VMworld2005 virtualizenow

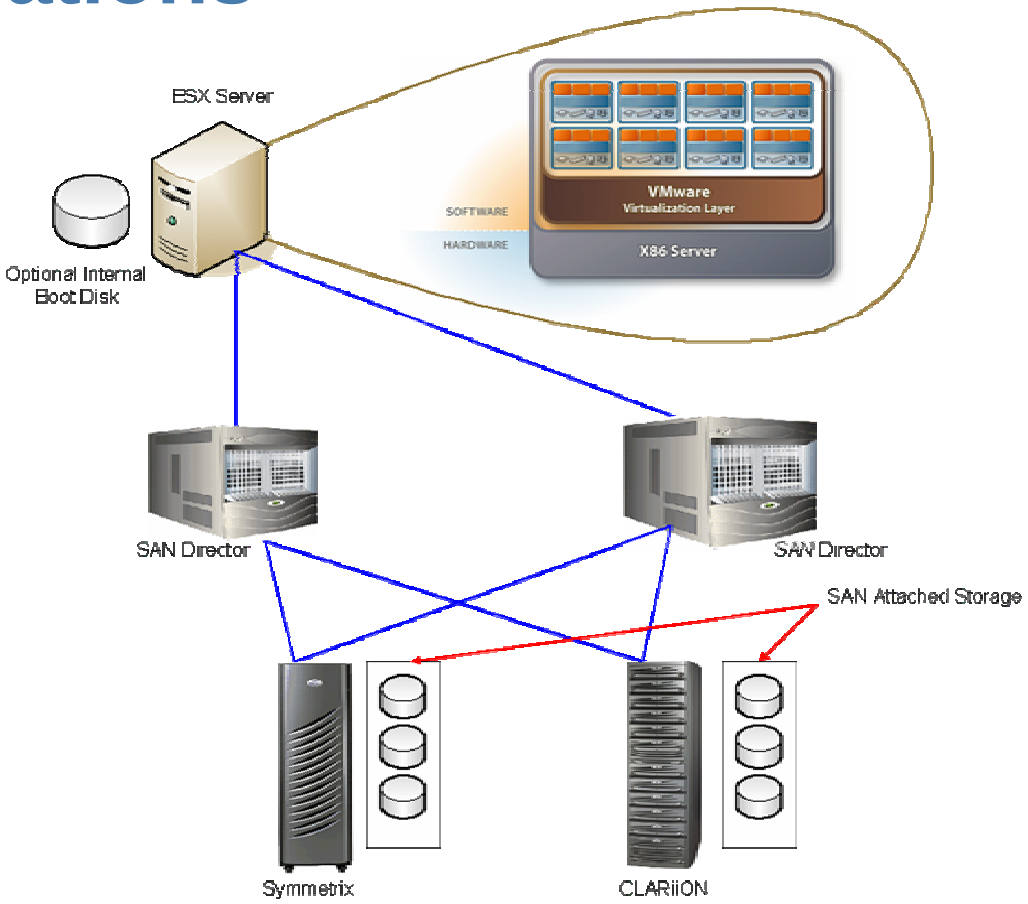
Storage Best Practices for VMware ESX Server 2.x

Bala Ganeshan
Corporate System Engineer
EMC Database Solutions Team

Agenda

- Introduction
- ESX Servers on SAN
- Booting ESX Servers of EMC Storage Arrays
- Using ESX Server v2.x on EMC Storage Arrays
- Using ESX Server v2.x with EMC Symmetrix Arrays
- Backup and recovery considerations
- Related sessions
- Question and answer session

SAN Attached ESX Server Configurations



Best Practices For Using EMC Storage With VMware ESX Server

- Recommendations to allow scalability
- Address potential performance limiting characteristics in the future
- May not be appropriate for all environments
- Recommendations enable optimal balance between performance, management and functionality

SAN Boot of ESX Servers

Booting ESX Server Off the SAN

- ESX Server 2.5 and later supports booting off the SAN
- Refer to “EMC Host Connectivity Guide for ESX Server 2.x” for configuration details
- ESX Servers can be booted off both Symmetrix and CLARiiON storage arrays
  - Advantages
    - Boot device protected on enterprise storage
    - x86 Servers become appliances (can be replaced in case of failures)
    - Leverage best practices for backing and restoring off the SAN

Booting ESX Server Off the SAN

- Disadvantages
  - Dedicated HBAs needed if RDM are to be used
    - Two different type of HBAs would be needed
    - Impacts VMkernel heap usage
    - ***Not supported by EMC***
  - Can impact management functions if HBAs are shared with service console
  - Potential performance impact if swap files are on the SAN disk
    - If internal disks are available, the swap file can be on protected internal disks.

Using ESX Servers with EMC Storage Arrays

Storage Layout Considerations

Storage Layout Considerations

- Do not present SAN storage to ESX Server farm as one big SCSI disk
  - Use multiple meta volumes or meta LUNs to present the storage
  - Plan on presenting storage as 8-10 meta volumes or meta LUNs
- Both meta volumes and meta LUNs can be grown non-disruptively
- Use of spanned VMFS with dynamic LUN growth mitigates risk

Storage Layout Considerations, cont.

	Storage as Single LUNs	Storage as Multiple LUNs
Management	▪ Easier management ▪ Storage can be over provisioned ▪ One VMFS to manage	▪ Slightly harder management. ▪ Storage provisioning has to be on demand ▪ One VMFS to manage (spanned)
Performance	▪ Can result in poor response time ▪ No manual load balancing	▪ Multiple queues ensure minimal response times ▪ Manual load balancing
Scalability	▪ Limited # of virtual machines due to response time issue ▪ Limited # of IO intensive virtual machines since one VMFS	▪ Multiple VMFS allows more virtual machines per ESX Server ▪ Response time of limited concern (can optimize)
Functionality	▪ All virtual machines share one LUN ▪ Cannot leverage ALL available storage functionality	▪ Use VMFS when storage functionality not needed ▪ Judicious use of RDM vs. VMFS

Storage Layout Considerations, cont.

- Use RDM (raw disk mapping) instead of VMFS if advanced storage functionality is desired
  - Required if using CLARiiON storage array application
- Use physical disk compatibility mode for RDMs
  - Enables VMotion while allowing virtual machines to recognize storage disk characteristics

Storage Layout Considerations, cont.

- If practical present multiple VMFS per ESX Server
  - Ideally 4-5 VMFS per ESX Server
  - Separate boot disk and application data
  - Separate log and data volumes
  - Separate classes of applications
- With ESX Server 2.x, VMFS can span multiple physical disks and/or partitions
  - Can be done using “vmkfstools” or MUI

Storage Layout Considerations, cont.

- Span VMFS across SAN volumes if needed
  - Advantages
    - Allows ESX Server administrator to still manage storage as one entity
    - Allows VMkernel to schedule multiple IOs to underlying disks
    - Provides flexibility to both ESX Server and storage administrator in assigning storage

Storage Layout Considerations, cont.

- Disadvantages:
  - Limits the number of IO intensive virtual machines on a ESX Server farm
    - Approximately 32 IO intensive virtual machines per VMFS
  - **Potential availability issue. Loss of one member of spanned VMFS will cause disruption and loss of all data on that VMFS**
    - Not a major issue with enterprise storage
    - Ensure VMFS does not span different storage class
  - Filesystem is spanned. IOs not balanced across VMFS “physical extents”

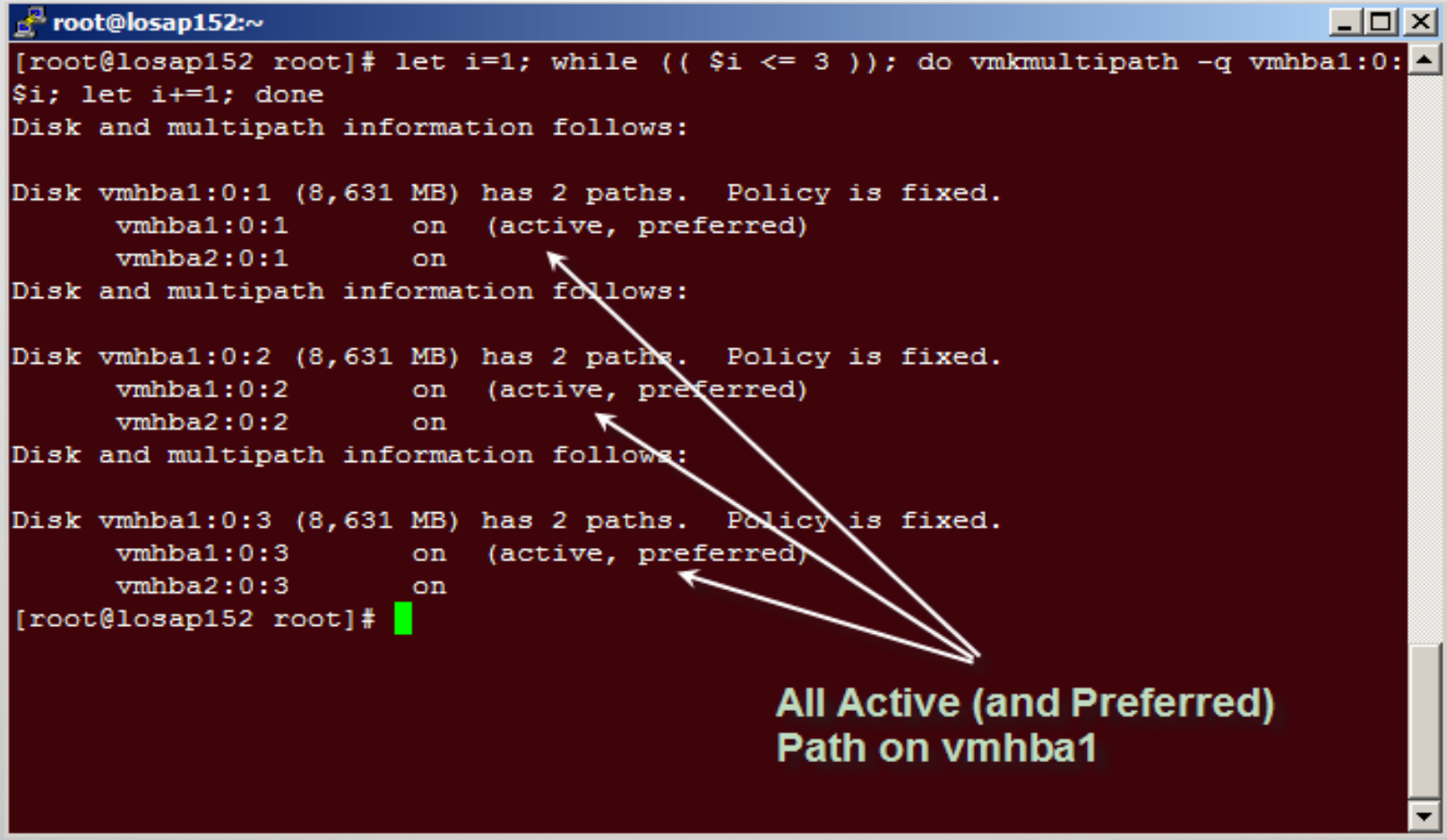
Using ESX Servers with EMC Storage Arrays

Path Management Considerations

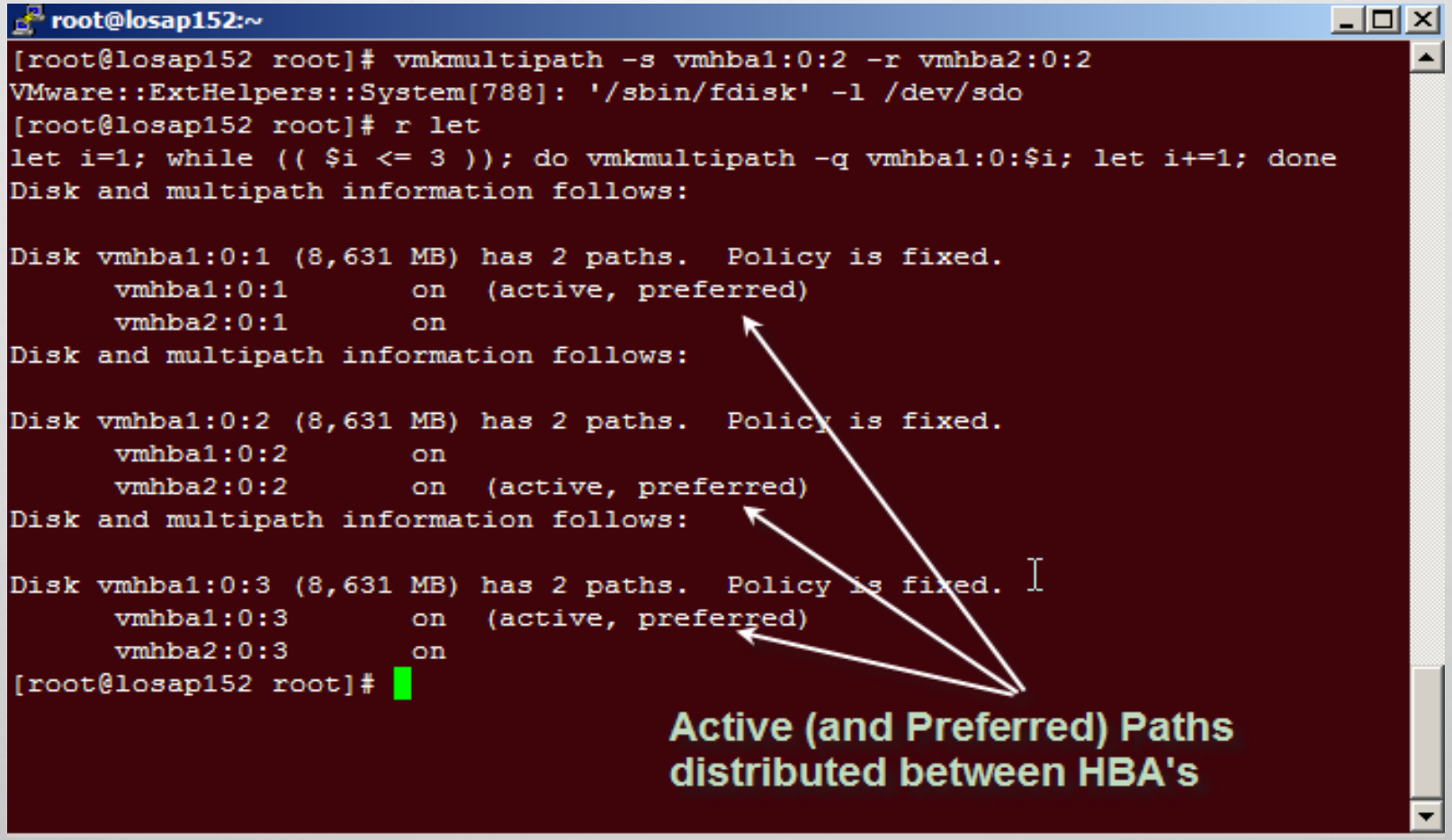
Path Management Considerations

- PowerPath is not supported
  - If applicable, uninstall PowerPath from virtual machine after P2V is run
- ESX Server 2.x does not support active load balancing. Only path failover is available
- Set preferred path to enable some level of load balancing across HBAs
- Statistics will continue to be assigned on the first path discovered by Vmkernel
  - However, IOs will be directed to the active path

Path Management Considerations, cont.



Path Management Considerations, cont.



Using ESX Servers with EMC Storage Arrays

Partition Alignment

Alignment Issue on ESX Server

- ESX Server exhibits the same alignment problems seen on ALL Intel based platforms
- Legacy BIOS code from IBM PC
  - Used Cylinder, Head and Sector addressing instead of LBA addressing
  - Cylinder 0 – 1023, head 0-254, sector 1-63
  - Our reported geometry is 15 heads and 64 sectors
  - HBA BIOS code use mapping: e.g. 30 heads and 32 sectors

LBA	BIOS Mapping
Cyl 0, head 0, sec 1-64	Cyl 0, head 0, sec 1-63 AND cyl 0 head 1, sec 1
Cyl 0, head 1 , sec 1-64	Cyl 0, head 1, sec 2-63 AND cyl 0, head 2, sec 1-2
Cyl 0, head 2, sec 1-64	Cyl 0, head 2, sec 3-63 AND cyl 0, head 3, sec 1-3

Alignment Issue on ESX Server

- First track is reserved for boot code
- Means first partition starts at cyl 0 head 1 sector 1
- This is LBA 63 and is therefore, unaligned
- **As environments grow it can become performance limiting**
- To prevent this misalignment:
  - Offset the starting point of the partition to specified block
  - Create an aligned dummy partition
- Presentation shows how this can be done
  - Offset method on Linux
  - Creating dummy partition on Windows
- Either methodology can be used in both operating systems

Alignment Issue on ESX Server

- Default block size for VMFS is 1 MB
  - Use fdisk on service console to ensure virtual disks are aligned
  - Guest OS can misalign the IOs. Use diskpar or fdisk to align on the guest
- Use diskpar if RDM is presented to Windows as guest OS
- Use fdisk to align partitions on RDM presented to Linux guest OS

Using `fdisk` on Service Console to Align VMFS

- By default, ESX Server will create VMFS that are misaligned
- Execute the following steps to align VMFS
  1. On service console, execute “fdisk /dev/sd<x>”, where sd<x> is the device on which you would like to create the VMFS
  2. Type “n” to create a new partition
  3. Type “p” to create a primary partition
  4. Type “1” to create partition #1
  5. Select the defaults to use the complete disk
  6. Type “x” to get into expert mode
  7. Type “b” to specify the starting block for partitions
  8. Type “1” to select partition #1
  9. Type “128” to make partition #1 to align on 64KB boundary
  10. Type “r” to return to main menu
  11. Type “t” to change partition type

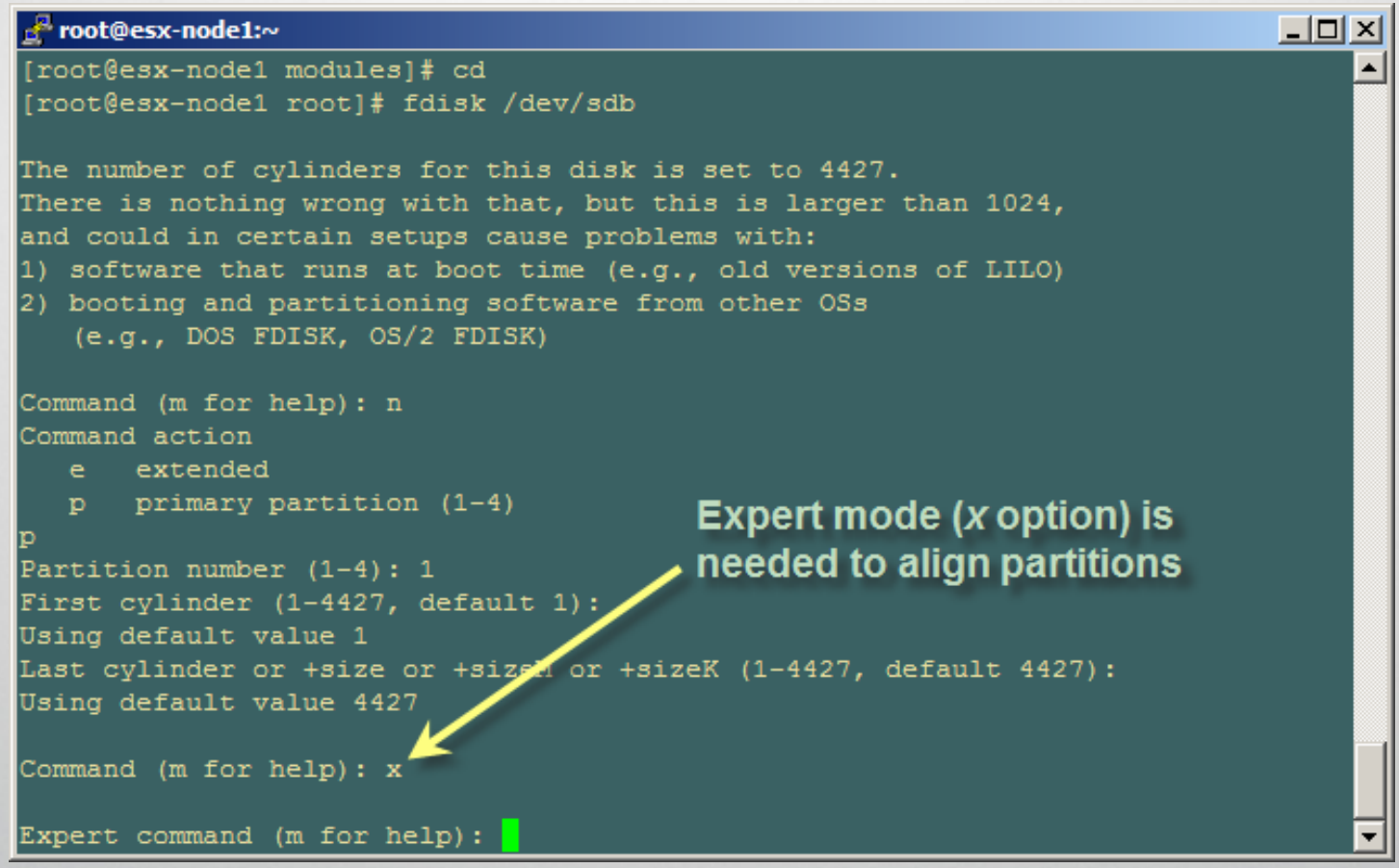
Continued on next page

Using `fdisk` on Service Console to Align VMFS

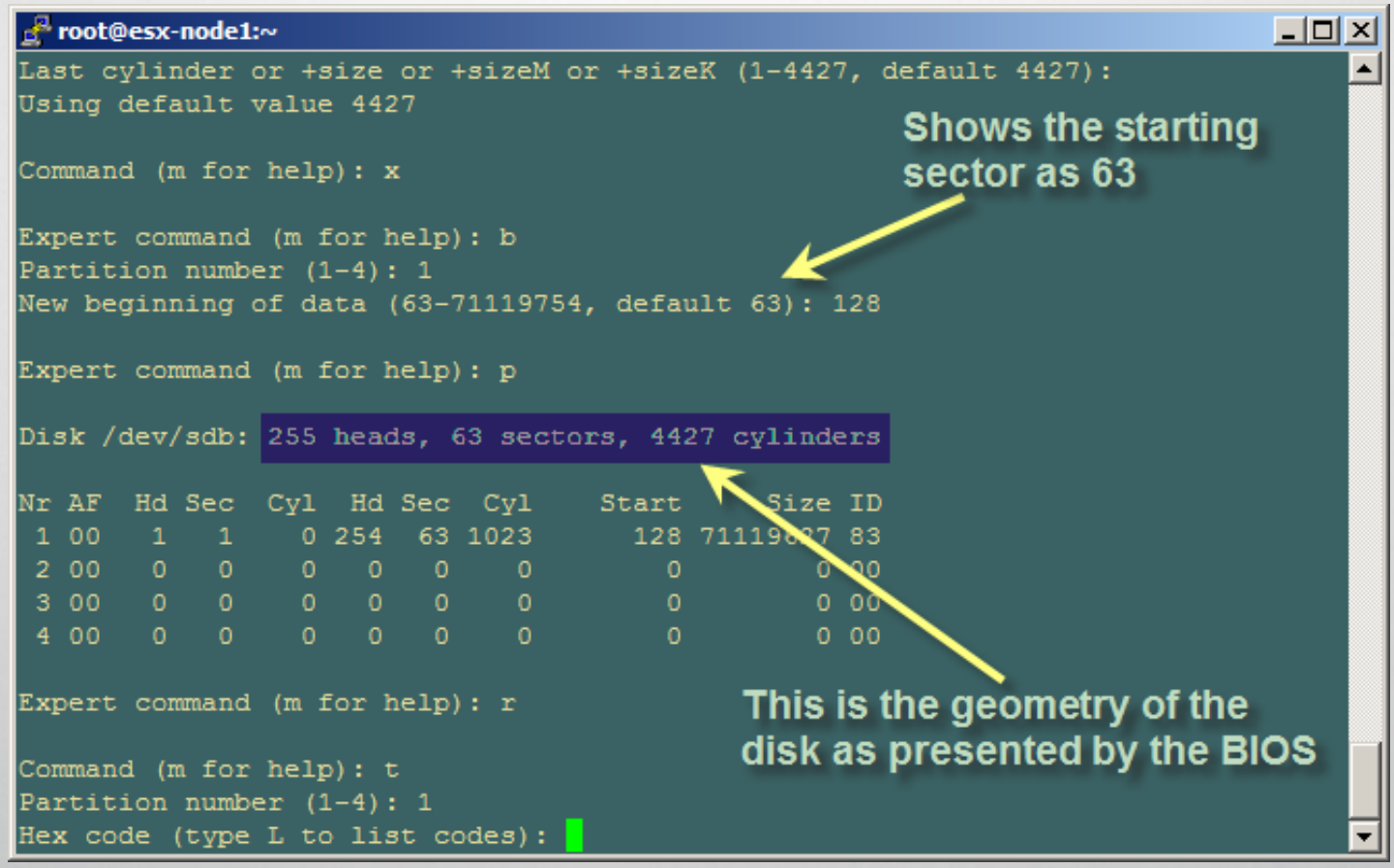
12. Type “1” to select partition 1
13. Type “fb” to set type to fb (VMFS volume)
14. Type “w” to write label and the partition information to disk

- You are now ready to create VMFS on the partition
- Use MUI or vmkfstools to create the VMFS
- The virtual disks on the VMFS will now be aligned
- However, guest OS will misalign IOs inside the virtual disk
- For Linux guest OS follow the procedure listed above
- Procedures for Windows hosts is presented after service console screen shots

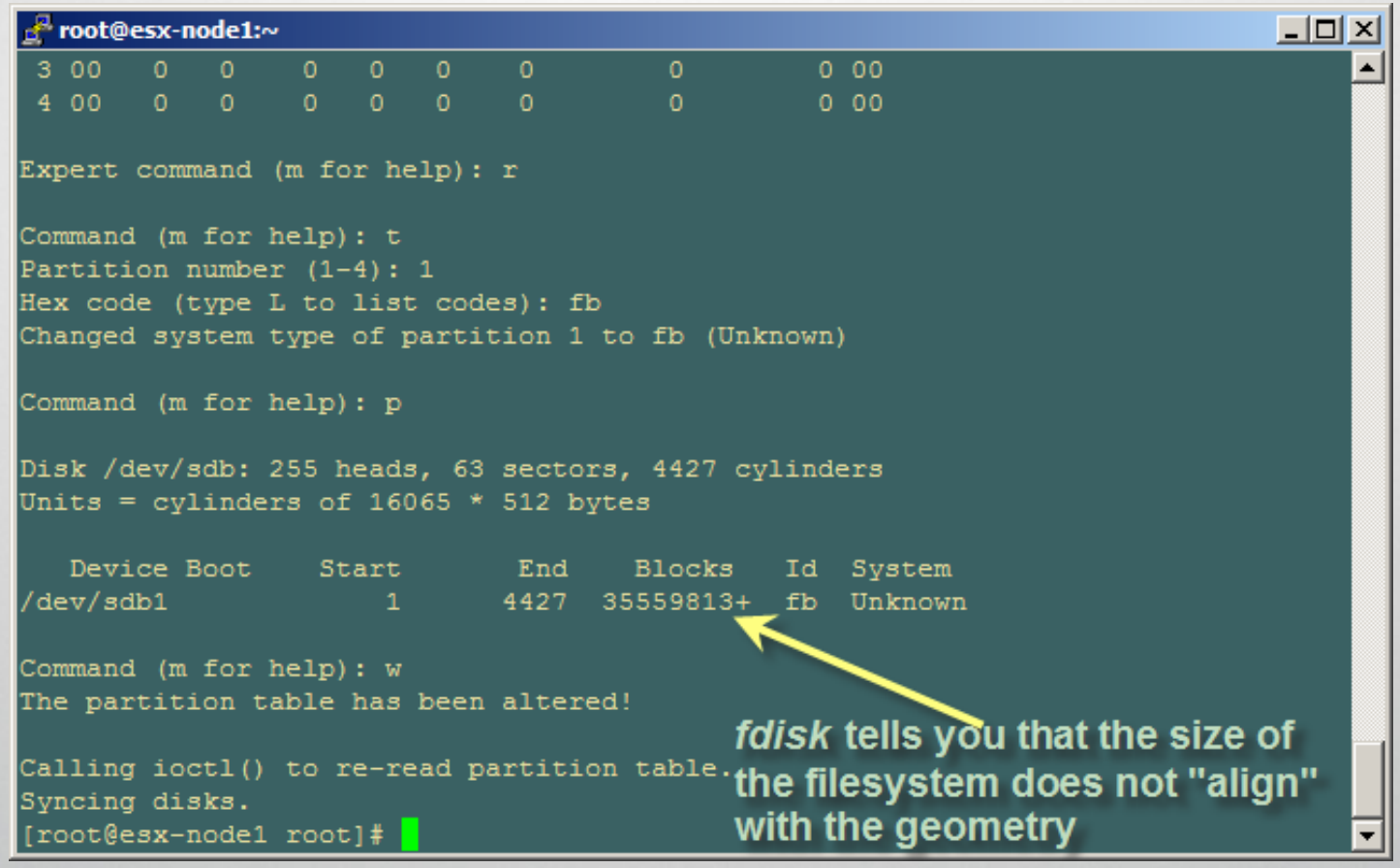
Using `fdisk` on Service Console to Align VMFS



Using `fdisk` on Service Console to Align VMFS



Using `fdisk` on Service Console to Align VMFS

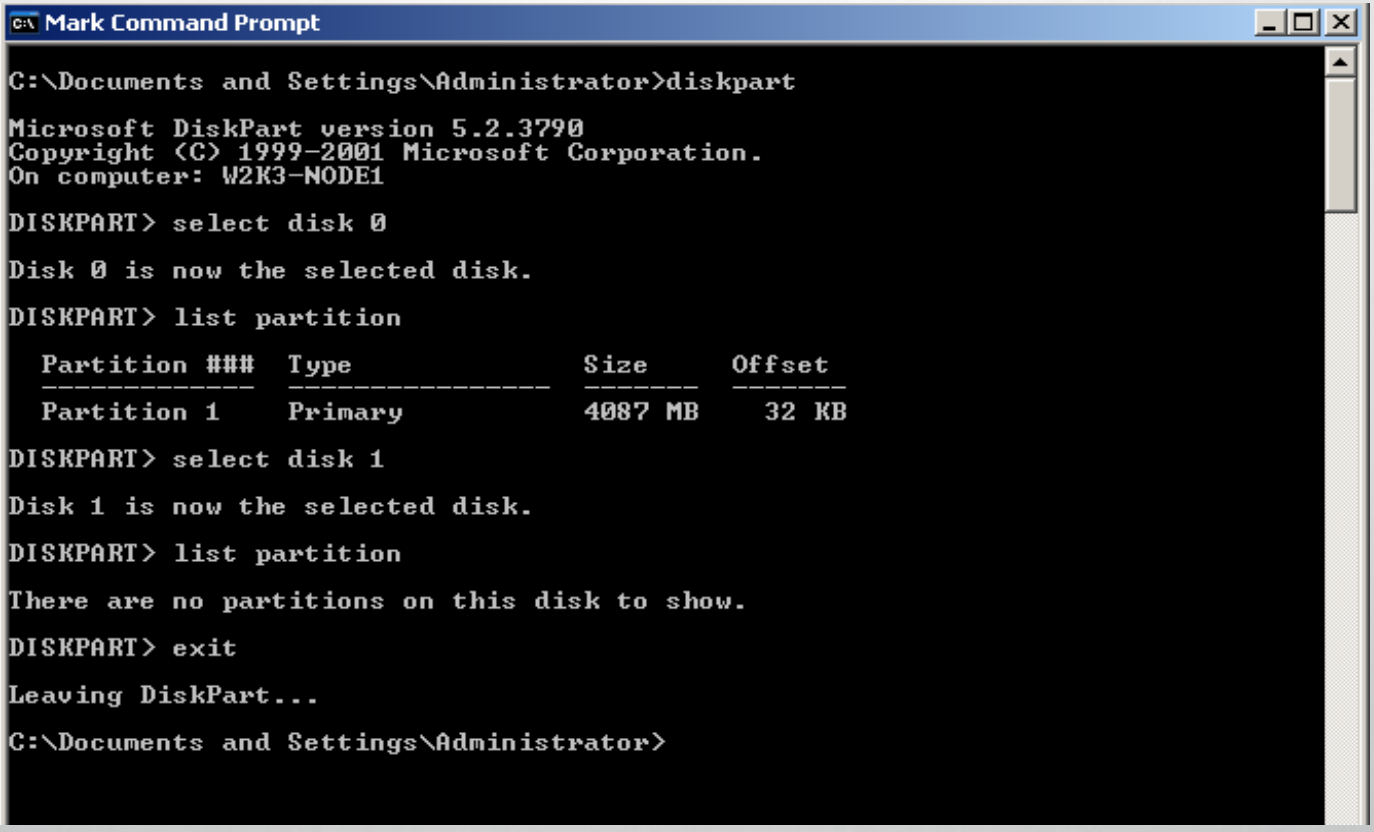


Aligning Partitions on Windows Hosts (Guest OS)

- The procedure applies to both disks that are presented as System LUN/Disk and virtual disks (vmdk files on VMFS)
- Windows 2000 and later “lies” about alignment when using diskpart
  - Partitions are offset by 32256 and not 32768 bytes
- First Partition by default is at LBA 63. This misaligns IOs
- Use diskpar.exe (and not diskpart.exe) to align the usable partitions

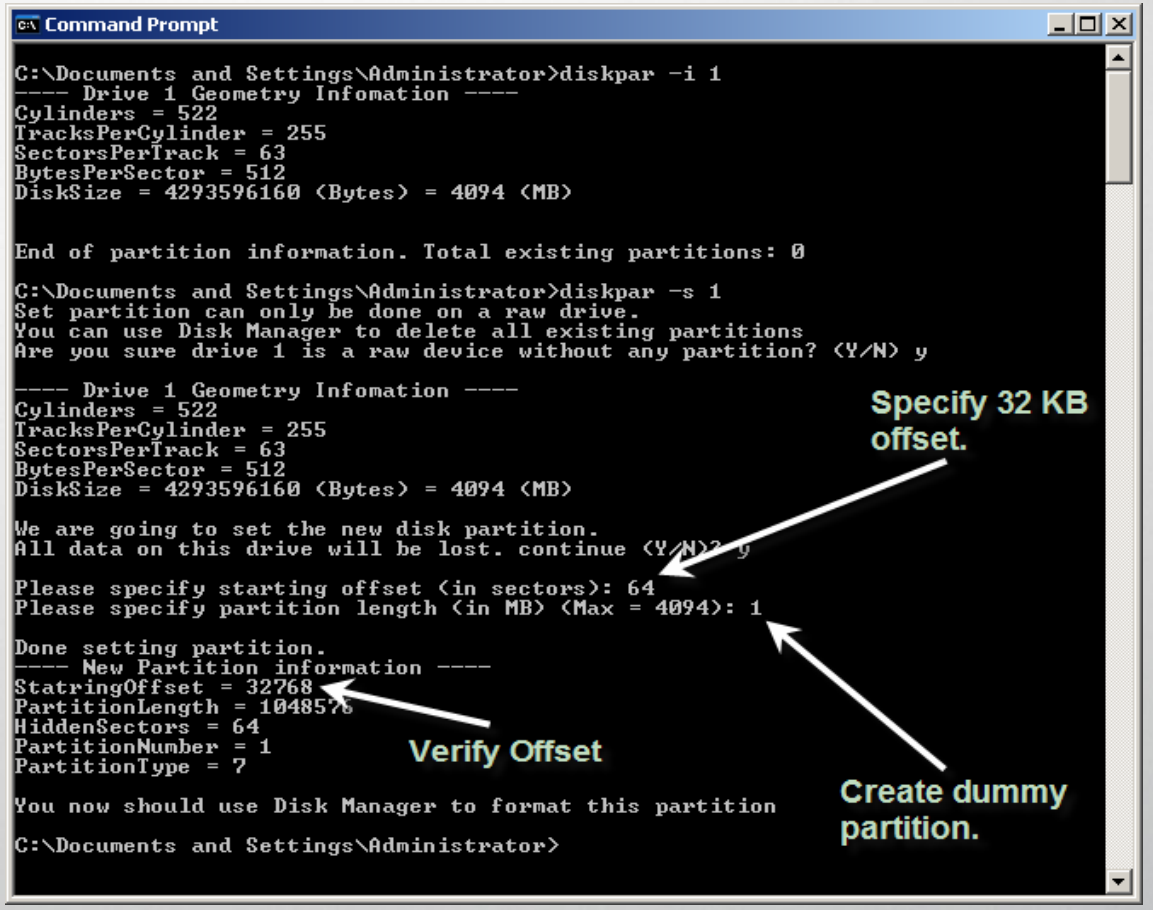
Output From DiskPart

- Diskpart will show partitions as aligned. In reality they are not



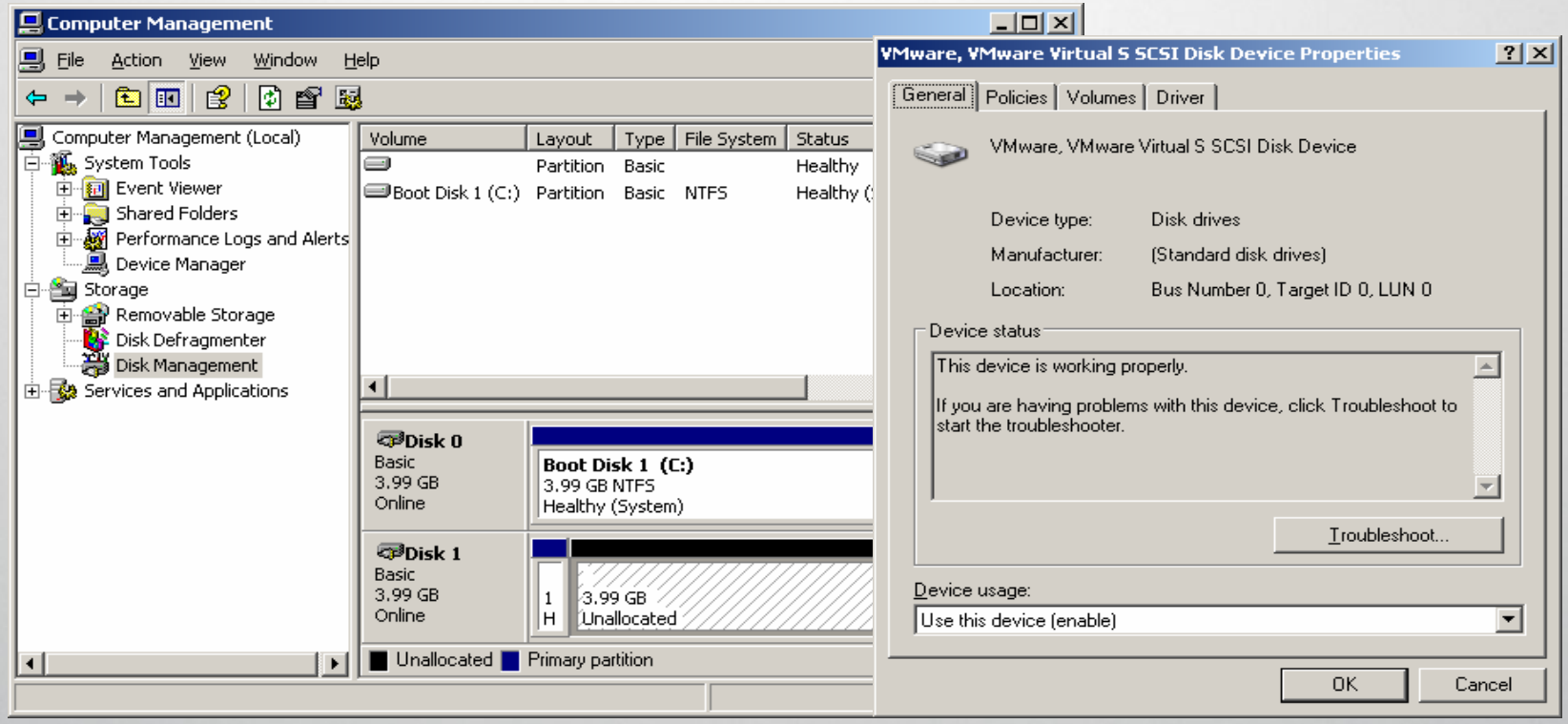
How to Use Diskpar to Create Aligned Partitions

- Create a 1 MB "dummy" partition
- The partition is aligned on 32 KB boundary
- All subsequent partitions will be aligned



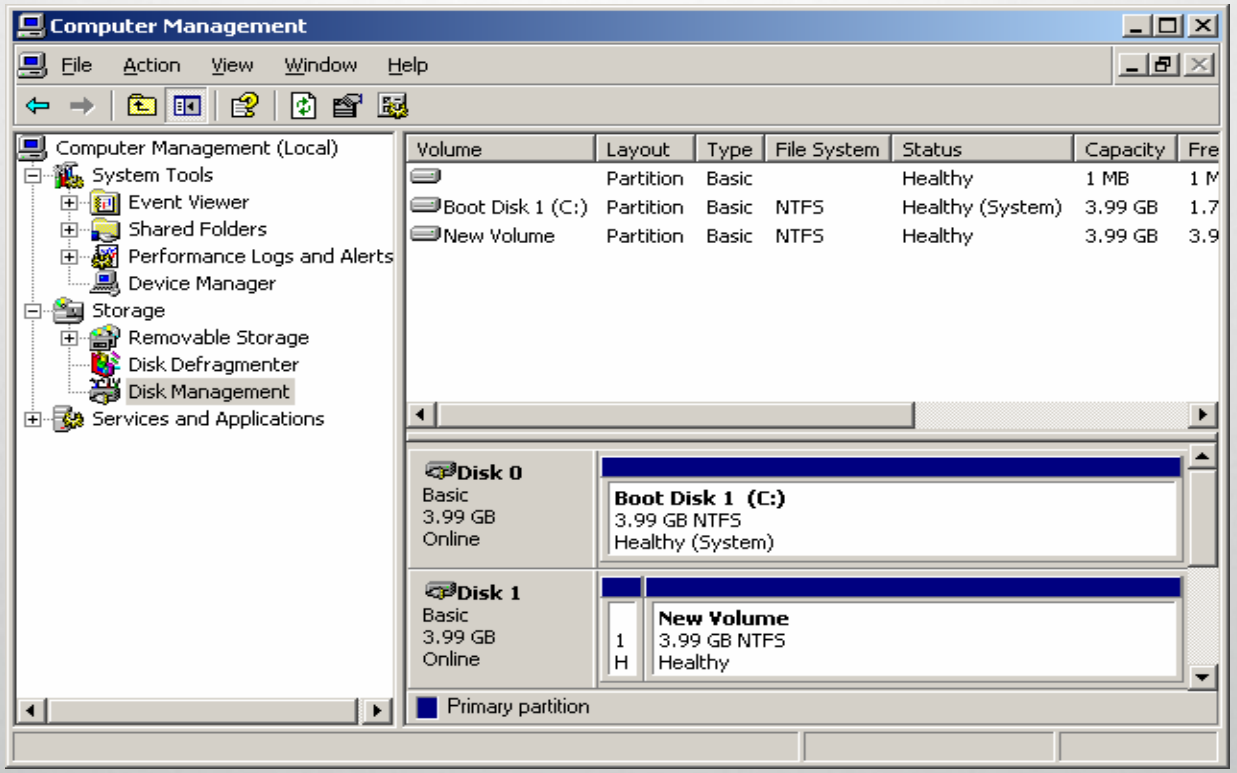
How to Use Diskpar to Create Aligned Partitions

- 1 MB Partition that was created in the previous step allows rest of the disk to have aligned IOs



How to Use Diskpar to Create Aligned Partitions

- Create partition and format NTFS with allocation unit of 32KB
- This will align all IOs on 32KB boundaries



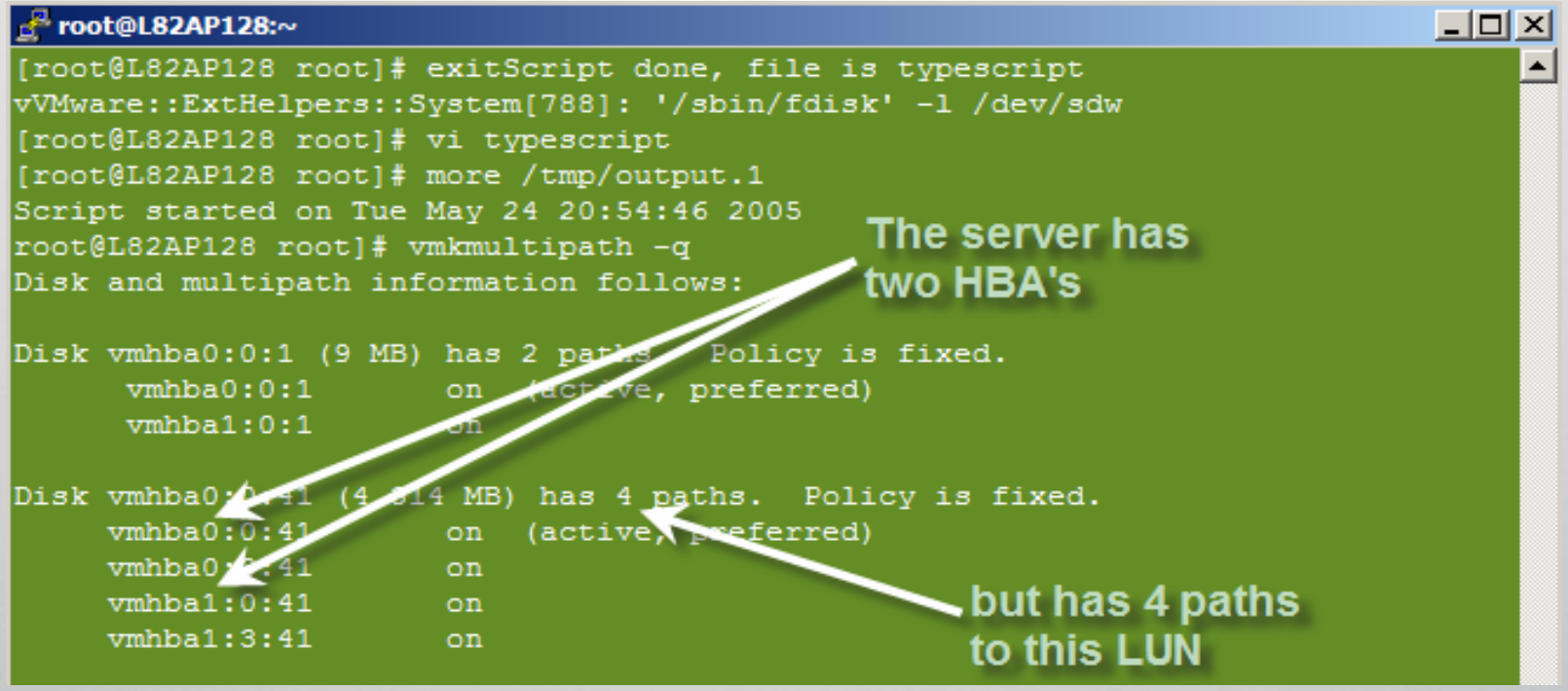
Using ESX Servers With Symmetrix Arrays

Best Practices For Using EMC Symmetrix Arrays With VMware ESX Servers

- Use striped meta volumes
  - Follow EMC best practices when creating meta volumes
    - Use the standard split size if it exists
    - Ideally 4, 8 or 16 way meta volumes
    - Meta volumes should not have members on the same physical disks
    - Stripe size should be two cylinders
- For optimal performance and scalability meta volume should not be much larger than 200 GB
  - Balance between performance versus management
  - With 8 to 10 LUNs per ESX Server farm, this provides approximately 2 TB of storage
  - Can be grown as appropriate

Best Practices For Using EMC Symmetrix Arrays With VMware ESX Servers

- Use multi-target zoning if possible



Best Practices For Using EMC Symmetrix Arrays With VMware ESX Servers

- Solutions Enabler 5.4 and higher is supported
  - Do not install Solutions Enabler under guest operating system unless RDM is used (no value add)
  - Installation of extraneous software on service console is NOT recommended. SE 6.0 will NOT run on service console
- Use GNS to enable propagation of information across multiple ESX Servers (farms) and guest operating systems

Best Practices For Using EMC Symmetrix Arrays With VMware ESX Servers

- Use of RDM is recommended if advanced functionality is needed
  - Leverage consistency technology
  - Storage functionality can be coordinated across multiple OS and applications
- Gatekeepers need to be provided and mapped to guest OS to leverage storage functionality

Backup and Recovery Considerations

Backup and Recovery Considerations

- ESX Server and Virtual Machines/Applications
  - Native tools provided by VMware
    - vmsnap, vmsnap_all, vmres etc.
  - Symmetrix Layered Applications
    - TimeFinder/Mirror, TimeFinder/Clones, TimeFinder/Snap, SRDF/S, SRDF/A and Open Replicator
  - CLARiiON Layered applications
    - SnapView, MirrorView, SAN Copy
  - CLARiiON Disk Library

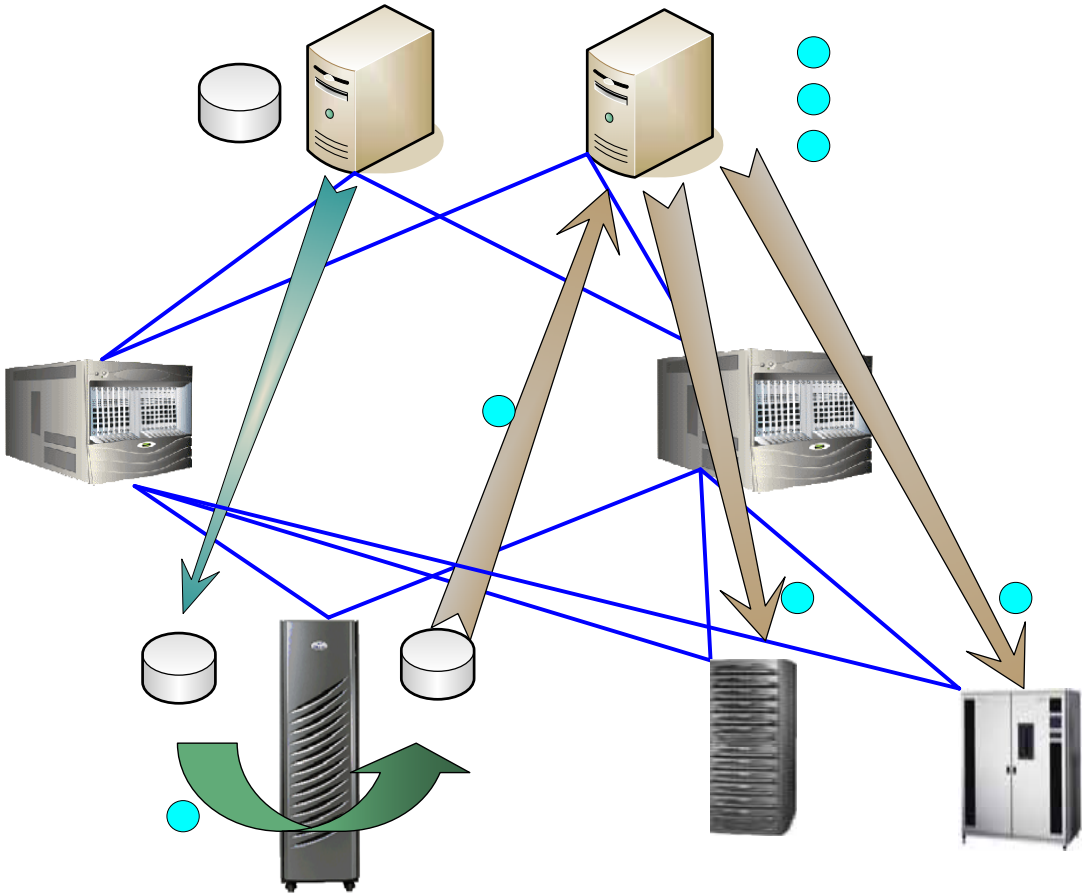
Backup and Recovery Considerations

- Backup Software
  - VMware tests major backup software. Refer to matrix http://www.vmware.com/pdf/esx2_backup_guide.pdf
  - Legato NetWorker has been tested.
    - "Storage Node" in Virtual Machines (Linux and Windows)
    - "NetWorker Client" in Virtual Machines
    - "Storage Node" or "NetWorker Client" for Linux on Service Console
  - EMC recommends configuring "Storage Node" or "NetWorker Client" in Virtual Machines
    - Provides granular backup and restores in this configuration
    - Add on modules (Oracle, Exchange etc.) can be used
- VMware add-on tools (vmware-mount.pl & vmware-loop) maybe needed for restores of individual files

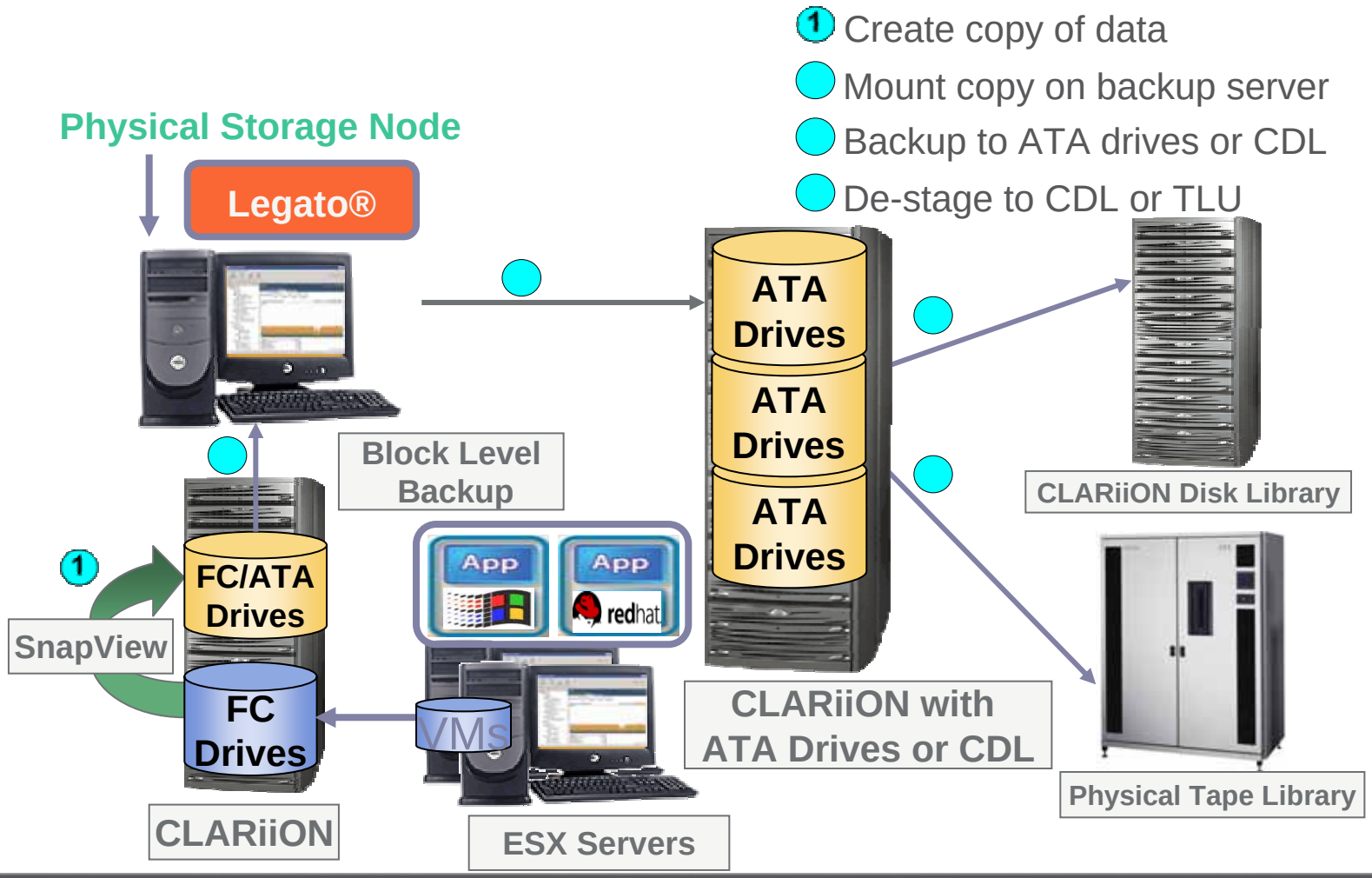
Backup and Recovery Considerations

- On the Symmetrix side, extensive testing has been done with TimeFinder/M and SRDF/S
  - Refer to “Integrating VMware ESX Server with EMC Symmetrix Remote Data Facility” white paper
  - Solution Guides discussing integration of ESX Server with Symmetrix are also available
- All CLARiiON layered applications listed above are fully supported
  - Some of the restrictions are discussed in the previous slides
  - Refer to the white paper “CLARiiON Integration with VMware ESX Server” for further details

Backup and Restore Using TimeFinder



Backup using CLARiiON Disk Library

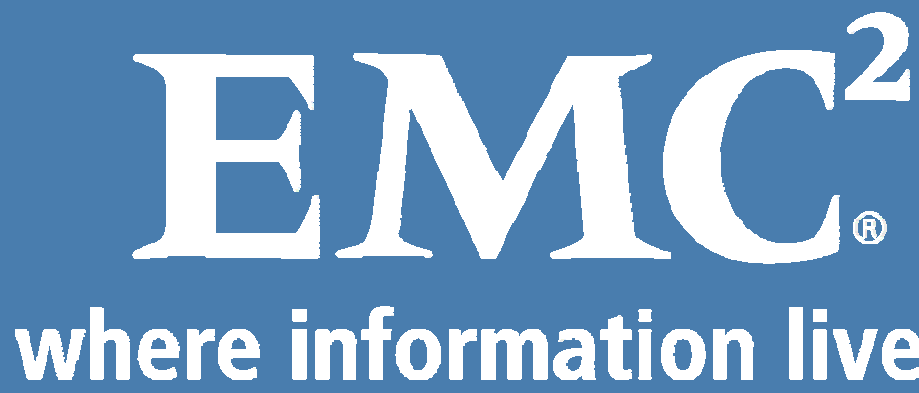


Summary

- Storage should be presented as multiple disks
- If using VMFS separate log and application data on separate file system
- Use static load balancing
- Use multi target zoning if possible
- Ensure partitions and file systems are aligned
- Follow best practices for creating metavolume and metaLUNs
- Use Symmetrix GNS if there are multiple ESX servers sharing storage
- Using CLARiiON layered applications require RDM
- When possible leverage layered applications to save precious CPU cycles

Related Sessions

- **SLN056** – VMware ESX Server Workload Analysis: How to Determine Good Candidates for Virtualization
- **SLN381** – VMware with CLARiiON


EMC²
where information lives®

las vegas • october 18-20, 2005