CONVERGE
TECHNOLOGY SOLUTIONS

FBA, CKD, FCP, FICON, NVMe

What are they?
What are the differences?
Is one better than the other?

John Wolfgang
Tom Corrado
20 June 2024

Agenda

- Disk Formats
  - CKD
  - FBA
- Interface Types
  - SCSI
  - NVMe
- Data Transmission Protocols
  - FICON
  - FCP

- Compare and Contrast
  - Cost
  - Performance
  - Configuration Differences
  - Advanced Capabilities
  - Replication Considerations

Caveats

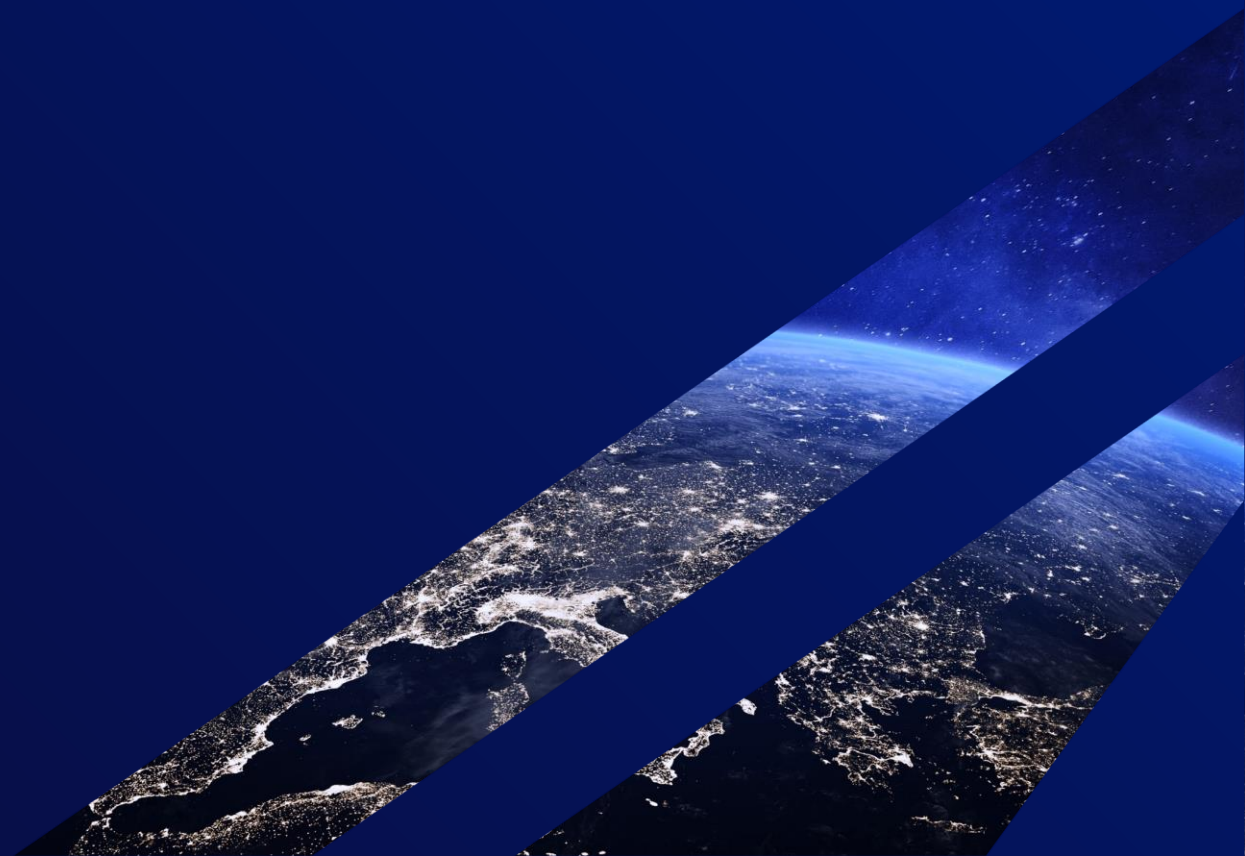
- Block Storage-Centric
- Storage System Dependent
- High-level Overview
- General Concepts
- No Step-by-Step or specific commands

Does any of this warrant detailed analysis?

CONVERGE
TECHNOLOGY SOLUTIONS

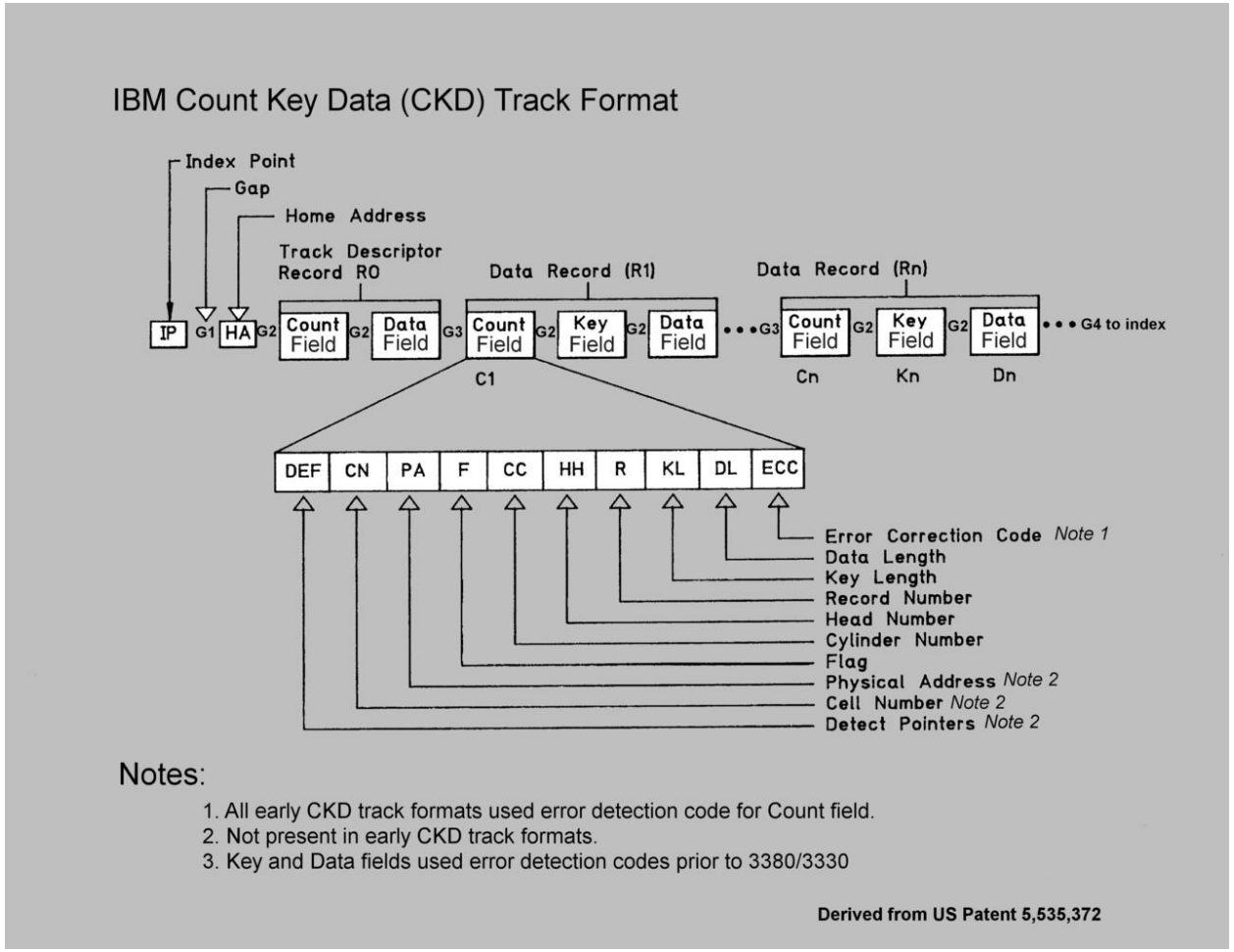
Disk Formats

CKD & FBA



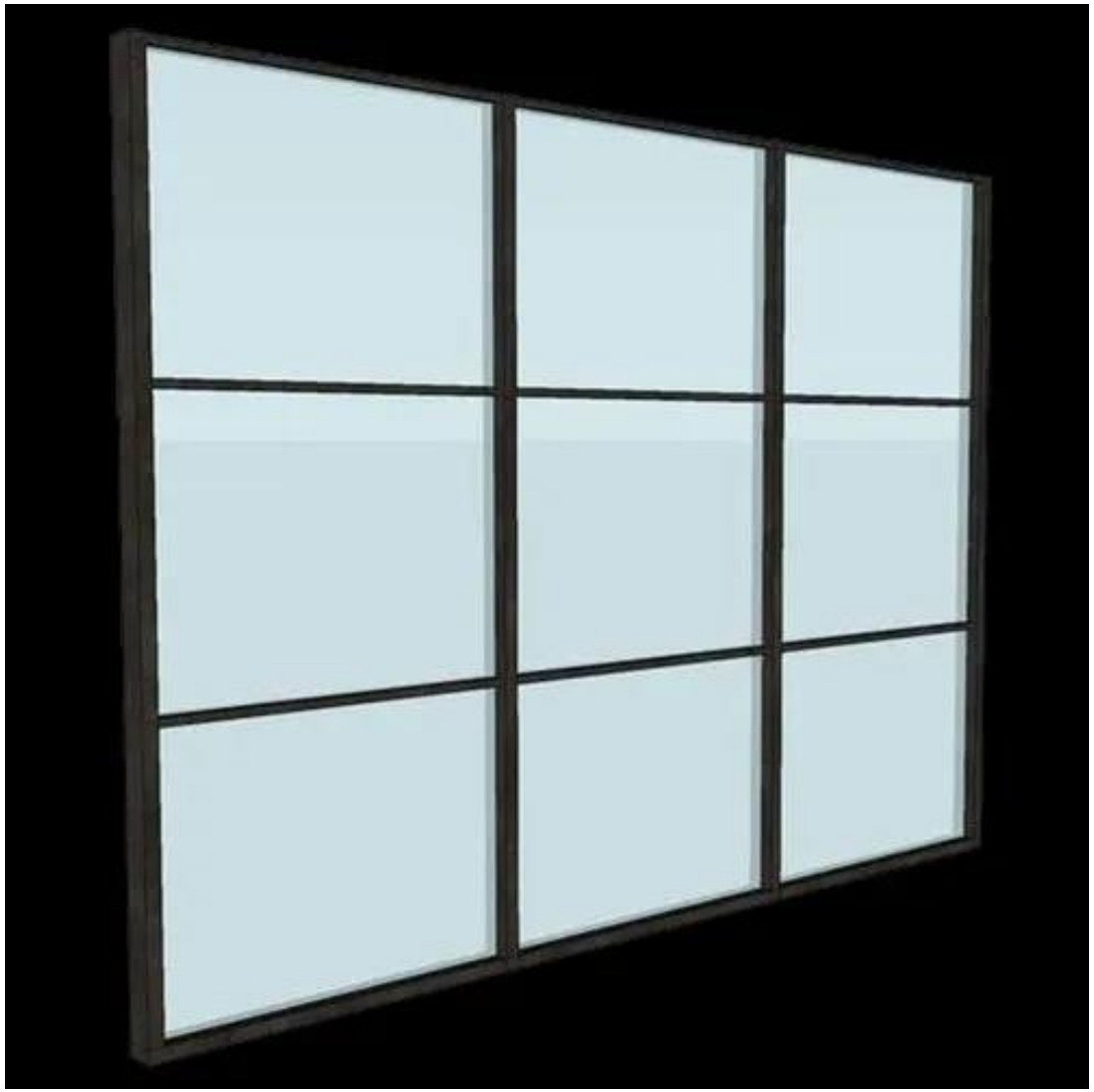
Disk Format: CKD

- CKD (**C**ount **K**ey **D**ata)
- Variable-length architecture– sector size is defined within the "count" area of each data record
- Required for IBM mainframes
- Introduced by IBM in 1964 with System/360
- CKD also refers to mainframe channel commands for using CKD-format devices



Disk Format: FBA

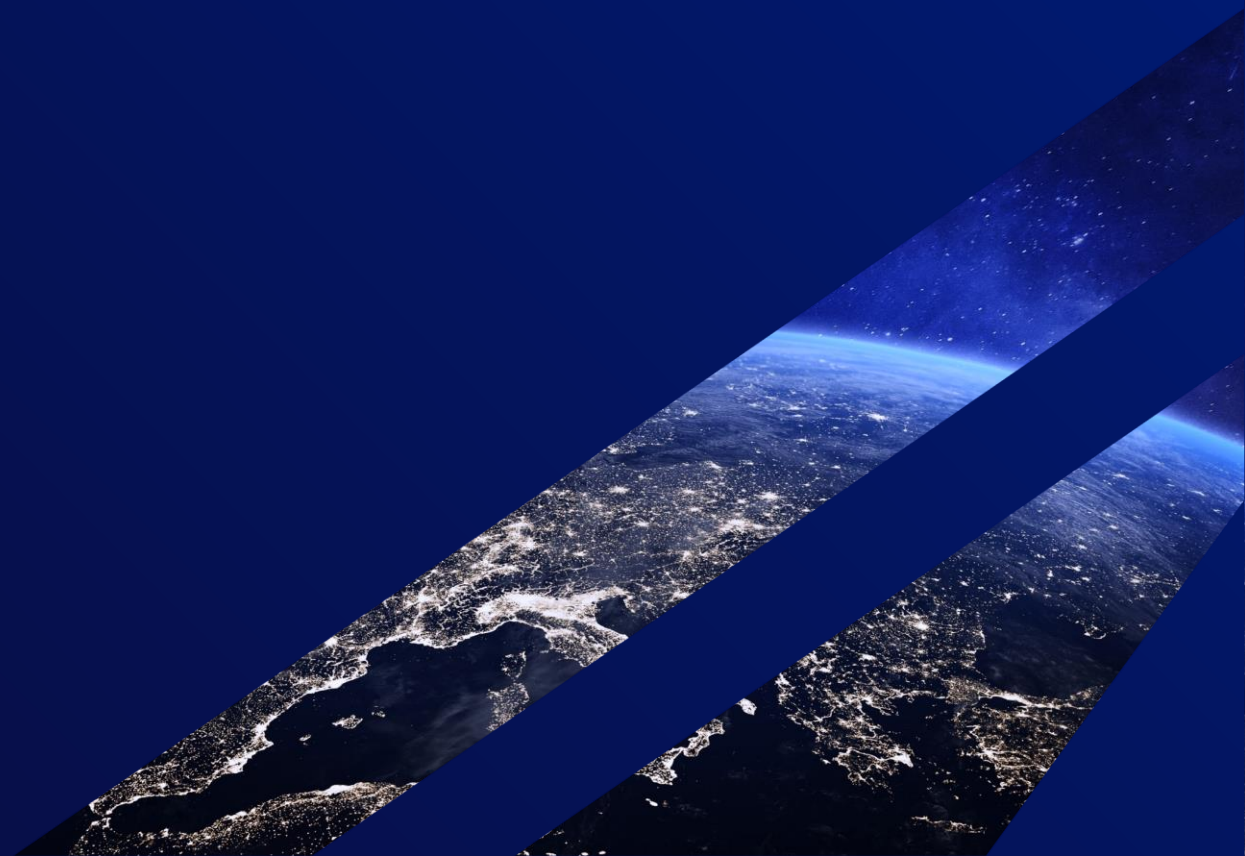
- FBA (**F**ixed **B**lock **A**rchitecture)
- Each addressable sector is the same size
- Term coined by IBM in 1979 after moving away from its variable-length mainframe architecture
- Available on mainframe storage devices as the “open systems” option
- Disk model on which SCSI is predicated



CONVERGE
TECHNOLOGY SOLUTIONS

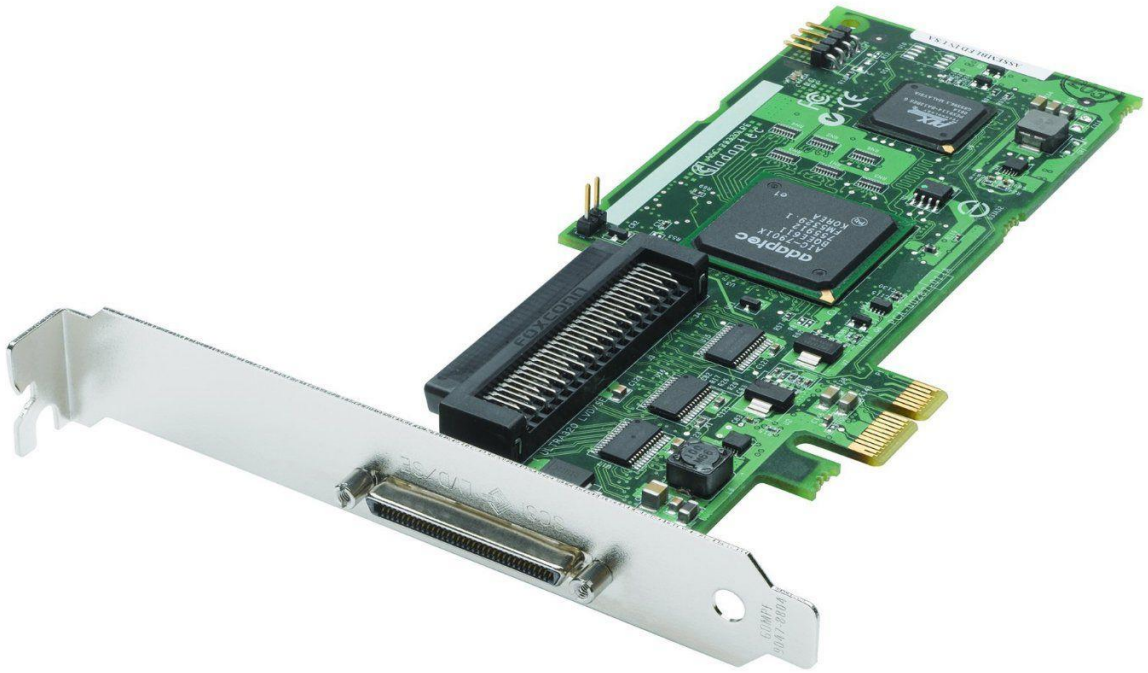
Interface Type

SCSI and NVMe



Interface: SCSI

- SCSI (**S**mall **C**omputer **S**ystem **I**nterface)
- Set of standards for physically connecting and transferring data between devices
- Best known for its use with hard drives
- Introduced in 1980s and continually updated



Interface: NVMe

- NVMe (**N**on-**V**olatile **M**emory **E**xpress)
- Allows host hardware/software to capitalize on the low latency and internal parallelism of solid-state storage devices
- Used for accessing NAND Flash memory
- Architecture physically contained within the storage media, updated by updating storage media
- No processor overhead
- Open-source interface specification

NVMe vs SCSI

SCSI was designed for slower hard disk drives (HDDs) and tape drives

NVMe was developed for use with memory-based technology and flash storage

- NVMe has a streamlined register interface and command set
- NVMe reduces CPU overhead
- NVMe lowers latency and improves performance

Features	Legacy Interface	NVMe
Max Command Queues	1	65536
Max Queue Depth	32	65536

Internal Flash Storage "NVMe"

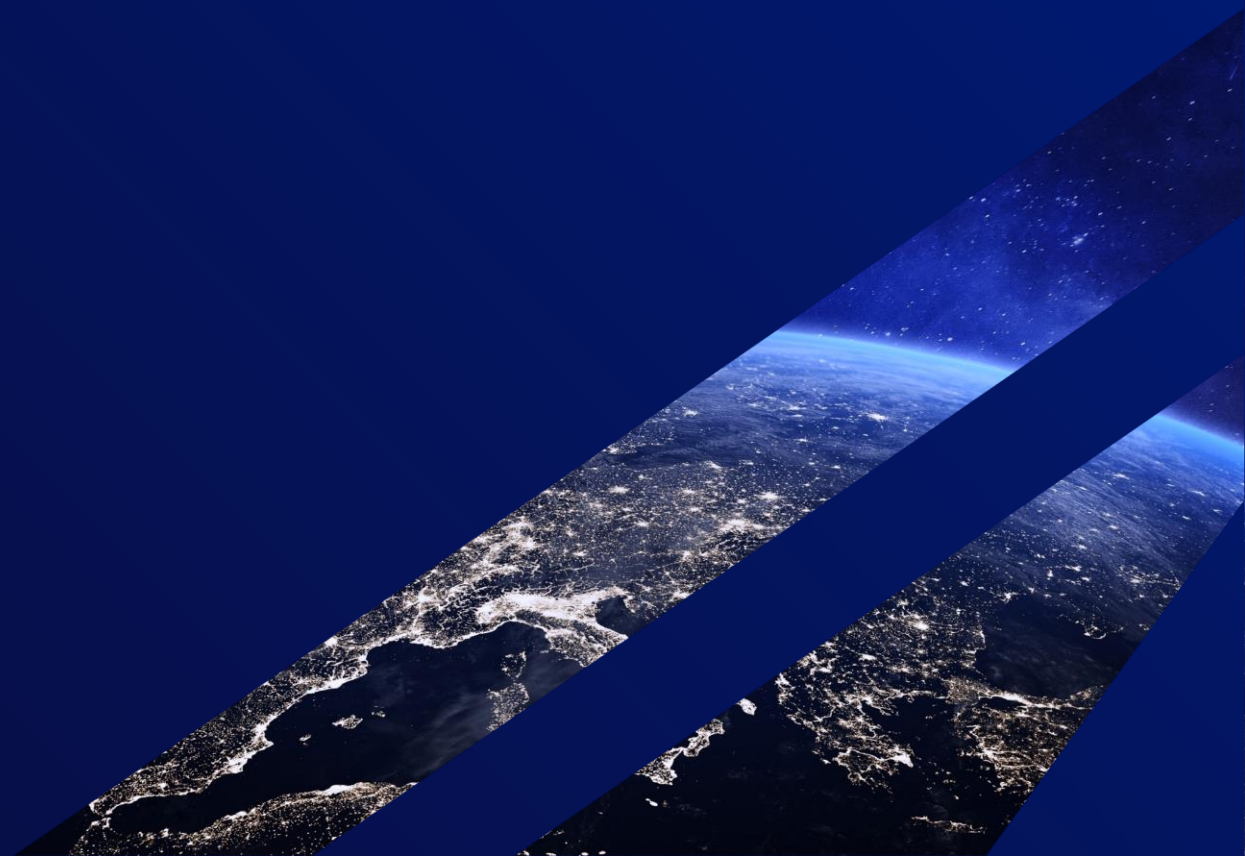
- Z Systems mainframes can have internal storage directly on the PCIe bus
- Sometimes referred to as "NVMe"
- Not to be confused with standard NVMe protocol

The term "NVMe" as used in this presentation refers to the standard NVMe protocol and not the internal flash storage

CONVERGE
TECHNOLOGY SOLUTIONS

Data Transmission Protocols

FICON, FCP, NVMe over FC



Protocol: FICON

- FICON (**Fi**ber **Con**nection)
- Allows for use of IBM's channel-to-control-unit connection with existing fiber channel infrastructure
- Required for native mainframe-to-storage connections
- Introduced in 1998, replaced bus and tag/ESCON in 2013
- Data rates up to 64Gbps at distances up to 10 km

Protocol: FCP

- FCP (**F**ibre **C**hannel **P**rotocol)
- SCSI interface protocol using a fibre channel connection
- High-speed data transfer mechanism for connecting hosts, storage devices, displays and more
- One standard for networking, storage, and data transfer
- Used to connect mainframe host/storage to “open systems”
- Data rates up to 64 Gbps at distances up to 10km

Protocol: NVMe over FC

- Allows for the performance benefits of NVMe between devices via fiber channel
- Provides interface-level performance at the network layer, similar to how FCP = SCSI over fiber
- Simplifies the NVMe command sets into basic FC protocol instructions
- Extremely low latency, higher performance, scalability, and parallel I/O in transferring data using the NVMe command set

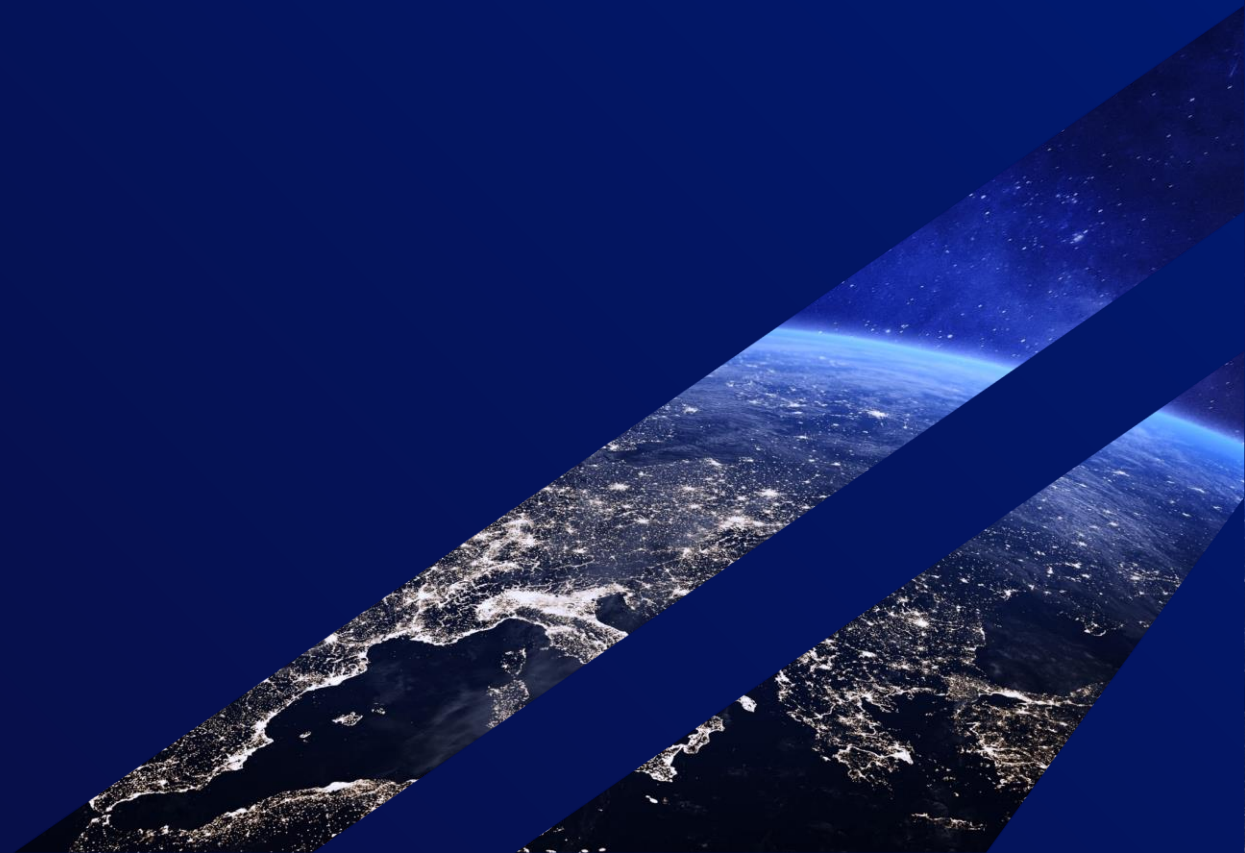
NVMe over Fibre Channel can coexist on your FC SAN along with your existing FCP or FICON traffic

Putting it All Together

- FBA LUNs are accessed via FCP or NVMe
- CKD volumes are accessed via FICON

CONVERGE
TECHNOLOGY SOLUTIONS

Compare & Contrast



Example Storage Systems

CKD & FBA Hybrid Storage Systems

Dell PowerMax 2500 & 8500
Hitachi Vantara VSP 5000 series
IBM DS8900F family

FBA-Only Storage Systems

Dell PowerVault & PowerStore series
Hitachi Vantara VSP One & VSP E-series
IBM FlashSystem series
PURE FlashArray series
Many other options

Compare & Contrast: Cost

Costs vary drastically and are very specific to configuration

In General:

- Systems that provide CKD will be more expensive
  - Highest-end enterprise systems that provide the most performance, reliability, flexibility, and features
  - Environmental costs are typically higher as well
  - If a user wants/needs CKD, this is the only option
  - These systems can do both CKD/FBA, so you can consolidate
- Users opt for the FBA-only systems whenever possible for a lower price point

If you need CKD anyway, no price advantage
If you can use an FBA-only system, FBA will be less expensive

Compare & Contrast: Performance

Native FICON v. FCP

Native FCP provides better performance than native FICON

Performance is Storage System Dependent

zHPF

High Performance FICON for z Systems

Parallel Access Volumes (Aliases)

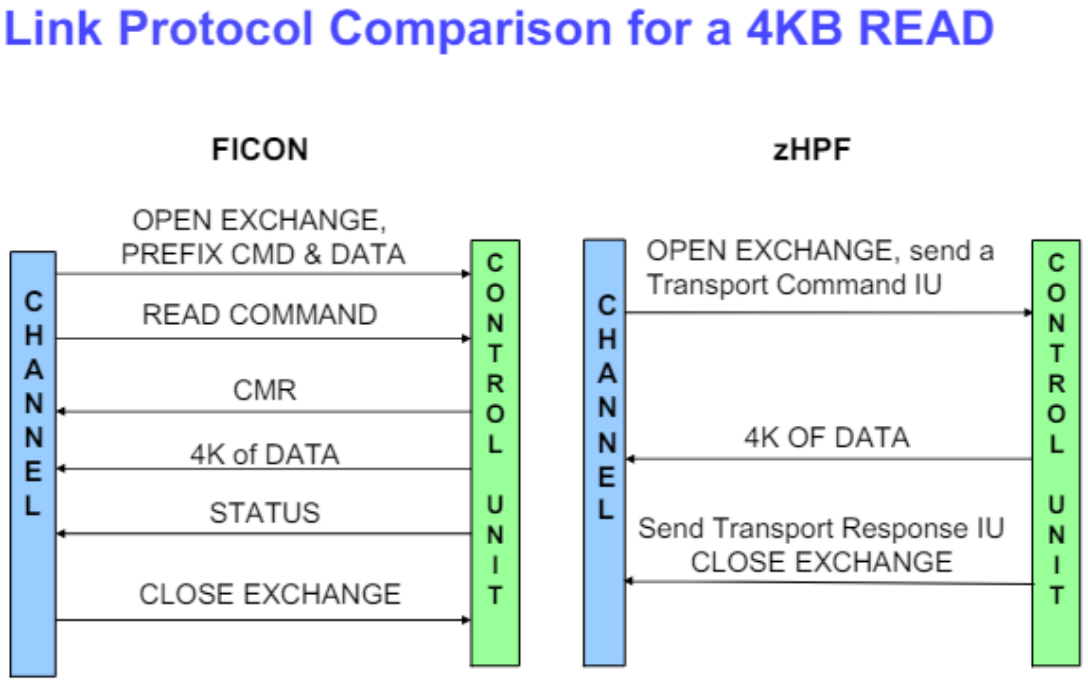
Static
Dynamic
HyperPAVs
SuperPAVs

zHyperlink

CKD only
z/OS only

High Performance FICON (zHPF)

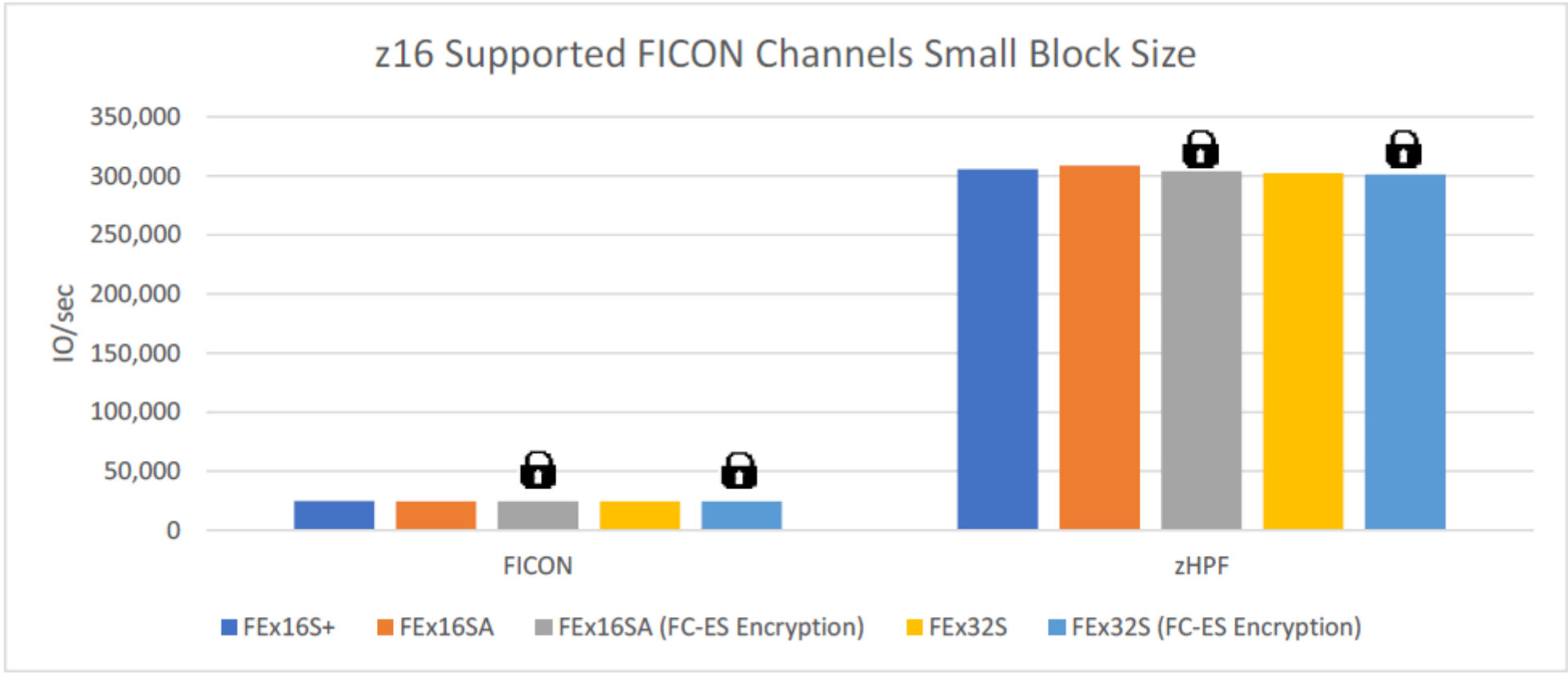
- zHPF is an extension to the FICON architecture designed to improve the execution of small block I/O requests.
- zHPF streamlines the FICON architecture
- Improves the way channel programs are written and processed to reduce channel overhead
- Sends multiple channel commands as a single entity instead of multiple separate commands
- Increases the number of active open exchanges



IBM z16™ FICON Express32S Performance
May 2023

In effect, zHPF is FICON acting more like FCP to achieve more FCP-like performance

zHPF: Drastic Improvement



IBM z16™ FICON Express32S Performance
May 2023

Parallel Access Volumes

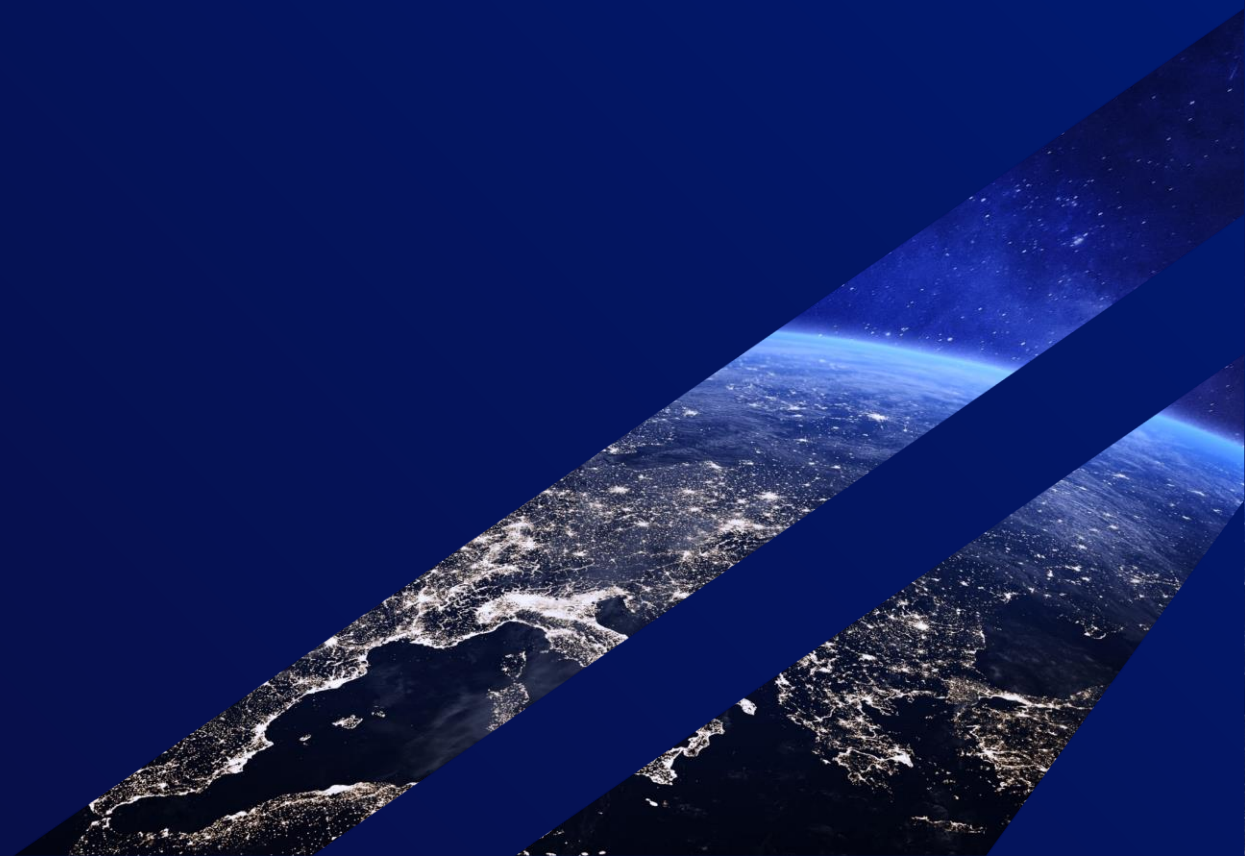
- Accessing a CKD volume via FICON is a single I/O operation at-a-time experience
  - Causes queuing delays
- PAVs enable a single server to process multiple I/O operations to the same logical volume simultaneously
- Aliases are created and temporarily assigned to the base volume that is being accessed
- Static, Dynamic, Hyper, Super
- FBA volumes use queues (QDIO devices) so they don't have this same issue

Combine zHPF & HyperPAVs/SuperPAVs
to get the most "FCP-like" performance

CONVERGE
TECHNOLOGY SOLUTIONS

Compare & Contrast

Configuration Differences



Data Access Control
Multiple Image Facility (MIF)

CONVERGE TECHNOLOGY SOLUTIONS

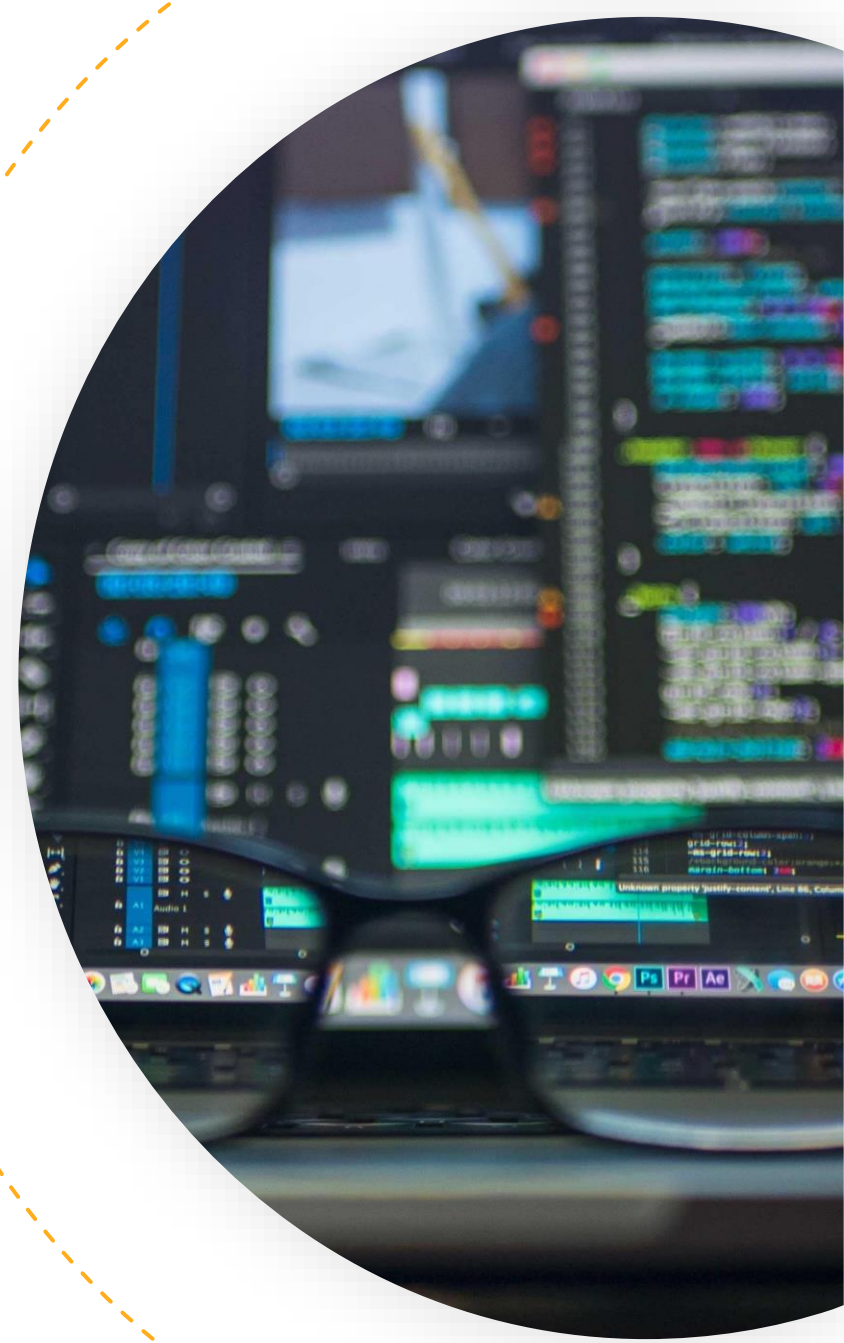
- FICON relies on MIF to manage shared channels & devices
  - Provides ultra-high access control and security of data
  - Ensures one OS image and its data requests cannot interfere with another

- MIF also allows FCP channels to be shared between Linux Logical Partitions and z/VM Logical Partitions with Linux guests

- FCP does not exploit the data access control and security functions of MIF resulting in the following limitations:
  - OS images share a WWPN and are indistinguishable from each other within the fabric
  - LUN access is first come, first served

Data Access Control
Node Port ID Virtualization

- Allows for a single FCP channel to be presented as multiple channels from multiple operating systems
- Each OS sharing an FCP channel receives a unique WWPN (**W**orld**w**ide **P**ort **N**ame) which it uses on the SAN
- FCP traffic can therefore be isolated by WWPN despite using the same physical port
- WWPN can be used for:
  - Device-level access control in storage controllers (LUN masking)
  - Switch-level access control (zoning)
- Requires a switch that supports NPIV



Data Access Control Switch Topology

- FICON SAN topology is limited to a two Director, single hop configuration
- FCP channels support full fabric connectivity, meaning that several directors/switches can be used between a System Z system and the device

HCD/IOCP Differences

FICON definitions require Control Unit layout to be defined in OS Configuration section of HCD

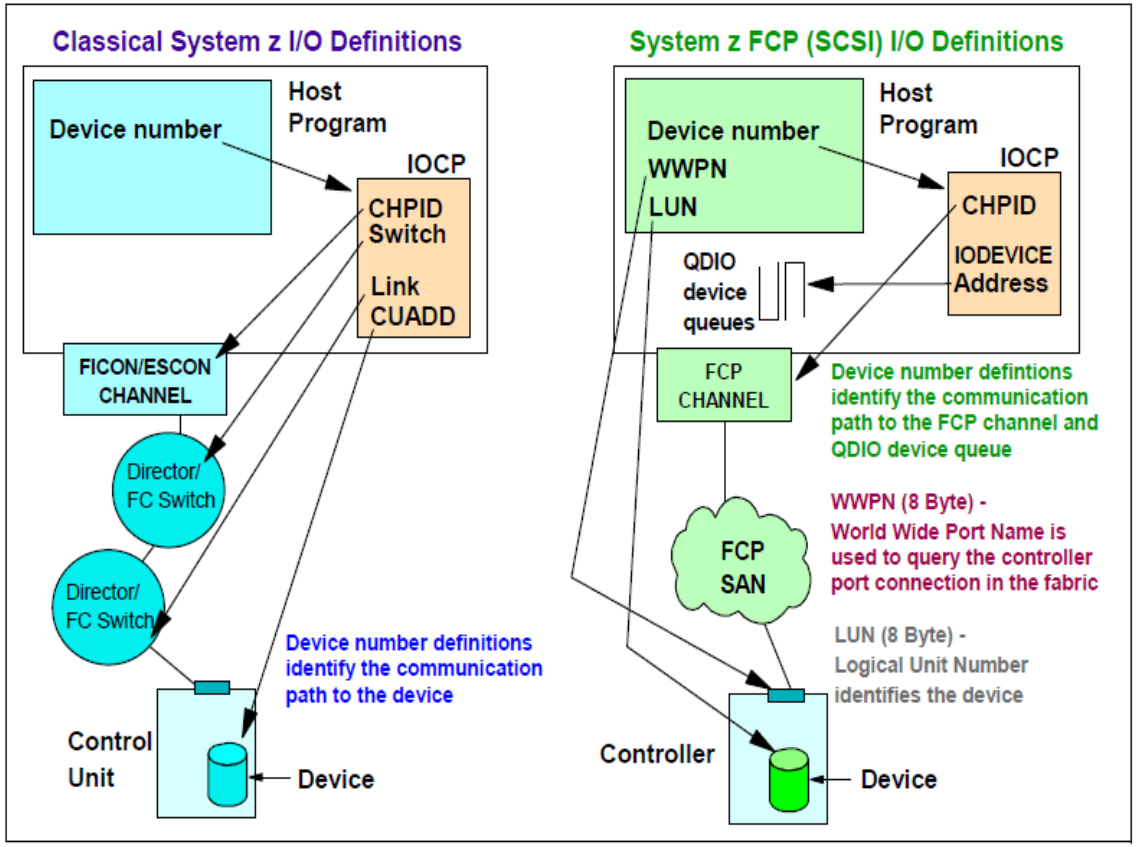
- Base & alias volumes

FICON addressing uses CHPID, Director Port, and Control Unit Address

For FCP, only the channel type and QDIO data devices are defined in the HCD/IOCP

FCP devices are addressed using the World Wide Names (WWNs) and Logical Unit Numbers (LUNs)

- Configured in the OS – NOT HCD/IOCP

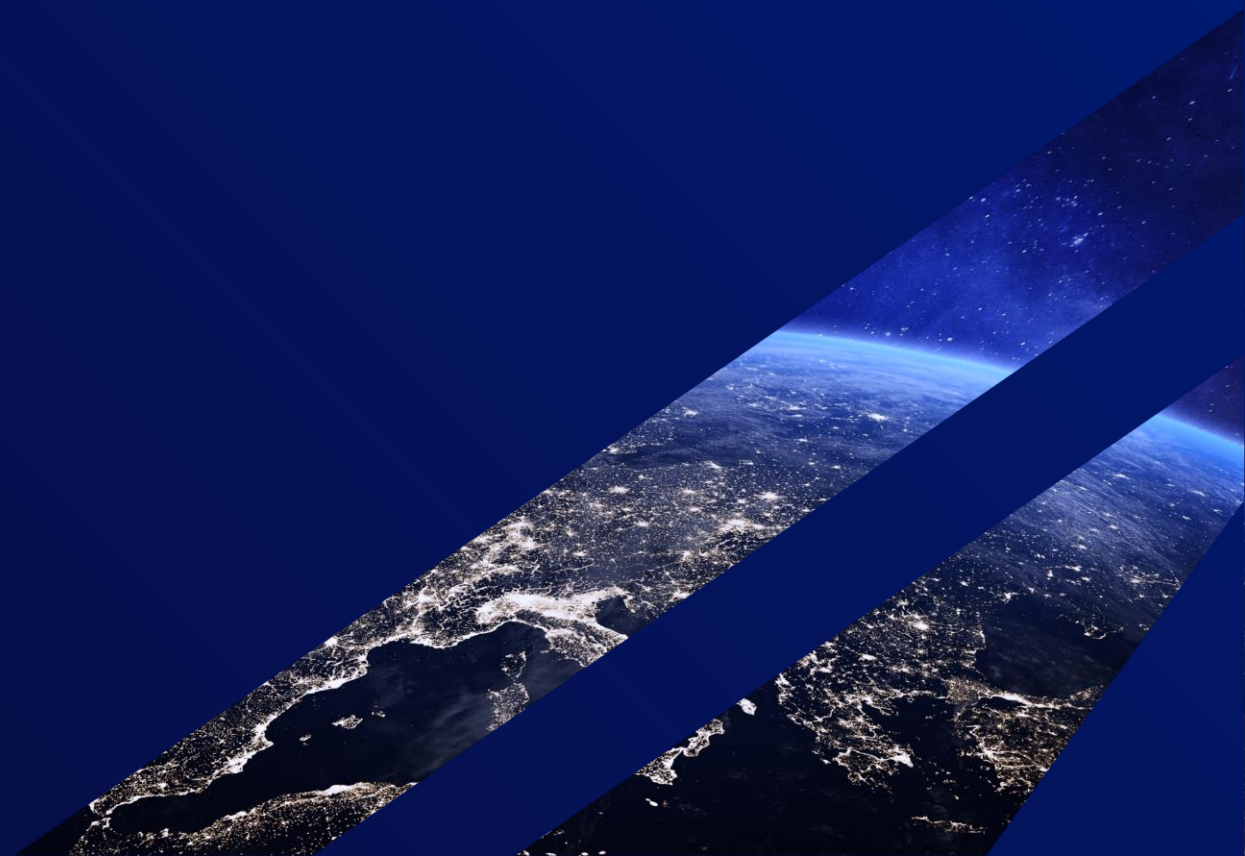


IBM z16™ FICON Express32S Performance
May 2023

CONVERGE
TECHNOLOGY SOLUTIONS

Compare & Contrast

Advanced Capabilities



Multipathing Differences

CKD multipathing is handled invisibly by the operating system

- z/OS is presented a single device
- Multipathing happens under channel subsystem control

FCP multipathing is managed at the Linux system level

- Each path to a LUN appears to the OS as a separate device
- Four paths to a LUN means Linux sees four SCSI devices.
- Multi-pathing implementation varies with the Linux distribution

Single System Image

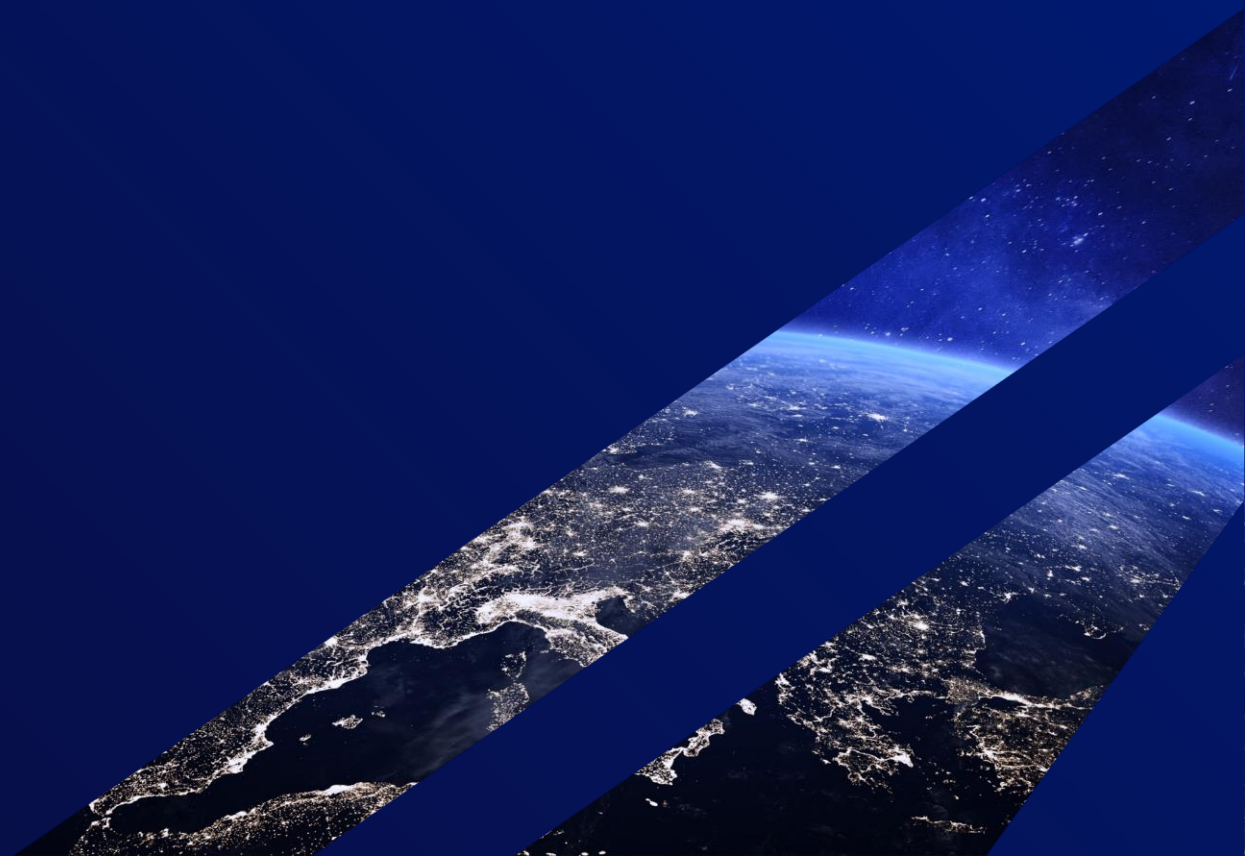
- Multiple systems can be clustered together to appear as a single system
- Enables multiple z/VM systems to share and coordinate resources within a Single System Image structure
- Ability to relocate Linux LPARs seamlessly
- zLinux SSI enables Live Guest Relocation (LGR)

Requires z/VM running on CKD volumes
(Guests can be on FBA)

CONVERGE TECHNOLOGY SOLUTIONS

Compare & Contrast: Replication

- **Replication Overview**
- High-Availability
- Replication Management Software



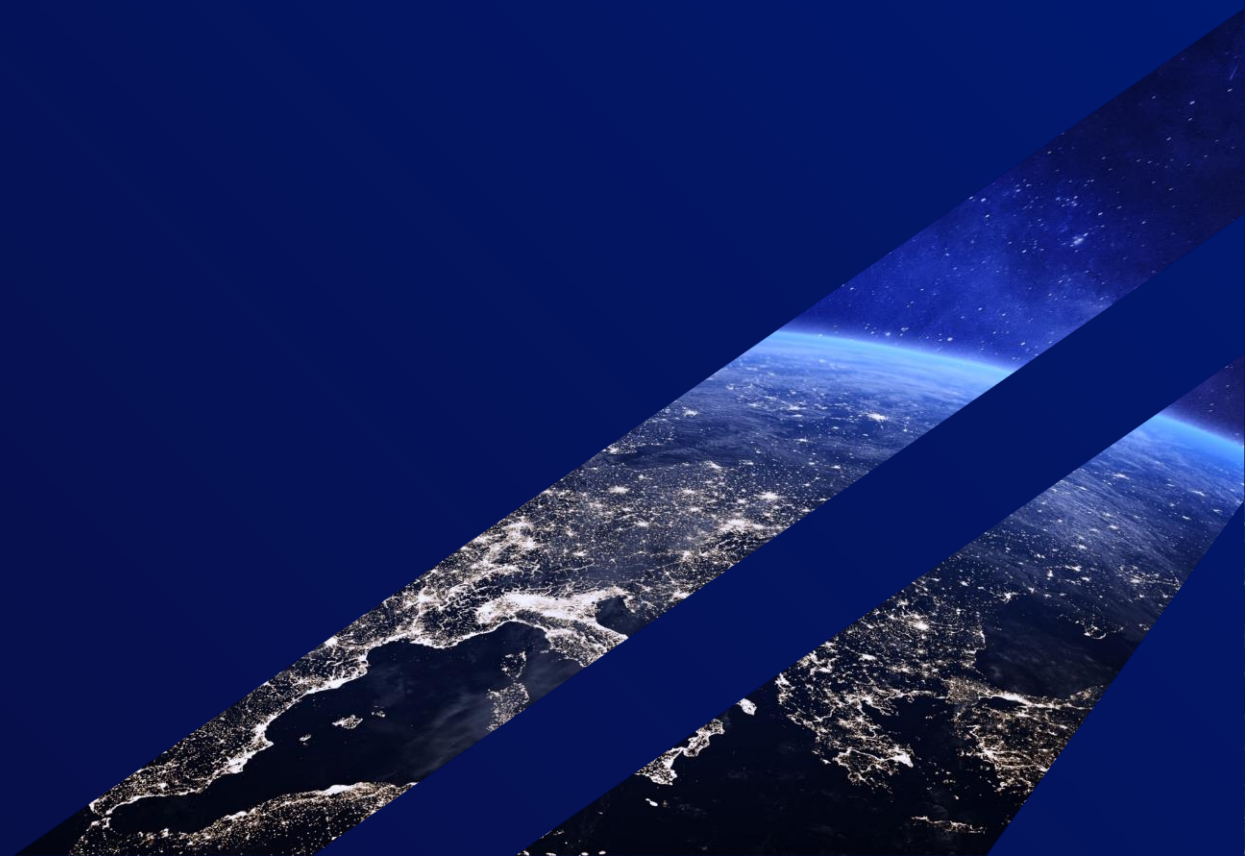
Replication Overview

- No Differences in Basic Replication Capabilities
  - Synchronous, Asynchronous, Point-in-time all available for both
- FBA provides larger volume capability in some storage systems
  - IBM DS8900F Replication Max: 4 TiB FBA vs. 1 TiB CKD
  - IBM FlashSystems provide much larger FBA capacity LUNs
  - Of course, there may be OS restrictions on usable LUN sizes

CONVERGE TECHNOLOGY SOLUTIONS

Compare & Contrast: Replication

- Replication Overview
- **High-Availability**
- Replication Management Software



High-Availability

Dell

Dell AutoSwap

IBM

IBM HyperSwap
IBM FlashSystem HyperSwap

Hitachi Vantara

IBM CSM Basic HyperSwap
IBM GDPS HyperSwap

Dell AutoSwap

- Swaps workload from one set of volumes to another set in different storage systems with no interruption of operations
- Uses standard z/OS services
- Used for both planned and unplanned swaps

Dell AutoSwap is CKD ONLY

IBM FlashSystem HyperSwap

- Provides dual-site access to a volume – FBA ONLY
- HyperSwap volumes have a copy at one site and a copy at another site. Data that is written to the volume is automatically sent to both copies.
- If one site is no longer available, the other site can provide access to the volume.
- The system automatically provisions change volumes to provide consistency protection.
- The synchronization process is managed automatically by the system.

IBM System z HyperSwap

Runs on the IBM DS8000 storage family

Switches all UCBs of the primary volume to point to the secondary volume and redirects all I/O transparently to running applications

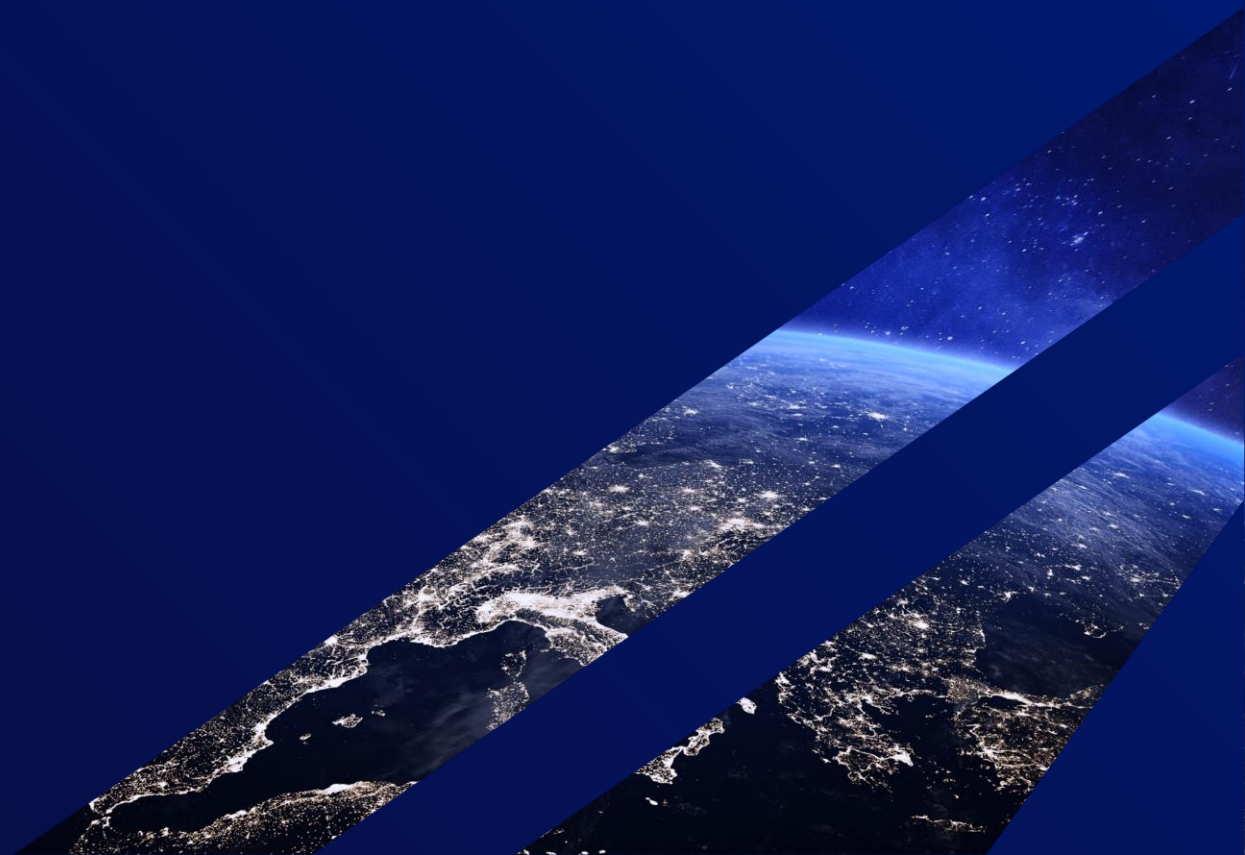
Requires synchronized Metro Mirror replication relationships between the volumes being swapped

Used for both planned and unplanned swaps

IBM System z HyperSwap can operate on both CKD & FBA depending on Management Software

Compare & Contrast: Replication

- Replication Overview
- High-Availability
- **Replication Management Software**
  - **Dell GDDR**
  - **IBM GDPS**
  - **IBM CSM**



Dell Geographically Dispersed Disaster Restart (GDDR)

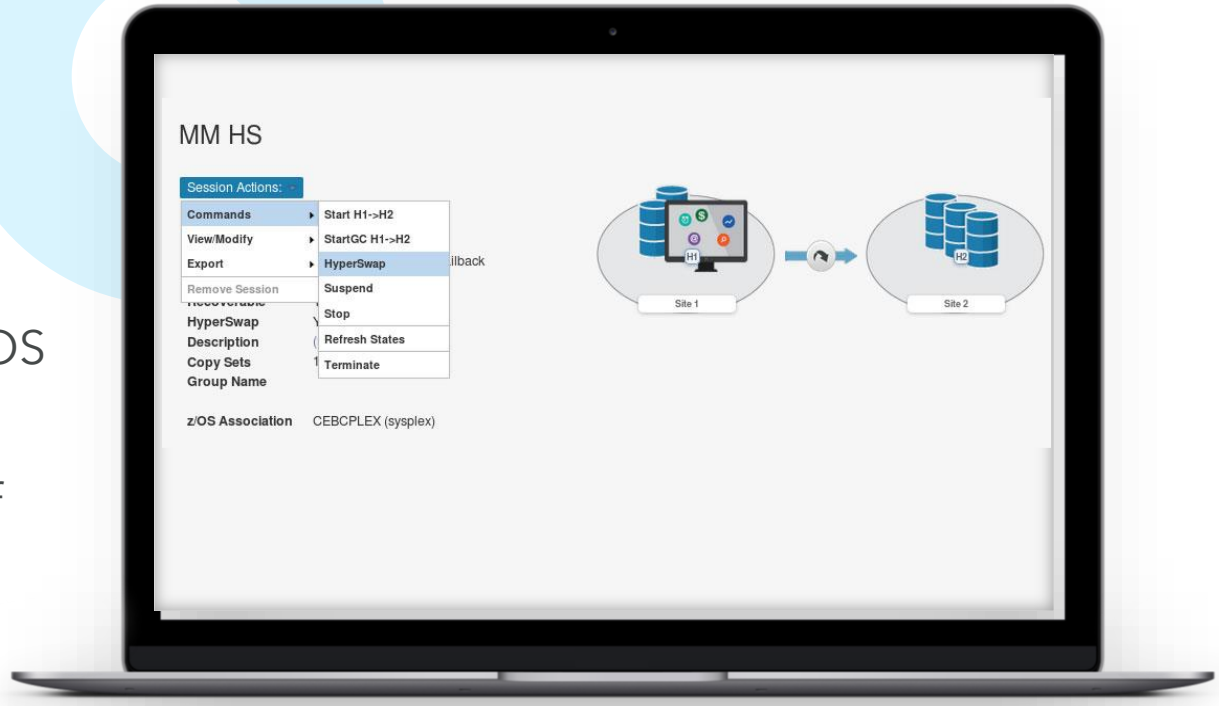
- GDDR automates business recovery following both planned outages and disaster situations, including the total loss of a data center
- Provides monitoring, automation, and quality controls to many Dell and third-party hardware and software products required for business restart
- Supports an intermix of CKD and FBA volumes
- Does NOT support automated AutoSwap of FBA volumes
- During an AutoSwap, the source FBA disks are made “Not Ready” which causes application timeouts on the attached servers.
- **Installed on z/OS**

IBM Geographically Dispersed Parallel Sysplex (GDPS)

- Family of software products for disaster recovery and resiliency
- Manages storage replication across heterogenous platforms
- Automates IBM Parallel Sysplex operational tasks & Performs Failure Recovery
- Supports replication management of both ECKD and FBA volumes
- Provides Data consistency across the IBM Z and distributed applications including HyperSwap via xDR managed FBA disk
- Has extended support for the new Single System Image (SSI) 8-way cluster capability
- GDPS Metro Linux® in LPAR mode provides support for running Linux natively in an LPAR on IBM Z® hardware
- Supports Hitachi Vantara systems for HyperSwap
- **Installed on z/OS**

IBM Copy Services Manager (CSM)

- IBM CSM controls Copy Services in heterogeneous storage environments
- Can be installed on open systems servers, DS8900F HMCs, or z/OS
- Manages replication of both CKD & FBA volumes
- Can manage CKD volumes via FICON connectivity
- When installed on z/OS or can communicate with z/OS IOS component
- Via the z/OS connectivity, can manage HyperSwap of z/OS CKD volumes only
  - Basic HyperSwap or full HyperSwap capability
  - Supports Hitachi Vantara systems for HyperSwap



Summary

- Disk Formats
  - CKD
  - FBA
- Interface Types
  - SCSI
  - NVMe
- Data Transmission Protocols
  - FICON
  - FCP

- Compare and Contrast
  - Cost
  - Performance
  - Configuration Differences
  - Advanced Capabilities
  - Replication Considerations
    - Replication Overview
    - High-Availability
    - Replication Management Software

CONVERGE
TECHNOLOGY SOLUTIONS

Thank you

john.wolfgang@convergetp.com
tom.corrado@convergetp.com

References

- *IBM z16™ FICON Express32S Performance*, May 2023
- Dr. Steve Guendert , Brocade Communications, *Understanding NPIV and the Performance of Channels with zLinux,* SHARE Boston 2013
- John Crossno, *Understanding the Benefits of SCSI for Linux on z Systems*, SHARE Orlando 2015
- *IBM System z Connectivity Handbook*, 2013