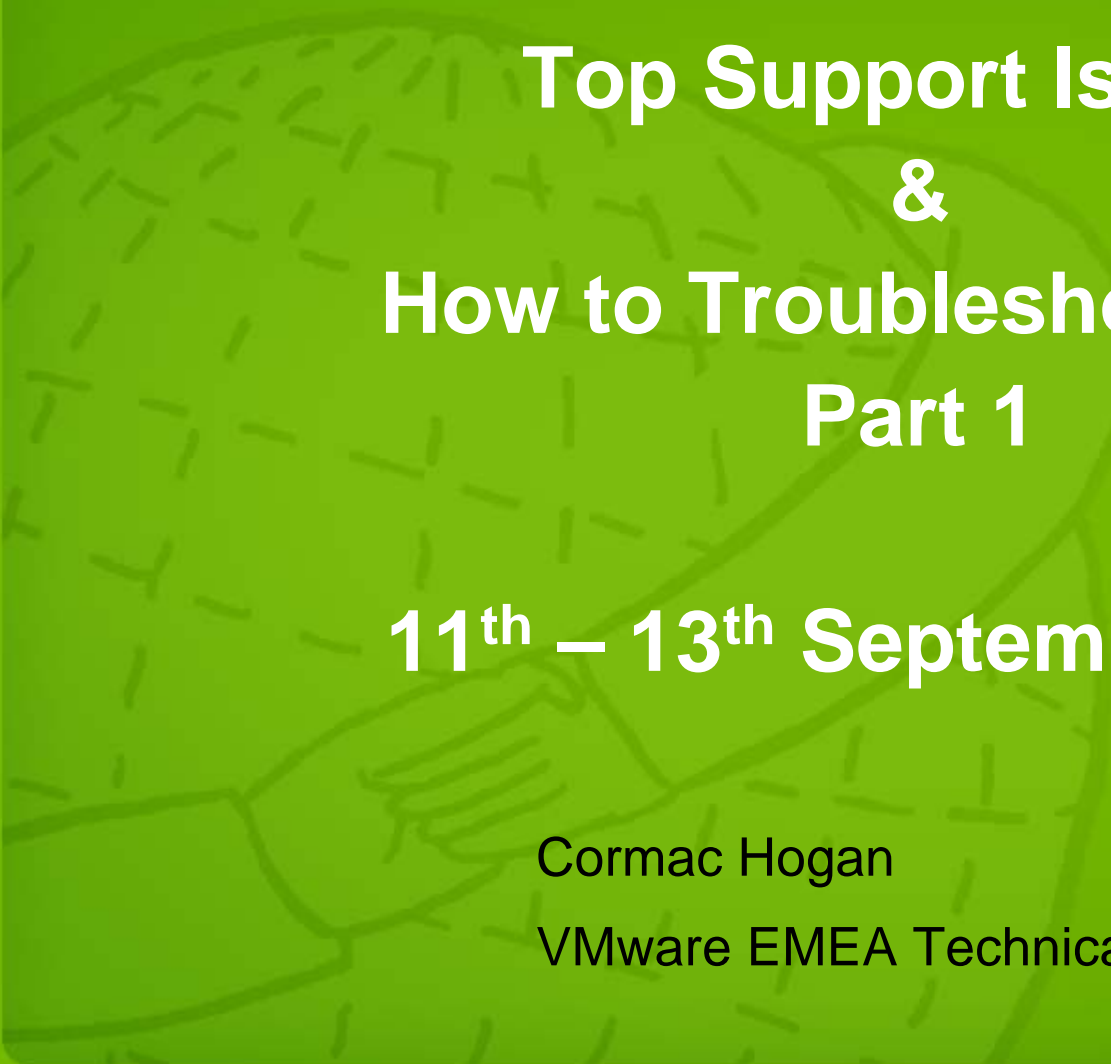# VMWORLD 2007

EMBRACING YOUR VIRTUAL WORLD

Top Support Issues
&
How to Troubleshoot them
Part 1

11th – 13th September 2007

Cormac Hogan

VMware EMEA Technical Support

Purpose of Presentation

- **An overview of the top support issues.**
  - SAN Multipathing – Best Practices.
  - Snapshots – Resignature or Allow Snapshots?
  - VMFS Upgrades – Common issues.
  - VCB – Common issues.
- **And how to identify, troubleshoot & resolve these issues.**

Issue #1

SAN multi-path Misconfiguration

Leads to Path Thrashing

Storage Arrays Type

○ **There are basically 2 types of storage array:**

➤ Active/Active

- LUNs presented on multiple Storage Processors

- **Fixed** path policy

  - Failover on **NO_CONNECT**

  - Preferred path policy

  - Failback to preferred path if it recovers

➤ Active/Passive

- LUNs presented on a single Storage Processor

- **MRU** path policy (Most Recently Used)

  - Failover on **NOT_READY**, **ILLEGAL_REQUEST** or **NO_CONNECT**

  - No preferred path policy

  - No failback to preferred path

Best Practices – Active/Passive SAN configuration

A/P Storage Array

SPA | SPB
0 1 | 0 1

FC Switch 1 | FC Switch 2

HBA | HBA
1 | 2
Server X

HBA | HBA
1 | 2
Server Y
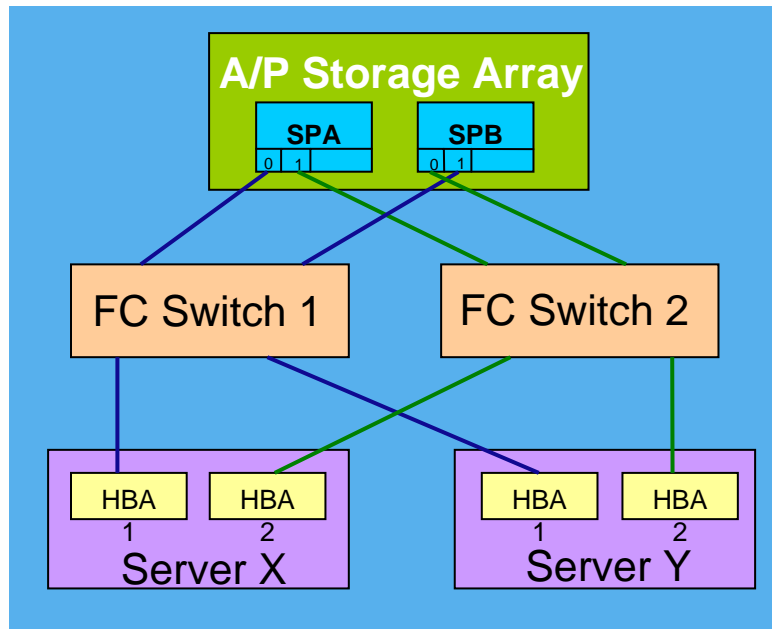
We could lose any single component in this configuration & both ESX servers would still access the LUN via the same SP & avoid path thrashing.

```
  Server X, HBA 1 zoned to Storage Array SP 'A' port 0
& Server X, HBA 1 zoned to Storage Array SP 'B' port 1
  Server X, HBA 2 zoned to Storage Array SP 'A' port 1
& Server X, HBA 2 zoned to Storage Array SP 'B' port 0

  Server Y, HBA 1 zoned to Storage Array SP 'A' port 0
& Server Y, HBA 1 zoned to Storage Array SP 'B' port 1
  Server Y, HBA 2 zoned to Storage Array SP 'A' port 1
& Server Y, HBA 2 zoned to Storage Array SP 'B' port 0
```

VMware recommends **4** paths from every ESX to an A/P storage array, commonly referred to as the **ABAB** configuration.

A good A/P multipath configuration

```
# esxcfg-mpath -q --lun vmhba1:0:1
Disk vmhba1:0:1 /dev/sdk (20480MB) has 4 paths and policy of Most Recently
Used
  FC 3:11.0 210000e08b1c8e42<->5006016030021b9df vmhba1:0:1 On active
preferred
  FC 3:11.0 210000e08b1c8e42<->500601683021b9df vmhba1:1:1 Standby
  FC 5:4.0  210000e08b1c7143<->500601613021b9df vmhba2:0:1 On
  FC 5:4.0  210000e08b1c7143<->500601693021b9df vmhba2:1:1 Standby
```

The two HBAs are each zoned to two ports on the storage array, giving us our **4 paths**:

- **HBA 1** (wwpn 8e42) is zoned to **SP A port 0** (wwpn 60) and **SP B port 0** (wwpn 68).
- **HBA 2** (wwpn 7143) is zoned to **SP A port 1** (wwpn 61) and **SP B port 1** (wwpn 69).

This meets our best practice requirement of having an **ABAB** path configuration.

A/P Failover behaviour – loss of HBA

- Active path is HBA1 -> SPA (**A**B**A**B)
- First HBA on Server A fails (**A**B**A**B)
- Check Unit Ready
  - LUN is READY on the HBA2 -> SPA path (**AB**A**B**)
  - ESX Server A fails over to this path
- No failover occurs on the storage side since SPA is still the LUN owner.
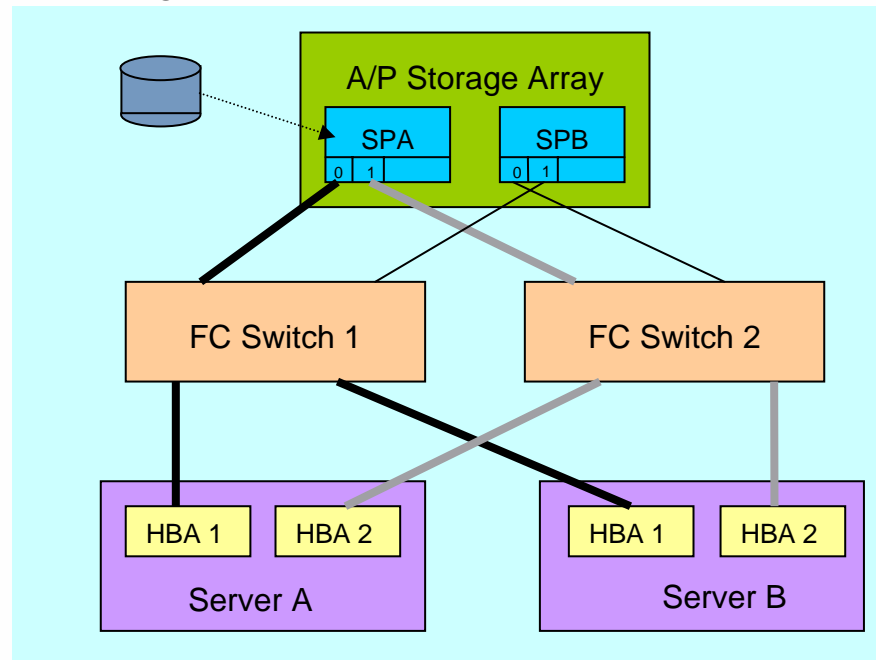
Log snippet – failover to standby HBA

```
WARNING: SCSI: 2812: Manual switchover to path vmhba2:0:1 begins.
SCSI: 2816: Changing active path to vmhba2:0:1
WARNING: SCSI: 2360: Did not switchover to vmhba2:0:1. Check Unit Ready
Command
returned READY instead of NOT READY for standby controller .
WARNING: SCSI: 2841: Manual switchover to vmhba2:0:1 completed successfully.
```
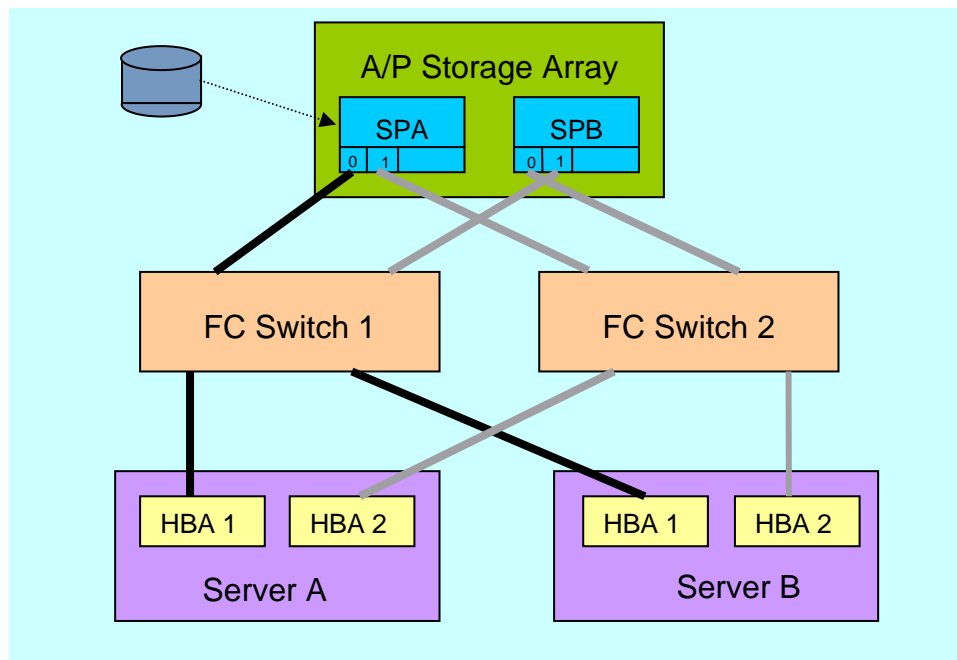
Note the message:

Check Unit Ready Command returned READY instead of NOT READY for standby controller.

- Indicates that the LUN returned a **UNIT READY** on path vmhba2:0:1.
- Therefore no need to initiate a *failover* on the storage array.
- There is a path via a *different* HBA to the *same* SP that currently owns the LUN.

A/P Failover behaviour – loss of SP

- Active path is from HBA1 -> SPA (**A**B**A**B) and SPA fails (**A**B**A**B)
- LUN no longer visible on any path to SPA, ESX server initiates path failover
- This transfers ownership of the LUN from SPA to SPB
- Active path is now from HBA1 -> SPB (**A**B**A**B)
- Failover is on the storage, no failover on host

Log snippet – failover to standby SP

- `WARNING: SCSI: 2812: Manual switchover to path vmhba2:1:1 begins.`
- `SCSI: 2816: Changing active path to vmhba2:1:1`
- `SCSI: 8816: vmhba2:1:1:0  Retry (unit attn)`
- `WARNING: SCSI: 2841: Manual switchover to vmhba2:1:1 completed successfully.`

The tell-tale sign of a *failover* having occurred on the array is the <u>absence</u> of the message:
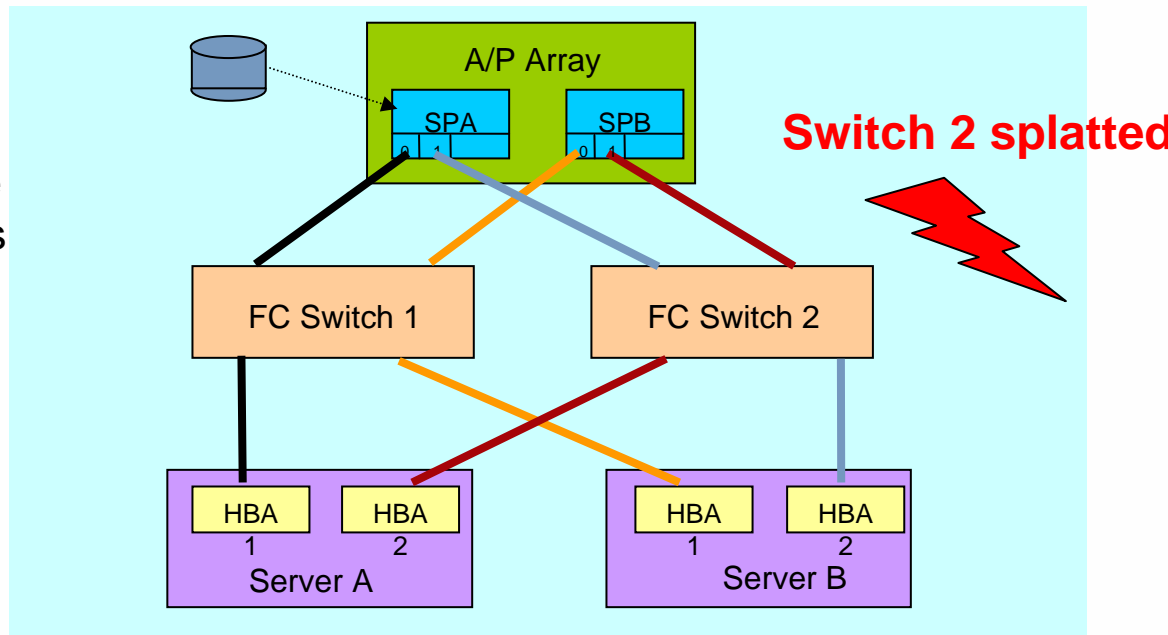
Check Unit Ready Command returned READY instead of NOT READY for standby controller

The ESX initiated a request to make the LUN active on the standby SP.

Retry (unit attn) indicates that the LUN had outstanding I/Os which were discarded, i.e. the LUN is reset.

What's wrong with this A/P configuration?

LUN shown
is shared
between the
ESX servers

A/P Array

SPA

SPB

Switch 2 splatted

FC Switch 1

FC Switch 2

HBA 1   HBA 2
Server A

HBA 1   HBA 2
Server B

Via FC Switch 1 { `Server A, HBA 1 zoned to Storage Array SP 'A' port 0`

&

Via FC Switch 2 { `Server A, HBA 2 zoned to Storage Array SP 'B' port 1`

Via FC Switch 1 { `Server B, HBA 1 zoned to Storage Array SP 'B' port 0`

&

Via FC Switch 2 { `Server B, HBA 2 zoned to Storage Array SP 'A' port 1`

Load Balancing

- **On Active/Active arrays**

  - Uses *fixed path* policy - implement load balancing using the *preferred path* where LUNs are accessed by the ESX via different Storage Processors/Controllers.

- **On Active/Passive arrays**

  - Uses *mru path* policy - *manually* load balance the LUN presentations from the array side, presenting different LUNs via different Storage Processors/Controllers.

  - This is perfectly acceptable, but in the event of a failover, all the load balancing is lost.

  - Why not just set **fixed** path policy on an **A/P** array and use the **preferred path** option?

The reason it is not a good idea - Failover conditions

- On A/P arrays, there are three SCSI return codes / check conditions which will initiate a failover:
  - NO CONNECT
  - NOT READY
  - ILLEGAL REQUEST

- On A/A arrays, there is only one SCSI return codes / check conditions which will initiate a failover:
  - NO CONNECT

- If we set a Fixed path policy and implement the preferred path on an A/P array, it is possible we get a check condition which should initiate a failover, but will not do so if the path policy is set to Fixed rather than MRU.

Load Balancing

> With the next release of ESX, a new load balancing path policy will be introduced.

> This is known as the Round Robin path policy and will allow switching to a new path:

- After a specific number of commands are sent on a path
- After a specific number of blocks of I/O are sent on a path

> The new path can be either selected by specifying:

- A different Storage Process.
- A different HBA.

> This policy is already available in ESX 3 but support is *experimental*.

Issue #2

VMFS Volumes & Snapshots

Cannot see my VMFS volumes

- **A common support issue – my VMFS volumes have disappeared!**
  - In fact, it is often the case that the volumes are seen as snapshots which, by default, are not mounted.
- **Why are VMFS volumes seen as snapshots when they are not?**
  - *ESX server A* is presented with a LUN on ID **0**.
  - Same LUN is presented to *ESX server B* on ID **1**.
  - VMFS-3 volume created on LUN ID 0 from server A.
  - Volume on server B will <u>not</u> be mounted when SAN is rescanned.
  - Server B will state that the volume is a snapshot because of LUN ID mismatch.
- **LUNs must be presented with the <u>same LUN IDs</u> to all ESX hosts.**
  - In the next release of ESX, LUN ID is no longer compared if the target exports NAA type IDs.

How does ESX determine if volume is a snapshot?

> When a VMFS-3 volume is created, the **SCSI Disk ID** data from the LUN/storage array is stored in the volume's LVM header.

> This contains, along with other information, the LUN ID.

> When another ESX server finds a LUN with a VMFS-3 filesystem, the **SCSI Disk ID** information returned from the LUN/storage array is compared with the LVM header metadata.

> The VMkernel treats a volume as a **snapshot** if there is a mismatch in this information.

How does this mismatched information get reported?

- **Device** vmhba2:2:2:1 **is a snapshot:**
- **disk ID: <type** 3, **len** 15, **lun** 2, **devType** 0, **scsi** 3, **h(id)** 1771423412675533879>
- **m/d disk ID: <type** 3, **len** 15, **lun** 2, **devType** 0, **scsi** 3, **h(id)** 9219142619163180480>
- **Device** vmhba2:2:2:1 **is a snapshot:**
- **disk ID: <type** 3, **len** 15, **lun** 2, **devType** 0, **scsi** 3, **h(id)** 1771423412675533879>
- **m/d disk ID: <type** 3, **len** 15, **lun** 2, **devType** 0, **scsi** 3, **h(id)** 9219142619163180480>
- **vmhba2:2:2:1** may be snapshot: disabling access. **See resignaturing section in SAN config guide.**
- .
- .
- **Device** vmhba2:2:9:1 **is a snapshot:**
- **disk ID: <type** 3, **len** 15, **lun** 9, **devType** 0, **scsi** 3, **h(id)** 1771423412675533879>
- **m/d disk ID: <type** 3, **len** 15, **lun** 9, **devType** 0, **scsi** 3, **h(id)** 9219142619163180480>
- **Device** vmhba2:2:9:1 **is a snapshot:**
- **disk ID: <type** 3, **len** 15, **lun** 9, **devType** 0, **scsi** 3, **h(id)** 1771423412675533879>
- **m/d disk ID: <type** 3, **len** 15, **lun** 9, **devType** 0, **scsi** 3, **h(id)** 9219142619163180480>

- This logging appears in the **/var/log/vmkernel** log file.
- The **disk ID** line is queried directly from the target.
- The line containing **m/d disk ID** that information is stored in the LVM header.
- In this case it is the **h(id)** data in the LVM header which is mismatched.

How should you handle Snapshots?

- **First of all, determine if it is really a snapshot:**
  - If it is a mismatch of LUN IDs across different ESX hosts, fix the LUN ID through array management software to ensure that the **same LUN ID** is presented to all hosts for a share volume.
  - Other reasons a volume might appear as a snapshot could be changes in the way the LUN is presented to the ESX:
    - HDS Host Mode setting
    - EMC Symmetrix SPC-2 director flag
    - Change from A/P firmware to A/A firmware on array

- **If it is definitely a snapshot, you have two options:**
  - Set **EnableResignature**
- **Or**
  - Disable **DisAllowSnapshotLUN**

LVM.EnableResignature

- **Used when mounting the original and the snapshot VMFS Volumes on the same ESX.**

  - Set **LVM.EnableResignature** to **1** and issue a rescan of the SAN.

  - This updates the LVM header with:

    - new **SCSI Disk ID** information

    - a new VMFS-3 **UUID**

    - a new **label**

      - Label format will be **snap-\<generation number\>-\<label\>**, or

        snap-\<generation number\>-\<uuid\> if there is no label, e.g.

        - Before resignature: **/vmfs/volumes/lun2**
        - After resignature: **/vmfs/volumes/snap-00000008-lun2**

  - Remember to set **LVM.EnableResignature** back to **0**.

LVM.DisallowSnapshotLUN

- **DisallowSnapshotLUN** will not modify any part of the LVM header
- To allow the mounting of snapshot LUNs, set:
  - **EnableResignature** to **0** (disable)
    - **and**
  - **DisallowSnapshotLUN** to **0** (disable)
- Do not use **DisallowSnapshotLUN** to present snapshots back to same ESX server
- **LVM.EnableResignature** overrides **LVM.DisallowSnapshotLUN**

LVM.EnableResignature

-- snapshot --

LUN 0 → LUN 1

Storage

SPA   SPB
0 1   0 1

FC Switch 1   FC Switch 2

HBA 1   HBA 2
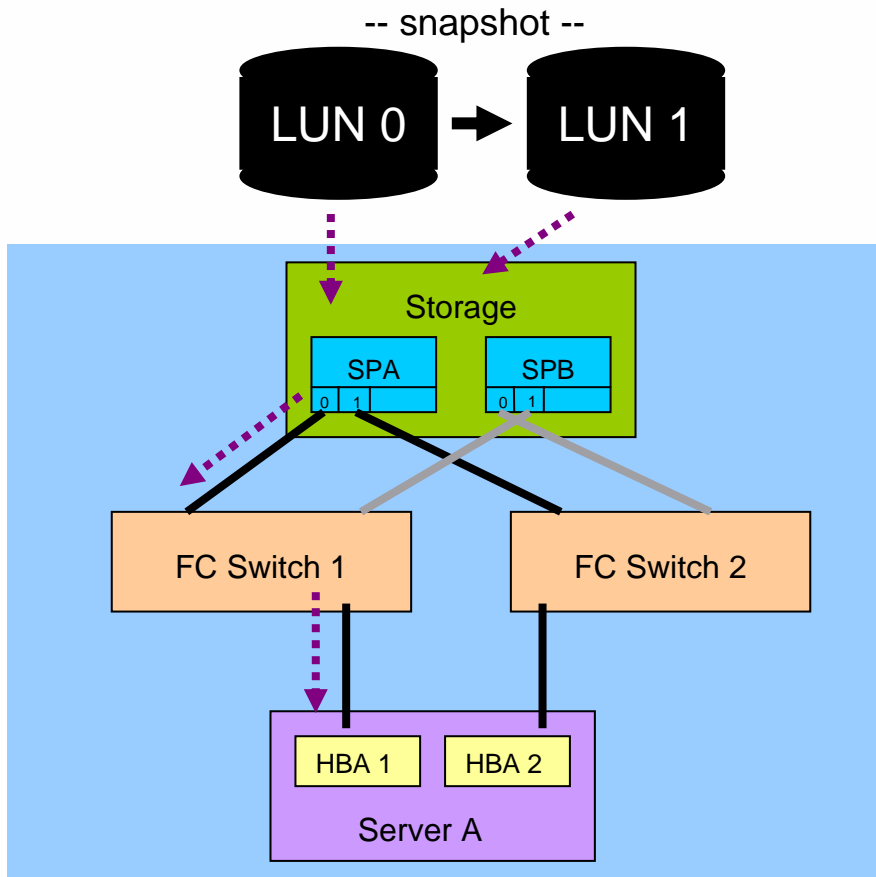
Server A

LVM.EnableResignature will have to be used to make the volume located on cloned LUN, LUN 1, visible to the same ESX server after a rescan.

Two volumes with the <u>same</u> UUID must **not** be presented to the same ESX server. Issues with data integrity will occur.

Enable Resignaturing & Rescan

- **[root@cork LVM]#** pwd
- **/proc/vmware/config/LVM**

- **[root@cork LVM]#** cat EnableResignature
- **EnableResignature (Enable Volume Resignaturing) [0-1: default = 0]:** 0

- **[root@cork LVM]#** echo 1 > EnableResignature

- **[root@cork LVM]#** esxcfg-rescan vmhba2
- **Rescanning vmhba2...done.**
- **On scsi6, removing: 0:0 0:1 0:2 0:3 2:1 2:10 2:11 2:12 2:13 2:14 2:15 2:16 2:17 2:18 2:19 2:2 2:20 2:21 2:22 2:23 2:24 2:240 2:241 2:25 2:26 2:27 2:28 2:29 2:3 2:30 2:31 2:32 2:33 2:34 2:35 2:36 2:37 2:38 2:39 2:4 2:40 2:41 2:42 2:48 2:49 2:5 2:6 2:7 2:8 2:9.**
- **On scsi6, adding: 0:0 0:1 0:2 0:3 2:1 2:10 2:11 2:12 2:13 2:14 2:15 2:16 2:17 2:18 2:19 2:2 2:20 2:21 2:22 2:23 2:24 2:240 2:241 2:25 2:26 2:27 2:28 2:29 2:3 2:30 2:31 2:32 2:33 2:34 2:35 2:36 2:37 2:38 2:39 2:4 2:40 2:41 2:42 2:48 2:49 2:5 2:6 2:7 2:8 2:9.**

- **[root@cork LVM]#** echo 0 > EnableResignature

Remote/Target side messages during Resignaturing

- Device vmhba2:2:2:1 is a snapshot:
- disk ID: <type 3, len 15, lun 2, devType 0, scsi 3, h(id) 1771423412675533879>
- m/d disk ID: <type 3, len 15, lun 2, devType 0, scsi 3, h(id) 9219142619163180480>
- Device vmhba2:2:2:1 is a snapshot:
- disk ID: <type 3, len 15, lun 2, devType 0, scsi 3, h(id) 1771423412675533879>
- m/d disk ID: <type 3, len 15, lun 2, devType 0, scsi 3, h(id) 9219142619163180480>
- **Device vmhba2:2:2:1 unsnapped**
- .
- .
- .
- Device vmhba2:2:9:1 is a snapshot:
- disk ID: <type 3, len 15, lun 9, devType 0, scsi 3, h(id) 1771423412675533879>
- m/d disk ID: <type 3, len 15, lun 9, devType 0, scsi 3, h(id) 9219142619163180480>
- Device vmhba2:2:9:1 is a snapshot:
- disk ID: <type 3, len 15, lun 9, devType 0, scsi 3, h(id) 1771423412675533879>
- m/d disk ID: <type 3, len 15, lun 9, devType 0, scsi 3, h(id) 9219142619163180480>
- **Device vmhba2:2:9:1 unsnapped**

- Snapshot LV <snap-1b47edf4-45b60b75-157d9671-088a-000423c5> complete on device vmhba2:2:9:1
- **Begin resignaturing volume** label: lun2, uuid: 45dd7f39-e69dd091-c5f4-000423c5a2ec
- **End resignaturing volume** label: snap-00000008-lun2, uuid: 45dd7563-b3f744de-12e0-000423c5a56c

LVM.EnableResignature OR LVM.DisallowSnapshotLUN

-- snapshot --
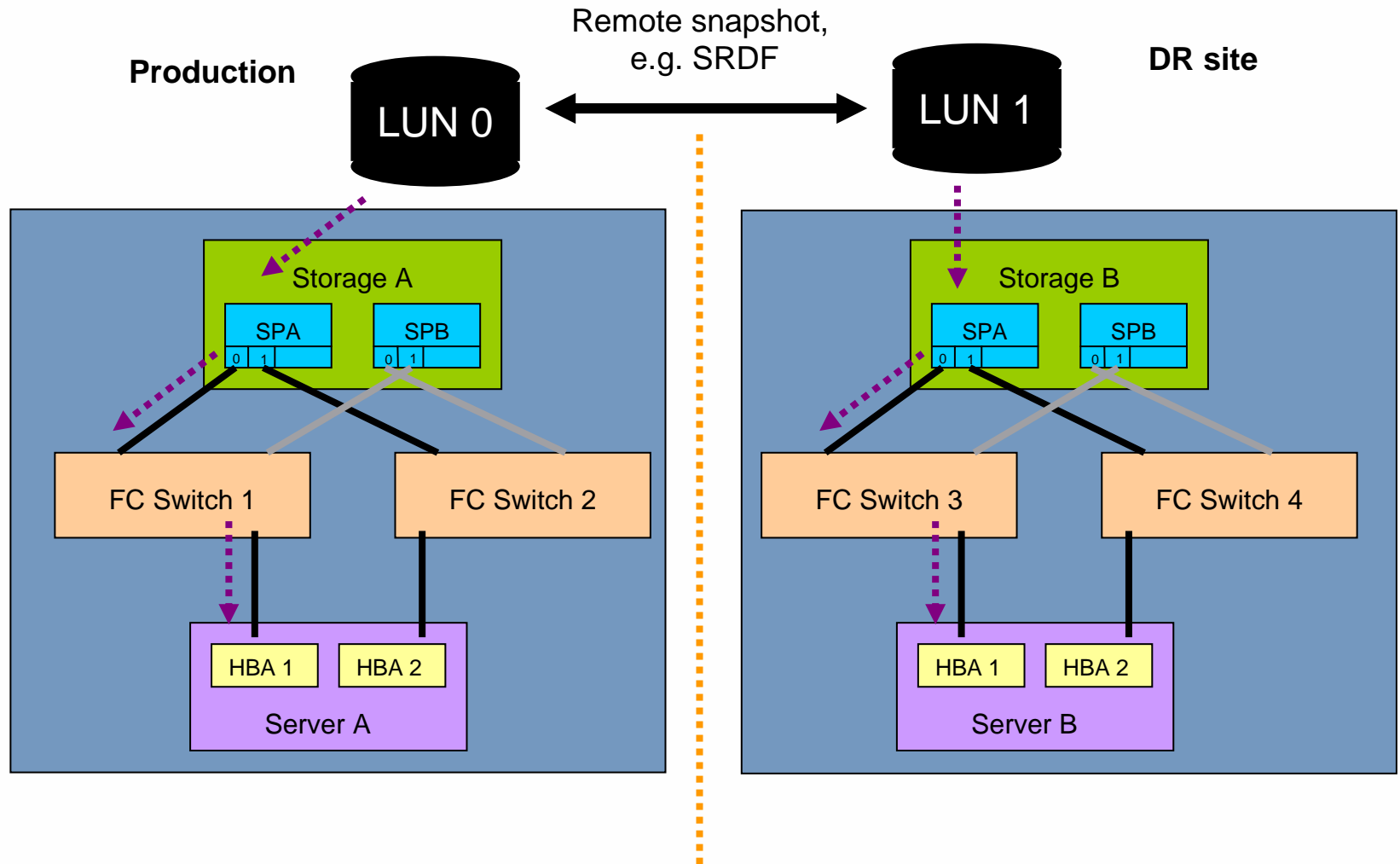
Snapshot LUN presented
to a different ESX server

LUN 0 → LUN 1
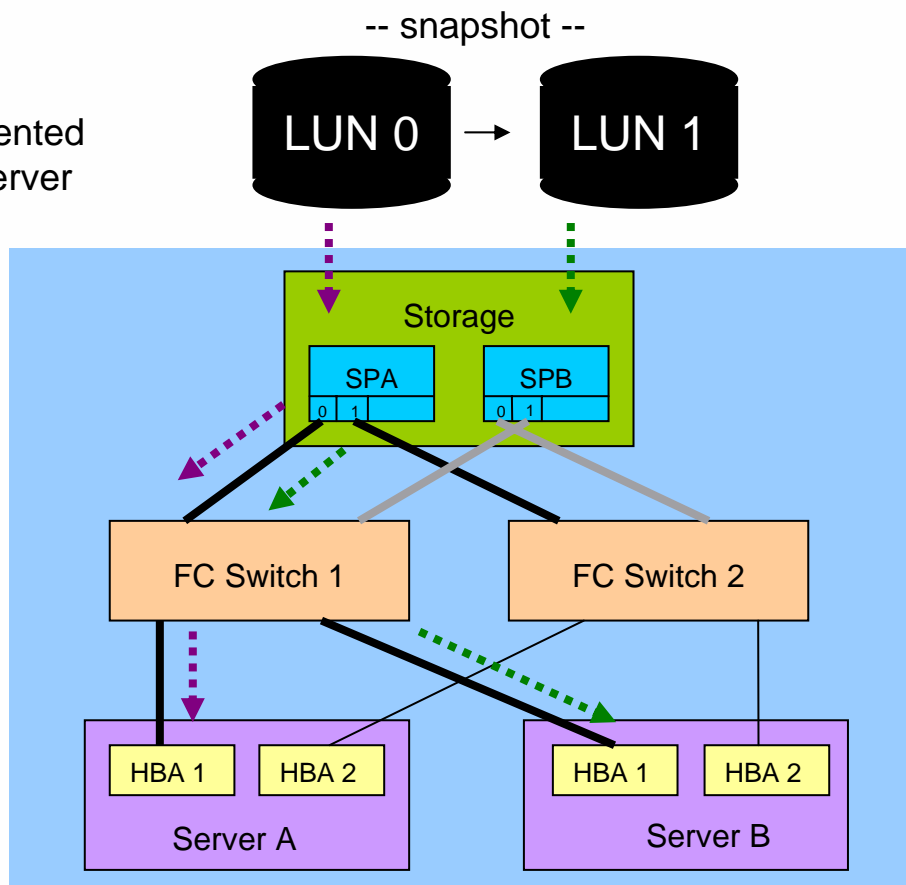
We can present the snapshot
LUN, LUN 1, using

DisallowSnapshotLUN = 0
on Server B as long as Server
B cannot see LUN 0.

Storage

SPA   SPB

0 1   0 1

FC Switch 1   FC Switch 2

HBA 1   HBA 2   HBA 1   HBA 2

Server A   Server B

If Server B can also see LUN 0, then we must use **resignaturing** since we cannot present
two LUNs with the same UUID to the same ESX server.

LVM.DisallowSnapshotLUN

Remote snapshot, e.g. SRDF

Production

DR site

LUN 0

LUN 1

Storage A

Storage B

SPA

SPB

SPA

SPB

0 1

0 1

0 1

0 1

FC Switch 1

FC Switch 2

FC Switch 3

FC Switch 4

HBA 1

HBA 2

HBA 1

HBA 2

Server A

Server B

Since there is not going to be a LUN with the same UUID at the remote site, one can **allow snapshots**.

Issue #3

VMFS-2 to VMFS-3 Upgrade issues

VMFS-2 to VMFS-3 Upgrades

A number of issues have been encountered with upgrades:

1. Unable to upgrade a VMFS-2 volume that has a block size greater **8MB**
   - VMFS-3 only supports block sizes of **1, 2, 4 or 8MB**

2. Cannot upgrade a VMFS-2 volume created with a 1MB block size containing files greater than **256GB**
   - VMFS-2 supports larger files than VMFS-3 on same block size

3. Must have **1.2 GB** free space on VMFS-2
   - We can workaround this using the **fsaux** driver

VMFS Upgrade 1 – block size incompatibility

- **Upgrading a VMFS-2 with file block size greater than 8MB fails**
  - Example:

    "WARNING: FSA: 1431: Block size 16777216 is beyond VMFS-3 limits (min 1048576, max 8388608)"

  - Cause:
    - VMFS-3 only supports file block sizes of 1MB, 2MB, 4MB or 8MB
  - Workaround:
    - Migrate VMs from the VMFS-2 with the file block size greater than 8MB to a VMFS-2 with a file block size less than or equal to 8MB before upgrading to VMFS-3

VMFS Maximum File Sizes

	VMFS2	VMFS3
1MB	456 GB	256GB
2MB	912 GB	512GB
4MB	1.78 TB	1 TB
8MB	3.56 TB	2 TB
16MB	7.13 TB	-
32MB	14.25 TB	-
64MB	28.50 TB	-

VMFS Upgrade 2 – file size incompatibility

- **Upgrading a VMFS-2 filesystem with a file block size of 1MB containing a file greater than 256GB fails**

  - Example:

    - "ConvertToFS3 failed with code 27: File too large and message Unable to upgrade filesystem"

  - Cause:

    - VMFS-2 supports larger files than VMFS-3 with the same file block size

  - Workaround:

    - Move out the files, re-create the volume as a VMFS-3 with a bigger block size and copy the files back in again

    - Create a brand new VMFS-3 with a bigger block size which can accommodate a file of this size and move the large files over to the new VMFS-3

Filesystem Access

	VMFS2	VMFS3
ESX 2.x	Read/Write	No Access
ESX 3.x	Read Only	Read/Write

VMFS Upgrade 3 – not enough free space

- **Consider this common scenario seen by Technical Support…**
  - You have upgraded your ESX from v2 to v3.
  - Now you attempt to upgrade VMFS-2 to VMFS-3.
  - This fails because you do not have the required 1.2GB of free space available on the volume to complete the upgrade.
  - VMFS-2 volumes are read-only to ESX v3 so you cannot delete files from the volume.
  - How do we resolve this?

By using the auxiliary FS driver, fsaux

- **Commonly used functions of this driver are:**
  - In-place upgrade from VMFS-2 to VMFS-3
    - **fsauxFunction=upgrade**

  - Removing files from a VMFS-2 volume (unlink mode)
    - **fsauxFunction=fs2unlink**

Demo

- Use the **fsaux** driver to do an in-place *upgrade* from VMFS-2 to VMFS-3

- Discover that there is not enough free space (1.2GB) on the volume that you are trying to upgrade

- Using the **fsaux** driver to place the VMFS-2 volume in *fs2unlink* mode to allow files to be moved or removed from it by an ESX 3 host

- Completing the *upgrade* using **fsaux** driver once again

Issue #4

Consolidated Backup (VCB)

Disabling Automatic Drive-Letter Assignment on Proxy

- All versions of Windows, except Windows 2003 Enterprise Edition & Datacenter Edition automatically assign drive letters to each visible volume.

- For VCB, you must change this behaviour so that volumes are <u>not</u> automatically mounted on the proxy.

- You do this you use the **diskpart** utility and set the options:
  - automount disable
  - automount scrub

- This will disable automatic drive-letter assignment to newly seen volumes and cleans out entries of previously mounted volumes in the registry.

- A complete step-by-step procedure is in the VCB documentation.

Most Common Configuration Issues

- **LUN ID on VCB proxy must match LUN ID on ESX server**

  Error: Failed to open the disk: One LUN could not be opened

  - This is the error one would see when the LUN is <u>not</u> presented with the same ID to both the ESX server and the proxy server.

  - Will not be a requirement in **ESX 3.0.2 / VCB 1.0.3** as we will use VMFS UUID rather than LUN ID for identifying volumes and LUNs.

- **VCB proxy does not support multipathing**

  Error: Failed to export the disk: The device is not ready

  - You <u>cannot</u> have *EMC Powerpath*, *Veritas DMP* or any other multi-pathing software installed on your VCB proxy.

  - If you have multiple paths to a SAN LUN showing up on the proxy, all the inactive paths on the proxy must be deactivated.

  - http://kb.vmware.com/kb/2195414

Disabling an Inactive Path

- In the "Computer Management" window, select "Disk Management" to display a list of all disks and visible to the system.

- If a disk is visible through multiple paths, there will be an entry for each path in this list.

- The *Inactive paths* to disks will be indicated by a small "No Entry" icon superimposed over the disk system.
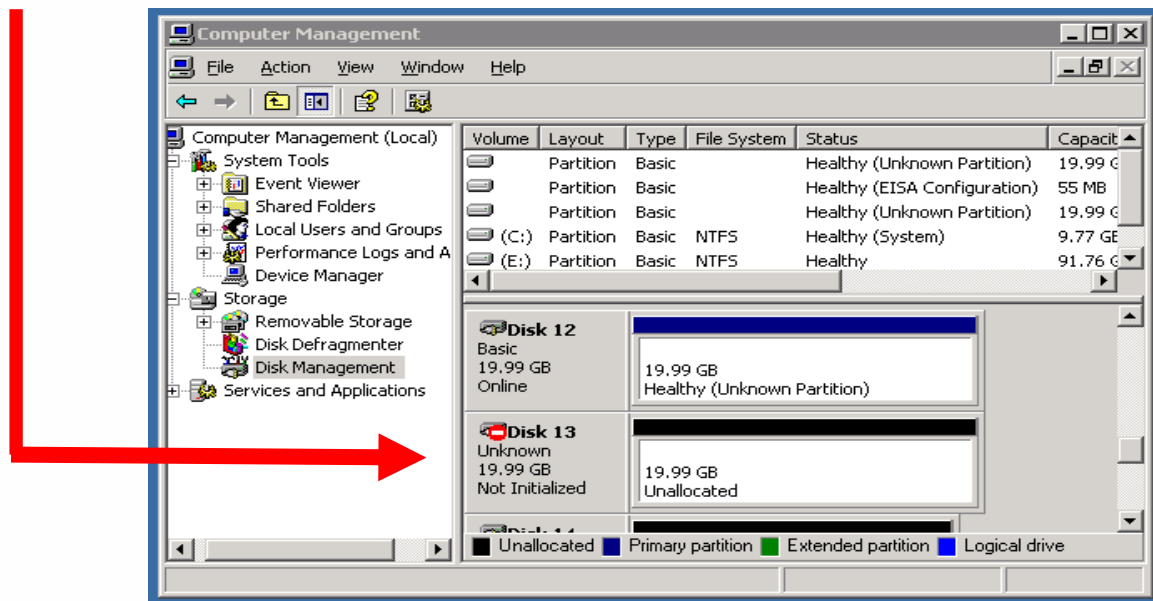
Disabling VCB Multipath configuration

> In order to disable an inactive path, right-click on its icon with the superimposed "No Entry" icon on it and select Properties from the context menu.

> In the Disk drive *Properties* General tab, change the value for Device usage to *Do not use this device (disable).*

> This will cause the entry for the disk to be removed from the list of devices presented by "Disk Management".

> Repeat for all inactive paths.

Unable to determine IP Address of the VM

- **To see if the ip address of the VM is being seen, use:**

- # vcbVmName -s powerstate:on

```
Found VM:
moref:vm-15
name:VC-Win2003
uuid:564dfaae-254f-0827-1d1e-9e1f243effbf
ipaddr:10.16.112.192
```
← IP address visible

```
[2006-12-06 14:55:57.268 'vcbVmName' 3076456576 warning] IP
   address not set.
Found VM:
moref:vm-28
name:iSCSI_server
uuid:564d4c25-d7dd-a806-5aa3-7db4c80568a0
ipaddr:
```
← IP address not visible

If VMware Tools is not installed in the Virtual Machine, the IP address cannot be discovered. The VM should be powered on at least once with VMware Tools to get the IP address.

Any questions on…

- **Multipathing?**

- **Snapshots?**

- **VMFS Upgrades?**

- **VCB Configurations?**

The End