

ESX Server Storage II

Tips and Tricks

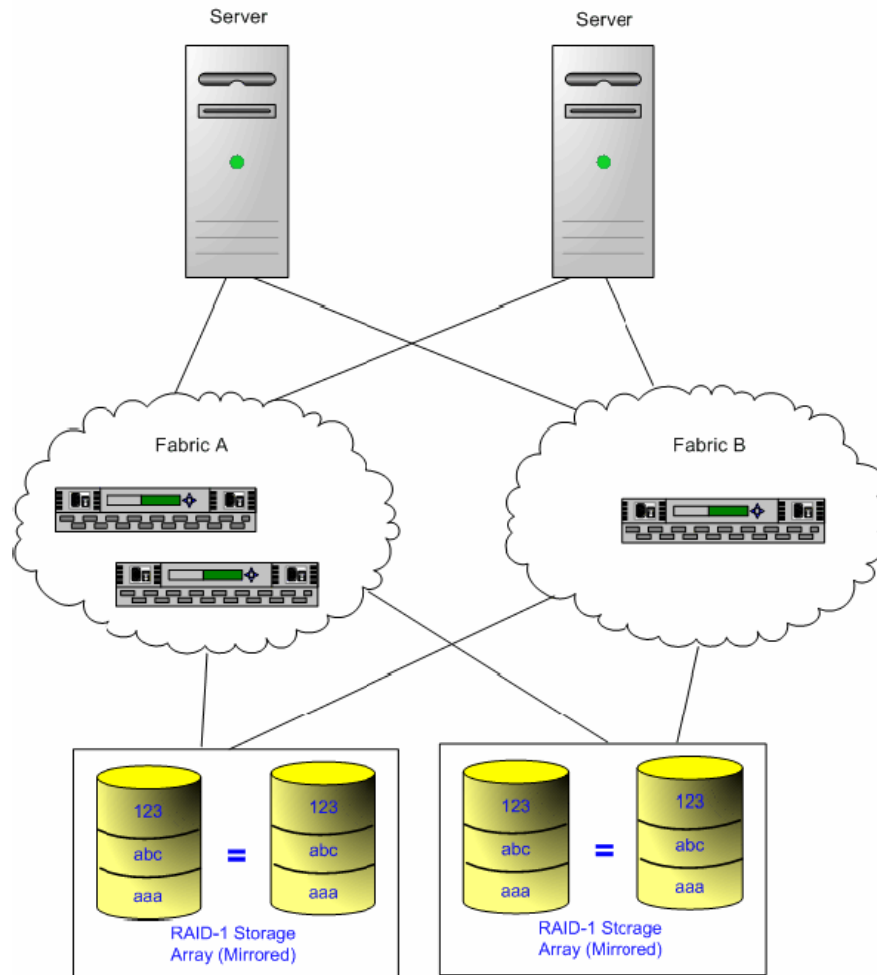
Raw Disk Mapping

Mostafa Khalil and Bob Slovick

What We Will Talk About

- SAN connectivity
- Fibre attached devices
- RAW LUNs
- Using RAW Device Mapping (RDM)
- RDM tips and tricks

SAN Connectivity



Fibre Attached RAW Devices

- Tape Devices
 - Fibre Attached Tape Devices are not supported by ESX Server 2.5.x
 - Possible future support in ESX Server 3.x
- SAN LUNs
 - LUNs presented to the server's HBAs are accessible by VMkernel and VMnix

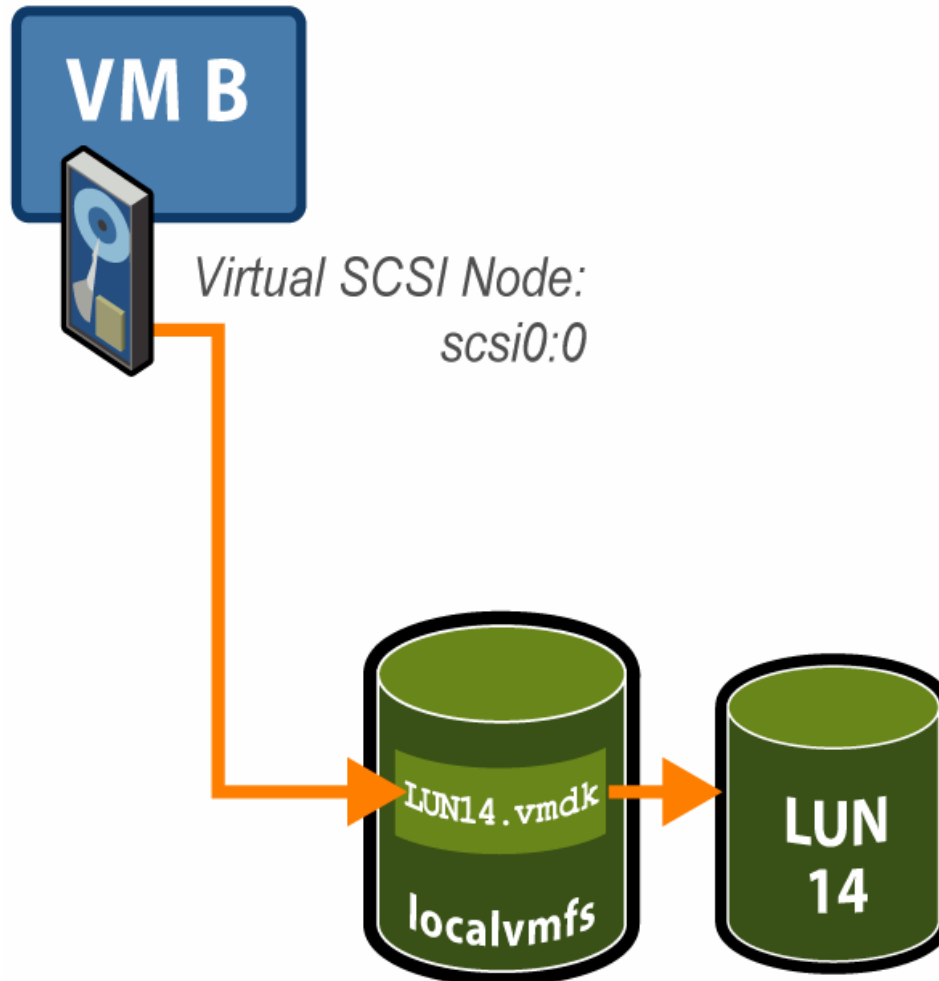
ESX Server RAW Device Mapping

- ESX Server 2.5.x
- Types of RDM
 - Physical compatibility mode
 - Virtual compatibility mode
- Not with shared HBAs

Physical Compatibility Mode

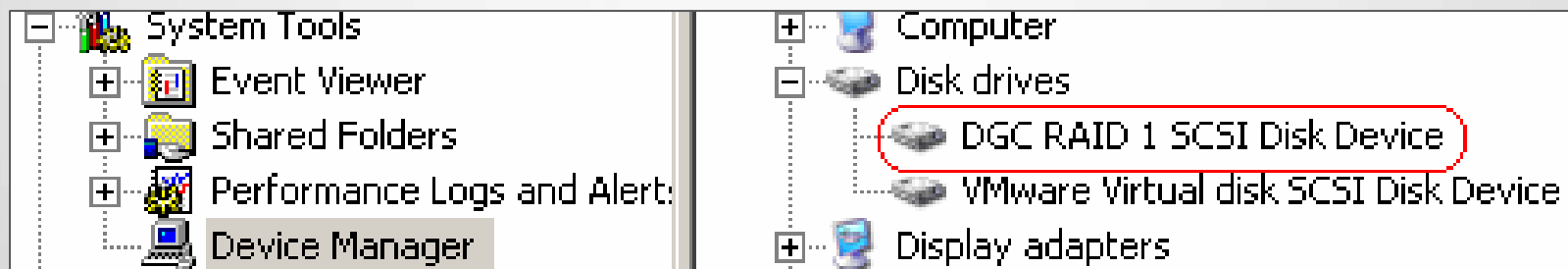
- AKA: SCSI-Passthru-rdm
 - Allows virtual machines to pass SCSI commands directly to the physical hardware
 - RAW Disk Mapping file is created on a VMFS volume. This mapping file, which has a vmdk extension, points to the raw device
 - RDM file lock is translated to SCSI Reservation on the LUN
 - Utilities like admsnap and admhost, installed on the virtual machines, can directly access the raw device/partition

RDM 'Pass-Thru' Mode



View From the Virtual Machine

- How a Passthru-rdm is seen in the guest OS



View From the ESX Server Console

- How an RDM file is seen on the VMFS volume
 - The file name points to the RAW LUN
 - The size reflects the LUN Size with Zero blocks used
 - Attribute is “raw disk”

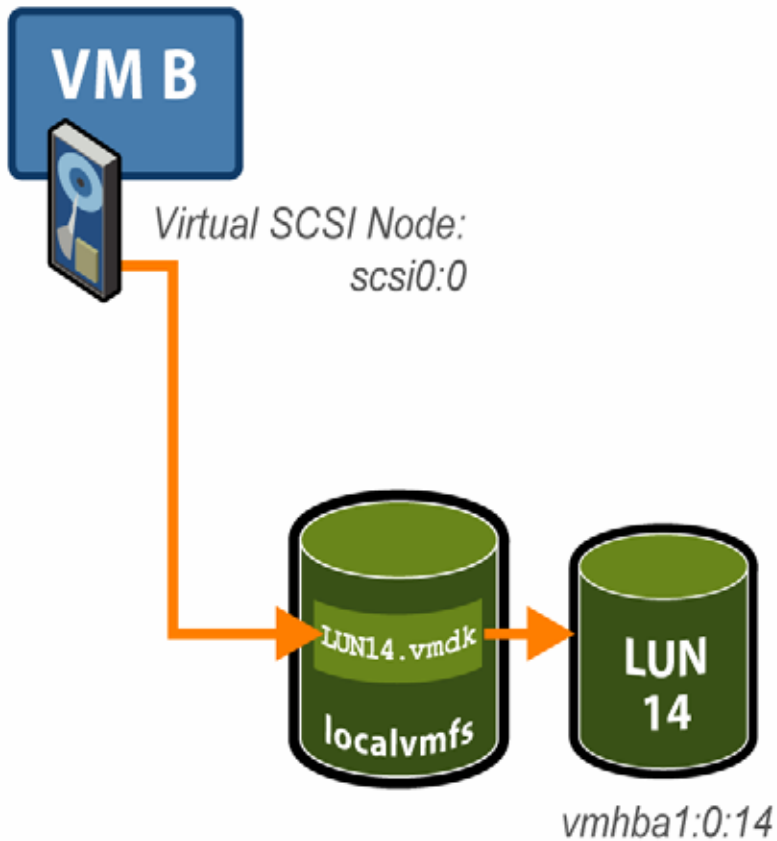
```
# vmkfstools -lM IBM-80
Name: IBM-80      (public)
Capacity 85896566784 (81910 file blocks * 1048576), 85888860160 (81910 blocks) avail
Permission  Uid   Gid   Attr   Bytes (Blocks) Last Modified Filename
rw-----   0     0   raw disk 10737418240 (  0) Oct  1 17:30 ntfs-rdm.vmdk -> vmhba2:0:2:0
```

Virtual Compatibility Mode

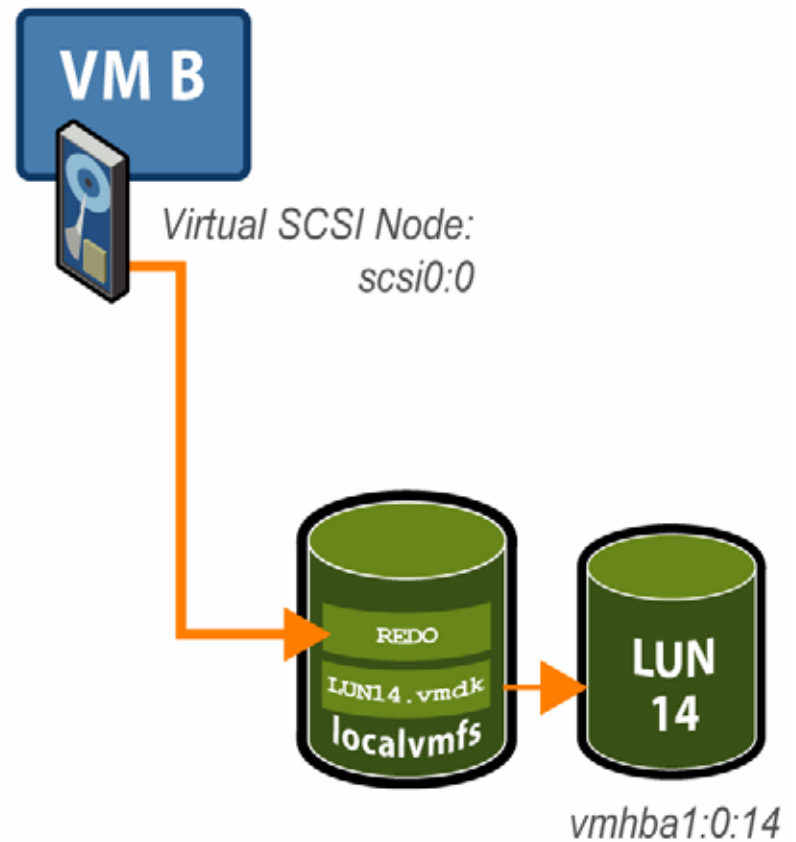
- AKA: SCSI-non-Passthru
- Full virtualization of the mapped device. It appears to the guest operating system exactly the same as a virtual disk file in a VMFS volume
- The real hardware characteristics are hidden
- VMFS advanced file locking for data protection and redo logs

RDM 'Non-Pass-Thru' Mode

Persistent Mode

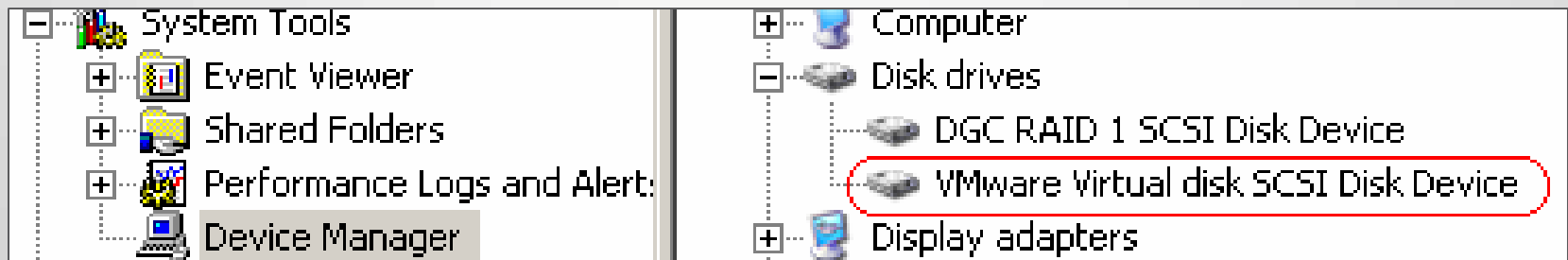


Undoable Mode



View From the Virtual Machine

- How a non-Passthru-rdm is seen in the guest OS



VMFS vs. RDM

RDM Advantages

- Virtual machine partitions are stored in the native guest OS file system format, facilitating “layered applications” that need this level of access
- As there is only one virtual machine on a LUN, you have much finer grain characterization of the “storage container”, and no I/O or SCSI reservation lock contention at a LUN level. The LUN can be designed and configured for optimal performance
- With “Virtual Compatibility” mode, virtual machines have many of the features of being on a VMFS, such as file locking to allow multiple access, redo logs, and VMotion

VMFS vs. RDM

RDM Advantages

- With “Physical Compatibility” mode, it gives a virtual machine the capability of sending almost all “low-level” SCSI commands to the target device, including command and control to a storage controller, such as through SAN Management agents in the virtual machine. An example of this is where a virtual machine tells a storage controller to invoke a snapshot
- Dynamic Name Resolution: Stores unique information about LUN regardless of changes to physical address changes due to hardware or path changes

VMFS vs. RDM

RDM Disadvantages

- Not available for block or RAID devices that do not report a SCSI serial number
- Not available if FC-HBA shared with Service Console
- No redo logs in “Physical Compatibility” mode, only available in “Virtual Compatibility” mode

VMFS vs. RDM

- If you think you may need the added functionality of RDM, and can live with the downsides, consider deploying RDM early on
- It can be difficult to migrate from VMFS to RDM and vice versa

Identifying RAW Devices

- Locate the target
 - `wwpn.pl -v`

```
WWPN 1.02 Copyright VMware 2003
Display WW port names and VMHBA information for fibre channel adapters
For each vmhba here are the corresponding Qlogic and Emulex WW Port Names
Adapter   WWPN                               PCI (decimal)
vmhba2: 210000e08b19fa44 (Qlogic) 6:1:0 /proc/scsi/qla2300/0
vmhba2:0: 5006016030204db2 scsi-qla0-port-0=50060160b0204db2:5006016030204db2;
vmhba2:1: 5006016830204db2 scsi-qla0-port-1=50060160b0204db2:5006016830204db2;
vmhba2:2: 200600a0b8174edc scsi-qla0-port-2=200600a0b8174edb:200600a0b8174edc;
vmhba2:3: 200700a0b8174edc scsi-qla0-port-3=200600a0b8174edb:200700a0b8174edc;
```

Identifying RAW Devices

- Locate the LUN's primary HBA
 - vmkmultipath -q

```
Disk and multipath information follows:
```

```
Disk vmhba0:0:0 (208,378 MB) has only 1 path.
```

```
Disk vmhba2:0:0 (81,917 MB) has 2 paths. Policy is mru.
```

```
vmhba2:0:0          on (preferred)
```

```
vmhba2:1:0          on (active)
```

```
Disk vmhba2:0:2 (10,240 MB) has 2 paths. Policy is mru.
```

```
vmhba2:0:2          on (active, preferred)
```

```
vmhba2:1:2          on
```

```
Disk vmhba2:2:1 (81,917 MB) has 2 paths. Policy is mru.
```

```
vmhba2:2:1          on (active, preferred)
```

```
vmhba2:3:1          on
```

Identifying RAW Devices

- Locate the LUN's Proc Node
 - #ls /proc/vmware/scsi/vmhba2

```
# ls /proc/vmware/scsi/vmhba2  
0:0 0:2 2:1 2:2 2:3 stats
```

Identifying RAW Devices

- List the LUN's partitions
 - #cat 0:2

```
Vendor: DGC          Model: RAID 1          Rev: 0217
Type:   Direct-Access          ANE
Id: 60 6 1 60 f9 a0 11 0 ba c4 20 98 d2 32 da 11 52 4
Size:   10240 Mbytes
Queue Depth: 32
```

```
Partition Info:
Block size: 512
Num Blocks: 20971520
```

num:	Start	Size	Type
1:	63	20964762	7

Identifying RAW Devices

- Virtual machine configuration


```
scsi0:1.present = "TRUE"  
scsi0:1.name = "IBM-80:ntfs-rdm.vmdk"  
scsi0:1.deviceType = "scsi-passthru-rdm"
```

How to Create RDM






- Using MUI
- Using Virtual Center
- Using vmkfstools

Creating an RDM Using the MUI

- Use the “Configure Hardware” option to add a disk


**Add Device**
What type of hardware do you want to install?

Device Type

 Hard Disk	Create a new virtual disk, use an existing virtual disk or access a VMFS volume directly.
 Network Adapter	Attach a new network adapter to a virtual network of your choosing.
 DVD/CD-ROM Drive	Create a new virtual DVD or CD-ROM drive to access a system drive or an ISO image.
 Floppy Drive	Create a new virtual floppy drive to access a system drive or a floppy image file.
 Generic SCSI Device	Create a new virtual device that directly accesses a system SCSI device.

Creating an RDM Using the MUI

- Select “System LUN/disk” option

**Virtual Disk**
Add a hard disk to your virtual machine.

Which type of virtual disk would you like to add?

Blank	Create a new virtual disk.
Existing	Attach an existing virtual disk to your virtual machine.
System LUN/Disk	Give your virtual machine direct access to a SAN LUN.

Creating an RDM Using the MUI

- The choice is yours!

Virtual Disk
Add a hard disk to your virtual machine.

Edit Virtual Disk Configuration

System LUN/Disk

Target LUN/Disk: vmhba2:0:2: 10.0 G (Partitions: 1)

Disk Metadata

Use Metadata:

Metadata File Location: IBM-80 (vmhba2:2:1:1): 80.0 G free

Metadata File Name: ntfs-rdm.vmdk

Virtual Device

Virtual SCSI Node: 0:1

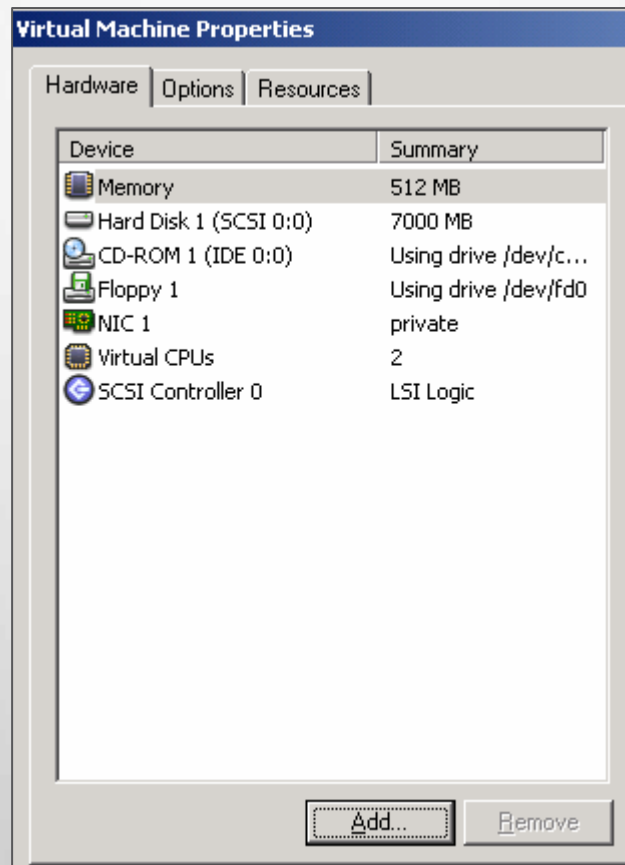
Compatibility

Physical Allow the guest operating system to access the underlying hardware directly.

Virtual Allow the virtual machine to take advantage of disk modes and other features of virtual disks.

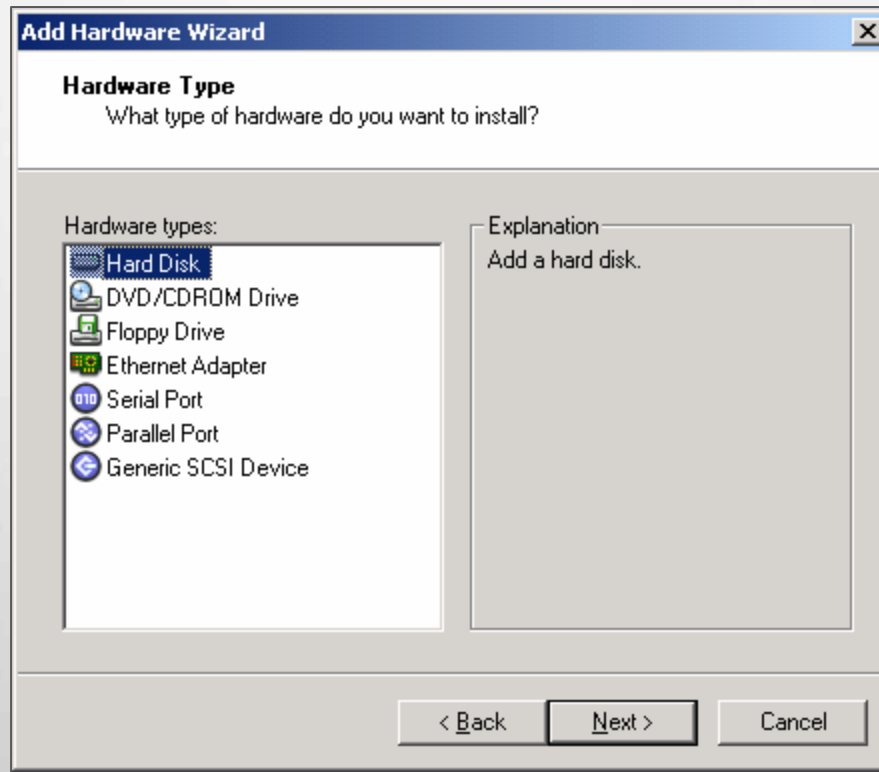
Creating an RDM Using the MUI

- Select “Add” option from “Edit Properties” menu



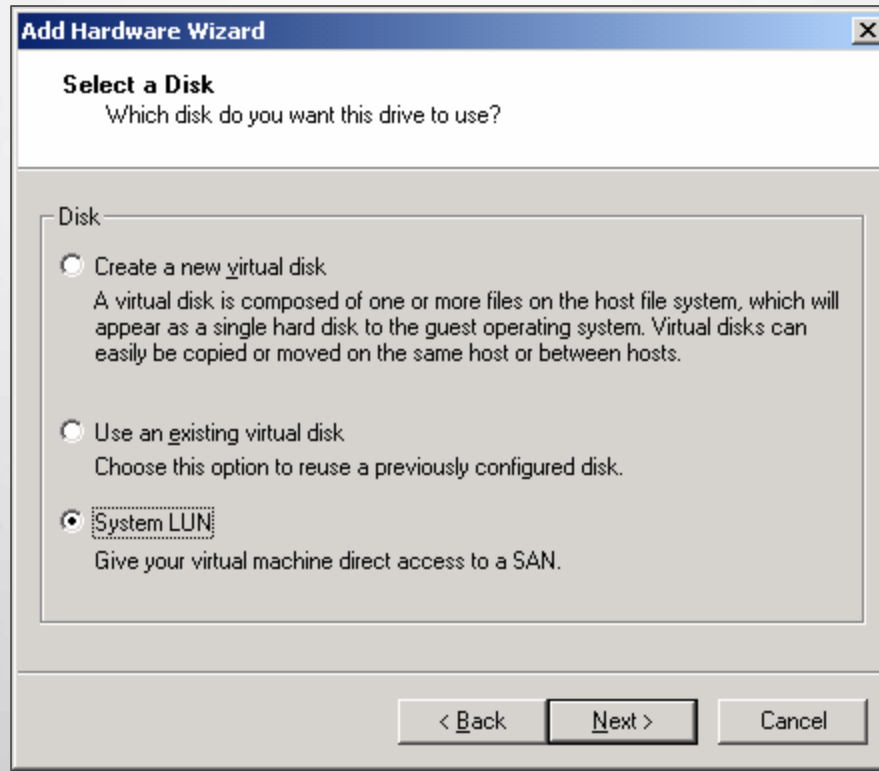
Creating an RDM Using the MUI

- Select “Hard Disk”



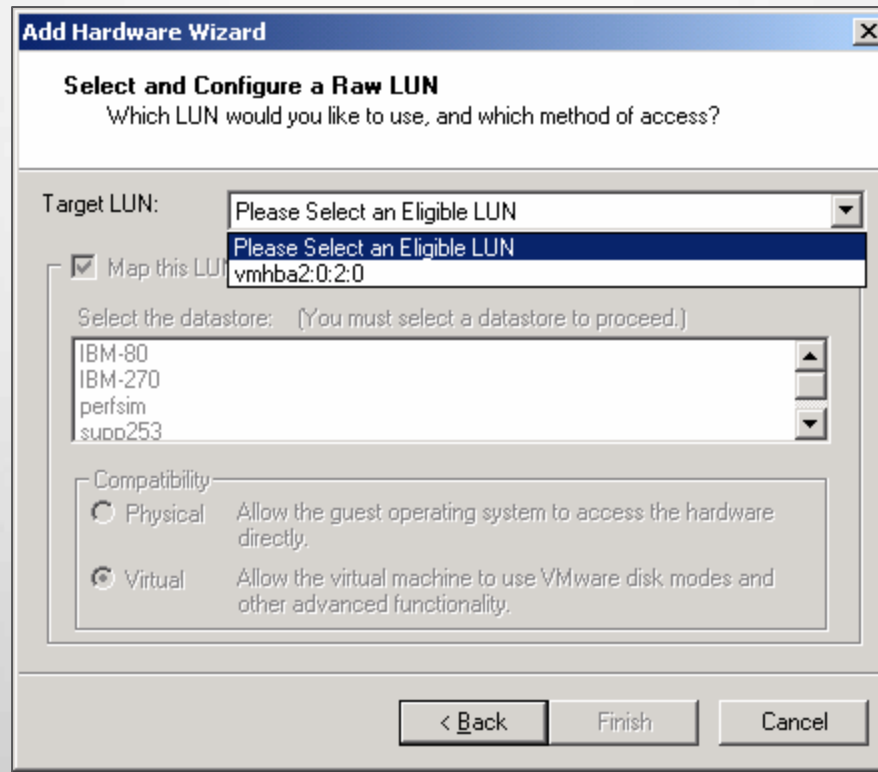
Creating an RDM Using the MUI

- Select “System LUN”



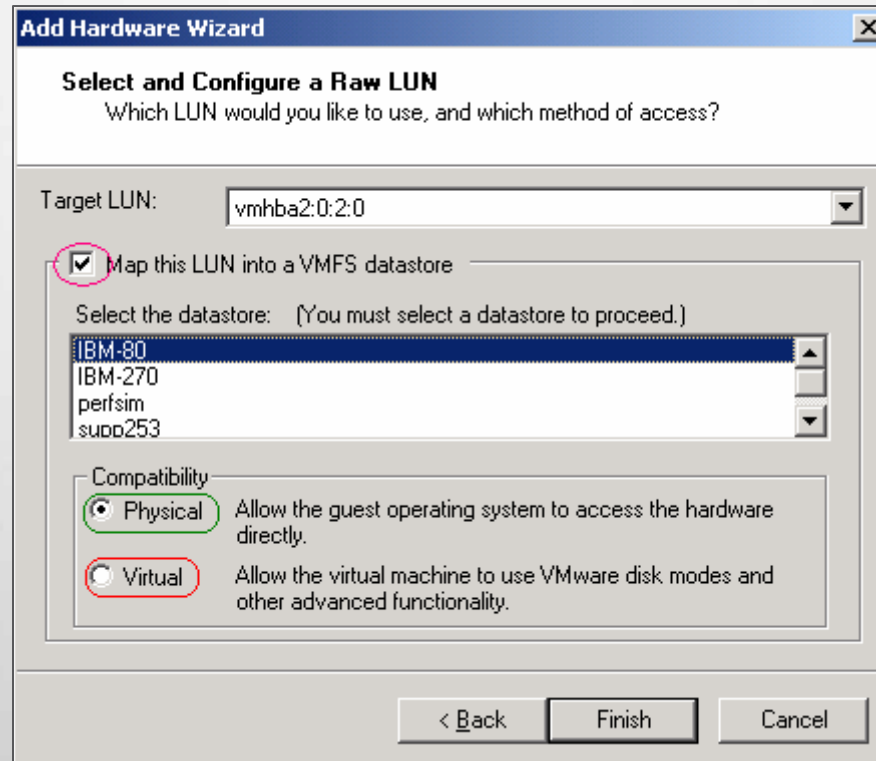
Creating an RDM Using the MUI

- Select the RAW LUN from the “Target LUN” menu



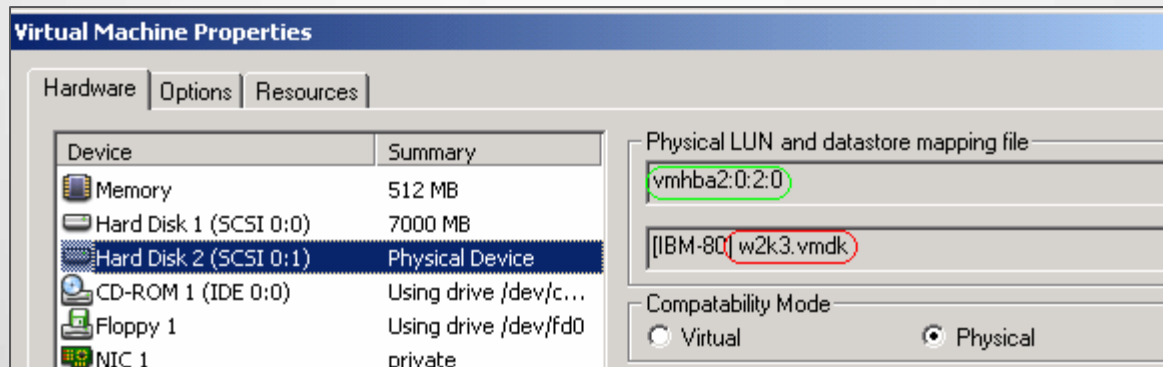
Creating an RDM Using the MUI

- Again! The choice is yours!



Creating an RDM Using the MUI

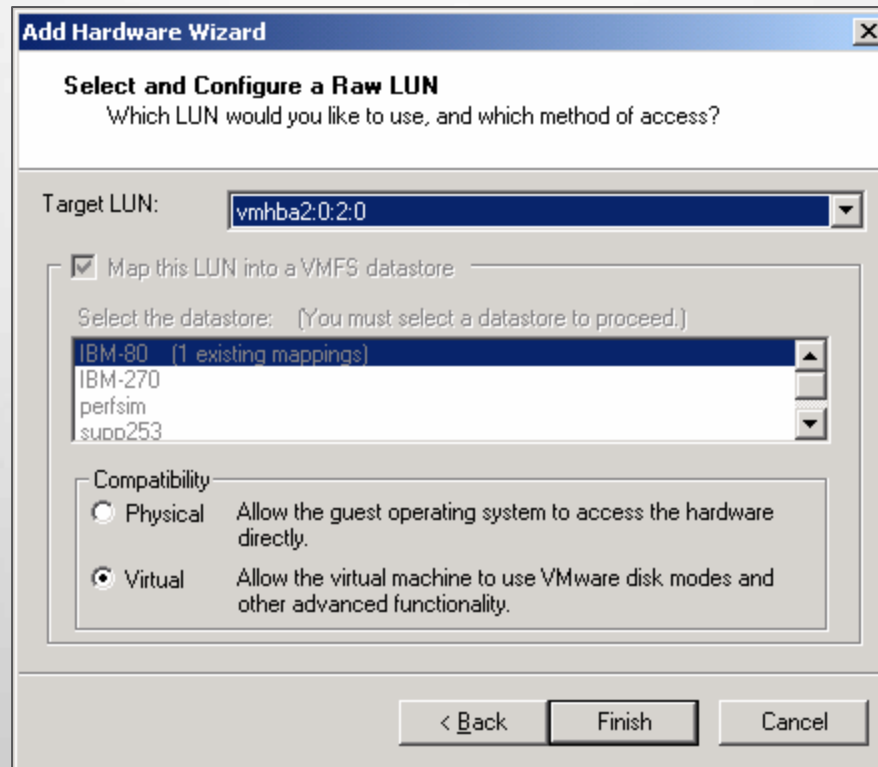
- Note that you do not have the choice to specify the RDM file name



You can locate the RDM file name here

Creating an RDM Using the MUI

- If the RDM file already exists, you only have a choice of compatibility mode



Creating an RDM Using vmkfstools

```
vmkfstools -r <raw-device> <vmfs-  
volume>:<rdm-name>.vmdk
```

```
# vmkfstools -r vmhba2:0:2:0 IBM-80:ntfs-rdm.vmdk  
# vmkfstools -lM IBM-80  
Name: IBM-80      (public)  
Capacity 85896566784 (81910 file blocks * 1048576), 85888860160 (81910 blocks) avail  
Permission Vid   Gid      Attr      Bytes (Blocks) Last Modified Filename  
rw-----    0      0      raw disk 10737418240 (    0) Oct  3 09:48 ntfs-rdm.vmdk -> vmhba2:0:2:0
```

Check the VMFS volume for the resulting file

```
vmkfstools -lM <vmfs-volume>
```

Then, you need to modify the virtual machine to use that RDM file.

Use With Clustering

	VMFS	RDM	Bus sharing
Cluster in a box	Public	Virtual	Virtual
Cluster across boxes (opt.1)	Shared	Physical	Physical
Cluster across boxes (opt. 2)	Shared	Virtual	Physical
N+1 Cluster	Public	Physical	Physical

Do not use C.A.B. option 2

Use With Clustering

- To add an RDM for N+1 configuration, use this command sequence:

```
vmkfstools -L lunreset vmhbax.y.z.0; vmkfstools -r  
vmhbax.y.z.0 path_of_new_rdm
```

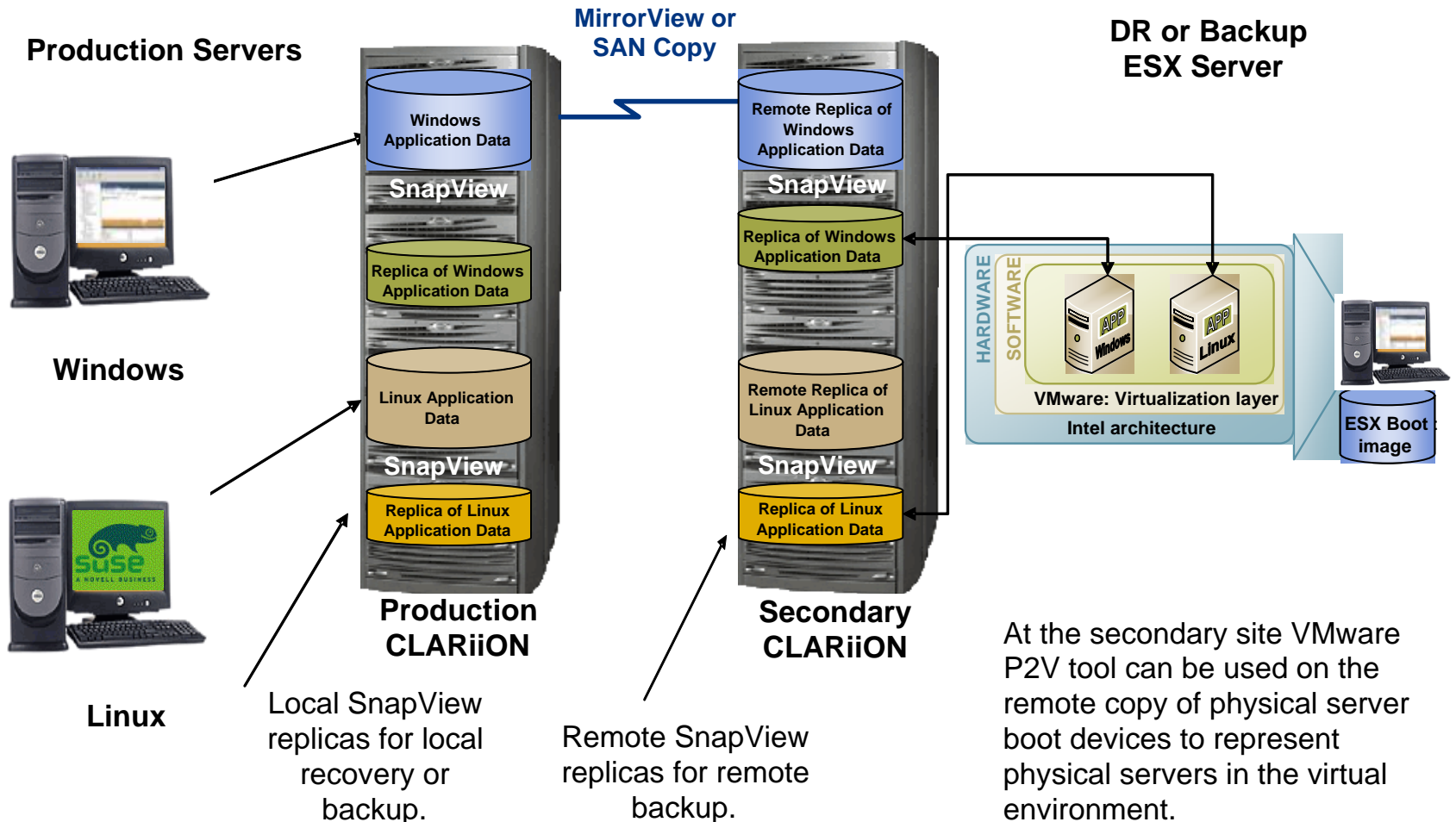
All on the same line.

This temporarily resets the SCSI reservation on the LUN to allow you to create the RDM file.

Use With Multilayered Applications

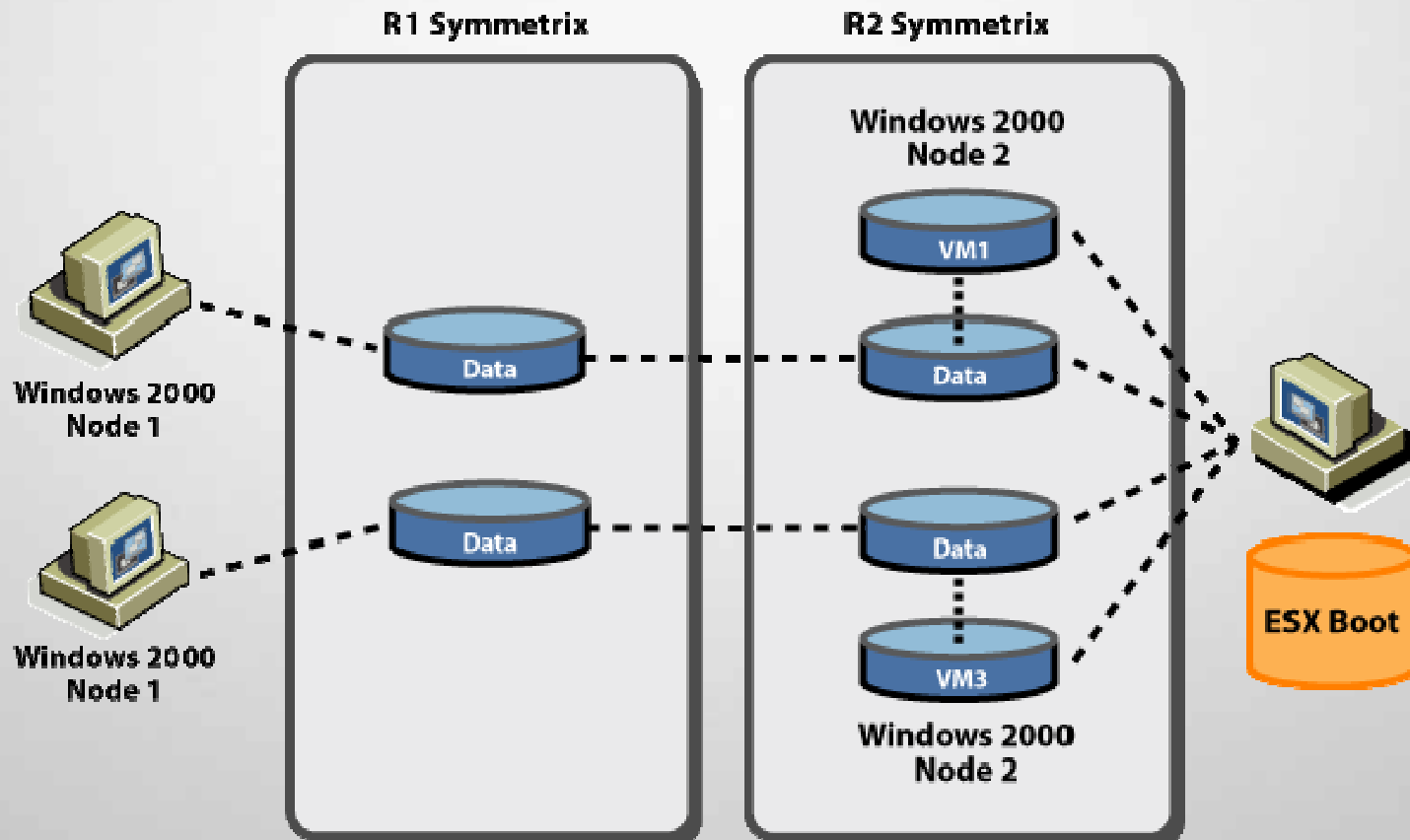
- Requires “Physical Compatibility” mode
- Clariion Requires FLARE 19
- White Paper at:
http://www.vmware.com/pdf/clariion_wp_eng.pdf
- Check out SLN012 session. Dan Weiss will discuss using Layered App with Exchange Server using RDM

Using RDM For DR



Use With SRDF

SRDF/CE for MSCS Physical to Virtual



Tips and Tricks

- How to list RDM files

```
root@supp253 root # vmkfstools -lM /vmfs/IBM-80
Name: IBM-80      (public)
Capacity 85896566784 (81910 file blocks * 1048576), 85888860160 (81910 blocks) avail
Permission  Uid   Gid   Attr   Bytes (Blocks) Last Modified Filename
rw-----   0     0   raw disk 10737418240 (    0)  Oct  1 17:30 ntfs-rdm.vmdk -> vmhba2:0:2:0
```

Tips and Tricks

- Importing a virtual disk to RDM

```
vmkfstools -i <virtual-disk> <target rdm>
```

Example:

```
vmkfstools -i /vmfs/localvmfs/win2000.vmdk  
/vmfs/san01/w2k-rdm.vmdk
```

Imports win2000.vmdk to w2k-rdm.vmdk which maps to a
RAW LUN

Upon completion of import, you get prompted with the disk
geometry to use in the virtual machine configuration

Tips and Tricks

- Using public volume with C.A.B
 - Use vmkfstools to create 2 RDM files mapped to the same LUN (one for each cluster node)

```
# vmkfstools -r vmhba2:0:2:0 IBM-80:ntfs-rdm.vmdk
# vmkfstools -lM IBM-80
Name: IBM-80      (public)
Capacity 85896566784 (81910 file blocks * 1048576), 85888860160 (81910 blocks) avail
Permission  Vid   Gid   Attr   Bytes (Blocks) Last Modified Filename
rw-----   0     0   raw disk 10737418240 (  0) Oct  3 09:48 ntfs-rdm.vmdk -> vmhba2:0:2:0
# vmkfstools -r vmhba2:0:2:0 IBM-80:ntfs-rdm2.vmdk
# vmkfstools -lM IBM-80
Name: IBM-80      (public)
Capacity 85896566784 (81910 file blocks * 1048576), 85888860160 (81910 blocks) avail
Permission  Vid   Gid   Attr   Bytes (Blocks) Last Modified Filename
rw-----   0     0   raw disk 10737418240 (  0) Oct  3 09:48 ntfs-rdm.vmdk -> vmhba2:0:2:0
rw-----   0     0   raw disk 10737418240 (  0) Oct  3 10:01 ntfs-rdm2.vmdk -> vmhba2:0:2:0
```

Tips and Tricks

- How to delete an RDM

```
# rm -rf /vmfs/IBM-80/ntfs-rdm.vmdk
```

Make sure you are deleting the correct file

Tips and Tricks

- Booting guest from an RDM
- Use P2V to reconfigure it

Where to Get More Information

- http://www.vmware.com/pdf/clariion_wp_eng.pdf
- http://www.vmware.com/pdf/esx_SAN_guide.pdf
- http://www.vmware.com/pdf/esx25_rawdevicemapping.pdf
- http://www.vmware.com/pdf/SRDF_wp_eng.pdf
- http://www.vmware.com/support/esx25/doc/admin/esx25admin_cluster.html#998201

Backup Slides

VMFS vs. RDM

VMFS Advantages

- Several virtual machines and ESX Servers can share a common LUN. This means less LUNs to administer, and there is a limit of no more than 128 LUN presented to an ESX Server. This also supports VMotion
- A virtual machine's partition is represented as a single file, making it very "portable"
- They use file level locks to prevent concurrent access to the same file
- They use SCSI reservation locks to protect file system metadata integrity

VMFS vs. RDM

VMFS Disadvantages

- Several virtual machines and ESX Servers can share a common LUN, leading to possible I/O contention
- A virtual machines partition is represented as a single file that “layered applications”, that need to access data at a native file system level, cannot
- Difficult to characterize disk I/O for a particular application, as all the virtual machines share the same “storage container”. Random I/O can be significant as virtual machines are scheduled. Read cache becomes less effective, due to poor “data locality”. Write cache can become saturated
- Redo logs can generate a great deal of “SCSI reservation lock” activity, as metadata needs to be locked to “grow the file”...

Questions?

VMworld2005

virtualize^{now}

las vegas • october 18-20, 2005