Best Practices for deploying VMware ESX 3.x and 2.5.x server with EMC Storage products

Sheetal Kochavara

Systems Engineer,
EMC Corporation
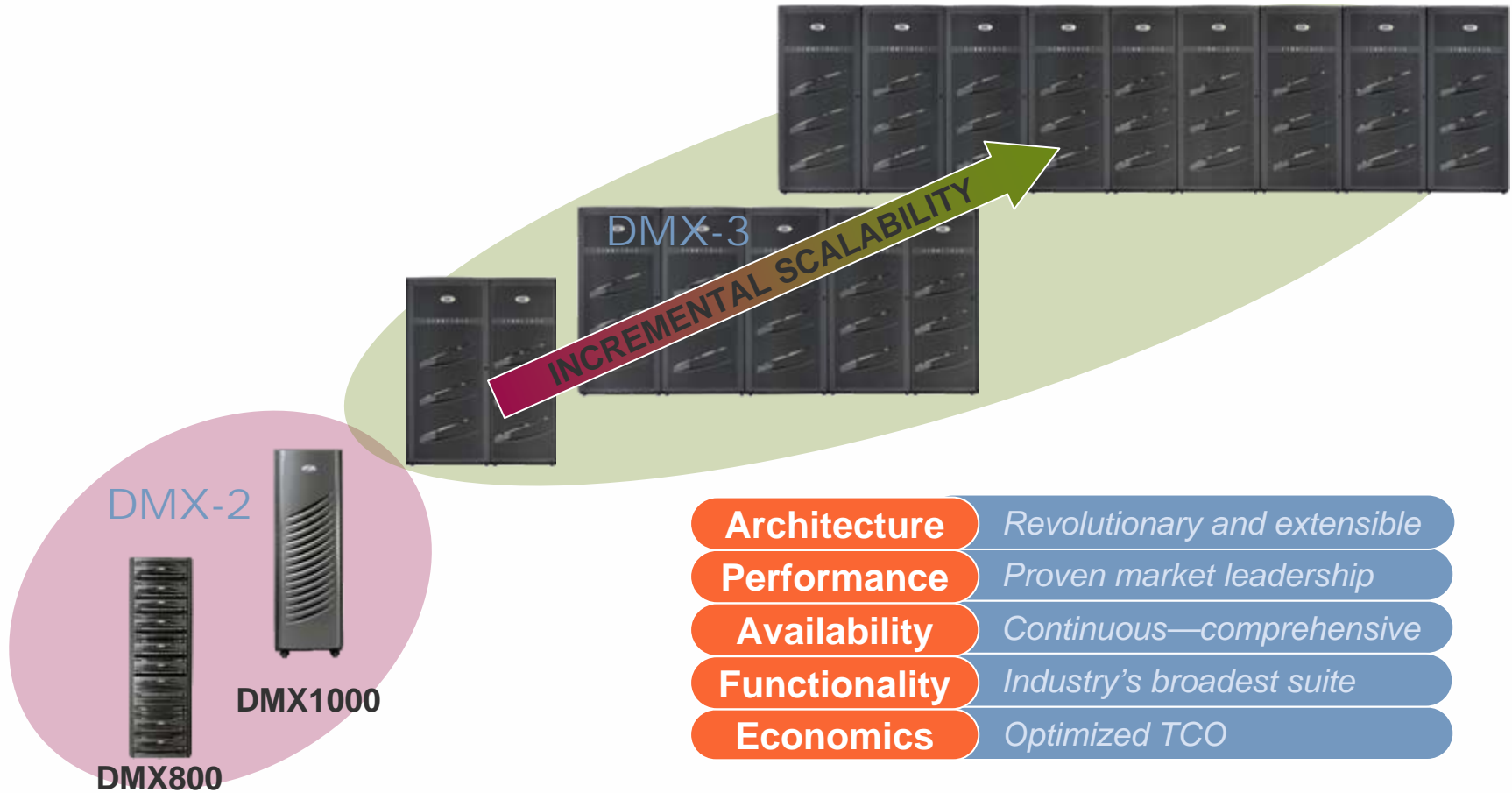
VMWORLD 2006

Agenda

- Overview of EMC Hardware and Software

- Best practices with VMware ESX 3.0.x and ESX 2.5.x
  - Connectivity
  - Performance
  - Business Continuity and Backup

- Customer Case Study

EMC Symmetrix DMX Hardware Series

Industry's Broadest and Most Powerful High-end Storage Family

DMX-3

INCREMENTAL SCALABILITY

DMX-2

DMX1000

DMX800

Architecture	*Revolutionary and extensible*
Performance	*Proven market leadership*
Availability	*Continuous—comprehensive*
Functionality	*Industry's broadest suite*
Economics	*Optimized TCO*

EMC CLARiiON Hardware Series

2006 CLARiiON FAMILY
COMPLETE RANGE OF AFFORDABLE AND EASY-TO-USE STORAGE PLATFORMS

CLARiiON Disk Library	CLARiiON AX Series	CLARiiON CX Series	CLARiiON CX3 UltraScale Series
DL210 DL7X0	AX150 AX150i	CX300 CX500 CX700 CX300i CX500i	CX3-20/ CX3-20C CX3-40/ CX3-40C CX3-80

EMC Celerra Hardware Series

Broadest Range of NAS Products

| NS350 NS500 NS700 NS40	NS704 NS80	NS500G NS700G NS40G	NS704G NS80G	NSX
High availability	Advanced clustering	High availability	Advanced clustering	Advanced clustering
One or two Data Movers	Four Data Movers	One or two Data Movers	Four Data Movers	Four to eight X-Blades
Upgradeable to gateway	Upgradeable to gateway	NAS gateway to SAN	NAS gateway to SAN	NAS gateway to SAN
Integrated CLARiiON	Integrated CLARiiON	CLARiiON, Symmetrix	CLARiiON, Symmetrix	CLARiiON, Symmetrix

EMC Array Software

	Symmetrix	CLARiiON	Celerra
Management	■ ControlCenter ■ Symm CLI ■ Performance Analyzer	■ Navisphere Manager ■ Navisphere CLI ■ Navisphere Analyzer ■ Navisphere Quality Server Manager	■ Celerra Manager - Performance Monitoring
Aggregate software	■ Metavolumes ■ Optimizer	■ MetaLUNs ■ LUN Migration	■ Virtual provisioning ■ Volume Manager

Replication	■ TimeFinder Snap
	■ TimeFinder BCVs
	■ TimeFinder Mirror
	■ SRDF/S
	■ SRDF/A and SRDF/AR
	■ Open Replicator
		■ SnapView Snapshots
		■ SnapView Clones
		■ MirrorView/S
		■ MirrorView/A
		■ SAN Copy
			■ SnapSure
			■ IP replication
			- Celerra Replicator
			■ SAN replication
			  - SRDF\SRDF/A
			  - MirrorView/S

Agenda

- Overview of EMC Hardware and Software

- Best practices with VMware ESX 3.0.x and ESX 2.5.x
  - > Connectivity
  - > Performance
  - > Business Continuity and Backup

- Customer Case Study

Connectivity - Boot from SAN

- **ESX boot from SAN (Fibre Channel (FC))**
  - Use RAID 1/RAID 1/0 (highest level of availability)
  - Use RAID 5 (optimal capacity/performance ratio)
  - Boot LUN size (Service Console swap space)
  - iSCSI boot using hardware initiator (Q4-06/Q1-07)
  - RDM supported with ESX 3.x boot
- **VM boot from SAN (FC and iSCSI)**
  - CLARiiON and Celerra iSCSI
    - ESX iSCSI software-initiator supported
    - Support for hardware initiator Q4-06/Q1-07
  - OS/application
    - Use RAID 1/RAID 1/0(highest level of availability)
    - Use RAID 5 (optimal capacity/performance ratio)
    - Each VM has it's own swap space for ESX 3.x
      - Place swap with OS (default settings)

Virtual Machines (VMs)

ESX Boot

VM Boot

ESX Server
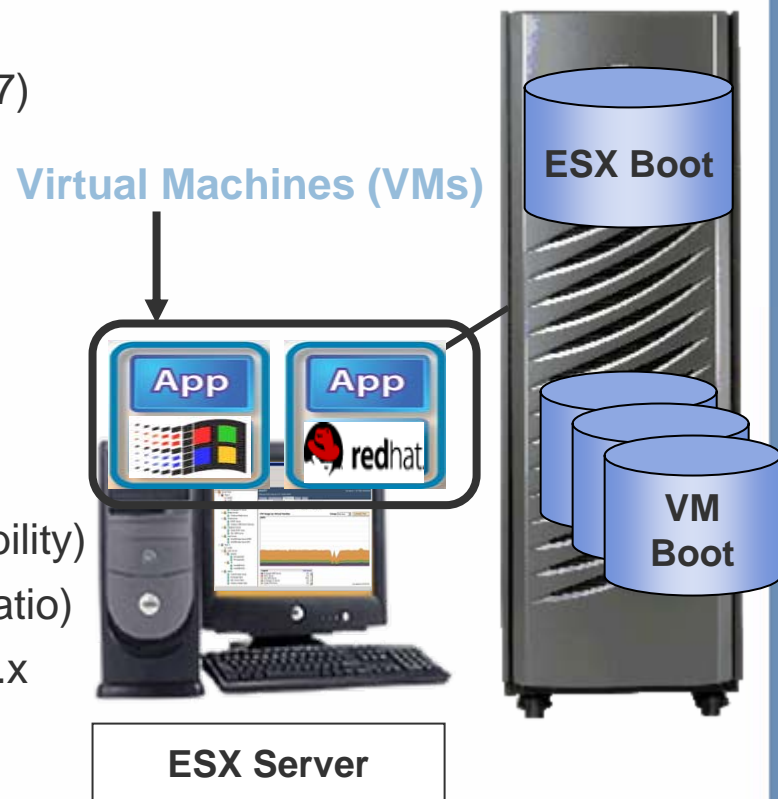
Connectivity - Boot from SAN (contd.)

- **Celerra NFS boot options**
  - Boot ESX from local device, SAN or iSCSI
  - Boot and provision storage for VMs from Celerra using NFS
    - Requirement for VMotion/DRS/VMware HA
    - NFS version 3 over TCP only (no version 2, no UDP)
    - Use RAID 5 or RAID 1
    - Place swap files on SAN, iSCSI or local storage
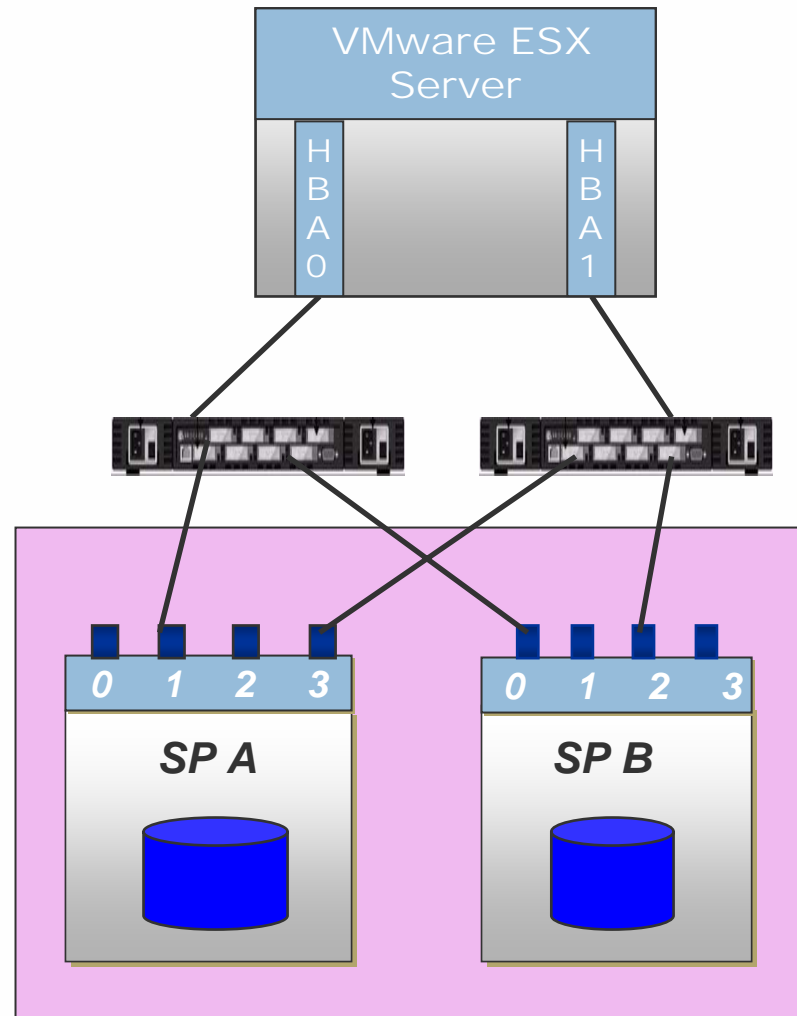- Be careful when using lower-tier (e.g. ATA) disks for boot
  - Targeted for sequential I/O workload
  - Swap file I/O characteristics is random
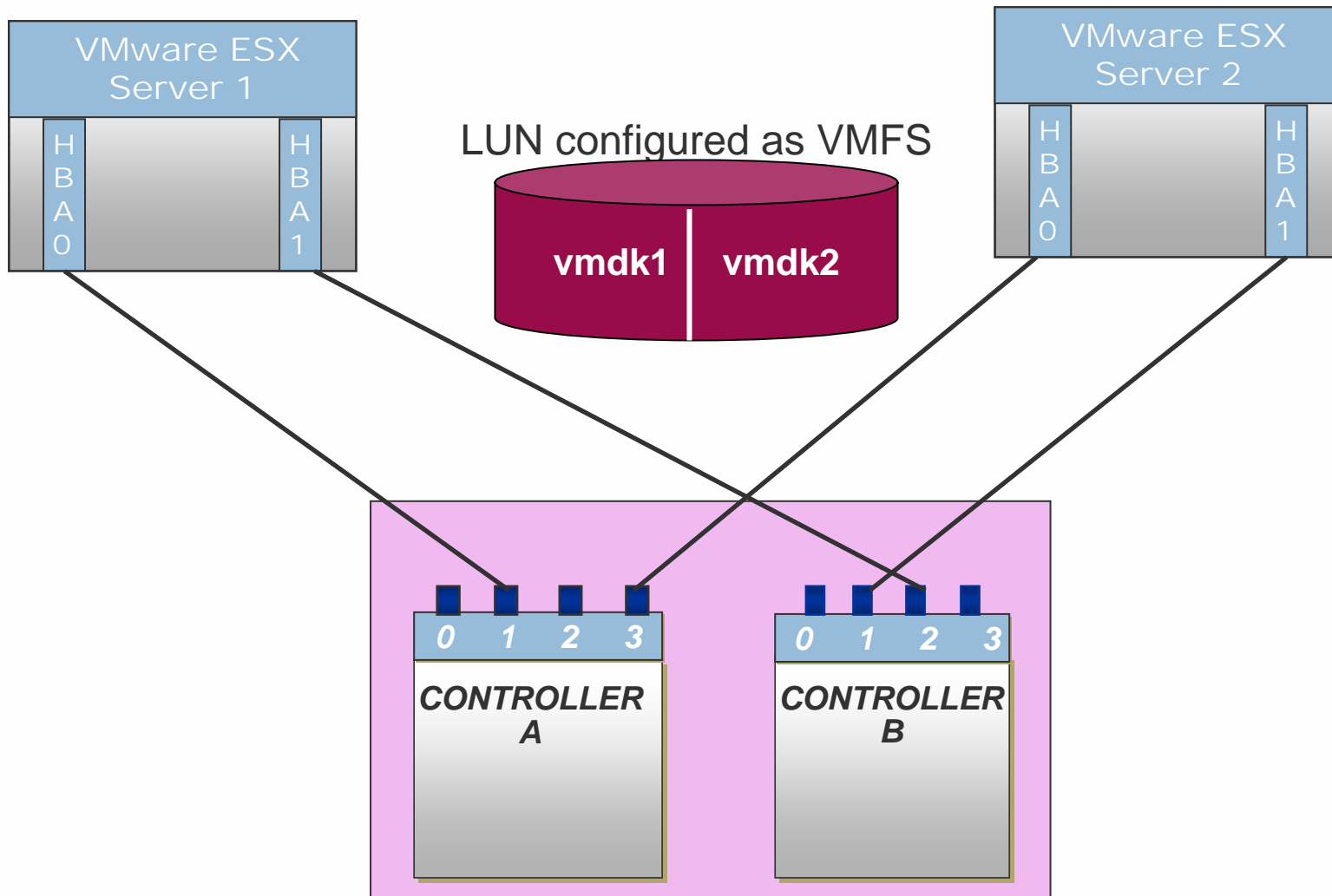- Separate LUN(s) for templates, ISO images
  - Can use iSCSI, NFS or ATA

Connectivity – Multipathing\Failover (Fibre Channel)

- VMware's native multipath software
  - Symmetrix (Fixed policy)
  - CLARiiON (MRU policy)
- Zoning
  - Single initiator (HBA) zoning
  - Static load balancing for Symmetrix
    - Balance load across FAs and HBAs
- Online Code Upgrades supported
- CLARiiON
 - Recommended that each HBA have access to both SPs

Single SP access to each HBA issue for active-passive arrays

VMware ESX Server 1

HBA0

HBA1

LUN configured as VMFS

vmdk1 | vmdk2

VMware ESX Server 2

HBA0

HBA1

0 1 2 3
CONTROLLER A

0 1 2 3
CONTROLLER B

Connectivity – Multipathing\Failover (CLARiiON iSCSI)

- VMware's native multipath software
  - CLARiiON (MRU policy)
- Enable NIC Teaming
  - Common termination point
  - Use dedicated Vswitches/VLANs
    - iSCSI traffic
- iSCSI Hardware initiator
  - Each HBA have access to both SPs
- Online Code Upgrades supported

NIC Teaming

VMware ESX Server

NIC 1   NIC 2

0 1 2 3   0 1 2 3

SP A   SP B
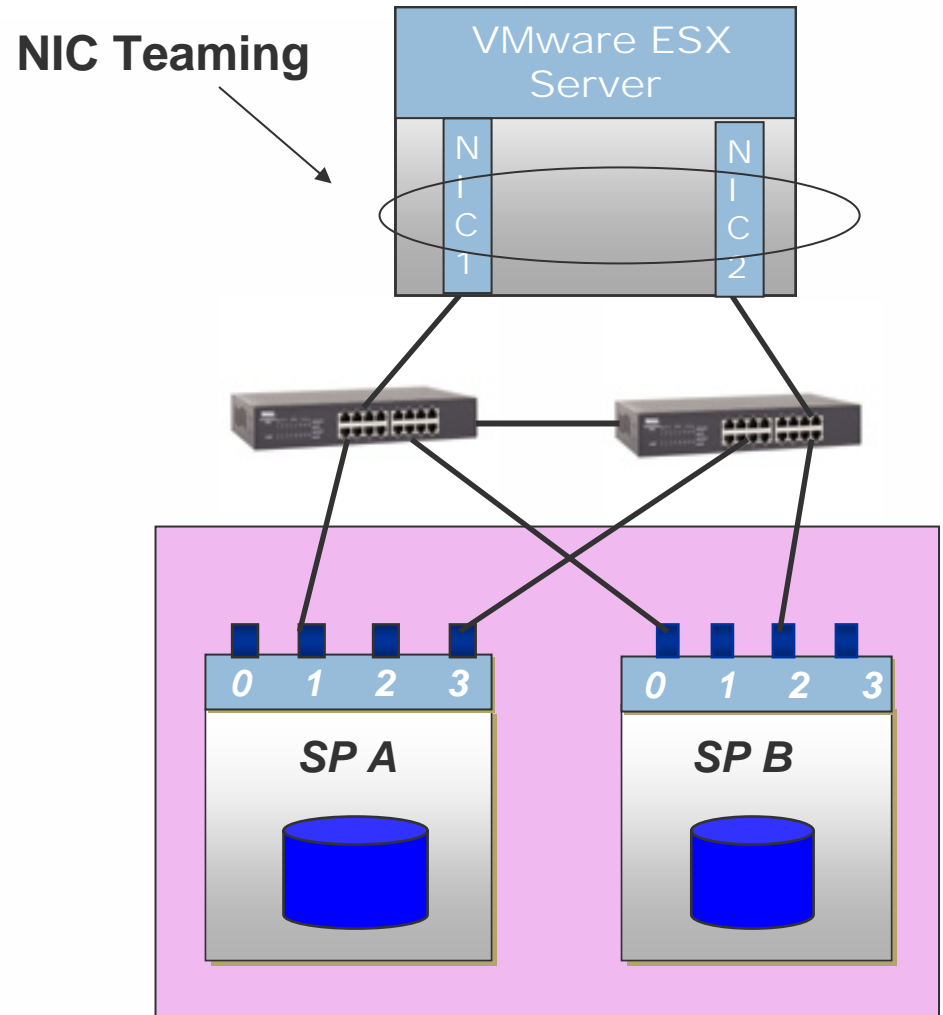
Connectivity – Multipathing\Failover (NFS and Celerra iSCSI)

- VMware's native multipath software
- Fail-safe network implementation
  - > iSCSi and NFS
- Use dedicated Vswitches and/or VLANs
  - > iSCSi traffic
- Internal locking protocol [NFS]
  - > Lease-based
  - > Protects against data corruption
- Security [NFS]
  - > Standard AUTH_UNIX/AUTH_HOST security

NIC Teaming

VMware ESX Server

NIC1

NIC2

Failsafe Network

0   1   2   3

Celerra Data Mover
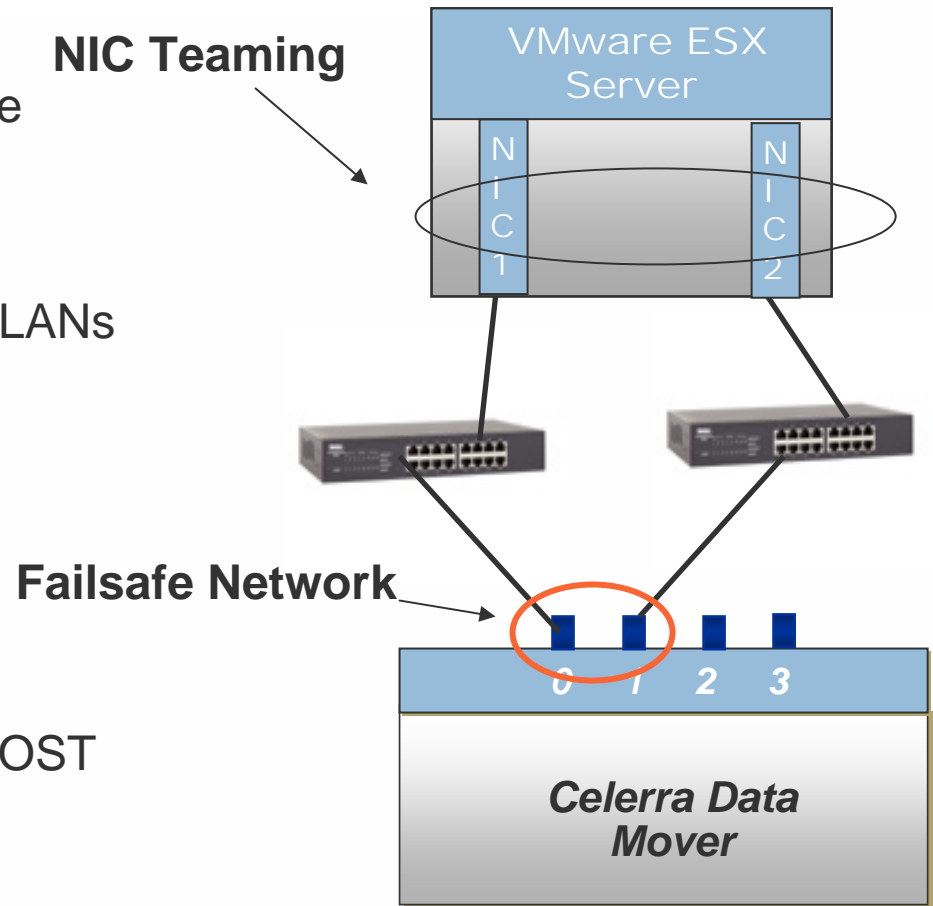
Performance Considerations (Fibre Channel and iSCSI)

- **Configuration for LUN Layout**
  - Avoid using same set of disks for different applications I/O characteristics
  - LUNs should be sized according to anticipated I/O workload
    - Recommended LUN size should be less or equal to 500GB
  - Option to spread load across multiple physical disks for capacity and expansion for application data volumes
    - Symmetrix Metavolumes, CLARiiON MetaLUNs or Celerra Striped Volumes
  - Maximum LUNs per VirtualCenter farm or cluster
    - Dependent on number of ESX servers and VM workload

# Performance Considerations (Fibre Channel and iSCSI) (Contd.)

- **Configuration for VM OS/application**
  - Use VMFS volumes to avoid increase in LUN count
  - Fibre Channel recommendations
    - Maximum 16 ESX servers and 32 I/O-intensive VMs per VMFS volume
    - Recommended: Each VMFS volume should not contain more that 10-15 VMs
      - To avoid SCSI-reservation issues and queuing
      - Factor in number and type of back-end disks
      - Reserve space for VMware snapshots
      - See replication best practices for VMFS volumes (slide 23)
  - iSCSI recommendations
    - Half the number of VMs that are recommended for Fibre Channel

# Performance Considerations (Fibre Channel and iSCSI) (Contd.)

- Managing data for I/O intensive application
  - Separate LUNs for OS/app and application data
  - Use RDM or present the entire VMFS volume (virtual disk) to the VM
  - **Follow best practice recommendations for a physical server**
  - Align application data disks at Virtual Machine Level for ESX 3.x
    - For VMFS and RDM virtual disks
    - Use VirtualCenter to create VMFS volumes for filesystem alignment

Tests conducted on misaligned and aligned VMFS volumes
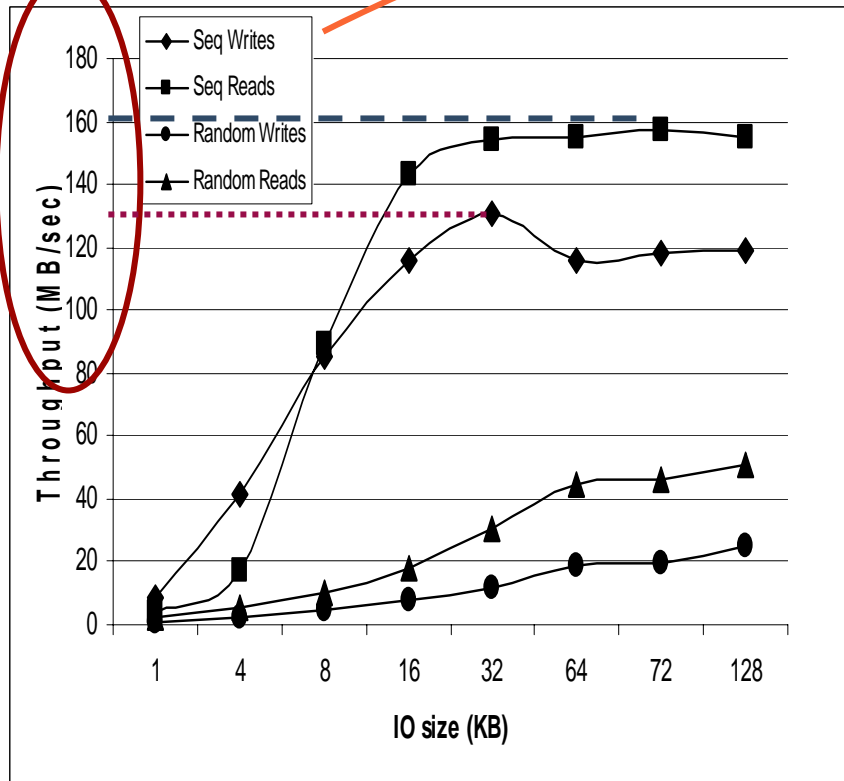
10-12% increase in throughout


VMFS Misaligned - RAID 5
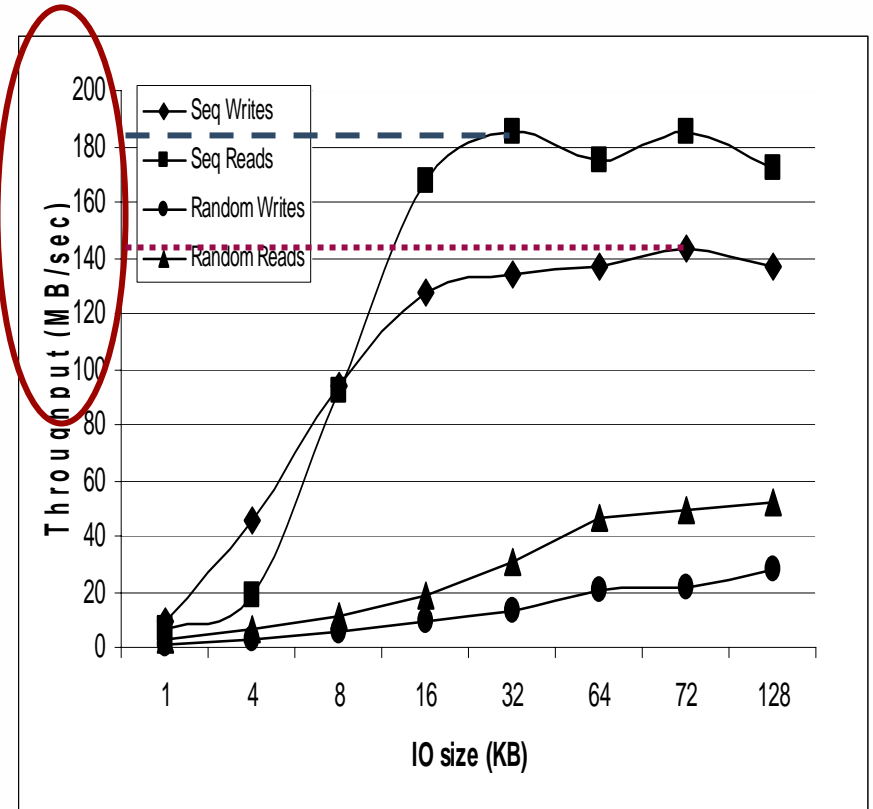
VMFS aligned to 64K - RAID 5

Celerra Performance Considerations (NFS)
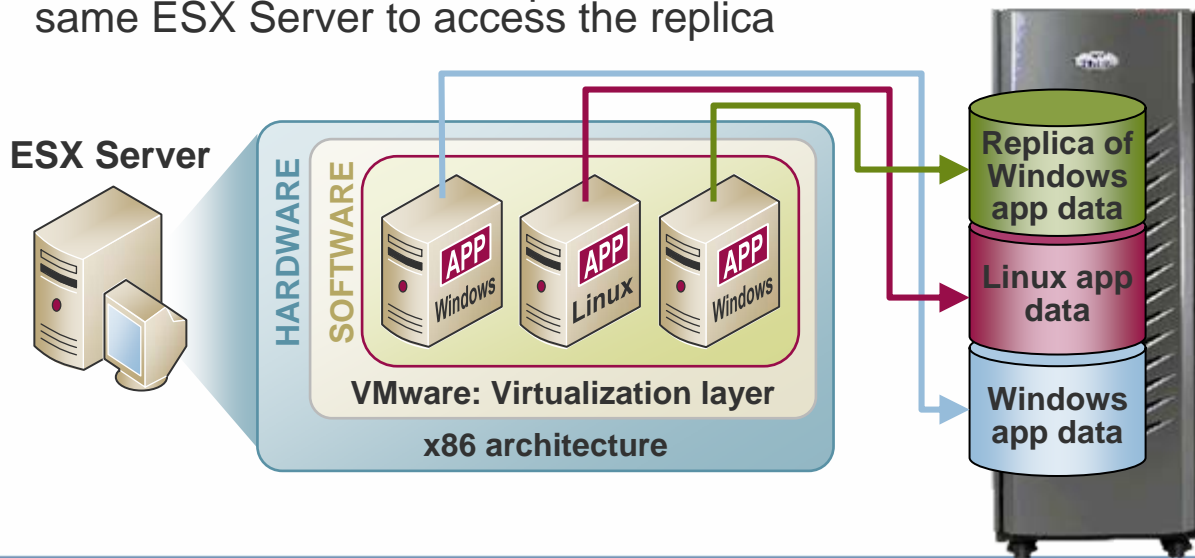
- Use TCP mount options with read and write buffer size=32k
- Use VLAN for ESX Storage interfaces
- Avoid VM swapping to NFS volumes
  - Edit VM config file to add sched.swap.dir
  - Place swap on SAN, iSCSI or local
- 8 NFS mounts per ESX Server allowed by default.
  - To increase: Select host from inventory, right-click Advanced Settings, select "NFS"
  - Adjust "NFS.MaxVolumes"
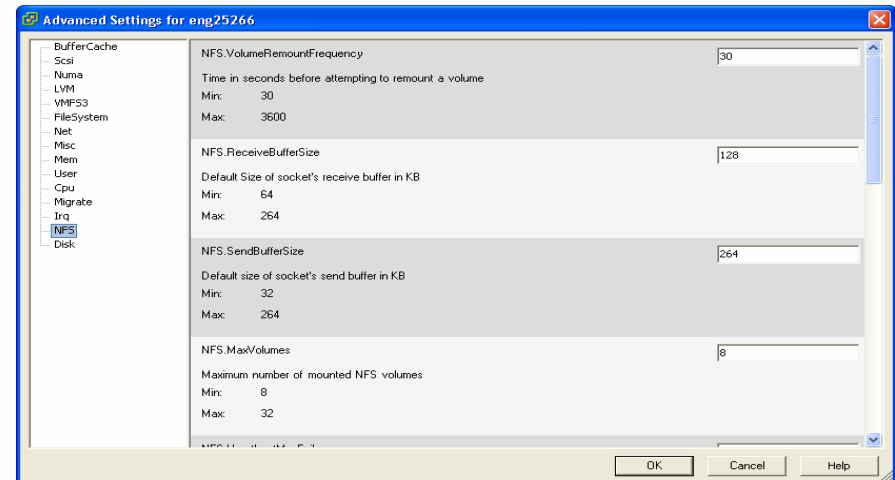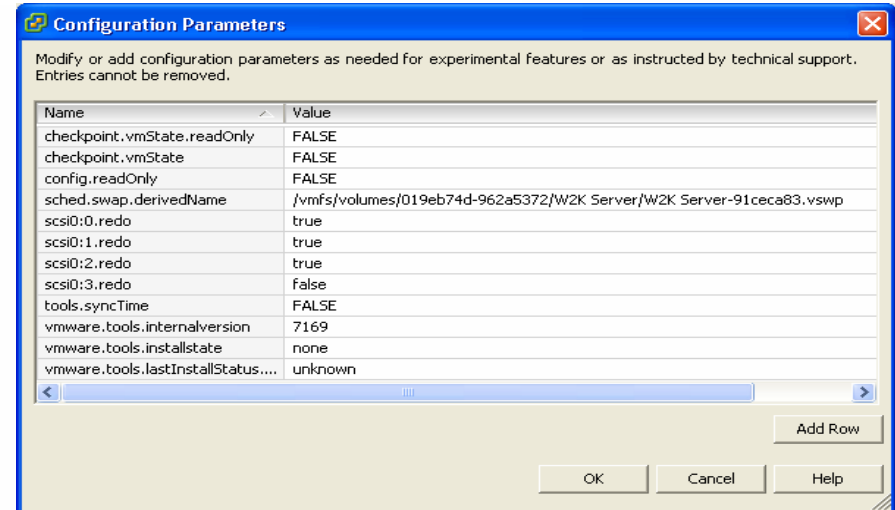
Array-based Replication and ESX Server Interoperability

- Symmetrix and CLARiiON replication products are supported with RDMs and VMFS volumes
  - Guest operating system images, as well as the application data, can be replicated
  - Both VMFS and RDM volumes, the replica can be presented to the same ESX Server with ESX 3.x
    - For ESX 2.5.x, only replicas of RDM volumes can be presented back to the same ESX server
    - Allows one VM to use the production LUN, and another VM on the same ESX Server to access the replica


ESX Server

HARDWARE
SOFTWARE

APP Windows
APP Linux
APP Windows

VMware: Virtualization layer

x86 architecture

Replica of Windows app data

Linux app data

Windows app data

CLARiiON

VMWORLD 2006

Array-based Fibre Channel Replication Considerations

VMFS or RDM for application data ?

- Use VMFS if:
  - You wish to administer less LUNs, easy provisioning of VM
  - Replicate entire VMFS volume along with all its virtual disks
    - Crash-consistent replica if not all virtual disks are application quiesced
  - Support for VMware snapshots is required
  - Using VMware Consolidate Backup (VCB)
  - Using VirtualCenter cloning

- Use RDM [physical compatibility mode] if:
  - Replicating application data to and from a physical to virtual infrastructure
  - SAN Replication software tools to run inside virtual machines
    - admsnap, admhost etc.
  - No support for VMware snapshots
  - For virtual machine clustering between physical servers
  - Online expansion using array-based expansion technology

Array-based Fibre Channel Replication Considerations (Contd.)

- Use array-based consistency technology
  - Groups of multiple LUNs
    - Replicate OS\application and application data LUNs
  - Applications distributed across many LUNs
    - e.g. databases, volume managers (VMFS spanning)
- Flush buffers
  - Host, disk, database buffers before replication
  - Use EMC-based tools to flush disk buffers
- Replication of VM OS images
  - Replica crash-consistent if VM not powered down
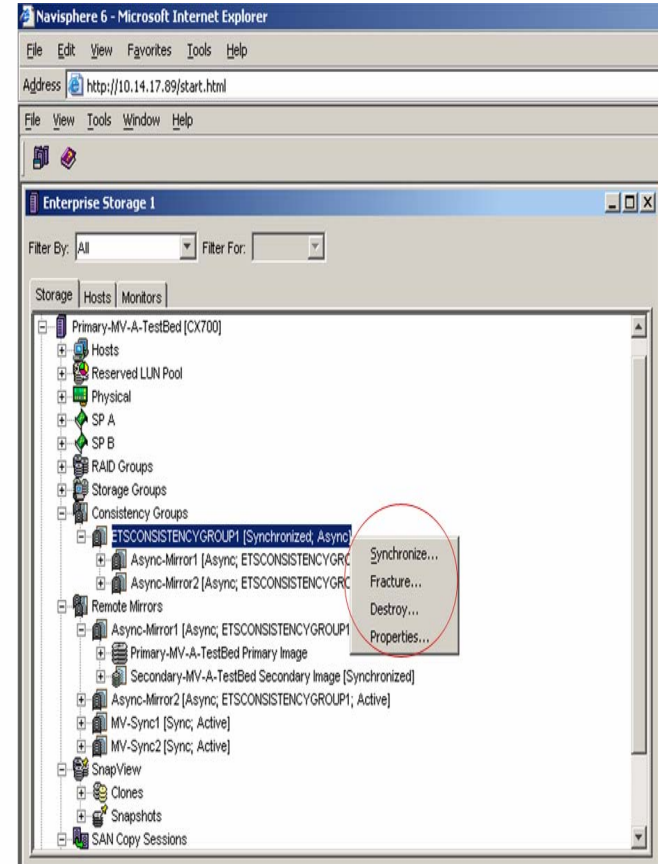
Array-based iSCSI Replication Considerations

- **CLARiiON iSCSI**
  - ESX iSCSI software and hardware (Q4-06/Q1-07) initiator
    - Replication granularity entire VMFS volume
    - Similar replication considerations as Fibre Channel
  - VM iSCSI software initiator
    - Replication granularity VM virtual disk
    - Replica presented to a different VM (same or different ESX server)
- **Celerra iSCSI**
  - ESX iSCSI software and hardware (Q4-06/Q1-07) initiator
    - Use Celerra Replicator (RDM volumes)
    - Support for Replication Manager/SE (RDM only)
  - VM iSCSI software initiator
    - Replication granularity VM virtual disk
    - Support for Replication Manager/SE

Array-based Replication NFS Considerations

- Integration with array replication technologies
  - Symmetrix SRDF/S, SRDF/A, CLARiiON MirrorView/S
- Celerra Replicator offers asynchronous volume based option
  - Replication granularity is entire NFS volume
- Snapsure can be used with NFS for checkpoints
  - Replication granularity is entire NFS volume

Implementing array-based replication of VMFS volumes with ESX 3.0

- ## Replica assigned to different ESX server
  - > For both local and remote array-based replication
    - **Recommended**
      - **Set the LVM.EnableResignature parameter to 0**
      - **Set the LVM.Disallowsnapshotlun to 0**
    - Within VirtualCenter, place destination ESX server in a different datacenter object

- ## Local Replication (application data volume)
  - > Ensure that the target VMFS is not accessed
    - No command to un-mount a VMFS volume

- ## Remote Replication
  - > Destination VM is not powered on unless there is a disaster

Agenda

- Overview of EMC Hardware and array-based software

- Best practices for implementing EMC storage with VMware ESX 3.0.x
  - > Connectivity
  - > Performance
  - > Business Continuity and Backup

- Customer Case Study

Case Study - Customer consolidation

- About "ABC" Corporation
  - Headquarters in Pennsylvania, US

- IT service department
  - Applications
    - SQL Server
    - Web servers
    - File Share Servers
    - Home grown applications

- Initial environment
  - CLARiiON running on VMware ESX 2.5.x
  - Host-level Virtual Machine Backup

Requirements

- Need for more Virtual Machines (servers) running of Windows 2003 and Windows 2000

- Required another SAN array for running other OSs and applications
  - Maxed out the capacity of their existing CLARiiON system

- Wanted to deploy ESX 3.0 in the environment
  - VMware Distributed scheduling
  - VMware High Availability

- Needed a remote replication solution for disaster recovery at a remote site
  - 20 miles apart from primary site

Implementation

- Three VMware ESX 3.0 servers
  - HP DL580 (G4) servers
  - Total: 10 Windows virtual machines
- Proposed CX3-20 at the primary site with 60 drives
  - Virtual Machines booting from CLARiiON
  - LUNs configured as VMFS volumes
    - Host LUN ID is identical across all ESX servers
    - Shared across the 3 ESX servers for VMotion, VMware HA, DRS
- MirrorView/A for disaster recovery to another CX3-20 system
  - 20 miles from primary site
  - Cisco 9216i over T3 line (FC over IP)
  - Replicating SQL Database
  - RPO ~ 4 hours

LUN Layout for VMware ESX

Data Volume: 668 GB

Log Volume 400 GB

Boot Volume: 133 GB

Raid Group Layout
Front View - RaidGroup/(Disk Size)

| Bus 0, Encl 3 | 101 (F146) r5 | 101 (F146) r5 | 101 (F146) r5 | 101 (F146) r5 | 101 (F146) r5 | 101 (F146) r5 | 102 (F146) r5 | 102 (F146) r5 | 102 (F146) r5 | 102 (F146) r5 | 103 (F146) r1_0 | 103 (F146) r1_0 | 6 (F146) r1_0 | 6 (F146) r1_0 | 238 (F146) hs |
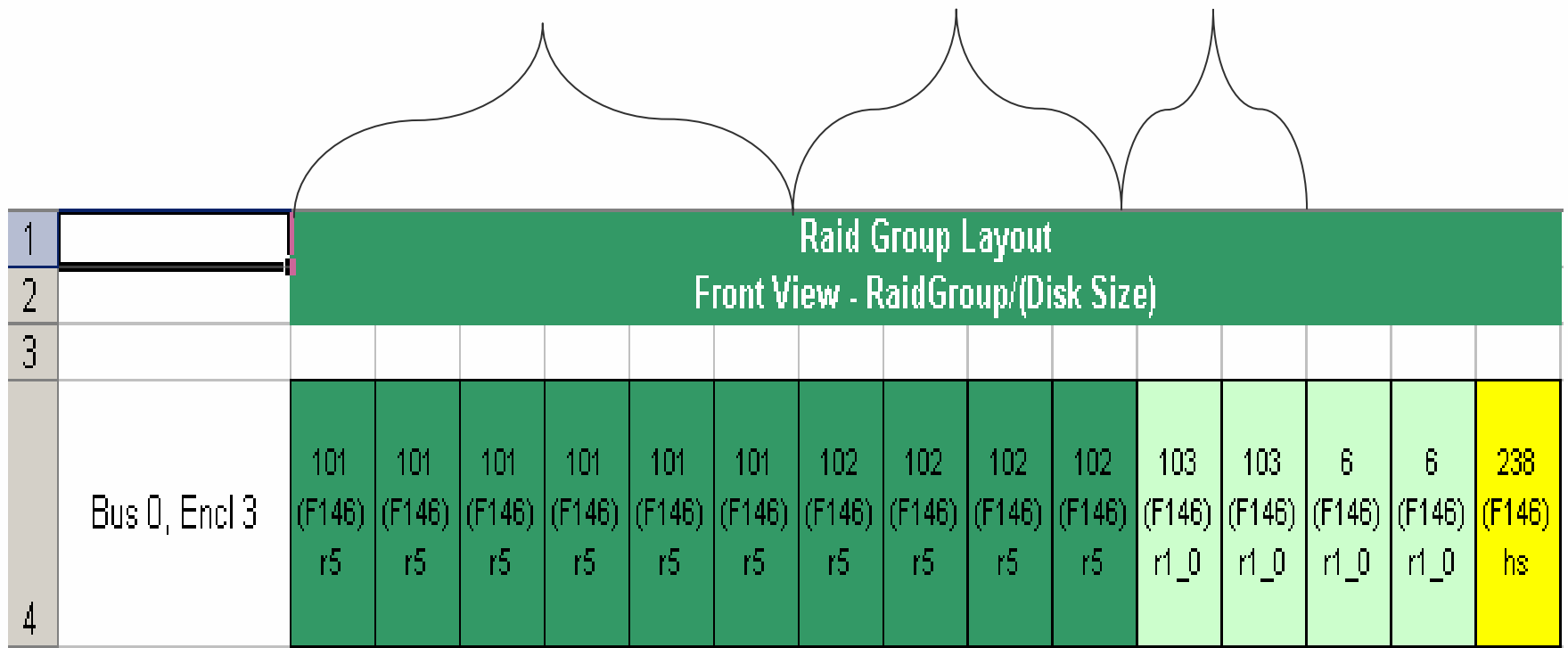
Final Implementation Diagram

CX3-20 Production CLARiiON

Production ESX servers

CX3-20 Secondary CLARiiON

Disaster Recovery ESX servers

MirrorView Asynchronous [20 miles]

Boot

Data

Log

Cisco MDS 9216i

Cisco MDS 9216i

Remote replica of Boot

Remote replica of Data

Remote replica of Log

VMware: Virtualization layer
Intel architecture

VMware: Virtualization layer

VMware: Virtualization layer

Consistency Group

Customer's Future Plans

- Current Backup strategy includes individual VM backup
  - Incremental Backup during the week
  - Full Backup on the weekend
- Looking at VMware Consolidated Backup

- Looking at installing SnapView on the remote CX3-20 system
  - Take Snapshots of their target MirrorView/A volumes
  - For test and development

References

- Data Migrations and Data Vaulting using EMC Open Replicator, EMC SAN Copy
  - 11:45 – 12:45 (11/08/2006)
- Introduction to Networked Storage Performance in a Virtualized environment
  - 3:15 – 4:15 (11/07/2006)
- VMware ESX server and Storage Architecture Best Practices for Performance, Backup and Disaster Recovery
  - 4:45 – 5:45 (11/08/2006)

- EMC Documentation on vmware.com
  - VMware ESX Server Using EMC Symmetrix Storage Systems
  - VMware ESX Server Using EMC CLARiiON Storage Systems
  - CLARiiON and VMware integration whitepaper
  - Celerra FAQ document

Please remember to complete your
session evaluation form
and return it to the room monitors
as you exit the session

The presentation for this session can be downloaded at
http://www.vmware.com/vmtn/vmworld/sessions/

Enter the following to download (case-sensitive):

Username: cbv_rep
Password: cbvfor9v9r