



# An Introduction to SAN and Fixed Block Disk for ECKD Users

Samuel D. Cohen  
IBM zSystems Consultant  
Levi, Ray & Shoup, Inc  
[sam.cohen@lrs.com](mailto:sam.cohen@lrs.com)  
(217) 862-9227



## Overview

- z/OS uses Variable-Block disk storage (aka Enhanced Count-Key-Data (ECKD)), while other IBM z operating systems can use ECKD and/or 512-byte Fixed Block (FB) disk.
- Accessing FB disk in a SAN uses Fibre Channel instead of FICON
  - Same hardware
  - Different communication protocols
- This presentation will compare/contrast accessing ECKD disk vs. FB disk
  - Part 1 – Hardware and Terminology
  - Part 2 – Example from a real system



## Background: IOCDs

- The I/O Configuration Dataset (IOCDs) defines connections between the z server and attached peripheral devices
- I/O Path Management is performed by the I/O Subsystem
  - Separate from the processor(s) used by the operating system
- Primary Entries in IOCDs
  - RESOURCE (LPAR Definitions and associated LPAR IDs (0-F))
  - CHPID (logical and physical ID and associated protocol)
    - Physical connection depends on the hardware location in the I/O cages and is identified via a Physical Channel Path ID (PCHID).
    - Up to 256 subchannels on a CHPID (aka channel), depending on attached device
  - CNTLUNIT (control unit to receive I/O for routing to device)
    - Connected to one or more CHPID(s)
    - May have multiple control units on one CHPID via CUADDR parameter
  - IOADDR (individual peripheral devices)
    - Connected to a one or more CNTLUNITs



## Background: Peripheral Devices

- Traditionally, peripheral devices are connected to a control unit, and control units have “channel” connections to the zSystems server
- Control Units can be directly connected to the server or go through a switch (also called a “director”). Control Units can also be daisy-chained on a single path (physically or logically, depending on the device)
- Addressing peripheral devices is based on a 4-digit (hex) I/O address. It used to be a concatenation of “channel” address with “unit” address, although that relationship went away in the 1990s. You may hear the term “cuu”, “ccuu” or “UCB”.



## Background: Peripheral Devices

Example of a Disk Subsystem with 1024 addressable devices:

- 4 Channels connected between server and tape subsystem

- 4 Control Units

- 256 disk devices defined per Control Unit (architectural limit)

The Control Units share the 4 channels by having unique Control Unit addresses within the subsystem. These address match the CUADDR parameter on each CNTLUNIT statement in the IOCDs

Different peripherals have different architectural limits



## Background: FICON

- Fiber Channel Connection (FICON) attachment uses fiber for communication between the IBM zSystem and the peripheral control unit
  - The IOCDs defines which FICON ports are used by a Control Unit, and which devices are attached to a particular Control Unit
  - Path management is determined by the I/O subsystem
  - Outbound communication is independent from inbound communication
  - Operating Systems (z/OS, VSE<sup>n</sup>, z/VM, z/TPF, Linux) are not involved in path selection; they send the message to the I/O subsystem for handling



## Background: FCP

- Fibre Channel Protocol (FCP) attachment is handled differently from FICON attachment, although the hardware is the same
  - FCP port is assigned a World-Wide Port Name (WWPN)
    - One WWPN per port
    - An FC Port is to a WWPN as an OSA Port is to a MACADDR
    - IBM z assigns WWPNs based on CPU Serial number and PCHID
    - Multiple subchannels are available but every subchannel sees the same traffic because traffic is routed between WWPNs on both sides of the connection
  - No path management by I/O subsystem
  - Path management by the operating system(s)



## Background: NPIV

- How do you keep traffic different FCP subchannels from seeing traffic on all other subchannels?
- Virtualization!!
- N\_Port ID Virtualization (NPIV) creates a virtual WWPN for each subchannel
  - Limited to 64 subchannels per FCP port in current hardware models
- Using NPIV, traffic between an FCP subchannel and a disk subsystem will not be seen by any other FCP subchannel
  - Traffic could be seen at the disk subsystem channel interface unless it also uses NPIV





## Background: SAN “Fabric”

- IBM z does not directly connect to FC HBAs
  - Must use a SAN switch that is certified for use with IBM z
- SAN provides the path management between FC-attached devices
- “Zoning” is the process of pairing these FC attachments
  - WWPNs are used in the zoning process
  - Not limited to a single point-to-point definition at each end
    - Can have 1:1, 1:many, many:1 or many:many
    - Pathing is managed by a multipath background process/started task/daemon in the host operating system
      - Configured by querying the SAN and devices attached at other end of the zone
- Usually want 2 separate fabrics for redundancy



## ECKD Storage Devices

- 1 or more Hardware Bus Adapters (e.g. FICON channel)
- Pre-defined Logical Control Units or Logical Subsystems (LSS number = CUADDR on IOCCS CNTLUNIT statement)
- Pre-defined Unit Addresses (0-255 per LCU)
  - Size of each logical disk is pre-defined
- IOCCS should have configuration statements matching pre-defined definitions in disk subsystem
  - IOCCS doesn't care about "base" vs. "alias" devices, HCD does
    - For z/VM to see alias disk devices when running alongside z/OS, ensure that the alias devices in HCD are in channel subsystem 0 (z/OS default is channel subsystem 1)



## FB Storage Devices

- 1 or more Hardware Bus Adapters (e.g. FC channel)
- Pre-defined 16-digit Host Addresses (WWPNs)
- **No** Logical Control Units
- Pre-defined Logical Units (LUNs)
  - Size of each LUN is pre-defined
- SAN Zones providing the linkage between IBM zSystems and the disk Host Bus Adapters (HBAs; e.g. channels)
- Disk subsystem definitions for the IBM zSystems WWPNs that will be accepted and matched with local LUNs



## SAN Zoning

- Independent of both server and storage
- Provides pathing for I/O
- Must be configured and activated before use
  - Many fabric administrators expect to see WWPNs before they are in use by the server...this is not necessarily true for IBM z
  - Connectivity issues are usually due to misconfiguration
- Zoning involves WWPNs only
  - LUNs are managed at the disk subsystem
- Disk subsystem may require pre-definition of incoming WWPNs in addition to SAN zoning



## Steps in SAN Zoning

- Identify the WWPNs you want to connect from both ends of the connection
- Define an alias for the WWPNs at each side of the connection
  - Suggestion: If using NPIV, put all the virtual WWPNs for one subchannel (across all LPARs) in the same alias
- Create a zone containing the aliases for each side of the connection
- Add the new zone to the zone configuration
- Activate the zone configuration



# Lost Yet?

## Let's take a break....



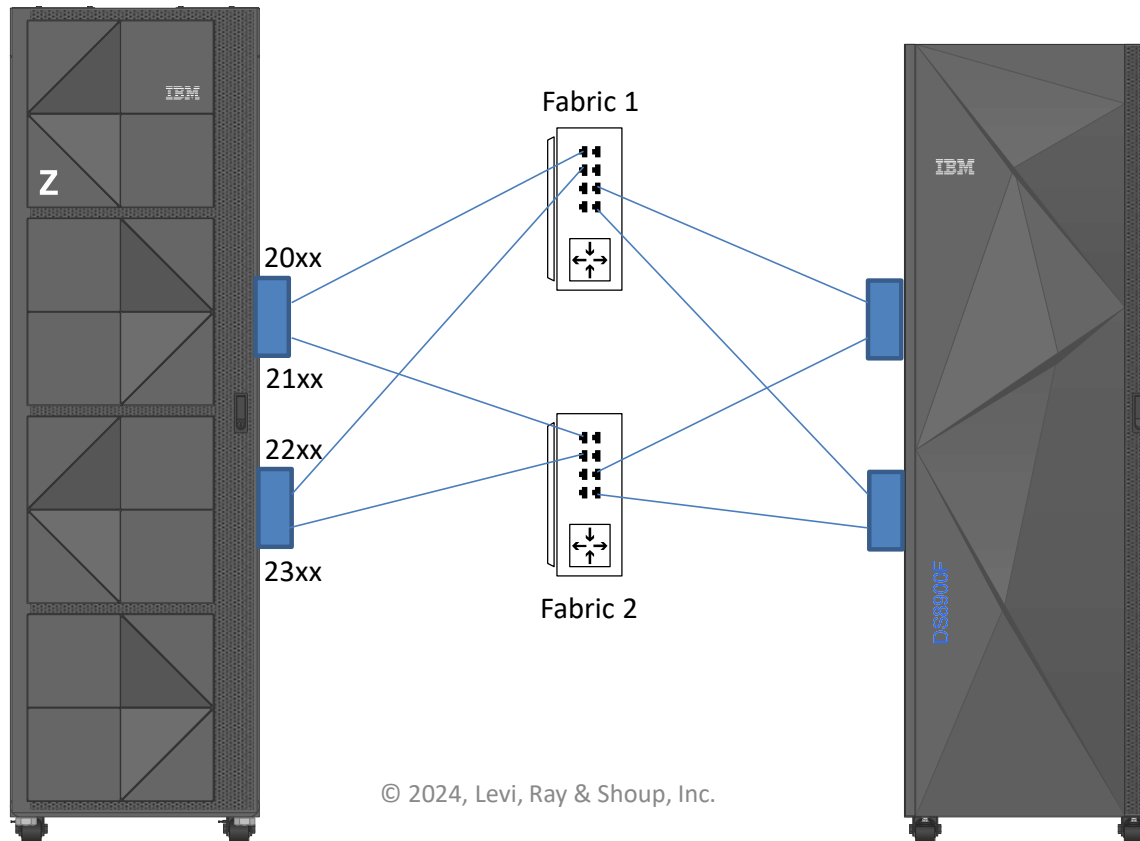


# Examples from a real system

zSystems

SAN Routers

Storage





## Examples from a real system

- IOCCDS source for an FCP channel

```
CHPID20  CHPID PATH=(CSS(0),20),TYPE=FCP,PART=((PROD,TEST)),PCHID=100
CU6600   CNTLUNIT CUNUMBR=2000,PATH=20,UNIT=FCP
DEV2000  IODEVICE ADDRESS=(2000,64),CUNUMBR=(2000),UNIT=FCP
```

- z/VM WWPNN Displays with NPIV active (2 LPARs)

- LPAR 1

```
q fcp wwpn 2000
FCP 2000      NPIV WWPNN C05076D691800380
      CHPID 20  PERM WWPNN C05076D691801141
      ATTACHED TO LNXUTILS
```

- LPAR 2

```
q fcp wwpn 2000
FCP 2000      NPIV WWPNN C05076D691800400
      CHPID 20  PERM WWPNN C05076D691801141
      FREE
```





# Examples from a real system

- SAN Fabric Definitions
  - Aliases: Giving Names to WWPNs

Zone Configurations   Zones   **Zone Aliases**   Preferences   z\_FCP2000

Name

2 Items   Members

<input type="checkbox"/>	Members	Type	Vendor	
<input type="checkbox"/>	c0:50:76:d6:91:80:03:80	WWN	-	▼
<input type="checkbox"/>	c0:50:76:d6:91:80:04:00	WWN	-	▼



# Examples from a real system

- SAN Fabric Definitions
  - Aliases: Giving Names to WWPNs

Zone Configurations   Zones   **Zone Aliases**   Preferences   z\_FCP2000

Name

2 Items   Members

<input type="checkbox"/>	Members	Type	Vendor	
<input type="checkbox"/>	c0:50:76:d6:91:80:03:80	WWN	-	▼
<input type="checkbox"/>	c0:50:76:d6:91:80:04:00	WWN	-	▼

LPAR 1 →



# Examples from a real system

- SAN Fabric Definitions
  - Aliases: Giving Names to WWPNs

Zone Configurations   Zones   **Zone Aliases**   Preferences   z\_FCP2000

Name

2 Items   Members

<input type="checkbox"/>	Members	Type	Vendor	
<input type="checkbox"/>	c0:50:76:d6:91:80:03:80	WWN	-	▼
<input type="checkbox"/>	c0:50:76:d6:91:80:04:00	WWN	-	▼

LPAR 1 →  
LPAR 2 →



# Examples from a real system

- SAN Fabric Definitions
  - Zones: Linking Aliases to create a path

Zone Configurations   ZONES   Zone Aliases   Preferences   z\_FCP00\_FS5030f

Name

Type Standard

4 Items   Members

<input type="checkbox"/>	Members	Type	
<input type="checkbox"/>	FS5030f_node1_p1_NPIV	ALIAS	▼
<input type="checkbox"/>	FS5030f_node2_p1_NPIV	ALIAS	▼
<input type="checkbox"/>	z_FCP2000	ALIAS	▼
<input type="checkbox"/>	z_FCP2200	ALIAS	▼



# Examples from a real system

- SAN Fabric Definitions
  - Zones: Linking Aliases to create a path

Zone Configurations   Zones   Zone Aliases   Preferences   z\_FCP00\_FS5030f

Name

Type Standard

Q 4 Items Members

<input type="checkbox"/>	Members	Type	
<input type="checkbox"/>	FS5030f_node1_p1_NPIV	ALIAS	▼
<input type="checkbox"/>	FS5030f_node2_p1_NPIV	ALIAS	▼
<input type="checkbox"/>	z_FCP2000	ALIAS	▼
<input type="checkbox"/>	z_FCP2200	ALIAS	▼

Previously Defined →

Add Remove

Save Delete Cancel



# Examples from a real system

- SAN Fabric Definitions
  - Zone Configuration: Set of Zones

Zone Configurations   Zones   Zone Aliases   Preferences   BPIC

Name

62 Items   Members

<input type="checkbox"/> Name ^	Type ^	Member Count ^	
<input type="checkbox"/> z_FCP0D_DEMOVM	Standard	4	▼
<input type="checkbox"/> z_FCP0E_DEMOVM	Standard	4	▼
<input type="checkbox"/> z_FCP0F_DEMOVM	Standard	4	▼
<input type="checkbox"/> z_FCP00_FS5030f	Standard	4	▼

Effective



# Examples from a real system

- Storage Subsystem
  - Host: Defining who can connect

The screenshot displays the IBM FlashSystem 5000 management console. At the top, the breadcrumb navigation shows 'Hosts' and the current host is 'z\_Channel\_00'. The main table lists host details:

Name	Status	Host Type	# of Ports	Host Mappings	Host Cluster ID	Host Cluster Name
z_Channel_00	Online	Generic	8	Yes		

Below the main table, three detailed views are shown:

- Host Details: z\_Channel\_00 (Port Definitions):** A table listing 8 FC (SCSI) ports with their IDs, types, and statuses (Active or Offline).
- Host Details: z\_Channel\_00 (Mapped Volumes):** A table listing 7 volumes (z\_Volume\_00 to z\_Volume\_06) with their SCST IDs, names, UIDs, and caching/I/O settings.
- Properties for Volume (z\_Volume\_00):** A detailed view of the first volume, showing its name, ID, state (Online), capacity (10.00 GiB), and various performance and security settings.



# Questions?







**So, how do I use this?**



# Booting an Operating System First Level

▼ Load - P00298E8:ZICP ⓘ

CPC: P00298E8  
Image: ZICP

Load type  
 Standard load  
 SCSI load  
 SCSI dump

Clear the main memory on this partition before loading it

Store status

Load address + 02000

Load parameter

Time-out value 60 60 to 600 seconds

Worldwide port name 23456789ABCDEF

Logical unit number 0010000000000000

Boot program selector 0

Boot record logical block address 00000000000000C8

Operating system specific load parameters cons=SYSG

OK Reset Cancel Help



## z/VM Use of FB Disks

- Emulated FBA (EFBA)
  - Define a “dummy” FBA address linked to an FCP channel+WWPN+LUN
  - Can define multiple FCP channel+WWPN+LUN combinations
  - z/VM then does multipathing, but only if initial channel is busy

Example:

```
SET EDEVICE 3000 TYPE FBA ATTR FLASH ,  
    FCP_DEVICE 2001 WWPN 0123456789ABCDEF LUN  
0001000000000000 ,  
    FCP_DEVICE 2101 WWPN 0123456789ABCDF0 ,  
    FCP_DEVICE 2201 WWPN 0123456789ABCDEF ,  
    FCP_DEVICE 2301 WWPN 0123456789ABCDF0
```



## z/VM: LGR Support

- If you will be attaching FCP subchannels to a guest that could be relocated to another z/VM system, define EQIDs for each subchannel and use them for attaching FCP to the guest:
- Assuming 4 FCP channels start at 2000, 2100, 2200 and 2300:
  - In SYSTEM CONFIG:  
Rdevice 2000 EQID FCP00 Type FCP  
Rdevice 2100 EQID FCP00 Type FCP  
Rdevice 2200 EQID FCP00 Type FCP  
Rdevice 2300 EQID FCP00 Type FCP
  - In VM Directory for a guest:  
COMMAND ATTACH EQID FCP00 TO \* AS 2000  
COMMAND ATTACH EQID FCP00 TO \* AS 2100  
COMMAND ATTACH EQID FCP00 TO \* AS 2200  
COMMAND ATTACH EQID FCP00 TO \* AS 2300



# z/VSE and VSE<sup>n</sup> use of FB disks

- SCSI Definitions in ASIPROC
  - Can have multipathing defined
    - Only used if initial path is busy
  - Limited to LUN size of approx. 24GB
- FBA Definitions in ASIPROC
  - Multipathing done at the z/VM Level
  - Standard 9336 processing
  - Limited to 2GB LUN



# Linux use of FB disks

- Enable Multipath Daemon and FCP Devices
  - SLES:
    - Use YaST to configure devices during initial installation
  - RedHat:
    - Run `/sbin/mpathconf` to create multipath config, then enable `multipathd`
    - Define FCP device addresses, WWPNs and LUNs in `/etc/zfcp.conf`
      - May need to run `cio_ignore -r FCP_addresses` to let FCP channels come online
- Multipath Daemon may use round-robin for I/O distribution, but depends on `multipath.conf` settings
  - Defaults are usually sufficient



# Who Should Multipath? z/VM or Guest?

## z/VM Multipathing

- Guest doesn't change if storage hardware changes
- Multipathing means more CP processing

## Guest Multipathing

- Each guest must change its SCSI definitions if storage hardware changes
- Multipathing means more guest processing