VM-Linux on System z Storage Technologies

Steve Wilkins

March 2009

IBM Systems

Agenda

- DASD Overview
- PAV Support
- SCSI Disk Support
- DS8000 Support
- Tape Overview
- Tape Encryption Support
- Device Support Statements
- Dynamic I/O Support

IBM Systems

The Presenter

Steve Wilkins

(607)429-4058

wilkinss@us.ibm.com

http://www.vm.ibm.com/devpages/WILKINSS/

IBM Systems

DASD Overview

IBM Systems

DASD Support

- Main device for VM system and guest data

- Can be dedicated to guest operating system with CP ATTACH command or DEDICATE user directory statement

- Can be attached to system with CP ATTACH rdev TO SYSTEM command -or- automatically attached at IPL with system configuration file statements (e.g., SYSTEM_RESIDENCE, CP_OWNED, or USER_VOLUME_LIST)

- System attached DASD provides:
  - ► Checkpoint data
  - ► Warmstart data
  - ► Paging and Spooling data
  - ► Guest minidisks (MDISK User Directory Statement)
  - ► Guest temp disks (CP DEFINE T3390 Command)

IBM Systems

IBM System Storage Disk Support
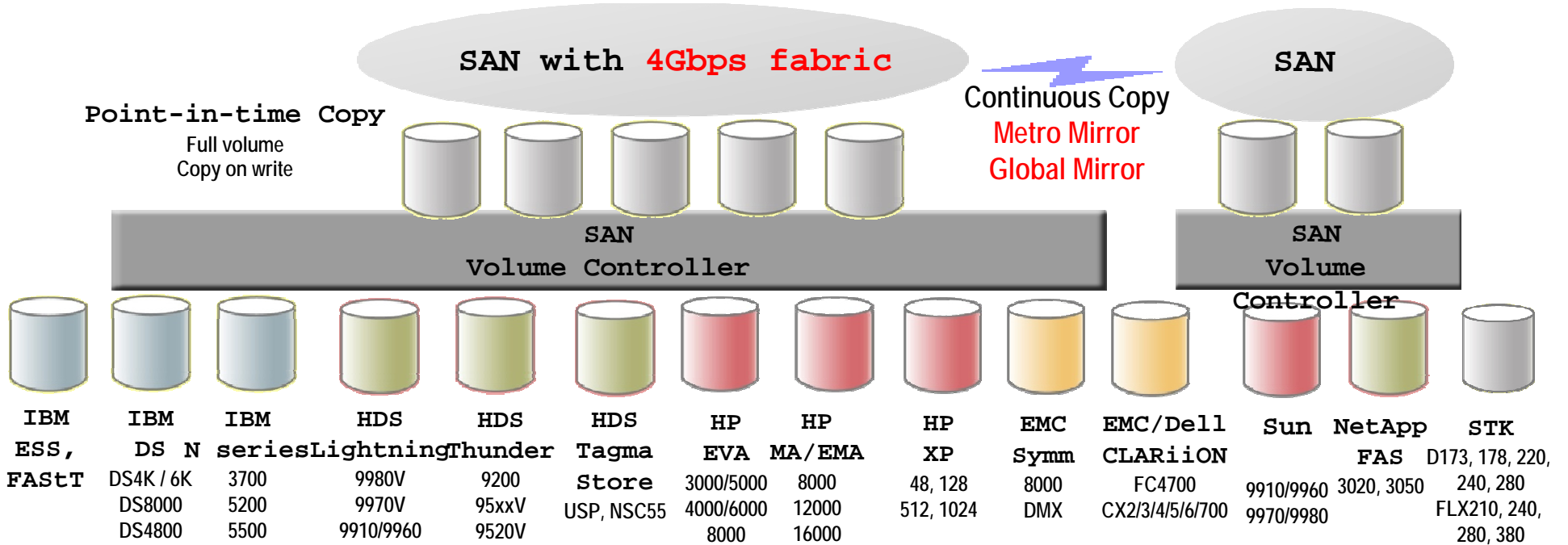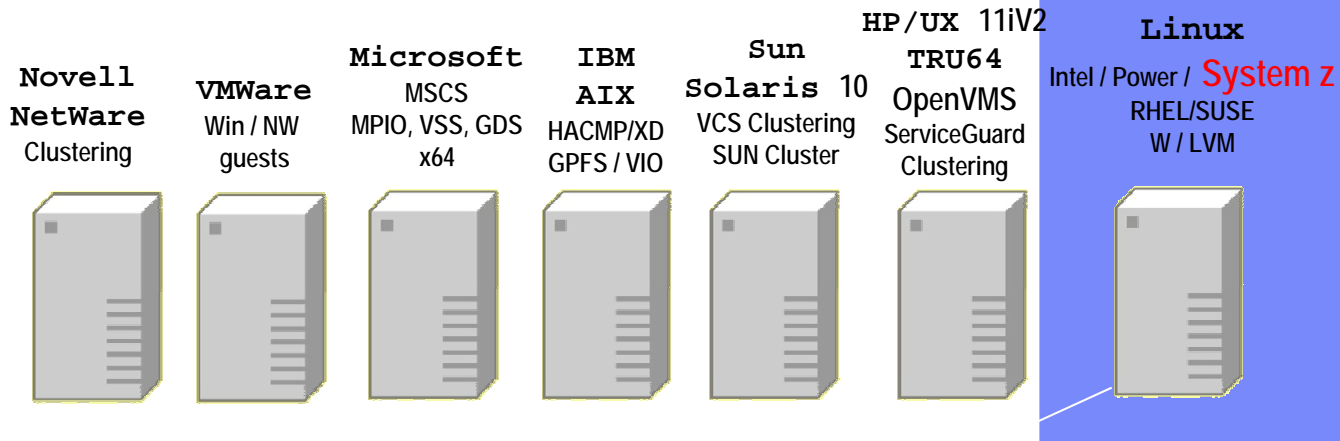
- **Models supported (for direct attachment)**
    - IBM Enterprise Storage Server (ESS) 750 & 800
    - IBM System Storage DS6000 Series
    - IBM System Storage DS8000 Series

- **Supported connections:**
    - ESCON (excluding DS6000)
    - FICON
    - FCP

- **ESCON/FICON supported as 3990 Model 3 or 6 Controller with:**
    - 3390 Model 2, 3, 9, 27 and 54 DASD (Model A support pending)
    - 3390 Model 2 and 3 DASD in 3380 track compatibility mode
    - Native controller modes supported: 2105 CU for ESS models, 1750 for DS6000, and 2107 for DS8000

- **FCP supported as emulated 9336 model 20 DASD**

IBM Systems

SAN Volume Controller (SVC) Support

- SVC allows attachment of many low cost SCSI disks to System z including the DS4000 and XIV
  - Presents devices as generic SCSI, multi-path disks
  - Device dependent drivers are hidden in the SVC hardware
  - SVC device type is 2145

- Minimum software levels:
  - SVC 4.1.0
  - Linux SLES 9 SP3
  - Linux RHEL 04
  - z/VM 5.3.0

- Complete "hardware supported" list:

`http://www-03.ibm.com/systems/storage/software/virtualization/svc/interop.html`

IBM Systems

SAN Volume Controller Supported Environments

Novell
NetWare
Clustering

VMWare
Win / NW
guests

Microsoft
MSCS
MPIO, VSS, GDS
x64

IBM
AIX
HACMP/XD
GPFS / VIO

Sun
Solaris 10
VCS Clustering
SUN Cluster

HP/UX 11iV2
TRU64
OpenVMS
ServiceGuard
Clustering

Linux
Intel / Power / System z
RHEL/SUSE
W / LVM

SAN with 4Gbps fabric

SAN

Point-in-time Copy
Full volume
Copy on write

Continuous Copy
Metro Mirror
Global Mirror

SAN
Volume Controller

SAN
Volume
Controller

| IBM ESS, FAStT	IBM DS N	IBM series	HDS Lightning	HDS Thunder	HDS Tagma Store	HP EVA	HP MA/EMA	HP XP	EMC Symm	EMC/Dell CLARiiON	Sun FAS	NetApp	STK
	DS4K / 6K	3700	9980V	9200	USP, NSC55	3000/5000	8000	48, 128	8000	FC4700	9910/9960	3020, 3050	D173, 178, 220, 240, 280
	DS8000	5200	9970V	95xxV		4000/6000	12000	512, 1024	DMX	CX2/3/4/5/6/700	9970/9980		FLX210, 240, 280, 380
	DS4800	5500	9910/9960	9520V		8000	16000						

IBM Systems

FlashCopy

- Instant copy of source disk to target
- Guest support via CCW channel program support
  - Restricted to dedicated or full-pack minidisks
- Native support via CP FLASHCOPY command
- FlashCopy Version 1 (v1) hardware
  - No concurrent copies (one at a time per source/target pair)
  - No extent relocation (full volume only)
- FlashCopy Version 2 (v2) hardware
  - Dataset Relocation (data on different track offsets when moved to target)
  - Multiple Relationships (source track can have many target tracks, but target track can have only one source track)
  - Cross LSS (source and target can be in different logical control units)

source → instant copy → target

IBM Systems
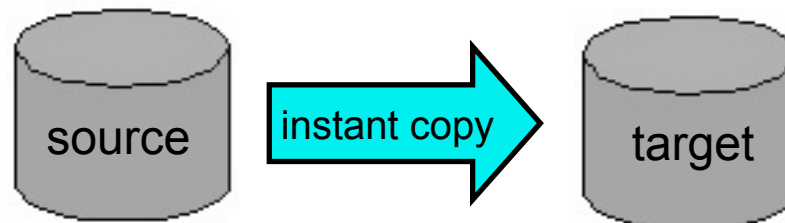
Peer-to-Peer Remote Copy (PPRC)

- Continuous copy of source disk changes to target
- Guest support via CCW channel program support
  - Restricted to dedicated or fullpack minidisks
- Native support via ICKDSF Release 17 running in CMS virtual machine
- PPRC hardware
  - **Metro Mirror** (changes from source to target synchronous to initiating I/O)
  - **Global Copy** (changes from source to target asynchronous to initiating I/O; allows extended distances between source and target)
  - **Metro/Global Copy** (3 volume PPRC link; synchronous link between source and intermediate volume; asynchronous link between intermediate and target volume; allows extended distances with added data integrity)
  - **Global Mirror** (asynchronous link between source and target; software managed consistency groups allow failover to point-in-time consistent copy in case of disaster)
- CP QUERY DASD DETAILS command displays PPRC volume status (Primary, Secondary, or Cascading Intermediate) when a PPRC link is active for a disk

source → Continuous copy → target
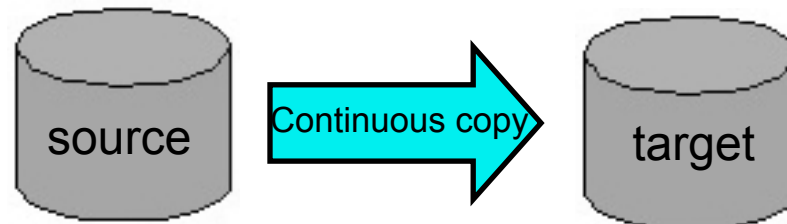
MiniDisk Cache (MDC)

- MiniDisk Cache is a host based, full track, write through cache for partial pack and full pack minidisks (linked by guest virtual machines)

- When one record of a track is read, the entire track is brought into the cache

- Records are kept in the cache based on *last reference*. Records not referenced are aged out of the cache.

- A *Fair Share* insertion algorithm prevents one user from monopolizing the cache
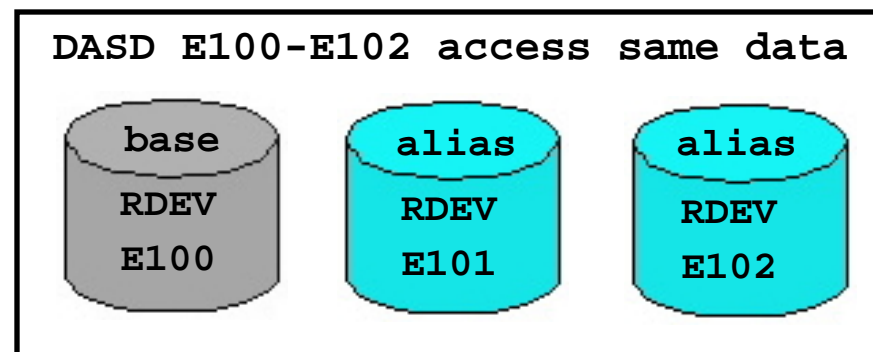
- MDC is active by default

IBM Systems

Virtual Disk in Storage (VDISK) Support

- A VDISK is a fully simulated FBA 9336 model 20 DASD. All data resides in main memory.

- Since the VDISK data is in volatile storage, it should be used for applications that make use of temporary files (e.g., assembly, VSE lock file, or sort applications).

- Any guest OS that supports FBA DASD can use VDISKs

- A VDISK can be created via the CP DEFINE VFB-512 command or by the MDISK user directory statement. VDISKs created by the MDISK statement are shareable by multiple guests.

- CP SET VDISK command or FEATURES VDISK system configuration file statement can be used to control the amount of memory used for VDISKs

IBM Systems

PAV Support

IBM Systems

Classic PAV Overview

- z/VM provides support for the Parallel Access Volumes (PAV) feature of IBM System Storage subsystems.

- With PAV, a real DASD volume is accessed through a Base subchannel (device) and one or more Alias subchannels

  ► Volume (represented by Base) shadowed by 1 or more Aliases

  ► Looks like multiple separate, real DASD to host operating system

```
DASD E100-E102 access same data

   base          alias          alias

   RDEV          RDEV           RDEV

   E100          E101           E102
```

IBM Systems

Classic PAV Overview …

- z/Architecture allows only 1 active I/O to a single ECKD DASD

- Aliases overcome this restriction providing the ability to have multiple concurrent I/O operations on a DASD

- Allows higher I/O throughput by reducing I/O queuing

- Control unit provides data serialization

- Each I/O request specifies cylinder range:

  - Controller provides shared access for read cylinder ranges
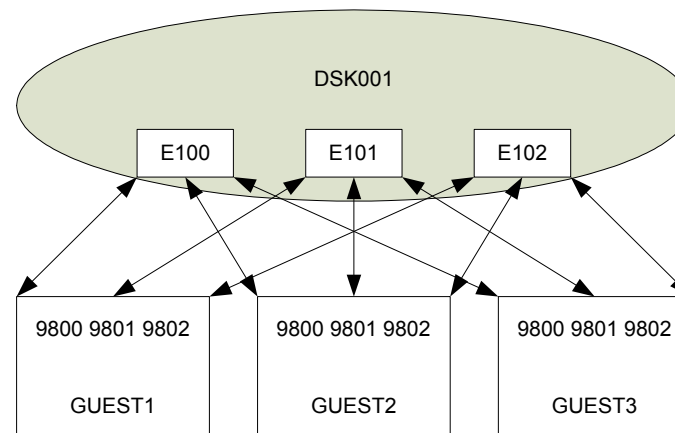 - Controller provides exclusive access for write cylinder ranges

IBM Systems

VM Dedicated DASD Configuration

- Configured with CP ATTACH command or DEDICATE user directory statement

- For guests that exploit the PAV architecture

DSK001

| E100 | E101 | E102 |

9800 9801 9802

GUEST1

IBM Systems
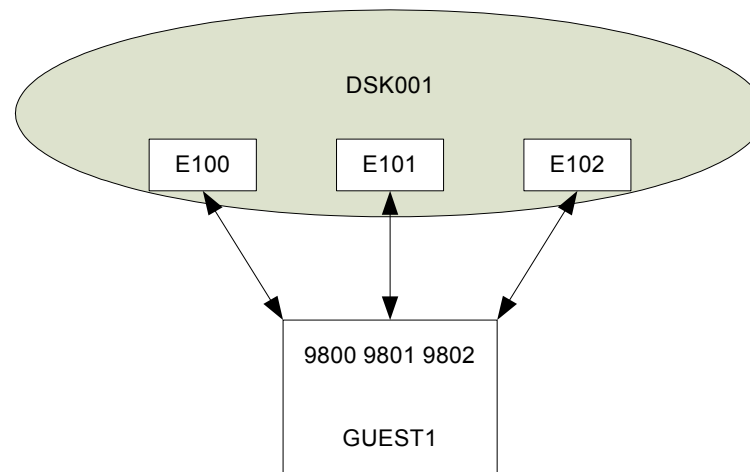
VM Minidisk Configuration

- **VM provides linkable minidisks for guests that exploit PAV (e.g., z/OS and Linux)**
  - ► Base minidisks are defined with the MDISK or LINK user directory statements (LINK command also supported)
  - ► Aliases are defined with PAVALIAS parameter of the DASDOPT and MINIOPT user directory statements or with the CP DEFINE PAVALIAS command

- **VM also provides workload balancing for guests that don't exploit PAV (e.g., CMS)**
  - ► Real I/O dispatcher queues minidisk I/O across system attached Aliases
  - ► Minidisks are defined without PAVALIAS parameter

DSK001

| E100 | E101 | E102 |

| 9800 9801 9802 | 9800 9801 9802 | 9800 9801 9802 |
| GUEST1 | GUEST2 | GUEST3 |

IBM Systems

DEFINE PAVALIAS

Privilege class G

```
>>-DEFINE—PAValias--vdev--.------.--BASE--basevdev--><
                         '-FOR-'
```

- The DEFINE PAVALIAS command is used to create virtual PAV Alias minidisks. Function can also be accomplished by using the DASDOPT and MINIOPT user directory statements.

- Newly defined virtual Alias is automatically assigned to a unique underlying real PAV Alias.

- The command will fail if no more unique real Aliases are available to be associated with the virtual Alias (per guest virtual machine).

IBM Systems

Setup example for Linux exploiting PAV minidisks

```
Base device predefined in user directory:
MDISK 200 3390 DEVNO E100 WR


q pav
Device E100 is a base Parallel Access Volume with the following
aliases: E101
Device E101 is an alias Parallel Access Volume device whose base
device is E100


attach E100 to system
DASD E100 ATTACHED TO SYSTEM WIL6 PAV BASE
attach E101 to system
DASD E101 ATTACHED TO SYSTEM WIL6 PAV ALIAS


define pavalias 201 for base 200
DASD 201 DEFINED


query virtual pav all
PAV BASE  0200 ON E100 WIL6 ASSIGNED E100
PAV ALIAS 0201 ON E101 WIL6 ASSIGNED E101 FOR BASE 0200
```

Configure Linux LVM to use virtual PAV Base 200 and Alias 201 as a single logical volume. For details, see Linux "How to Improve Performance with PAV" whitepaper at:
http://www-128.ibm.com/developerworks/linux/linux390/june2003_documentation.html
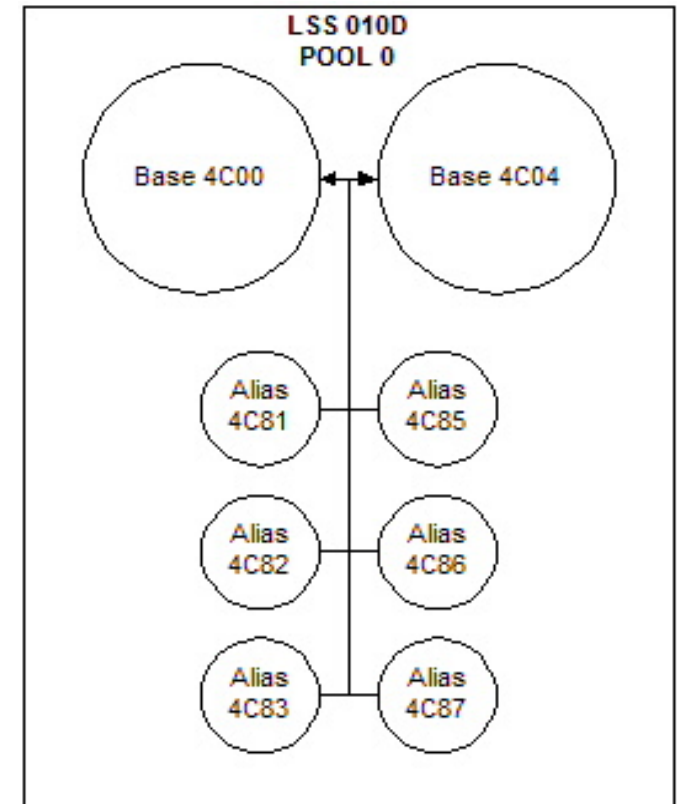
IBM Systems

Dynamic PAV

- Dynamic PAV is the ability to re-associate an Alias device from one Base to another

- Guest issued dynamic PAV operation to a dedicated Alias:
  - ► Real (and virtual) Alias to Base association will change as long as the new Base is dedicated to the same guest.   Otherwise, the dynamic PAV operation will fail.

- Guest issued dynamic PAV operation to an Alias minidisk:
  - ► Only the virtual configuration is altered if new virtual Base is the only minidisk on the underlying real Base and there is a unique real Alias available in which to associate the virtual Alias (per guest machine).   Otherwise, the Dynamic PAV operation fails.
  - ► The real Alias to Base association never changes for minidisks.

- Out-board (control unit) initiated dynamic PAV operations:
  - ► All Alias minidisks associated with a real system attached Alias will be detached from their guests.
  - ► A dedicated Alias will behave as if guest issued the dynamic PAV operation.

- Supported only for Full-pack minidisks or Dedicated DASD

IBM Systems

HYPERPAV Overview

- Removes the static Alias to Base binding associated with traditional PAVs

- Alias and Base volumes are *pooled* per each LSS. An Alias can be associated with any Base in the Pool; done by host OS on each I/O request.

- Makes traditional Dynamic PAV obsolete

- z/VM supports for Full-pack Minidisks and Dedicated DASD

- Not supported directly by Linux

IBM Systems

HYPERPAV Overview …

- **VM dedicated DASD support via CP ATTACH command or DEDICATE user directory statement**

- **VM Minidisk Support:**
  - workload balancing for guest's that *don't* exploit HyperPAV
  - linkable full-pack minidisks for guests that *do* exploit HyperPAV
  - New CP DEFINE HYPERPAVALIAS command creates HyperPAV Alias minidisks for exploiting guests
  - z/VM and z/OS are current exploiters of HyperPAV
  - Restricted to *full-pack* minidisks for exploiting guests

LSS 010D
POOL 0

Base 4C00   Base 4C04

Alias 4C81   Alias 4C85

Alias 4C82   Alias 4C86

Alias 4C83   Alias 4C87

IBM Systems

DEFINE HYPERPAVALIAS

Privilege class G

```
>>-DEFINE—HYPERPAValias--vdev----.-----.--BASE--basevdev--><
                                  '-FOR-'
```

- The DEFINE HYPERPAVALIAS command is used to create virtual HyperPAV Alias minidisks.

- A newly defined virtual Alias is automatically assigned to a unique underlying real HyperPAV Alias (in the same real hardware Pool as the Base).

- The command will fail if no more unique, real Aliases are available in the real hardware Pool to be associated with the virtual Alias (per guest virtual machine).

- There can only be 254 Aliases per Pool; and a limit of 16,000 Pools per image.

- Restricted to Full-Pack minidisks.

- Put in User Directory with COMMAND directory statement

 IBM Systems

SET CU

Privilege class B

```
              .-DASD--.                                      <----------'
>>-SET-CU-'-------'-.-HYPERPAV_allowed-.--.-ssid-------.--><
                    |-PAV_allowed------|   '-ssid-ssid--'
                    '-NOPAV_allowed----'
```

- The SET CU command is used to set the Parallel Access Volume function level of each applicable control unit (specified via controller's SSID).

- Default is either HYPERPAV_allowed or PAV_allowed depending on the installed capabilities of each control unit.

- HYPERPAV_allowed can't be set if capability is not available on the control unit.

- All Alias devices in the specified control unit (SSID) must be off-line when changing from *or* to the HYPERPAV_allowed setting.

- Command applies to only first-level VM images; error occurs otherwise.

- QUERY CU command displays the PAV and HYPERPAV capabilities of applicable DASD control units.

IBM Systems

Web Documentation

General z/VM setup and configuration info:

http://www.vm.ibm.com/storman/pav/

Linux LVM setup info:

http://www-128.ibm.com/developerworks/linux/linux390/june2003_documentation.html

IBM Systems

SCSI Disk Support

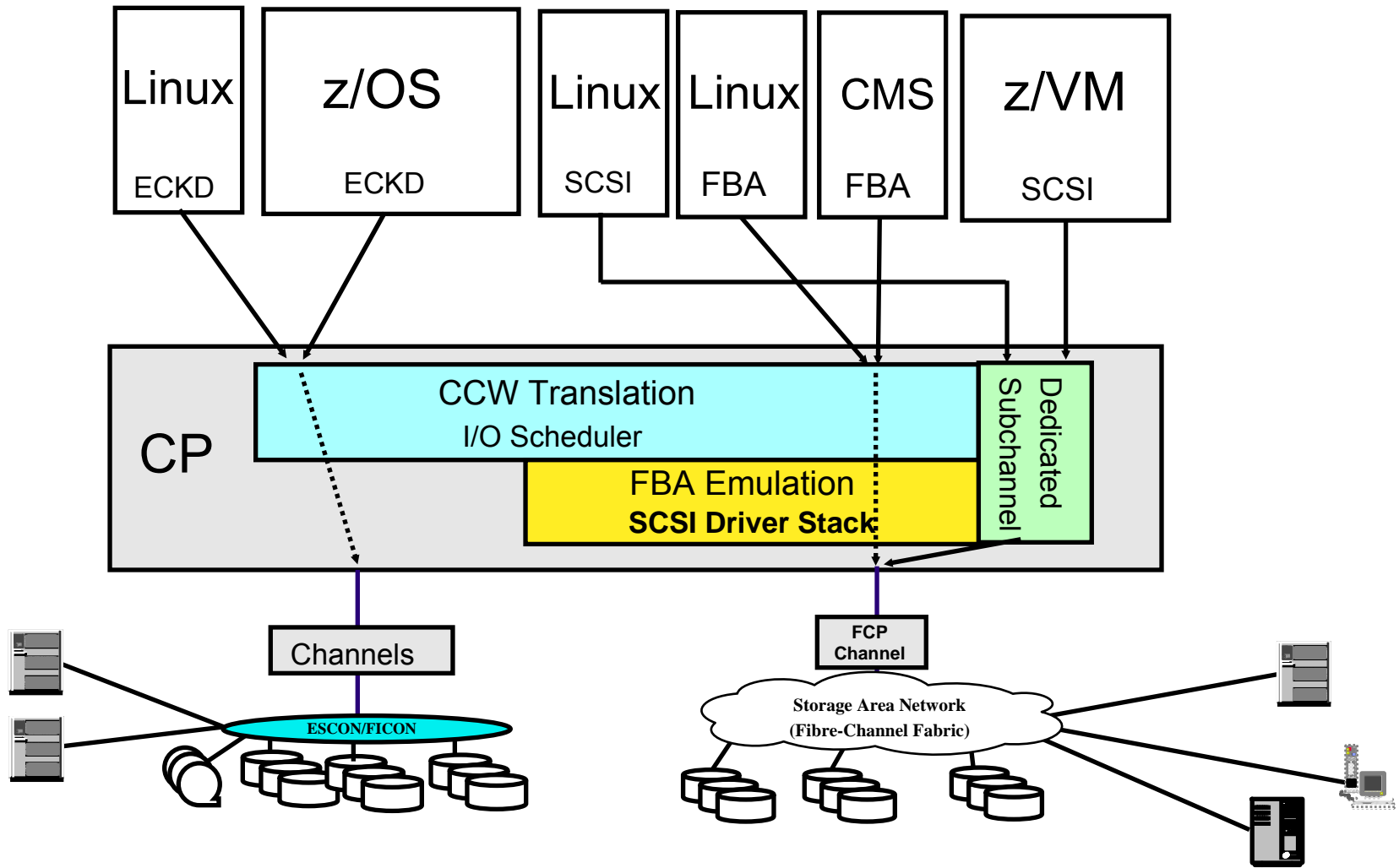© 2009 IBM Corporation

IBM Systems

Overview

- VM provides native support for SCSI disks for paging, spooling, and other system data

- Support is provided by configuring SCSI disk LUNs to VM as 9336 FBA 512-byte block DASD

- VM guests (such as CMS, GCS, and VSE) may keep data on SCSI disk LUNs without requiring the guest software to have SCSI disk drivers

- IPL, Dump, and Service from/to SCSI disk LUNs is provided to achieve a SCSI-only VM environment

- SCSI-only as well as mixed SCSI and ECKD environments are supported
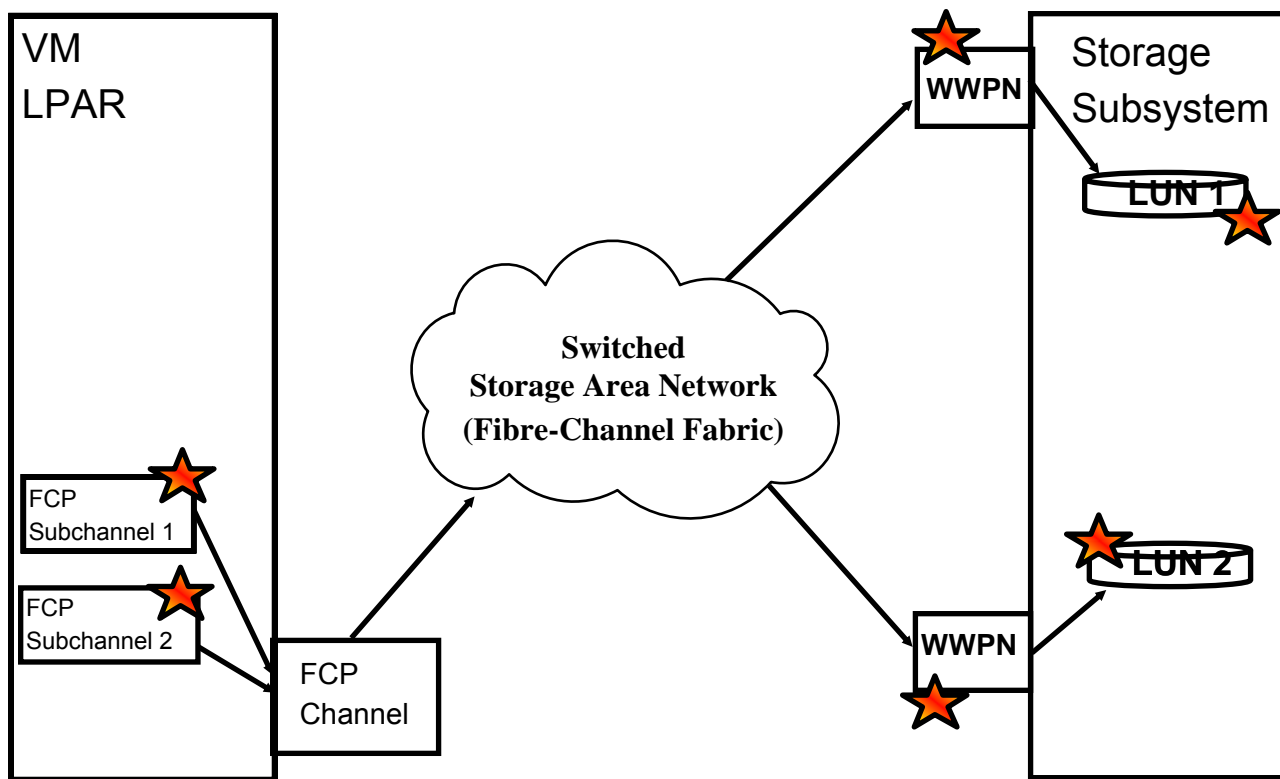
IBM Systems

FBA Emulation

- SCSI Disks are emulated as 9336 Model 20 FBA DASD

- FBA Emulation is used to reduce effort in supporting SCSI disks as well as allowing any operating system or application that supports a 9336 to utilize SCSI disks without change

- Emulated 9336 disk sizes supported are:
  - **1TB for CP** with the exception that PAGE, SPOL, and DRCT allocations must remain below the 64GB mark on a CP formatted volume since internal addressing of these slots is limited to 2^{24} 4K pages.
  - **381GB for CMS/GCS** including software functions dependent on CMS functions such as DIRMAINT MOVE, COPY, ERASE and DFSMS MOVE, COPY, CHECK.

- VM officially supports IBM System Storage 1750 (DS6000), 2105 (ESS), and 2107 (DS8000) SCSI disks as emulated 9336 DASD. However, other SCSI disks may also work since a generic SCSI driver is provided in addition to the IBM drivers.

IBM Systems

System Structure

IBM Systems

Configuration Elements

VM
LPAR

FCP
Subchannel 1

FCP
Subchannel 2

FCP
Channel

**Switched
Storage Area Network
(Fibre-Channel Fabric)**

WWPN

Storage
Subsystem

LUN 1

LUN 2

WWPN

⭐ VM configuration element

- SCSI disk is referred to as a LUN
- VM addresses a LUN via a path made up of 3 components
- A FCP subchannel (specified in IOCP), target WWPN on the disk controller, and a LUN name make up the 3 components
- VM allows 8 unique paths per LUN, each with the 3 components
- Figure shows 2 LUNs, each with a single path

IBM Systems

Commands

- SET EDEVICE command configures a SCSI disk LUN to the VM system as an emulated 9336 FBA DASD

- QUERY EDEVICE command obtains information related to SCSI disk LUNs defined to VM as emulated FBA DASD

- DELETE EDEVICE command removes a SCSI disk LUN from the VM system that has been previously defined as an emulated FBA DASD

IBM Systems

SET EDEVICE Command

Privilege Class: B

```
>>-SET--EDEVice--rdev----------------------------------------------->
 >-.-TYpe--FBA--ATTRibutes--.-1750-.---.-------------.-| Paths |-.-><
   |                          |-2105--|   |-ADD PATH----|          |
   |                          |-2107--|   '-DELete PATH-'          |
   |                          '-SCSI--'                            |
   '-CLEAR------------------------------------------------------'


Paths:

 <------------------------------------------------------------------<
|--FCP_DEV--nnnn--WWPN--wwwwwwwwwwwwwwww--LUN--llllllllllllllll-->
 >---------------------.-------------.-----------------------|
                       |-PREFerred----|
                       '-NOTPREFerred-'

Notes: You can specify a maximum of 8 "paths" to the device.
       PREF/NOTPREF only valid for 1750.
```

IBM Systems

QUERY EDEVICE Command

Privilege Class: B

```
                       <---------<
>>-Query-EDEVice-.-.-rdev--------.-.---------.-.->< 
                | '-rdev1-rdev2-' '-DETAILS-' |
                '-ALL--------------------------'
```

IBM Systems

QUERY EDEVICE Responses

q edev 607
EDEV 0607 TYPE FBA ATTRIBUTES 2105
Ready;

q edev 608 details
EDEV 0608 TYPE FBA ATTRIBUTES 2107
  PATHS:
    FCP_DEV: 8100 WWPN: 5005076300C604DA LUN: 5137000000000000
    FCP_DEV: 8200 WWPN: 5005076300C604DB LUN: 5137000000000000
Ready;

q edev 609 details
EDEV 0609 TYPE FBA ATTRIBUTES 1750
  PATHS:
    FCP_DEV: B100 WWPN: 4004066300C2023B LUN: 0011000000000000 PREF
    FCP_DEV: B200 WWPN: 4004066300C2023C LUN: 0011000000000000 NOTPREF
  Ready;

IBM Systems

DELETE EDEVICE Command

Privilege Class: B

```
>>--DELete--EDEVice--.-rdev--------.--><
                     '-rdev1-rdev2-'
```

Note: An EDEVICE must be varied off-line to use the DELETE command.

IBM Systems

Configuration File Statement

EDEVICE SYSTEM CONFIG file statement performing same function as SET EDEVICE command

```
>>-EDEVice--rdev---TYpe--FBA--ATTRibutes--.-1750--.--| Paths |--><
                                          |-2105--|
                                          |-2107--|
                                          '-SCSI--'
Paths:
 <------------------------------------------------------------<
|--FCP_DEV--nnnn--WWPN--wwwwwwwwwwwwwwww--LUN--llllllllllllllll-->
 >--------------------.--------------.----------------------|
                      |-PREFerred----|
                      '-NOTPREFerred-'

Notes: You can specify a maximum of 8 "paths" to the device.
       PREF/NOTPREF only valid for 1750.
```
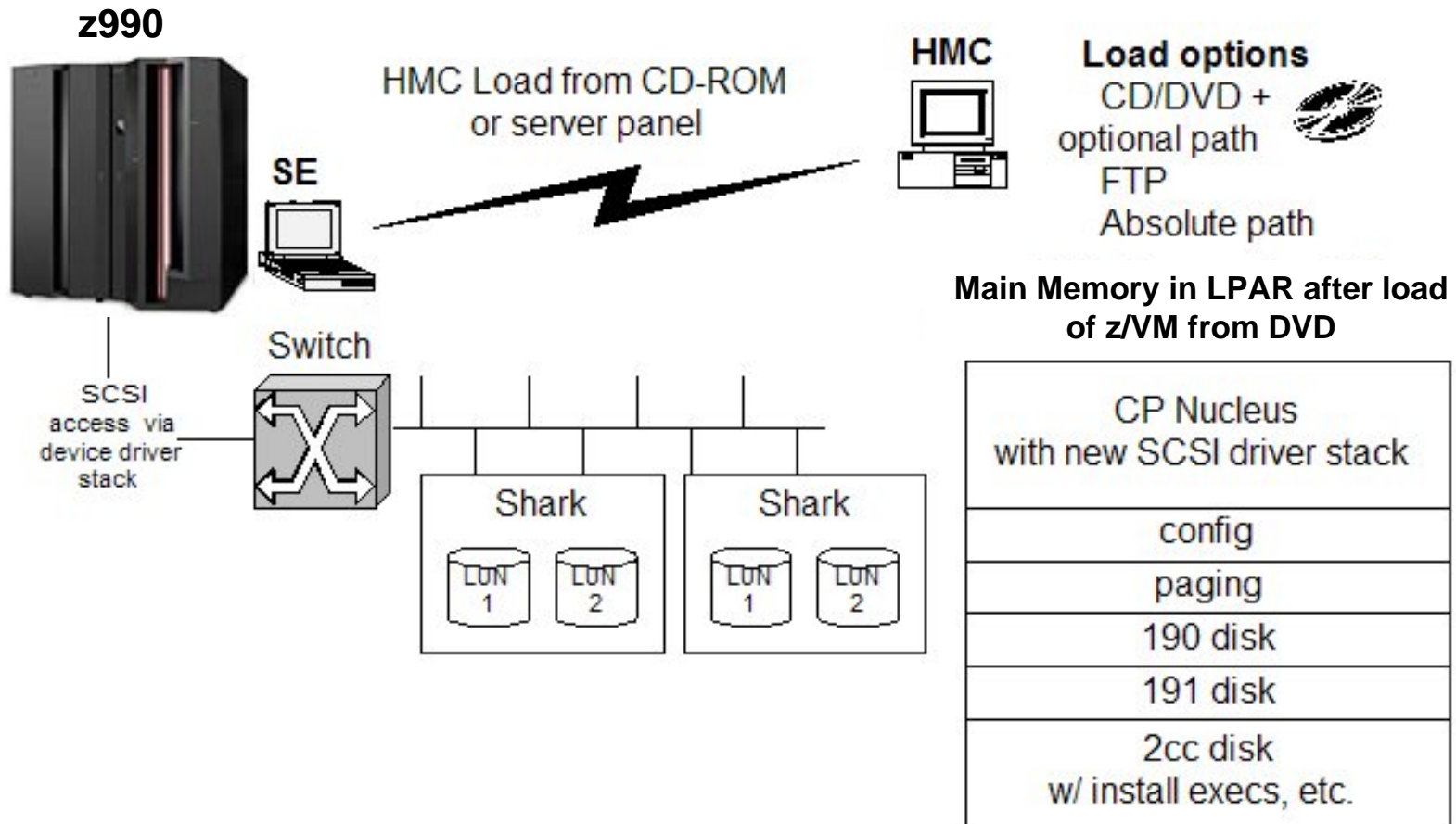
IBM Systems

Install

- Install to SCSI is done via DVD from the Hardware Management Console (HMC).

- HMC software with DVD Load and Integrated 3270 console support is required. A 512MB LPAR is also required.

- Install uses a special HMC hardware interface to bring in the VM starter system. The starter system uses a RAM disk to get itself and the install procedure started. Install writes the VM system and other files out to SCSI disk using FBA emulation (see next slide).

- Second-level install to SCSI is also supported. Requires FTP transfer of install files on the first-level system. Transfer is done automatically from the install DVD by the installation utility.

- Install from DVD is also available for 3390 models 3 and 9

IBM Systems

Install Illustration

z990

HMC Load from CD-ROM
or server panel

SE

HMC

Load options
CD/DVD +
optional path
FTP
Absolute path

SCSI
access via
device driver
stack

Switch

Shark

LUN 1   LUN 2

Shark

LUN 1   LUN 2

**Main Memory in LPAR after load
of z/VM from DVD**

CP Nucleus with new SCSI driver stack
config
paging
190 disk
191 disk
2cc disk w/ install execs, etc.

IBM Systems

IPL

- Once installed, VM can be IPLed from a SCSI LUN using the Load panel on the SE or HMC

- Stand-Alone Program Loader (SAPL) continues to be the VM IPL interface

- SAPL contains its own generic, stripped down SCSI stack so that it can read the VM load module from disk into memory. Once SAPL passes control to VM, FBA emulation is used to do all SCSI I/O.

- VM IPL parameter PDVOL must be specified for a SCSI IPL. It indicates the EDEVICE number of the SYSRES.
  - System stops with a wait state 6505 if not specified
  - Can be entered on the SAPL screen (PDVOL=xxxx) or defined when SAPL is set up ('burned in')

IBM Systems

IPL

```
Load
CPC: P000F12B
Image: ZFCP4
Load type:  ○ Normal   ○ Clear   ● SCSI   ○ SCSI dump
☐ Store status
Load address                    5C00
Load parameter                  SYSG
Time-out value                  060    60 to 600 seconds
World wide port name            5005076300800340
Logical unit number             5732000000000000
Boot program selector           0
Boot record logical block address  0000000000000C8
OS specific load parameters

OK   Reset   Cancel   Help
```

- Select SCSI radio button

- Load address is FCP subchannel

- Load parameter is console address for SAPL

- World wide port name (WWPN) and Logical unit number (LUN) address the target load volume

- Boot program selector and Boot record logical block address specify offset of IPL deck on load volume (For a CP formatted disk, use 0 and 0000000000000C8)

IBM Systems

Second-Level IPL

- Second-level IPL from a SCSI LUN works via existing support that went into z/VM 4.4.0 for Linux Guest IPL from SCSI

- Setup WWPN and LUN name of target load volume with CP SET LOADDEV command

- Setup Boot Program Selector and Boot Record Logical Block Address with SET LOADDEV command

- Use same Boot Program Selector and Boot Record Logical Block Address as first-level IPL for a CP formatted disk (i.e., 0 and 00000000000000C8). Add in start of RECOMP area for a CMS formatted disk.

- IPL virtual address of FCP subchannel to kick off the load

IBM Systems

Second-Level Example

```
att 50aa *
FCP 50AA ATTACHED TO STEVO 50AA
Ready; T=0.01/0.01 13:16:20


q v fcp
FCP  50AA ON FCP   50AA CHPID 40 SUBCHANNEL = 000E
50AA QDIO-ELIGIBLE      QIOASSIST-ELIGIBLE
Ready; T=0.01/0.01 13:16:24


set loaddev portname 50050763 00c20b8e lun 52410000 00000000
bootprog 0 br_lba 00000000 000000C8
Ready; T=0.01/0.01 13:16:33


q loaddev
PORTNAME 50050763 00C20B8E    LUN  52410000 00000000
BOOTPROG 0 BR_LBA   00000000 000000C8
Ready; T=0.01/0.01 13:16:38


IPL 50aa <Hit ENTER and the IPL starts...>
```
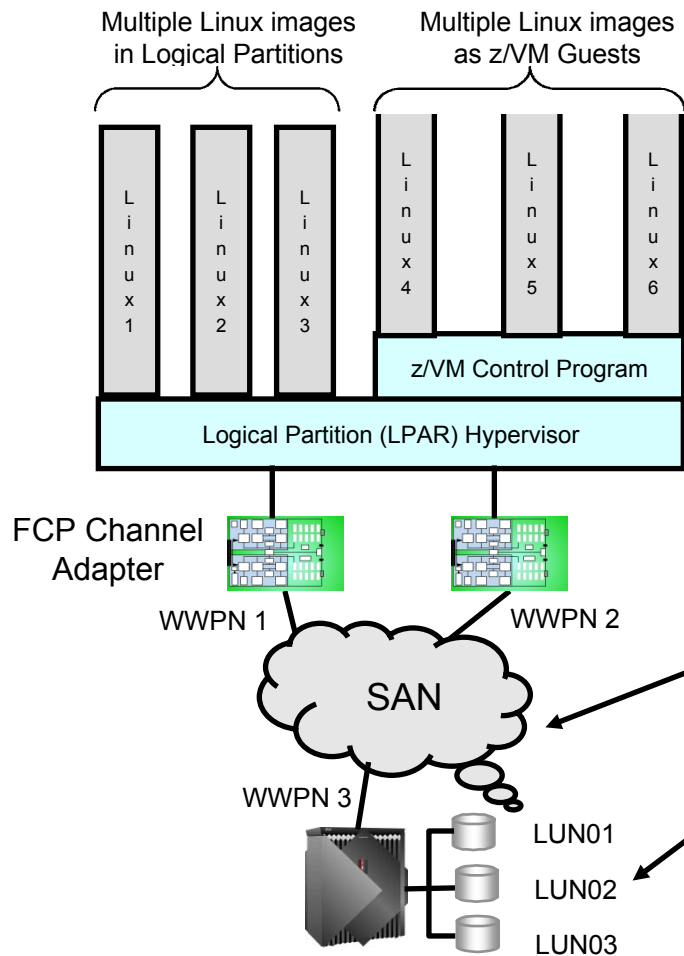
IBM Systems

Access Control

- Problem: Without Access Control
  - LUN Masking and Zoning only allow access control per WWPN of processor FCP Channel Adapter
  - No individual access control/protection for Operating Systems sharing the FCP Channel Adapter
- NPIV Solution:
  - Hardware creates unique virtual WWPNs for each FCP subchannel on Channel Adapter
  - Can use standard LUN masking and zoning

Multiple Linux images in Logical Partitions

Multiple Linux images as z/VM Guests

Linux 1 Linux 2 Linux 3

Linux 4 Linux 5 Linux 6

z/VM Control Program

Logical Partition (LPAR) Hypervisor

FCP Channel Adapter

WWPN 1

WWPN 2

SAN

WWPN 3

LUN01
LUN02
LUN03

With NPIV,
LUN Masking, and Zoning

Access to control unit ports and devices (LUNs) is controlled by zoning (in switches) and LUN masking (in controllers), based on identity of virtual FCP channels (WWPN).

LUN masking and zoning are industry standard methods of controlling access to target control units and LUNs

IBM Systems

NPIV (N-Port ID Virtualization)

- Virtual WWPNs are assigned to each FCP subchannel by the channel hardware and are managed at the Support Element (SE)

- A VM guest inherits the virtual WWPN assigned to a FCP subchannel when the subchannel is attached to the guest

- The VM system inherits the virtual WWPN when a FCP subchannel is attached to the system via the SET EDEVICE command

- Standard access controls such as zoning in switches and LUN masking in the controller can be used with the virtual WWPNs

- NPIV is supported on System z9 and z10 processors

- System z FCP Redbook (including chapter on NPIV):
  SC24-7266, Fibre Channel Protocol for Linux and z/VM on IBM System z

IBM Systems

DS8000 Support

IBM Systems

DS8000 Series

- **Models DS8100 & DS8300 deliver massive scalability and performance**
  - Scalable storage capacity from 1.1 TB to 1024 TB (384TB for 8100)
  - Support for 4-port 4Gbps FCP/FICON host adapters; up to 32 providing for 128 host ports (64 for 8100)
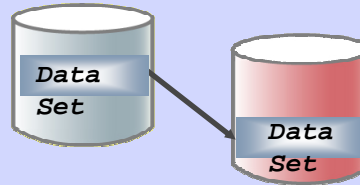 - Up to 256 GB of cache supporting high-capacity workloads

- **IBM Enterprise DASD functions limited to the DS8000**
  - Encryption (transparent to host)
  - HYPERPAV
  - FlashCopy Space Efficient (SE)

- **z/VM mod-54 service requirements:**
  - DIRMAINT APAR VM63700
  - DFSMS/VM APAR VM63664
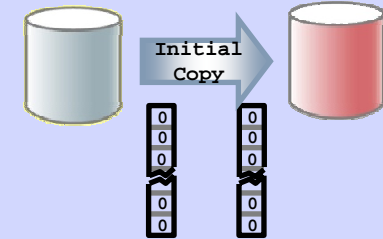
IBM Systems

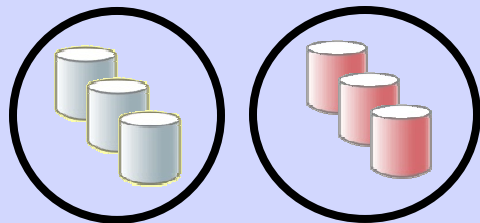# IBM System Storage FlashCopy Services
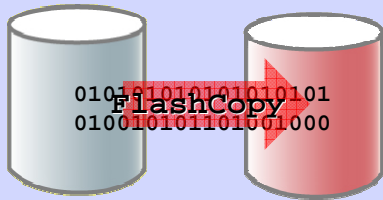
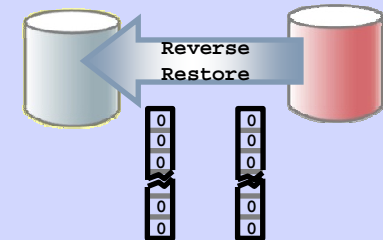FlashCopy Full Copy and NoCopy

Dataset Level FlashCopy

Incremental FlashCopy

FlashCopy *Consistency* Group

Persistent FlashCopy

FlashCopy *Reverse* Restore

IBM Systems

FlashCopy Space Efficient (SE)

- Persistent FlashCopy relationship where target is Space Efficient

- Definitions:

  - **Space Efficient Volume,** a disk whose backing storage is allocated out of a shared data repository where storage is provisioned only when necessary (e.g., unformatted records require no backing storage).

  - **Persistent FlashCopy,** a time-zero copy where a logical relationship exists in the control unit between the source and target (after the copy) until manually withdrawn; used to track changes from time-zero.

- FlashCopy SE time-zero operation maps SE target tracks to non-SE source tracks

  - SE repository space is only provisioned as source and target tracks are modified from time-zero; repository storage holds changed tracks.

  - SE volumes allow over-provisioning of storage, but only supported for Persistent FlashCopy.

  - Repository storage freed-up as each FlashCopy relationship is "withdrawn"

  - Perfect for testing (from production datasets) and back-up (to tape) scenarios

- z/VM 5.4.0 service upgrade (VM64449) provides FlashCopy SE support
  - CP FLASHCOPY command updated (for Persistent, FC SE, NOCOPY, Incremental, & Reverse)
  - CP QUERY command capabilities added to manage SE volumes and repositories

IBM Systems

Tape Overview

IBM Systems

Tape Support

- VM uses tape drives for many native system maintenance functions:
  - Backing up disk volumes with the DASD Dump Restore utility (DDR)
  - Backing up system spool data with the CP SPXTAPE command
  - Backing up CMS Minidisk files with the CMS TAPE command
  - Storing system trace data with the CP TRSAVE command
  - Managing volumes from a robotics tape library with the DFSMS/VM RMS Service Virtual Machine (SVM)

- A tape drive can be dedicated for use by a guest operating system with the SINGLEUSER option of the CP ATTACH command or DEDICATE user directory statement

- A tape drive can be *serially* shared by multiple guest operating systems with the MULTIUSER option of the CP ATTACH command (or DEDICATE statement)
  - Specifically for guests that manage their own assignment of tape drives, such as z/OS JES 2 Static Vary and Tape Auto-Switching
  - Not intended for CMS

IBM Systems

Supported Tape Units / Libraries

- 3480, 3490 Tape Subsystems

- 3590, 3592, TS1120 & TS1130 Enterprise Tape Subsystems

- 3494, 3495 Automated Tape Libraries

- 3494 Virtual Tape Server (VTS) Library

- TS3400 Tape Auto-Stacker

- TS3500 (3584) Tape Library

- Virtualization Engine TS7700 (7720,7740)Tape Library

- **z/VM tape library control functions provided by DFSMS/VM RMS. Library control functions are not directly supported by Linux.**

- **FICON connection supported by both z/VM & Linux - FCP by Linux only.**

Speed / Feed examples of different drive and cartridge types:

	3480	3490	3590E	3590H	3592 J1A	TS1120 (3592 E05)
Raw Capacity	200MB	800MB	20.5GB	30.8GB	300GB	500GB
Bandwidth	3MB/s	4MB/s	14MB/s	14MB/s	40MB/s	100MB/s

IBM Systems

Tape Encryption Support

IBM Systems

Tape Encryption Overview

- Enterprise class drives with encryption capability:
  - ✓ TS1120 (3592-E05) & TS1130 (3592-E06)

- z/VM uses a java-based external key manager running on a separate z/OS partition, Windows, AIX, or Linux machine

- z/VM 5.3.0 (and up) recommended since allows specification of Key Encrypting Key (KEK) labels; previous release allows only default keys defined to external key manager.

- Linux uses z/VM's guest transparency support (next slide)

- VM64260 provides ability to re-encrypt a cartridge with a different set of KEK labels.

Data encryption within the drive itself.
Now, why didn't somebody think of that?

IBM Systems

Tape Encryption Specifics

- CP ATTACH and SET RDEVICE commands accept a KEY parameter providing transparent encryption support for guest operating systems that haven't been coded to the new IBM tape encryption architecture

  ► Commands GIVE and SPXTAPE honor the ATTACH/SET RDEVICE settings

- z/VM I/O translation allows guest operating systems that exploit the new architecture to use the encryption hardware facilities

- DASD Dump Restore (DDR) utility supports tape encryption via parameter KEY

- CP QUERY TAPES and QUERY VIRTUAL TAPES commands display encryption capability and key settings

- SET KEYALIAS and QUERY KEYALIAS commands for management of KEK labels used with ATTACH and SET RDEVICE

- REKEY parameter provided on SET TAPE command

IBM Systems

Device Support Statements

IBM Systems

Linux for System z Hardware Support Statement

Unsupported	Do not attempt. Will not work. High risk. No IBM support line support	
Bronze	IBM Untested	No IBM testing No known issues
	3rd party supported for IBM compatible OEM devices	IBM Support Line supports only issues recreateable on "Gold level" environment
Silver	IBM Defect support	IBM testing not comprehensive No known issues IBM support Line limited support – e.g. PD/PSI risk is on customer
Gold	IBM Qualified/IBM RPQ	Thoroughly tested by IBM Fully Supported by IBM support line

- Testing may move environments from "Silver" to "Gold" level
  - ► "Silver" IBM defect support may mean, both
    - IBM is planning to qualifiy environment in the future, or
    - IBM has no plans to qualify environment
- To be sure in any "Silver" environment, raise an RPQ/One-Off in order to lift it to 'Gold' level
- Please confirm "Gold"-level "IBM RPQ" environments with IBM prior to running it.

IBM Systems

Linux FICON Disk Support (example)

Controller	z/VM 5.2	z/VM 5.3	SLES 8	SLES 9	RHEL3	RHEL4
IBM ESS E20/F20	IBM Qualified GA	IBM Qualified GA	Defect Support 2.4.21-261	Defect Support 2.6.5-7.191	Defect Support 2.4.21-27	Defect Support 2.6.9-11
IBM ESS 800	IBM Qualified 0501RSU	IBM Qualified GA	IBM Qualified 2.4.21-261	IBM Qualified 2.6.5-7.191	IBM Qualified 2.4.21-27	IBM RPQ 2.6.9-11
IBM DS6000	IBM Qualified 0502RSU	IBM Qualified GA	IBM Qualified 2.4.21-261	IBM RPQ 2.6.5-7.191	IBM Qualified 2.4.21-27	IBM RPQ 2.6.9-11
IBM DS8000	IBM Qualified 0502RSU	IBM Qualified GA	IBM Qualified 2.4.21-261	IBM Qualified 2.6.5-7.191	IBM Qualified 2.4.21-27	IBM Qualified 2.6.9-22.0.1
IBM compatible OEM Device	Vendor support 0502RSU	Vendor support GA	Vendor support 2.4.21-261	Vendor support 2.6.5-7.191	Vendor support 2.4.21-27	Vendor support 2.6.9-11

IBM Systems

z/VM Hypervisor Hardware Support Statement

- **z/VM has fully tested and supports all devices listed in Appendix C of the z/VM General Information manual (per given version/release).**

- **z/VM support line will provide support for hypervisor issues associated with 3$^{rd-party}$ IBM-Compatible OEM devices if re-creatable on supported IBM hardware.**

- **Important: Always check Hardware Interoperability statements from hardware providers (IBM and OEM) to ensure that hardware owner supports z/VM and Linux for System z.**
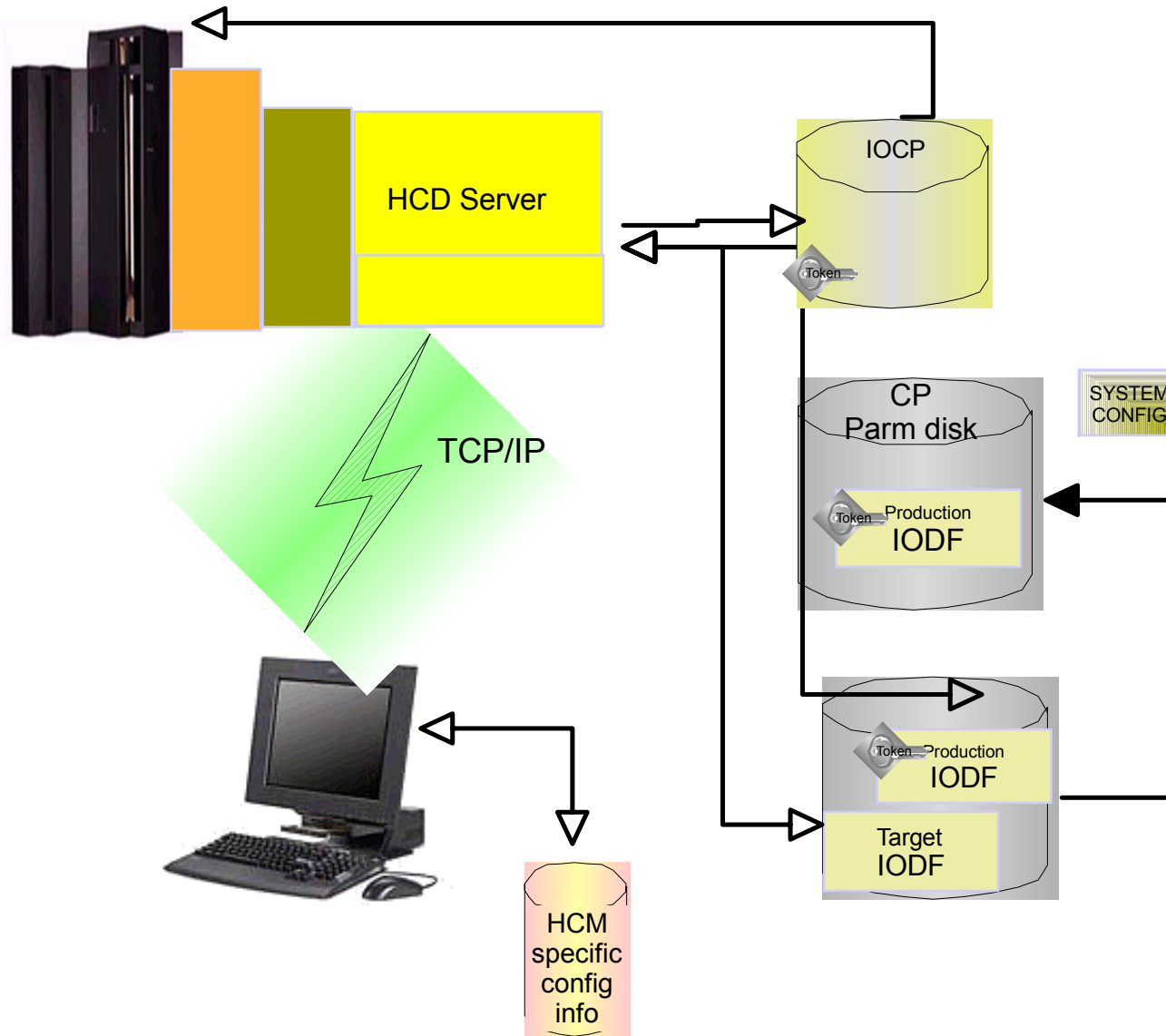
IBM Systems

Dynamic I/O Support

© 2009 IBM Corporation

IBM Systems

Dynamic I/O Configuration

- Function allows modification of real hardware I/O definitions (IOCP) without an outage

- Define, Modify, and Delete
  - ➤ CHPIDs (channel paths), Control Units, and Devices

- Manage I/O definitions for all LPARs on a CEC from one LPAR

- Controlled by CP's dynamic I/O commands (DEFINE, MODIFY, DELETE)
  - ➤ *Or,* HCD/HCM support

IBM Systems

VM HCD/HCM Support

- Hardware Configuration Definition (HCD) runs in a VM service virtual machine and acts as an I/O configuration server that the Hardware Configuration Manager (HCM) running on a PC can connect to via TCP/IP

- HCD uses an I/O Definition File (IODF) that contains the current hardware and software configuration information. This information is graphically displayed via the HCM terminal.

- VM's IPL code recognizes when HCD is controlling the I/O configuration (IODF system configuration file statement is specified) and interprets data from the HCD-supplied IODF file

- Dynamic changes to the configuration are made by updating information on the HCM screens to create a target configuration IODF file and then activating that target IODF file via the ACTIVATE command on the HCD service virtual machine console

- HCD processes the ACTIVATE request by comparing the active production IODF file with the target IODF file and creating a delta which represents the hardware and software changes that need to be executed. This delta is sent to CP to execute the changes.

IBM Systems

IOCP

Token

HCD Server

TCP/IP

CP
Parm disk

SYSTEM
CONFIG

Token   Production
IODF

Token   Production
IODF

Target
IODF

HCM
specific
config
info

Dynamic changes are only allowed if the token of the current production IODF matches a system token.

The SYSTEM CONFIG file identifies the IODF to be used.

IBM Systems

Trademarks

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

APPN*	GDPS*	POWERPC*	Tivoli*
CICS*	Geographically Dispersed Parallel Sysplex	PR/SM	Tivoli Storage Manager
DB2*	HiperSockets	Processor Resource/Systems Manager	TotalStorage*
DFSMSMVS	HyperSwap	QMF	Virtual Image Facility
DFSMS/VM*	IBM*	RACF*	Virtualization Engine
DirMaint	IBM eServer	Resource Link	VisualAge*
Distributed Relational Database Architecture*	IBM logo*	RMF	VM/ESA*
DRDA*	IBMlink	RS/6000*	VSE/ESA
e-business logo*	Language Environment*	S/390*	VTAM*
ECKD	MQSeries*	S/390 Parallel Enterprise Server	WebSphere*
Enterprise Storage Server*	Multiprise*	System 370	z/Architecture
Enterprise Systems Architecure/390*	On demand business logo	System 390*	z/OS*
ESCON*	OS/390*	System I	z/VM*
FICON*	Parallel Sysplex*	System p	z/VSE
GDDM*	Performance Toolkit for VM	System z	zSeries*
	POWER5	System z9	zSeries Entry License Charge

The following are trademarks or registered trademarks of other companies.

Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries

Linux is a trademark of Linus Torvalds in the united States and other countries..

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

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

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.

IBM Systems