- **Distance:** 10 km unrepeated; up to 100 km with a qualified DWDM. **More than 100 km requires RPQ 8P2781**
- Point-to-Point just like 1X and ISC-3; **CAN NOT** be utilized in a switched environment
- **Cabling:** Utilizes same 9 μ , Single Mode fiber type as 1X IFB and ISC-3 (9/125 micrometer SM optic)

Coupling Express LR I/O cards reside in the PCIe+ I/O Drawer



***The link data rates, do not represent the performance of the links. The actual performance is dependent upon many factors including latency through the adapters, cable lengths, and the type of workload*

z14 ZR1 Coupling Connectivity



z13 and z13s
ICA SR, CE LR

z14 ZR1
ICA SR, CE LR

Coupling Express LR (CE LR)
10 GBps, 10/100 km
z13, z13s, z14 to z13/z13s/z14 Connectivity ONLY

Integrated Coupling Adapter (ICA SR)
8 GBps, up to 150 m
z13, z13s, z14 to z13/z13s/z14 Connectivity ONLY



z196, z114, and older CPCs
CANNOT coexist in the same Parallel Sysplex or STP CTN with z14 ZR1 (no coupling connectivity)

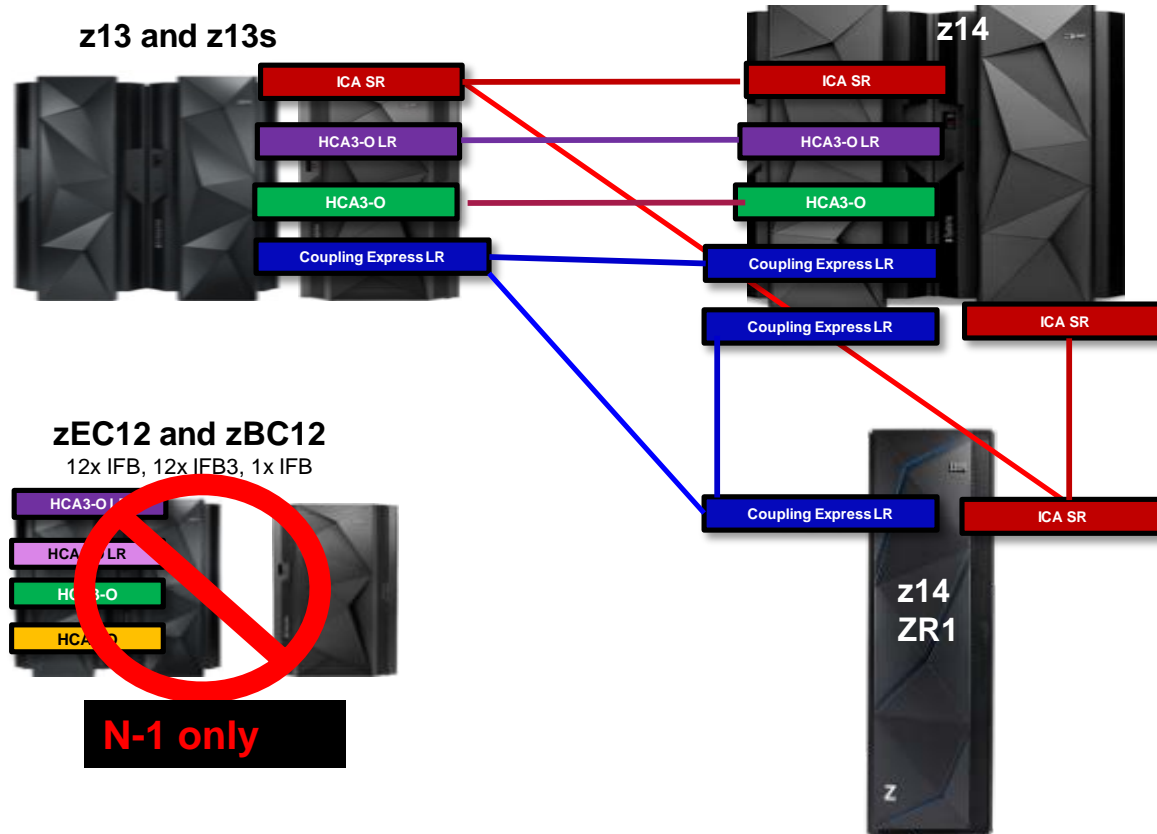
zEC12 or zBC12 can coexist in the same Parallel Sysplex with z14 ZR1 **only** if the CPC hosting the CFs has coupling connectivity to both the zEC12/zBC12 and z14 ZR1 CPCs

IC (Internal Coupling Link):
Only supports IC-to-IC connectivity

HCA2-O and HCA2-O LR and ISC-3 are NOT supported on z13, z13s and z14 M/T 3906
HCA3-O and HCA3-O LR are NOT supported on z14 ZR1 (M/T 3907)

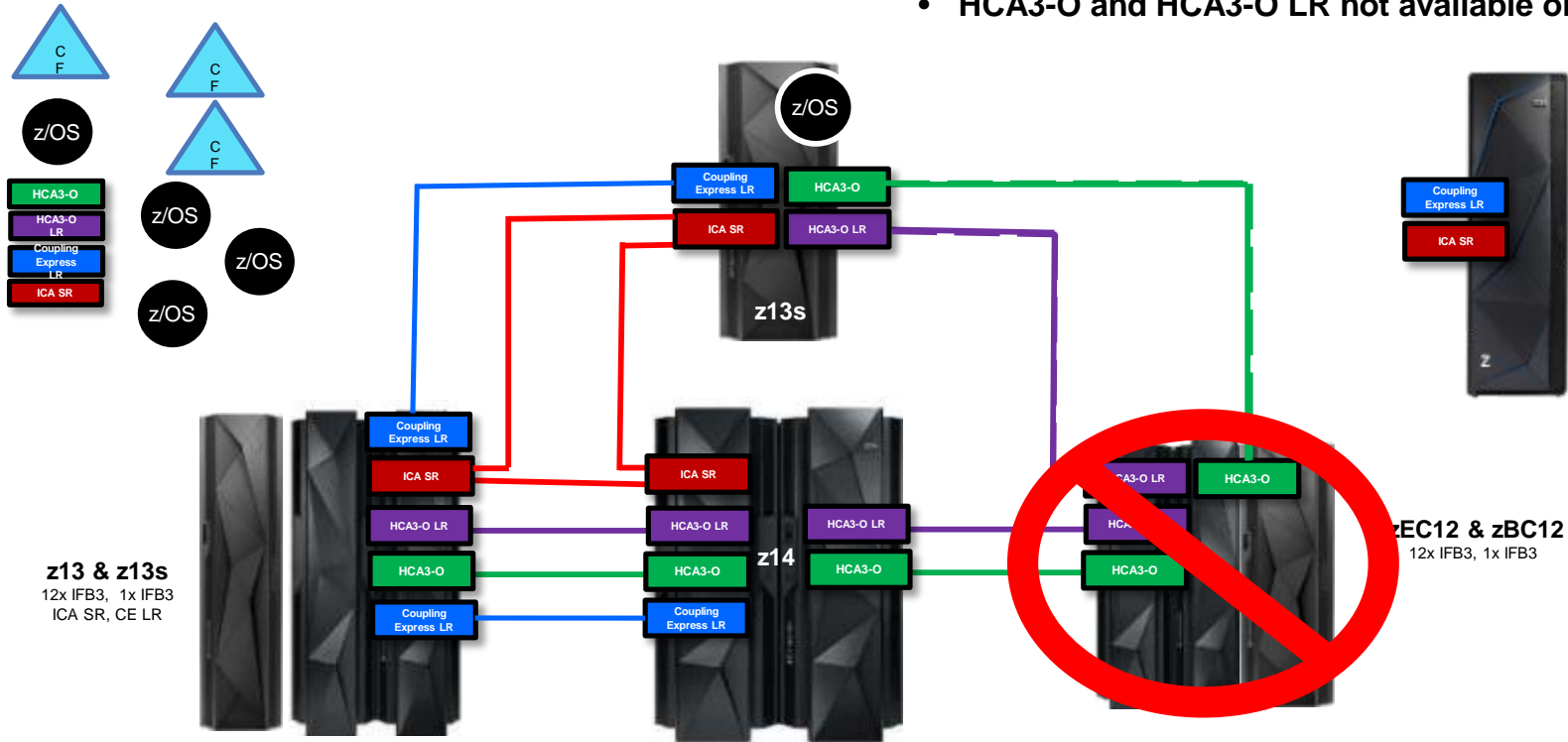
Note: The link data rates do not represent the performance of the links. The actual performance is dependent upon many factors including latency through the adapters, cable lengths, and the type of workload.

z14 ZR1 Coupling Considerations



z14 ZR1 Coupling Considerations

- z14 ZR1 cannot connect to zEC 12 or zBC12.
- HCA3-O and HCA3-O LR not available on ZR1



Coupling Link Considerations for z14 ZR1

- z14 ZR1 supports only PCIe-based coupling (ICA SR and CE LR)
 - Maximum number of physical ICA SR coupling links (ports) is 16 per CPC
 - However, long distance coupling requires PCIe+ I/O Drawer for hosting CE LR features, thus reducing the max. number of ICA SR features.
- While physical coupling links can be (and most often are) shared across images and sysplexes, there are some customers who configure dedicated physical CF links
 - “Stacked” consolidated sysplexes, on a set of physical CPCs, with dedicated connectivity provided for each sysplex
 - These kinds of dedicated/isolated configurations drive requirements for higher maximum limits on both physical links and logical CHPIDs
- CPCs that host standalone CFs tend to have the highest per-CPC consumption of coupling connectivity resources
 - CFs are “focal points” for both z/OS-to-CF and CF-to-CF sysplex connectivity, as well as STP timing roles
- z14 M/T 3906 is the last machine supporting the Infiniband coupling. Additional coupling link configuration complexity is expected for transitioning to ICA SR and CE LR coupling.

Migration Considerations - ICA SR

- **Planning required when ordering Integrated Coupling Adapter (ICA) SR**
 - The number of ICA SR features will impact the maximum number of I/O slots available.

ICA-SR	z14 ZR1	z13s	
	Max I/O slots	HCA	Max I/O slots
0	64	8	64
1	48		
2			
3			
4	32		
5	16		
6			
7			
8	0		

HCA	zBC12
	Max I/O slots
0	64
1	
2	
3	
4	
5	32
6	
7	0
8	

Sysplex Update

Asynchronous Cross-Invalidate 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/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 manager exploiters will be able to “opt in” for the CF to perform these XIs asynchronously (and then sync them up with the CF at or before transaction completion)
 - As long as all asynchronous XI signals complete by the time transaction locks are released, data integrity is preserved
 - Provides faster completion of cache update CF requests that generate XIs, *especially with cross-site distance involved*
 - Provides improved cache structure service times and coupling efficiency, as well as transaction elapsed time improvement
- Requires **z14 GA2 CFCC (coupling facility) CFLEVEL 23 support**, plus **z/OS PTFs on every exploiting system in the sysplex** (exploitation support rollback)
- Requires explicit **data manager exploitation/participation** – not transparent to the data manager
 - Initially, DB2 V12-based exploitation ... exploitation by other data managers is possible
- No SMF data changes for CF monitoring/reporting

Cross-Invalidation



- Many CF operations render exploiter's local copy of data invalid
 - Reading and/or writing entries and data
 - Reclaim (when CF must make room for new incoming entries)
 - Explicit invalidation requests, or deletion of data from the cache
- CF sends cross-invalidate (XI) signal to all exploiters with registered interest in invalidated data / directory entry
 - Signal delivered to exploiter's CPC
 - Firmware on target CPC processes signal and resets appropriate bit in exploiter's vector (0 = invalid)
 - XI considered complete when CF receives a signal response
- Multiple XI signals may result from a single cache request
 - A single named item may have multiple registrations of interest from different exploiters
 - Some IXLCACHE requests operate on multiple named items
 - Input list of names
 - Single input name with mask identifying names matching a pattern
 - One XI signal generated to send to each interested exploiter for each invalidated item
- Requestor recognizes invalidation of local copy by inspecting vector
 - IXLVECTR API used to test vector bit assigned to named item
 - Vector test performed under suitable serialization

Synchronous vs Asynchronous XI Processing



- Today, cache commands generating XIs cannot complete at the CF and return to z/OS until all XIs have been delivered (synchronous XI)
- If CPCs housing peer exploiters (XI recipients) are distant from the CF, this can significantly delay command completion for the “causing” command
 - Which directly influences the CF service times associated with those CF cache commands
- When multiple cache commands are issued within a single transaction, the only real requirement is that all XIs associated with all commands be delivered *by the time the transaction has to be committed*.
 - Transaction initiated under serialization
 - No other exploiter can be operating on the serialized data items, nor can they be inspecting their vector bits for those items
- So, we can save time by delivering XIs asynchronously, after command processing completes in CF and results returned to z/OS
 - Allows XIs for earlier commands to be delivered in parallel
 - Also benefit by reducing service time and thus CPU time / cost for individual commands, since XI delivery time no longer included in the “causing” command’s service time

Coupling Facility Hang Detect Enhancements



- CFCC dispatcher currently monitors its dispatched tasks for “hangs” where a dispatched CF task does not return to the dispatcher for 60 sec or more
 - When a hang is detected, the CF currently aborts, dumps, and reboots the CF image
- With this support, the CFCC dispatcher will:
 - Significantly reduce the CF hang detection interval to only 2 sec, allowing more timely detection and recovery from such hang problems
 - When a hang is detected, in most cases the CF will confine the scope of the failure to “structure damage” for the single CF structure the hung command was processing against, capture diagnostics with a non-disruptive CF dump, and continue operating without aborting or rebooting the CF image
 - Provides a significant reduction in failure scope and client disruption (CF-level to structure-level), with no loss of FFDC collection capability!
- Requires **z14 GA2 CFCC (coupling facility) CFLEVEL 23 support**; no z/OS SW support required

Why STP Split/Merge?



- Split
 - Prompted by an actual customer situation
 - One customer was selling a portion of their business to another customer
 - They were able to isolate the Sysplexes onto separate systems
 - One set of machines had to be shut down to remove them from the CTN and assign them to a different CTN.
 - This solution defines a system-assisted method for dynamic splitting of a CTN
- Merge
 - Prompted by an actual customer situation
 - Desire to merge two CTNs into a single CTN
 - Requires the times between the two CTNs to be closely synchronized prior to merge
 - This solution automates the time synchronization process
 - The solution manages the assignment of the roles within the new CTN
- These split/merge cases are not frequent, but they are very troublesome and error-prone when they do occur – simplify these situations.
- Requires **z14 GA2 STP support**; no new z/OS SW support required
 - **All affected CPCs in the STP network(s) must be at the z14 GA2 level to use the new support**
 - Makes use of existing z/OS detection and toleration support for CTN-related changes

Manage System Time - CTN Split and Merge Overview



- Added as new, advanced actions within HMC's Manage System Time task.
- Following the same User Experience paradigm as initial Manage System Time (introduced in z14 GA1).
 - IBM Z administrator is guided through a system time management workflow reducing need to refer to external documentation.
 - Inline definition of technical terms eliminates need to look up documentation to find out definitions.
 - Detailed instructions and guidelines are provided within task workflow.
- IBM Z administrator gets to see visual representation
 - current system time networks shown in topological display.
 - preview of any reconfiguration action is shown in topological display.
- IBM Z administrator's confidence for system time management operations is enhanced.
 - Potential errors are surfaced in the visualization. User can drill down to find out details.
 - Reconfiguration workflows clearly show potential pitfalls and errors before change is applied.
 - Workflows steps are design to prevent mistakes.
- System time management performed for all managed systems from a single task on HMC.

Manage System Time - Split CTN (1 of 2)



- When there is a need to split of one or more systems in to the separate CTN without interruption in the clock source then system administrator will need to perform “Split to new CTN” action.
- Launched under the “Advanced Actions”.
- Use CTN ID drop down to select target CTN for the action.
- If targeted CTN only has members with the roles then task launch will fail with error message.
- If targeted CTN has at least one system without any roles then task will launch.
- Informational warning for user to acknowledge sysplex workloads are divided appropriately.

Manage System Time - Split CTN (2 of 2)



Step 1: Create CTN ID for the newly created CTN

Step 2: Make selection of which system(s) will be split of in to the new CTN

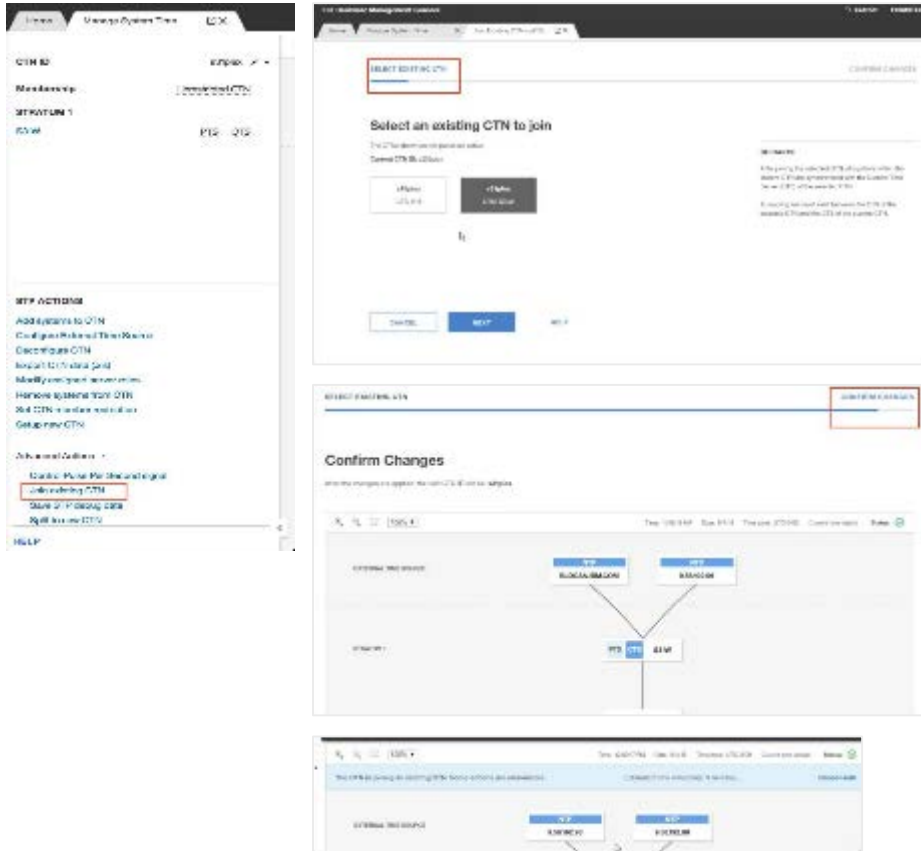
Step 3: From the selected systems choose one to be PTS.

Step 4: Examine proposed changes by toggling between topology representation for the target CTN and newly split CTN.

Step 5: After concluding that changes look good, click Apply and wait for the configuring process to complete.

- Once split action is complete, success message is displayed and newly created CTN will be available for viewing on the main task topology.
- New CTN will be listed in the CTN ID drop down.
- End result: one or more systems have been split in to the separate CTN without interruption in the clock source.

Manage System Time – Join existing CTN



- When there is a need to merge two separate CTNs in to the single CTN without interruption in the clock source then system administrator will need to perform “Join existing CTN” action.
- Launched under the “Advanced Actions”.
- Use CTN ID drop down to select target CTN for the action.
- An important thing to understand here is that after joining the selected CTN, all systems within the current CTN will be synchronized with the Current Time Server of the selected CTN. A coupling link must be in place connecting the CTS of the selected CTN and the CTS of the current CTN.
- Step 1: Select an existing CTN to join
- Step 2: View and confirm changes.
- When user clicks “Apply” the process of joining the two CTNs will begin, and “Join” wizard is closed.
- Back on the main task, each of the two CTNs will display a banner on top of the topology indicating this transitioning state.
- During the transition state most of the STP actions for the two affected CTNs are disabled.
- Once merge is completed then banner on the topology will be gone and STP actions on the left will be enabled again.

Dynamic IO for a Standalone CF – Background



- A z/OS Sysplex contains z/OS LPARs and Coupling Facility (CF) LPARs, usually across multiple CPCs for availability
 - It is generally not recommended to have the CF instance on the same CPC as the z/OS (or z/VM) instances because of the increased impact on the sysplex should there be a planned or unplanned outage of the underlying CPC or its surrounding infrastructure (power, networking, etc.).
 - So, many customers use CF-only CPCs (standalone CFs) to satisfy this recommended best practice for availability
- There is a need to occasionally update the I/O configuration on the servers to add/modify coupling links, add/modify images, etc.
- HCD (running in a z/OS LPAR) can drive software and hardware I/O updates.
 - Software update can be done sysplex-wide (across servers)
 - But, dynamic hardware updates are limited to the servers that have z/OS LPARs (and HCD instances) running on them.
 - Using HCD HMC-wide activate function, HCD allows also software updates on remote sysplexes and hardware updates on remote servers, if there is a z/OS or z/VM LPAR set-up to run the HCD service.

Dynamic I/O for Standalone CFs - Overview



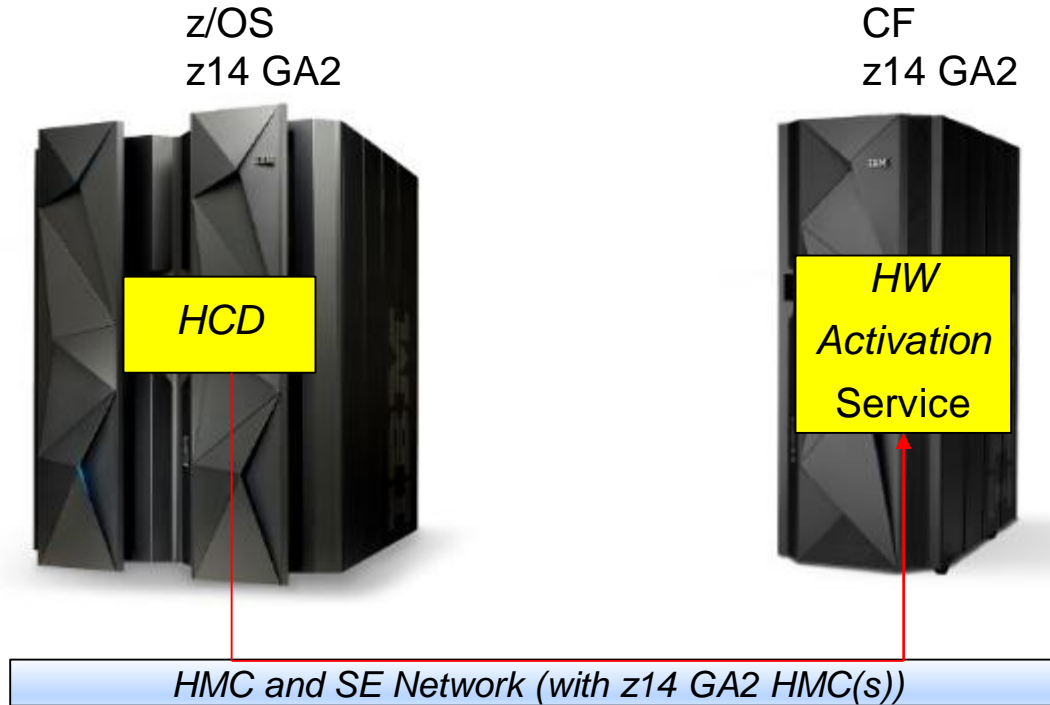
- Standalone CFs, by definition, have no co-resident z/OS (w/ HCD) images that can make hardware-only Dynamic I/O configuration changes on behalf of the CF partition(s)
 - Therefore, these I/O changes require disruptive IMLs of the standalone CF CPC, causing sysplex availability and complexity issues
- With new support, an MCS Linux-based HCD image will be started on the standalone CF CPC to perform this role
 - Simple, dynamic I/O changes with no IML requirement
- New firmware communication pathways from the “driving” HCD managing the IODF changes, via the HMC/SE, to the standalone CF CPC, and ultimately the MCS HCD
 - For passing the modified IODF
 - For driving the Dynamic I/O activate and associated recovery/management functions
- Remote Dynamic I/O hardware-only activations are performed on the Standalone CF CPC; CF image reacts to these changes just as if they’d been driven through a z/OS-based HCD
- **Requires z14 GA2 firmware support at both ends, plus z/OS PTFs at the “driving” z/OS system**
 - No new I/O adapter hardware requirements for the Standalone CF CPC; coupling links only

Dynamic I/O configuration for a Standalone CF – Solution

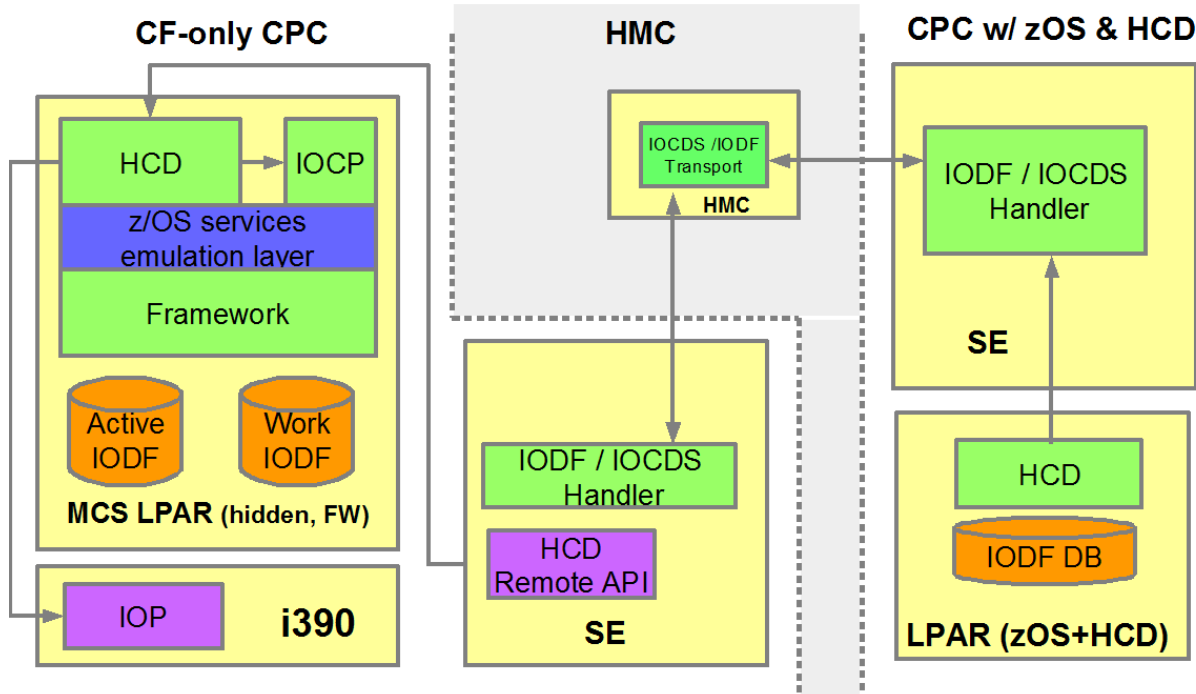


- With z14 GA2, dynamic activation of a new or changed IODF on a standalone CF CPC is supported:
 - Without requiring a POR/IML
 - Without requiring the presence (on the same CPC) of any z/OS or z/VM image running an HCD instance
- This is a base PR/SM solution; it does **not** require the use of Dynamic Partition Manager (DPM) mode
 - A new MCS LPAR (hidden, non-customer LPAR) is used, which is a firmware based appliance version of the HCD instance.
 - The MCS LPAR is a firmware LPAR
 - Fully managed by the z14 GA2 firmware
 - Included with the base firmware; no need to order a feature code
 - There will be a need to do a Power-on Reset with an IOCDs that includes and establishes the MCS LPAR on the standalone CF CPC before this new capability can be used.
 - Once this “last” POR is done on the standalone CF CPC, then all subsequent dynamic I/O changes can be done dynamically.
- The MCS HCD appliance LPAR on a z14 GA2 system will be driven by an updated HCD/HCM running in z/OS LPAR on a remote z14 (Driver Level 36) system

Dynamic I/O for a Standalone CF



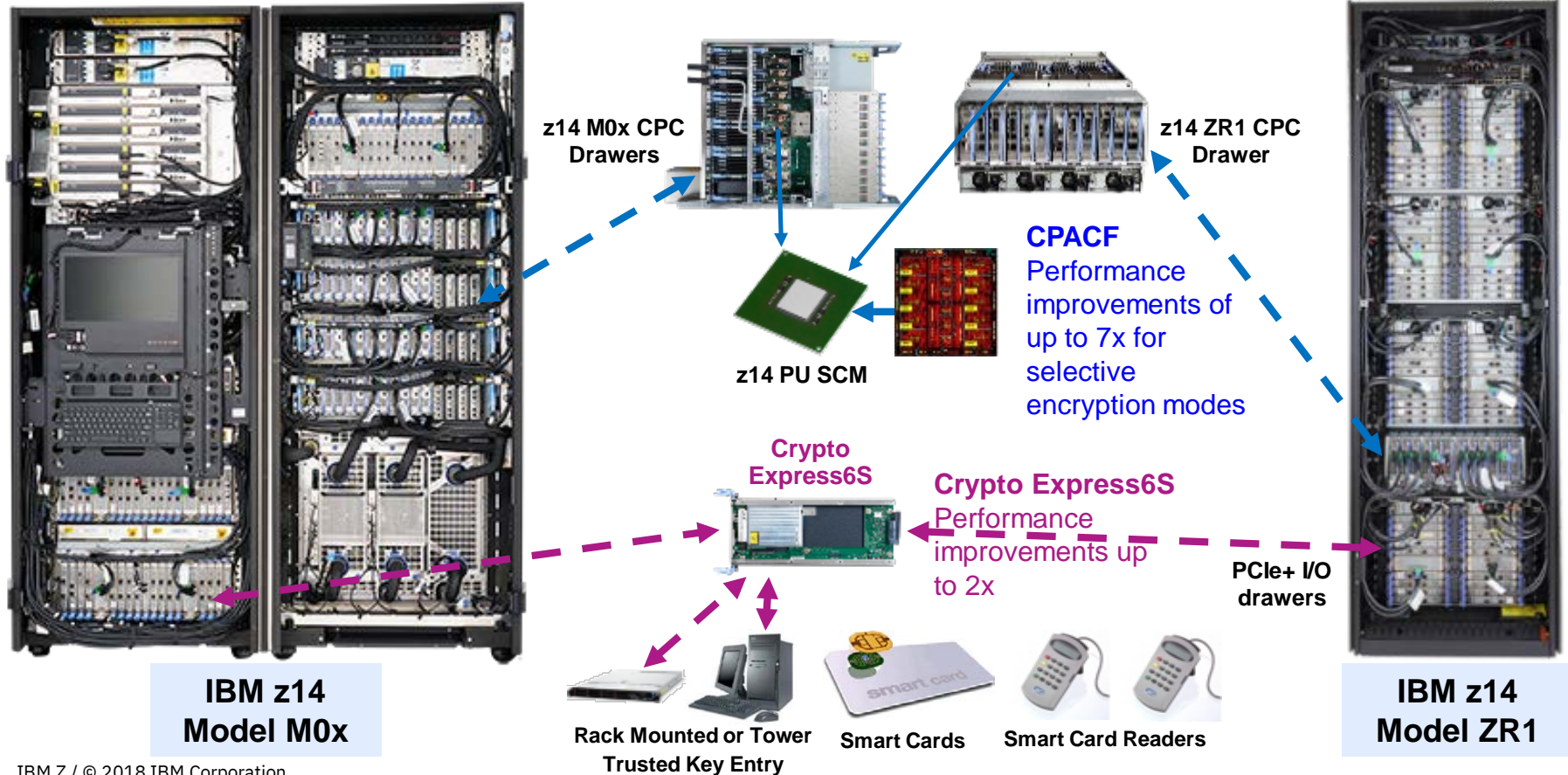
Dynamic I/O for a Standalone CF



Crypto and TKE 9.1

Hardware Crypto support in IBM Z Architecture

Industry exclusive
"protected key"
encryption

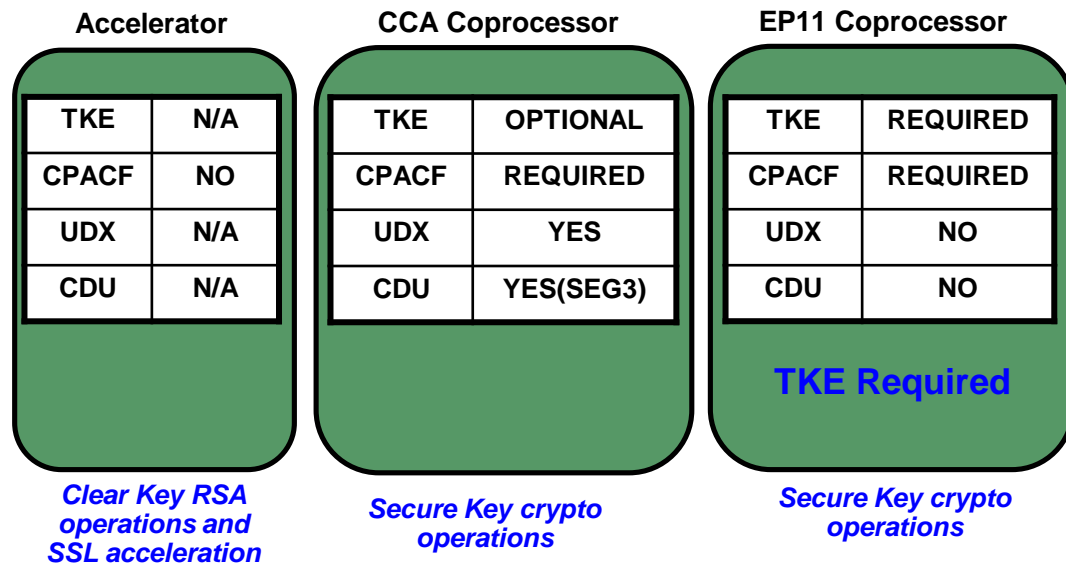


Crypto Express6S – FC #0893

- Enhanced Card Performance
 - Main PPC 476 (Qty 2 in lock-step) @ 1.2Ghz (vs. Crypto Express5S @ 800MHz)
 - Persistent Memory Management for faster boot time (FPGA)
 - New Miniboot Implementation for certification compliance
 - Allows easier transition of algorithms when certification requirements change
- Enhanced Public Key Cryptography Algorithms performance
- Upgraded secure module tamper detection technology
 - Improved thermal capabilities for increased performance
 - Continued support for temperature and voltage detection

Three Crypto Express6S configuration options

- Only one configuration option can be chosen at any given time
- Switching between configuration modes will erase all card secrets
- Exception: Switching from CCA to accelerator or vice versa



Security Certifications

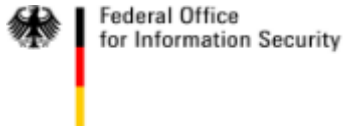
- Physical Security Standards in progress/planned:

- ✓ FIPS 140-2 level 4

- ✓ Common Criteria EP11 EAL4

- ✓ Payment Card Industry (PCI) HSM

- ✓ German Banking Industry Commission (GBIC, formerly E



Note: PCI-HSM certification is new for Crypto Express6S. The others also apply to Crypto Express5S.

Crypto Enhancements



CCA

- CCA 5.4/6.1
- CCA 6.2
- CCA 6.3

EP11

- EP11 Stage 4
- EP11 Concurrent Patch Apply
- eIDAS compliance

TKE

- TKE 9.1, EC smart cards

ICSF

- FIPS key wrapping
- Dynamic Service Update
- Early Availability

Misc

- RMF LPAR/Domain crypto statistics
- Crypto Adapter Serial # Tracking

TKE 9.1 LIC (FC 0880) and TKE additional Smart Cards (FC 0900)



- The Trusted Key Entry (TKE) 9.1 (FC 0880) level of Licensed Internal Code (LIC) is installed in a TKE workstation (#0080, 0081, 0085, 0086, or 0849).
- The TKE 9.1 LIC includes support for the Smart Card Reader (#0891) and (#0895)
- A new smart card for the Trusted Key Entry (TKE) allows stronger Elliptic Curve Cryptography (ECC) levels.
- TKE 9.1 License Internal Code enhancements for support EC521 strength TKE and Migration zones. An EC521 Migration zone is required if you want to use the migration wizard to collect and apply PCI-compliant domain information.
- TKE 9.1 also has a new family of wizards that makes it easy to create new EC521 zones on all of its smart cards. This simplifies the process of deploying a TKE for the first time or simplifies the process of moving data from a weaker TKE zone to a new EC521 zone.
- TKE 9.1 can be used to control IBM z13, z13s, zEC12, zBC12, z196, and z114 servers.
- Additional TKE Smart Cards (FC 0900, packs of 10, FIPS certified blanks) require TKE 9.1 LIC

Backup Material

Resource Groups and IFP

Review of the Integrated Firmware Processor (IFP)?

▪ Integrated firmware processor (IFP)

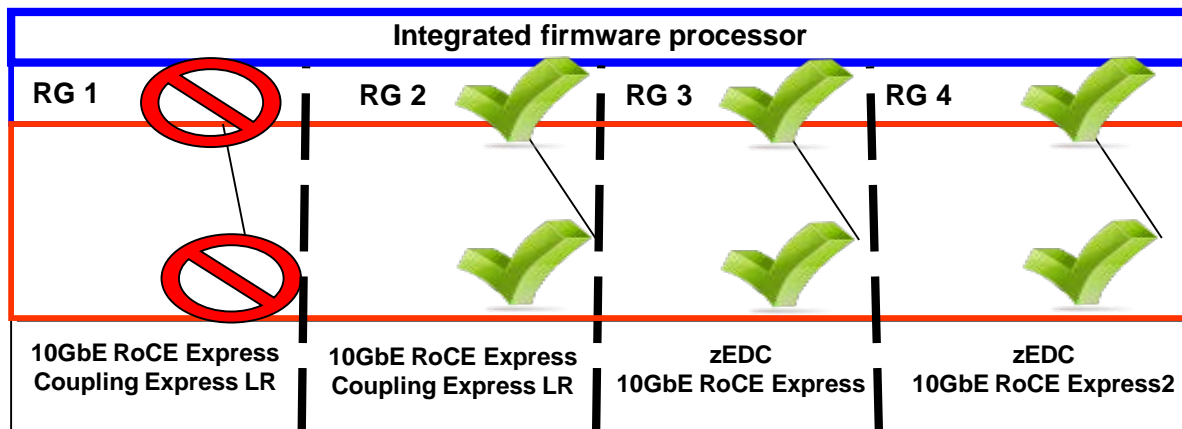
- The IFP is allocated from the pool of PUs available for the whole System
 - Unlike other characterized PUs, the customer doesn't pay for the IFP
 - It's a single PU dedicated solely for the purpose of supporting the native PCIe features and is initialized at POR if these features are present.
 - The z14 has four Resource Groups (RGs) which have firmware for:
 - 10GbE RoCE Express2
 - 10GbE RoCE Express
 - zEDC Express
 - Coupling Express LR
 - OSA-Express7S 25 GbE
 - 25GbE RoCE Express2
 - IBM Adapter for NVMe



IFP and Resource Groups – Basic Configuration

▪ Resource Groups (RG)


- Each Resource Group will handle 25% of the native PCIe features based on the plugging rules and purchases made in pairs of features
 - During firmware updates, error conditions, etc. that affects one RG, ALL the features attached to that RG will be unavailable
 - MCL update to Resource Group requires a RG outage of a few minutes



Concurrent Driver Upgrade

Concurrent Driver Upgrade

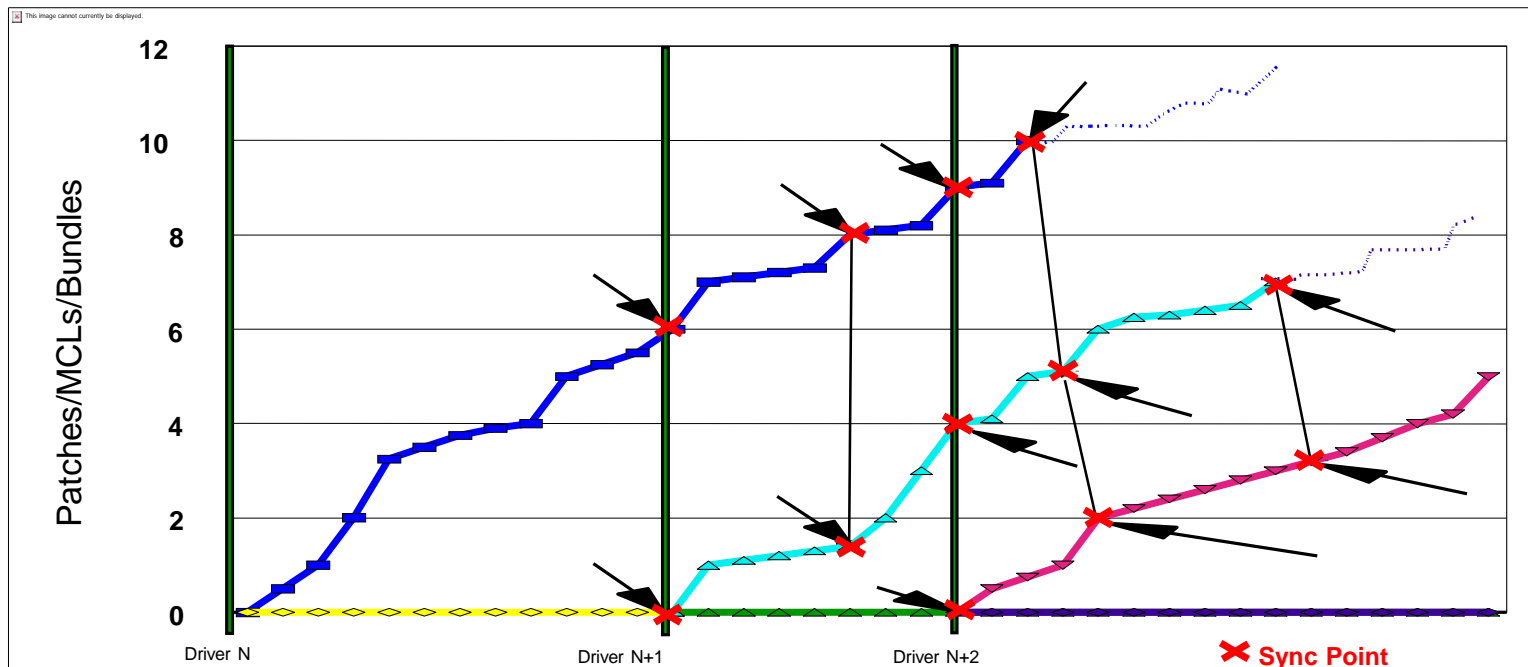
Also known as Enhanced Driver



Everything Old is New Again

- Need to apply GA MCL(s) to establish Sync Points in order to use EDM to upgrade from Driver Level 32 to Driver Level xx.
- All HMCs that have the CPC defined MUST be upgraded to Driver xx BEFORE the upgrade.
- Work through planning and execution with your SSR.
- The option to install new level of Driver code with a planned outage remains.

Concurrent Driver Upgrade – Sync-Points (min/max)

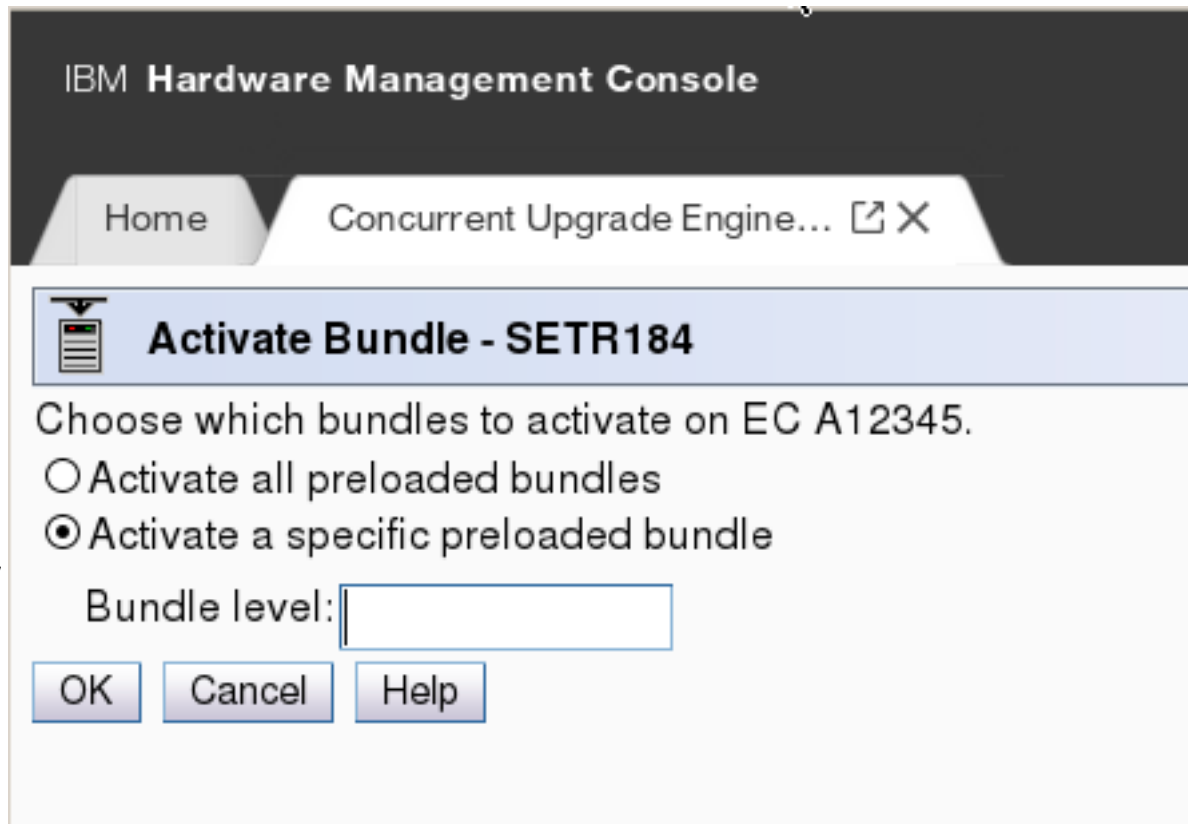


Customers are able to specify the target bundle.

Many clients want all of their machines to be at the same bundle level, while their enterprise is updated over a period of weeks.

Concurrent Driver Upgrade (CDU) in prior drivers automatically tried to load 'All MCL bundles'

Enhancement: Let the customer specify what bundle the machine should arrive at when the CDU has completed.



zEDC

Efficiency managing data movement to improve access time

On-chip compression coprocessor

- **Enhancements enable further compression of data** including Db2® indices, improving memory, transfer and disk efficiency
- In the future¹ Db2 plans to enable new **order-preserving compression for Db2 indices** using compression coprocessor to support index compression

zEDC

- **Compression further reduces cost to pervasively encrypt data** with less data to encrypt
- **More data active and effective compression** with a dedicated compression accelerator
- **Disk savings with improved utilization of storage tiers** with DFSMSdss™ use of compression


¹ IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.

Compression Coprocessor (CMPSC) vs. zEDC

Using the right hardware compression acceleration for each of your workloads

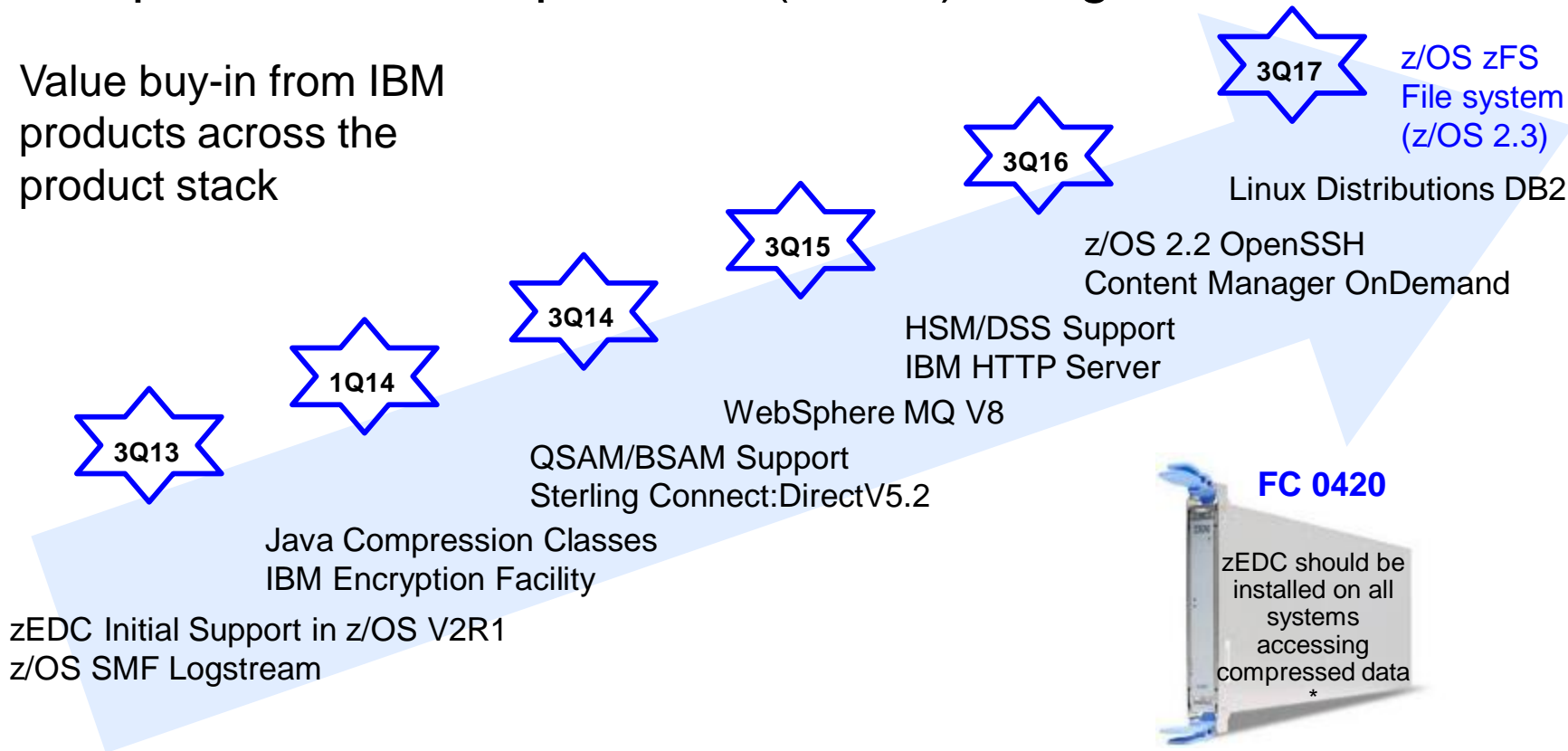
Compression Coprocessor

z Enterprise Data Compression

<p>On Chip</p> <p>In every IBM Z server</p> <p>Mature: Decades of use by Access Methods and Db2®</p> <p>Work is performed jointly by CPU and Coprocessor</p> <p>Proprietary Compression Format</p>	<p>PCIe Adapter</p> <p>Introduced with IBM zEnterprise® EC12 GA2 and IBM zEnterprise BC12</p> <p>Mature: Industry Standard with decades of software support</p> <p>Work is performed by the PCIe Adapter</p> <p>Standards Compliant (RFC1951)</p>	
<p>Use Cases</p> 		
<p><u>Small object compression</u></p> <ul style="list-style-type: none"> ▪ Rows in a database 	<p><u>Large Sequential Data</u></p> <ul style="list-style-type: none"> ▪ QSAM/BSAM Online Sequential Data ▪ Objects stored in a data base 	<p><u>Industry Standard Data</u></p> <ul style="list-style-type: none"> ▪ Cross Platform Data Exchange
<p><u>Users</u></p> <ul style="list-style-type: none"> ▪ VSAM for better disk utilization ▪ Db2 for lower memory usage ▪ The majority of customers are currently compressing their Db2 rows 	<p><u>Users</u></p> <ul style="list-style-type: none"> ▪ QSAM/BSAM for better disk utilization and batch elapsed time improvements ▪ SMF for increased availability and online storage reduction 	<p><u>Users</u></p> <ul style="list-style-type: none"> ▪ Java for high throughput standard compression via java.util.zip ▪ Encryption Facility for z/OS for better industry data exchange ▪ IBM Sterling Connect: Direct® for z/OS for better throughput and link utilization ▪ ISV support for increased client value

zEnterprise Data Compression (zEDC) Usage

Value buy-in from IBM products across the product stack





We want your feedback!

- Please submit your feedback online at
 - <http://conferences.gse.org.uk/2018/feedback/BG>
- Paper feedback forms are also available from the Chair person

- This session is BG



THANK YOU