IP43

Building Virtual Infrastructures With NAS

Peter Learmonth, Kim Weller
Consulting Systems Engineers, NetApp

Bud James
Data Center Operations, BEA

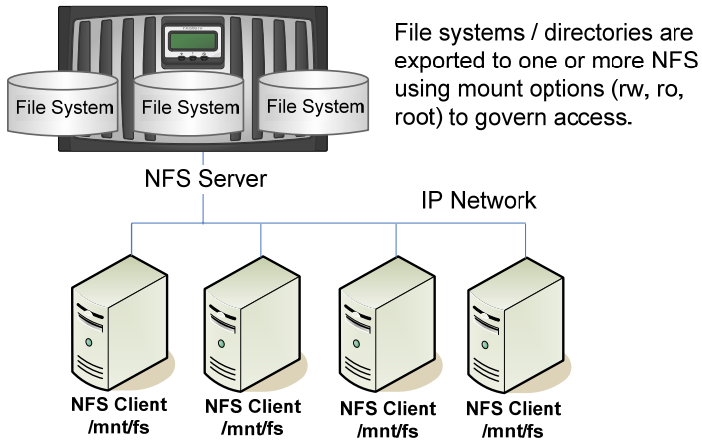
Agenda

- A Brief review of the NFS protocol
- Characteristics of ESX NFS Client and VI3 NFS Datastores
- Cost benefits of NFS VI3 Deployments
- Network Design with VI3 and NFS
- Configuring NAS & VI3
- Backup and Recovery
- NFS Performance
- BEA Implementation Profile

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Brief Overview Of The NFS Protocol

- **NFS Server**
  - Implements NFS daemon processes (nfsd) in order to make data generically available to clients
  - Administrator exports names and parameters of directories
    - /etc/exports configuration file and exports command
  - Server security-administration ensures that it can recognize and approve validated clients



- **NFS Client**
  - Requests access to exported data (with mount command)
  - Responsible for establishing parameters of mount
    - Transport: TCP/UDP
    - Error Handling: Hard / Soft / Interrupt
    - Write Size: 8k, 32k, etc.

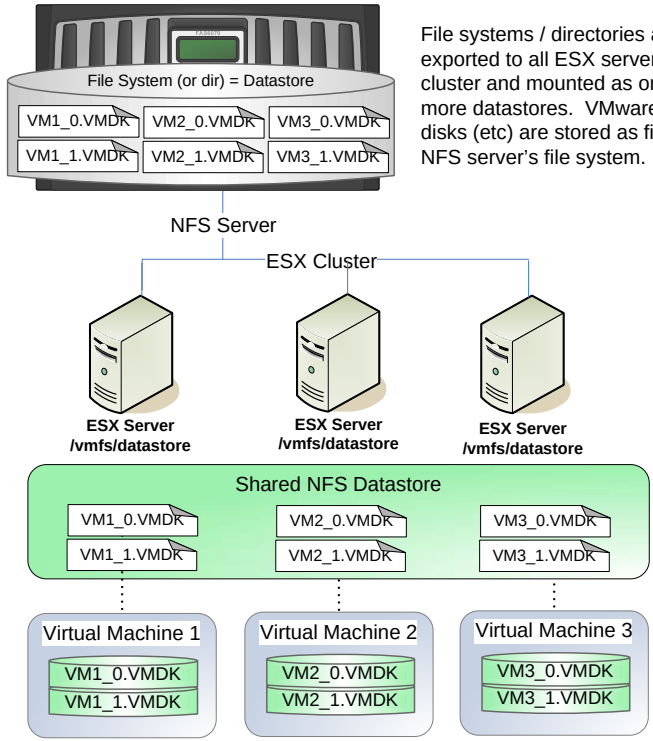
- **NFS Protocol**
  - Mature, standardized protocol as established in IETF RFCs 1094 1813 3530 (NFS v2, v3, v4)
  - Stateless protocol means that there is no need to do state recovery after a client, server, or network has become unavailable.
    - Clients can be configured to retry I/O to non-responding mounts indefinitely
    - Server can service requests as soon as it begins running.

Copyright © 2007 Network Appliance, Inc. All rights reserved.

VI3 NFS Implementation: Client and Datastores

- **Client characteristics:**
  - **NFSv3**
  - **Fixed Mount Options**
    - **TCP (retransmit only dropped packets)**
    - **Block Size**
  - **Ability to mount Read-Only datastores**
    - **Useful for read-only point-in-time copies or replication destination**

- **Datastores**
  - **8 default, 32 maximum**
  - **Mounted to /vmfs/volumes/datastore_name**
  - **Persistent across reboots**
  - **Dynamic Resizing**
  - **Leverages VMware NIC teaming / load balancing for path management**



File systems / directories are exported to all ESX servers in cluster and mounted as one or more datastores. VMware virtual disks (etc) are stored as files on NFS server's file system.

Note: The NFS client for datastore traffic is implemented at the VMKernel and is completely different/separate from the service console NFS client. In the absence of NFS datastores, the service console NFS client may be used to mount administrative exports containing application ISOs.

Copyright © 2007 Network Appliance, Inc. All rights reserved.

NFS Use Cases With VI3

- **ISO and Software Repository**
  - **Handy way to access single repository from all ESX servers**
- **Backup Target**
  - **ESX Ranger and similar used to backup VMFS to NFS**
- **Datastores containing VMs ★**
- **Other uses out of scope of this talk**
  - **Guest OS NFS client (in a VM) directly mounting NFS export**
  - **Traditional backup server backing up to NFS disk pool**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

VI3 NFS Implementation: Feature Support With VMs

- **Shared datastores – YES**
- **VMotion – YES**
  - **VMotion works with VM using VMDK on NFS datastore**
  - **ISOs on NAS shares prevent VMotion failure due to HW CD-ROM connected**
- **HA – YES**
- **DRS – YES**
- **VMware snapshots – YES**
- **VCB – NO**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Features Of An Enterprise-Class NAS Server

- **The BASICS**
  - **High Availability**
  - **Advanced cost-effective RAID Protection (High performing RAID-6)**
  - **Ability to resize (grow and shrink) datastores**

- **OPTIMIZING THE VALUE of NFS datastores**
  - **Ability to thin provision datastores**
  - **Storage (NFS server) based SnapShots of datastore**
    - **Restore individual virtual machine files without network copy**
    - **Restore datastores without network based data copy**
    - **Ability to create writable snapshots which may be mounted as non read-only datastores**
  - **Storage (NFS server) based cloning of datastore or VM**
    - **Ability to thin-provision clones**
  - **Storage (NFS server) based remote replication (DR)**
    - **Ability to create clones at replication destination**
  - **Deduplication / Single-Instance Storage for primary and backup storage**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Cost Benefits Of NFS With VI3

- **Standard, ubiquitous Ethernet Infrastructure**
  - **Switches**
  - **NICs**
- **Less specialized training (you already know how to do this!)**
- **Space optimization**
  - **One less layer of file system needing space for metadata and free space**
  - **VMDKs thin provision by default**
- **More flexible space management of datastores**
  - **Thin provision datastore ***
  - **Reclaim unused space by shrinking datastores ***

**** Not all NAS solutions can do this***

Copyright © 2007 Network Appliance, Inc. All rights reserved.

VI3: Current Caveats With NFS Datastores

- As with VMDK in VMFS, MSCS between VMs or VM and physical is not supported
- VMware Consolidated Backup works only with FCP or iSCSI protocol
  - Other ‘proxy’ backup options exist
- Maximum 32 NFS datastores
  - Default is 8

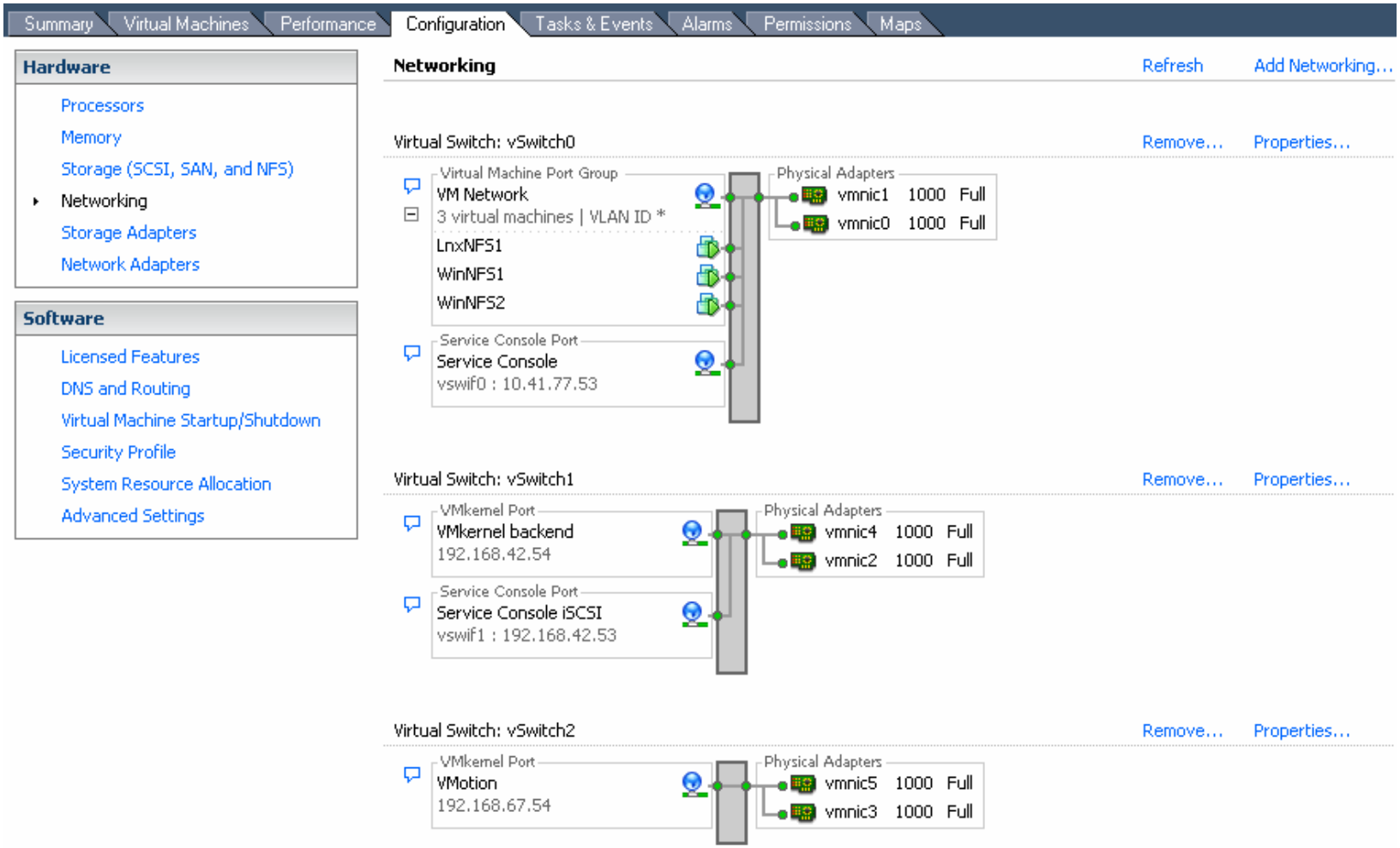
Copyright © 2007 Network Appliance, Inc. All rights reserved.

VI3 Network Design: Best Practices

- **Use Gigabit Ethernet (i.e. not 100Mb)**
  - **10GbE available on Network Appliance storage today (consult other vendors to determine their support)**
- **Build separate networks or VLANs**
  - **VM network**
  - **Storage**
  - **VMotion**
- **Achieve port redundancy and better throughput with NIC Teaming**
  - **Etherchannel (= VIFs on NetApp)**
  - **Match VMware load balancing and switch load balancing**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

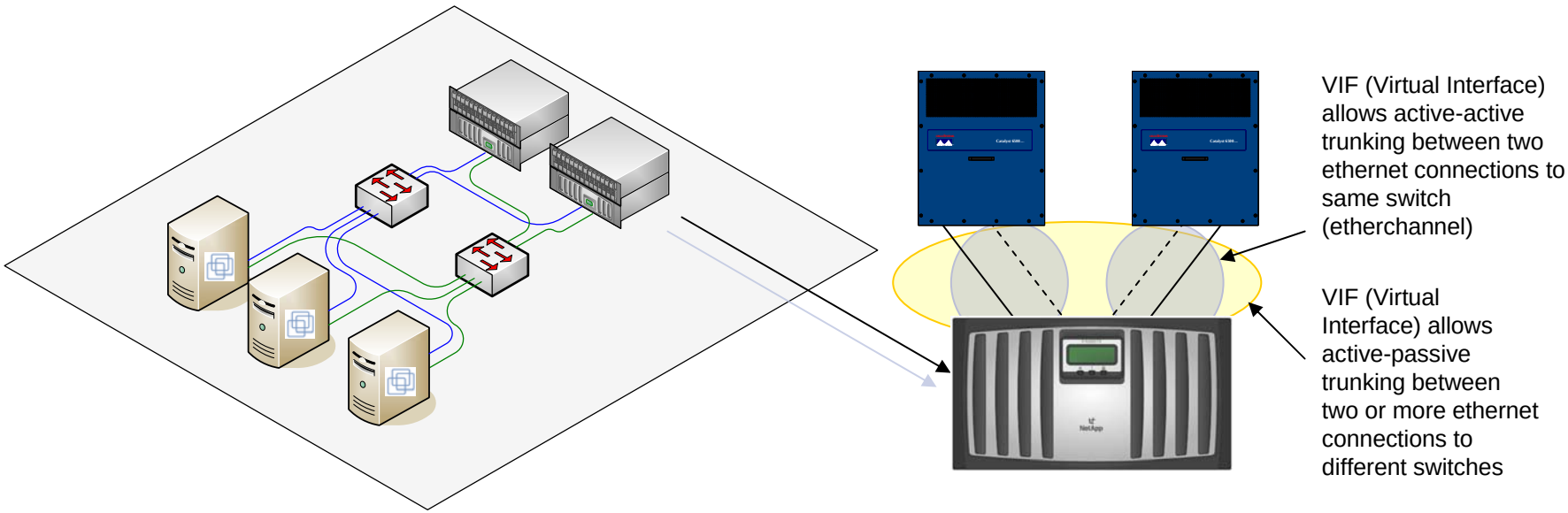
VI3 Network Design: Virtual Switches / Ports Groups



Copyright © 2007 Network Appliance, Inc. All rights reserved.

VI3 Network Design For NFS: Redundancy

- **Switches may be**
  - **Separate switches**
  - **Stackable switches (allows multimode VIF between switches on NetApp)**
  - **Separate blades in a high-end switch (Cisco Catalyst, etc.)**

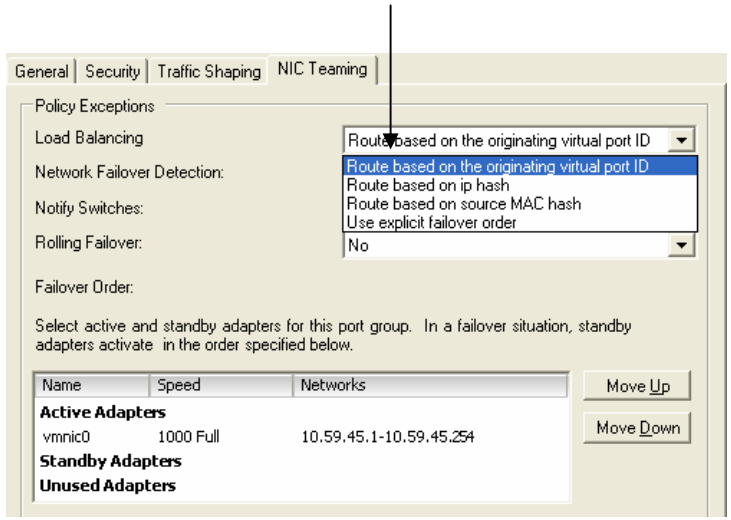


Copyright © 2007 Network Appliance, Inc. All rights reserved.

VI3 Network Design for NFS: Load Balancing

VI3 IP storage networks leverage NIC Teaming and Load Balancing policies

- **Originating virtual port ID:**
  - Choose uplink based on virtual port
- **IP Hash (Source-Destination IP):**
  - Choose uplink based on source+destination IP of each packet
- **Source MAC Hash:**
  - Choose uplink based on source MAC address
- **Explicit Failover Order:**
  - Always use the highest order uplink



CRITICAL: Ensure that virtual switch and physical switch use same load balancing algorithm :

Ex: IP Hash = Source / Dest IP

NFS traffic uses single VMKernel IP address per vSwitch / set of team NICs

This means that a single uplink will be selected for ALL traffic with:

- **Load balancing based only on source IP/MAC/Port**
- **Load balancing with source IP and a SINGLE DESTINATION IP**
- **Single datastore**

However, a bit of planning—and the right NFS server—can solve the problem of generating NFS traffic over multiple uplinks.

Copyright © 2007 Network Appliance, Inc. All rights reserved.

VI3 Network Design For NFS: Load Balancing (cont)

- **Network Configuration**
  - Single NFS server, separate physical networks
    - This will require more physical NICs and generally the use of additional network segments or VLANs.
  - Single NFS server, IP Aliases
    - Advanced feature of NAS storage which allows a single NIC or a group of teamed NICs (Virtual Interface--or VIF--on NetApp platforms) to have multiple IP addresses
    - To generate NFS traffic across multiple uplinks teamed on a single vSwitch, mount more than one NFS datastore using more than one IP addresses.
  - Multiple NFS servers
    - Most costly option. Self explanatory.
- **Datastore Configuration**
  - Single datastore per exported file system / directory
  - Multiple datastores per exported file system / directory
    - VI3 allows the same file system / directory to be mounted as multiple datastores
    - When this capability is leveraged with IP aliasing, physical resources are aggregated / leveraged.

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Setting Up The NAS

- **Create Volumes**
  - **GUI or CLI**
  - **Use UNIX permission/volume/qtree type (check this)**
    - NetApp is multiprotocol which also allows for CIFS access (or NFS access) directly from virtual machine or physical servers
- **Export**
  - **rw= and root= to ESX servers.**
    - rw only is OK for read only datastores
    - Export to all servers in the cluster
  - **List individual servers by back-end IP or whole back-end subnet**
    - Can use hostnames or FQDN if DNS is properly configured including hostnames for storage network interfaces

Copyright © 2007 Network Appliance, Inc. All rights reserved.

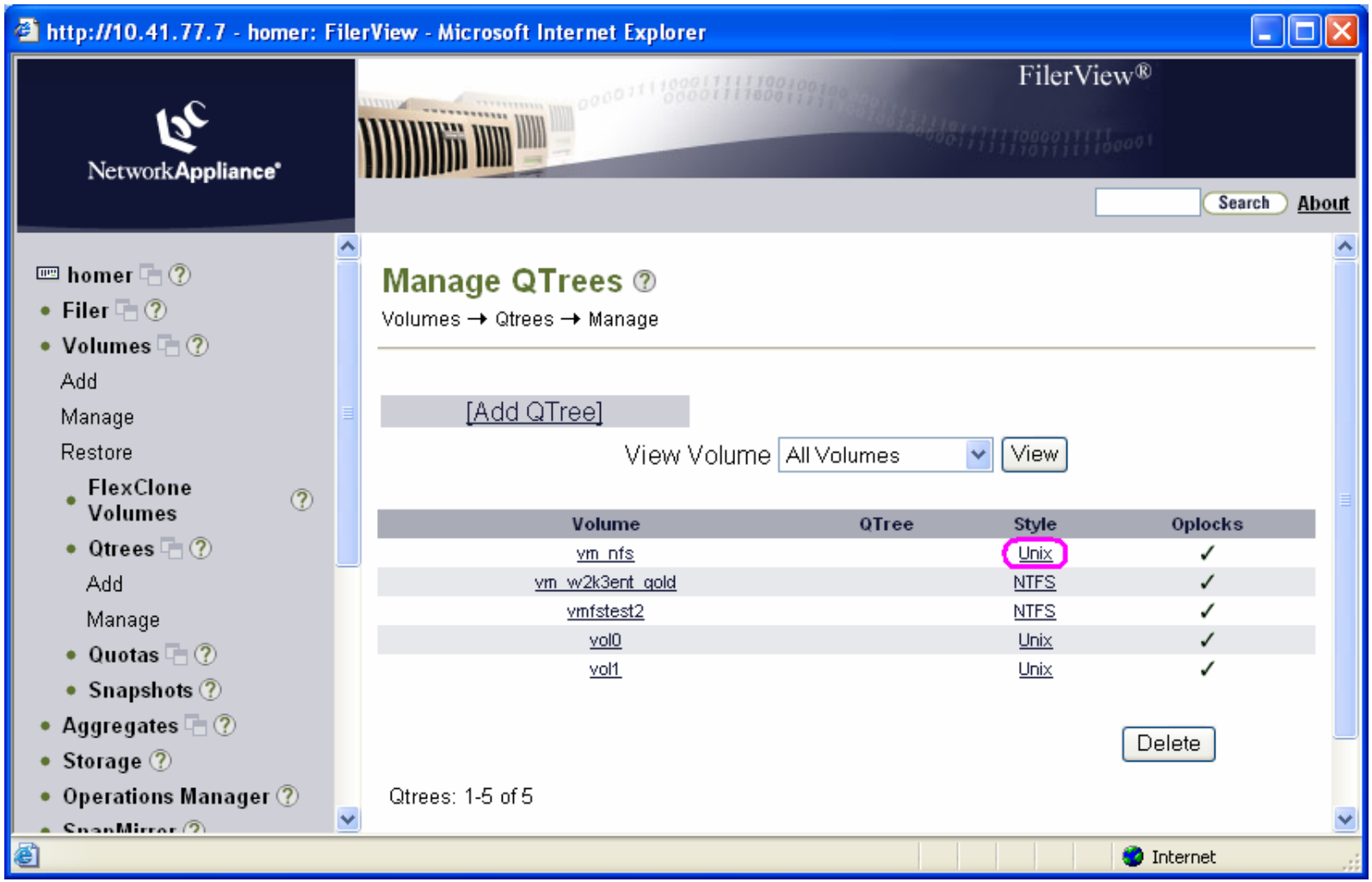
Create A Volume

- **FilerView GUI**
  - **Volumes → Add (Follow the wizard)**
- **Command Line**

```
vol create <volname> <aggregate> <size>
homer> vol create vm_nfs aggr0 200g
Creation of volume 'vm_nfs' with size 200g on
  containing aggregate 'aggr0' has completed.
```

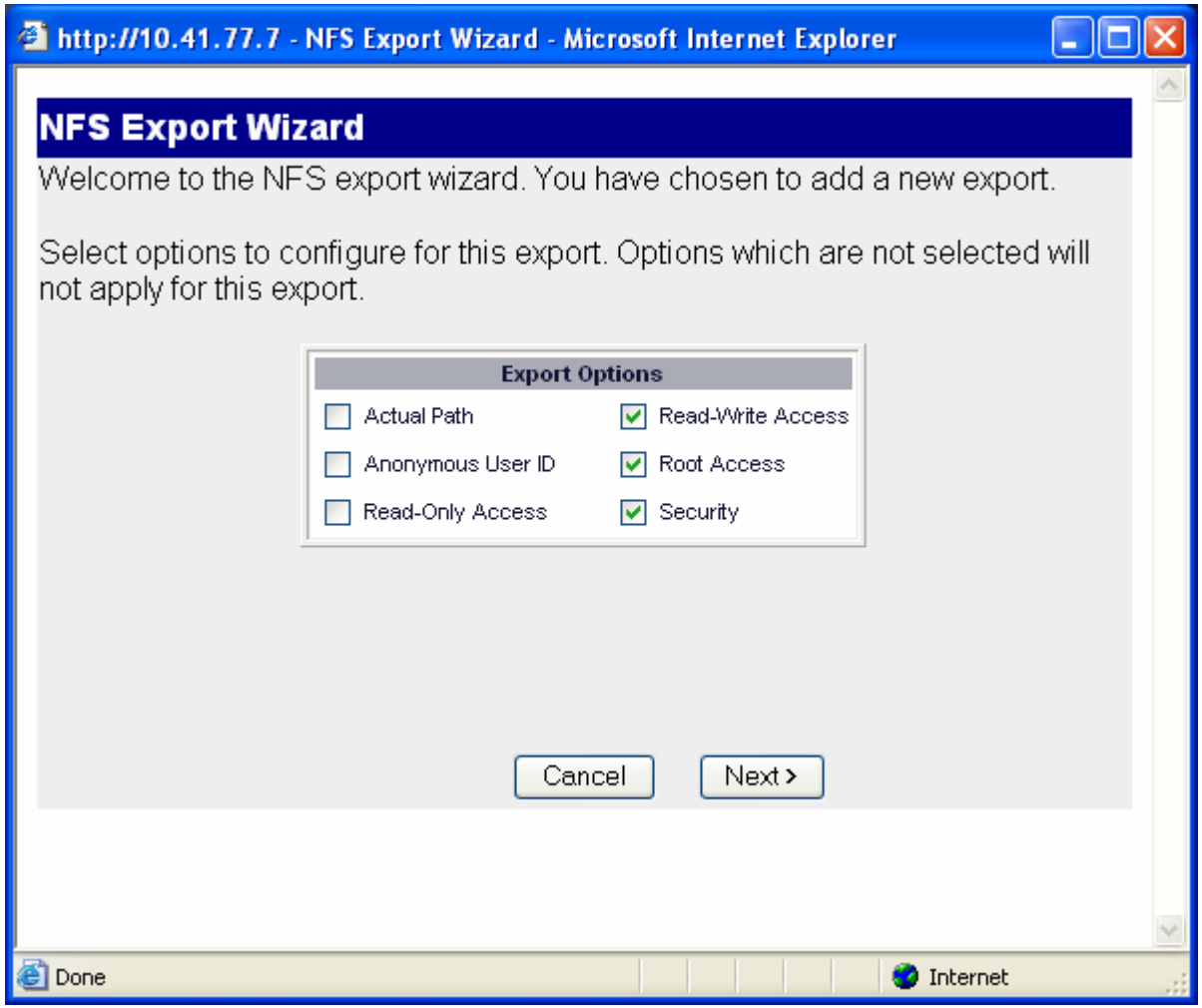
Copyright © 2007 Network Appliance, Inc. All rights reserved.

Set UNIX Permission Type



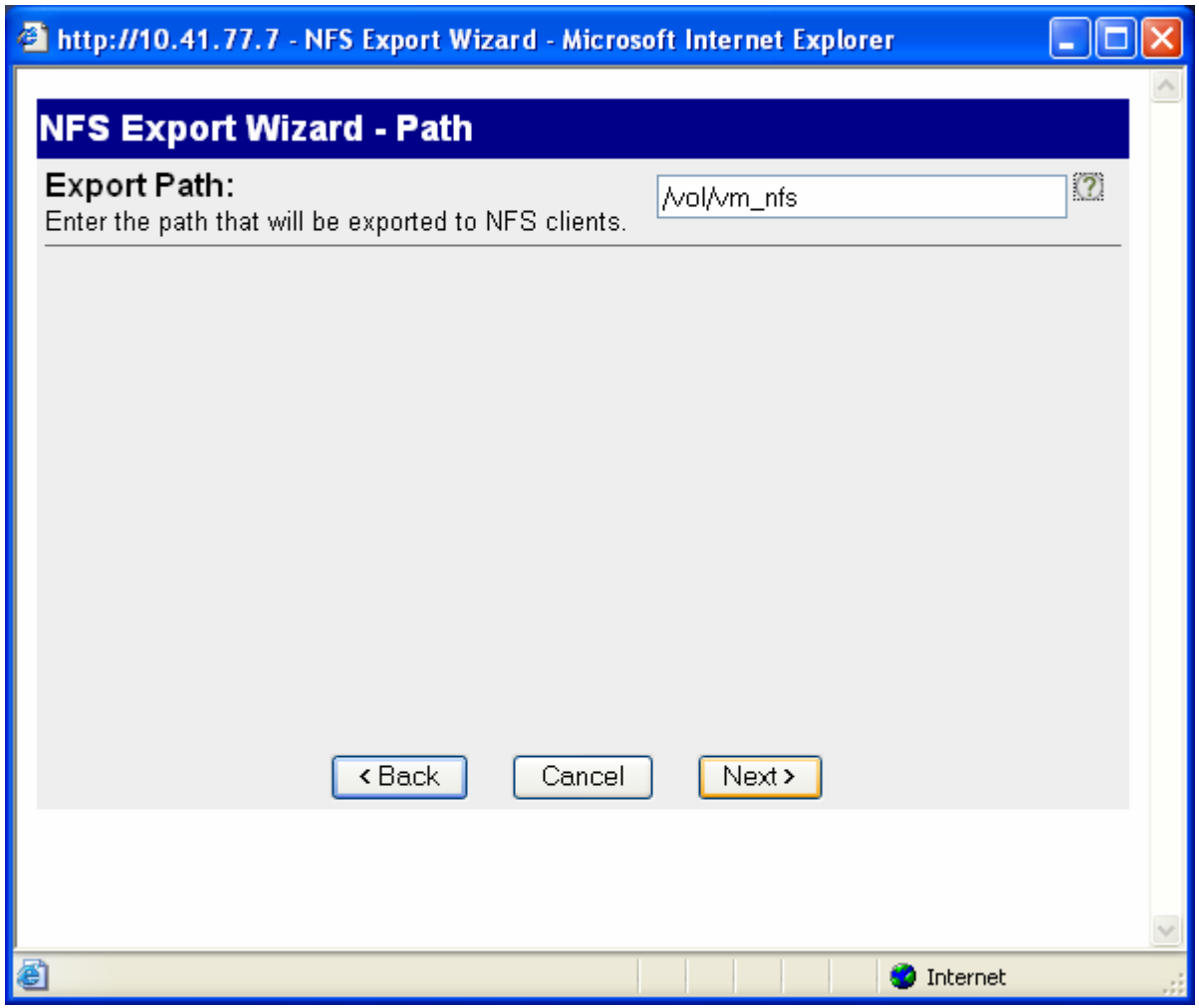
Copyright © 2007 Network Appliance, Inc. All rights reserved.

Export The Volume



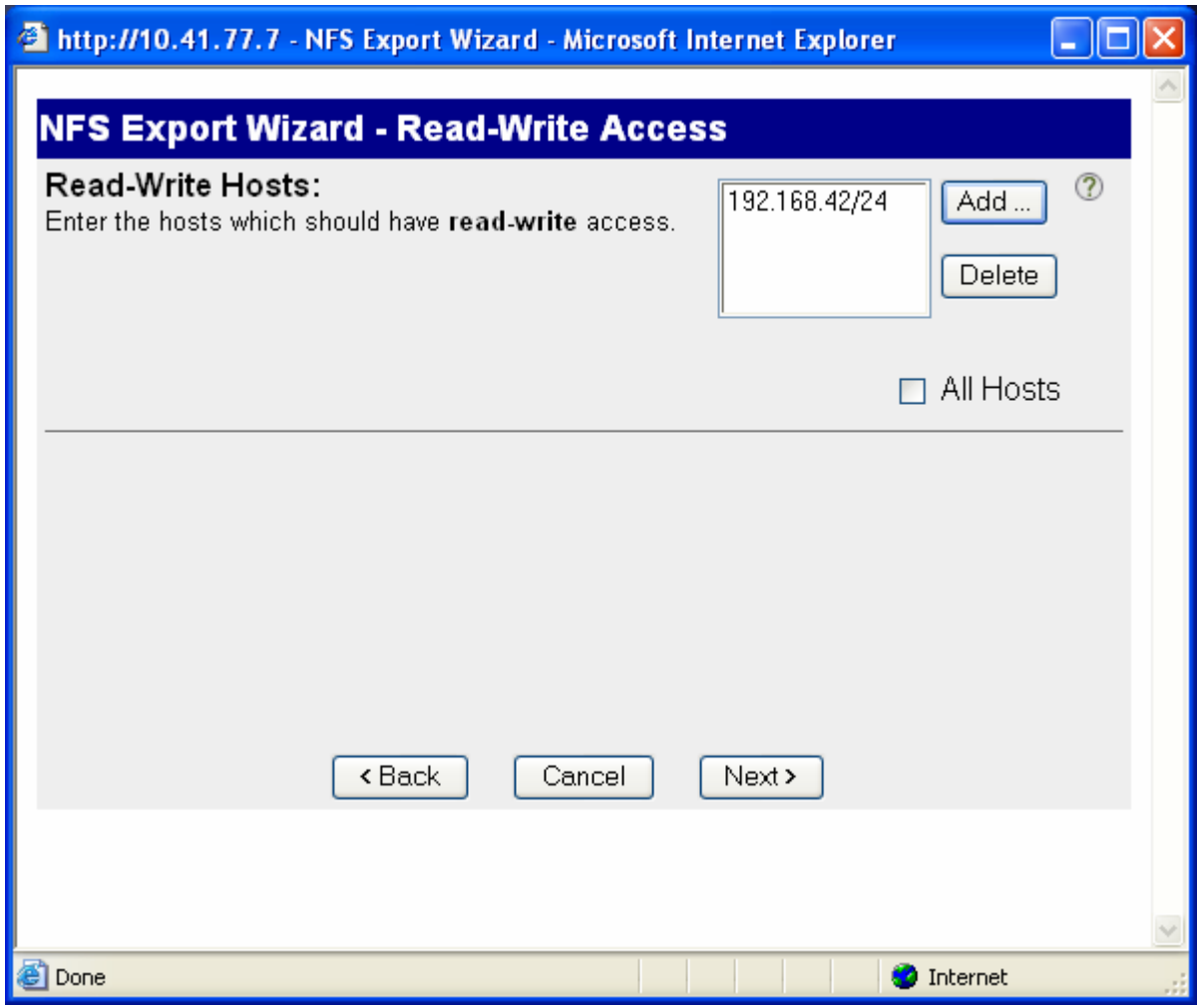
Copyright © 2007 Network Appliance, Inc. All rights reserved.

Export The Volume



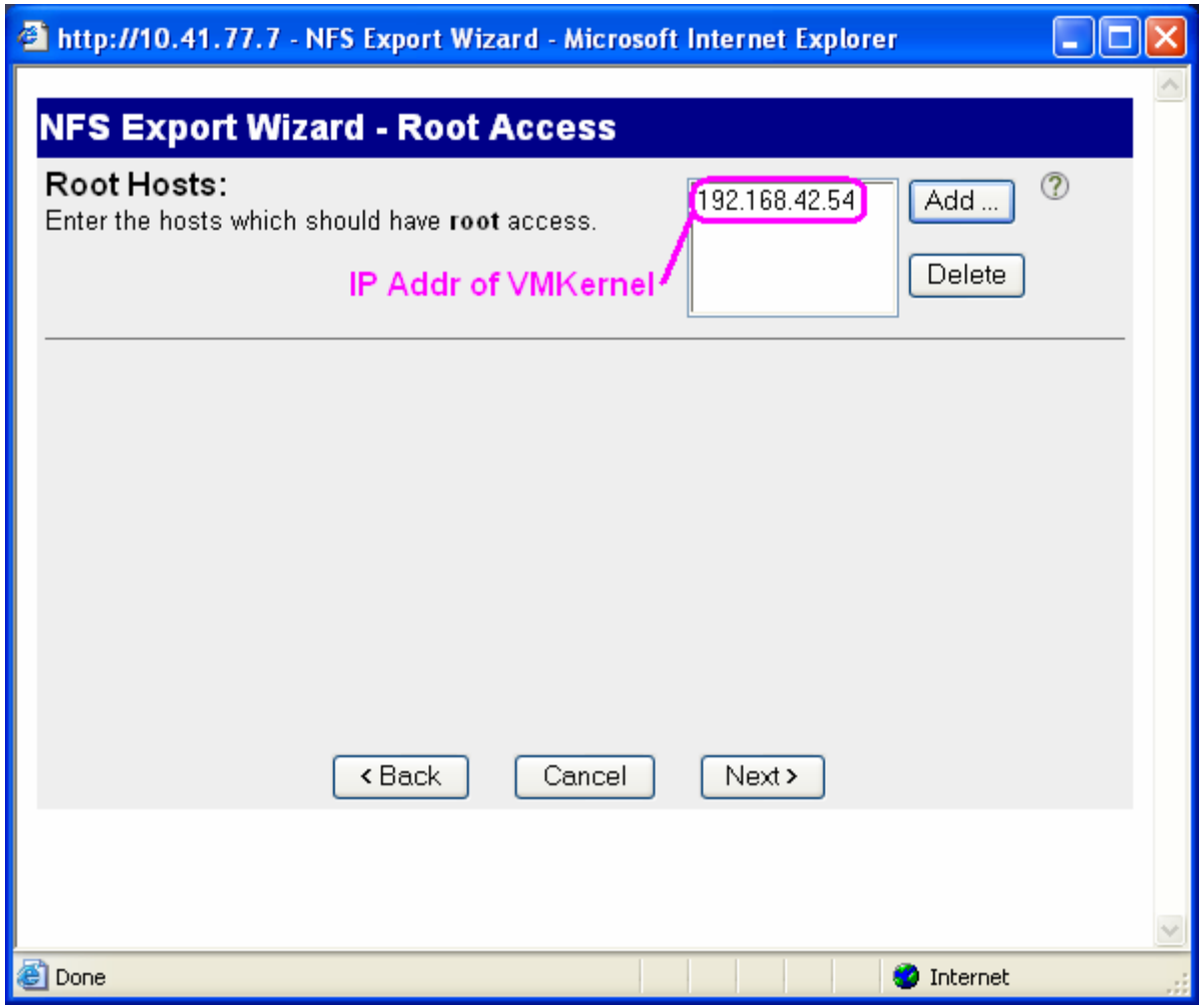
Copyright © 2007 Network Appliance, Inc. All rights reserved.

Export The Volume



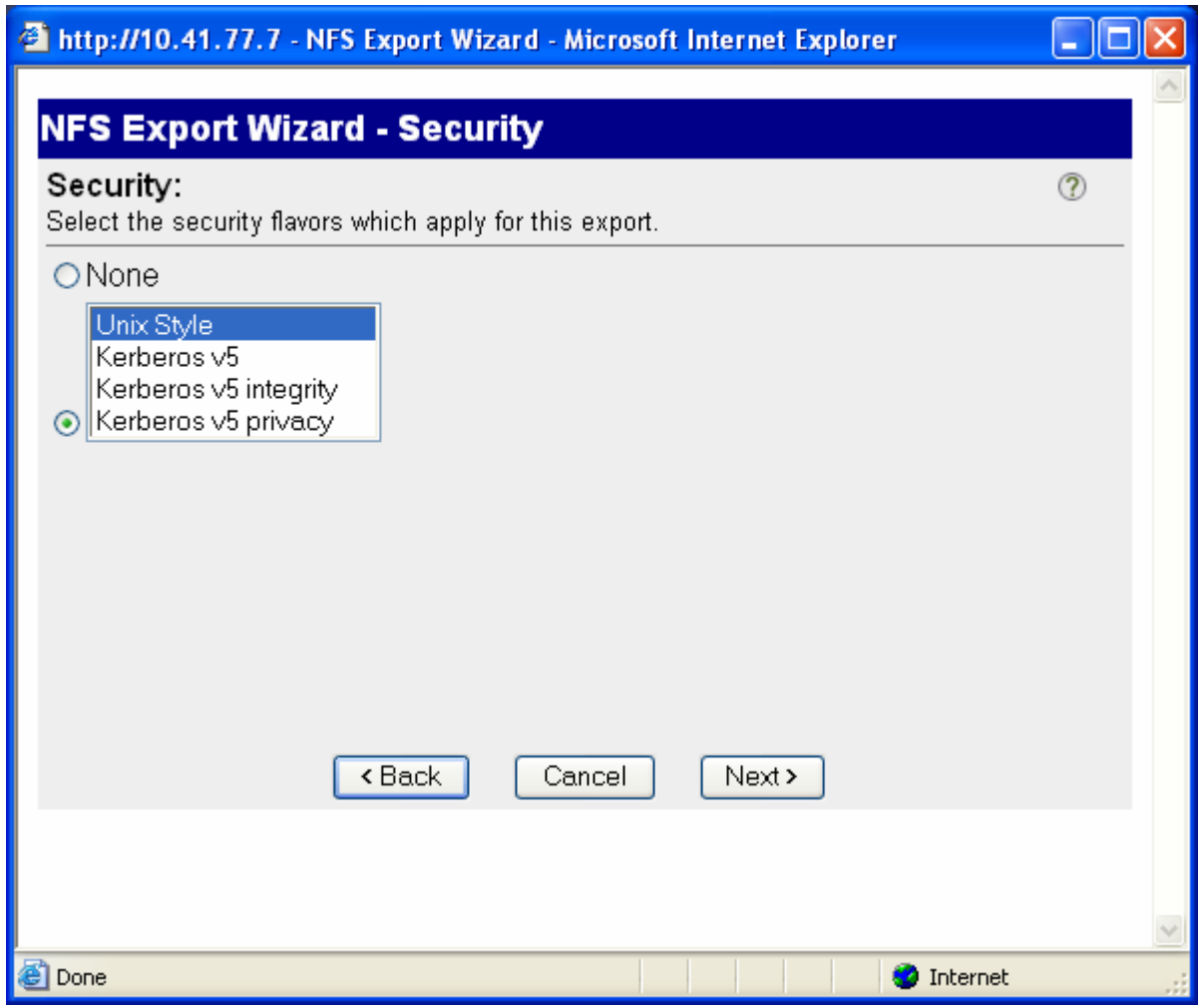
Copyright © 2007 Network Appliance, Inc. All rights reserved.

Export The Volume



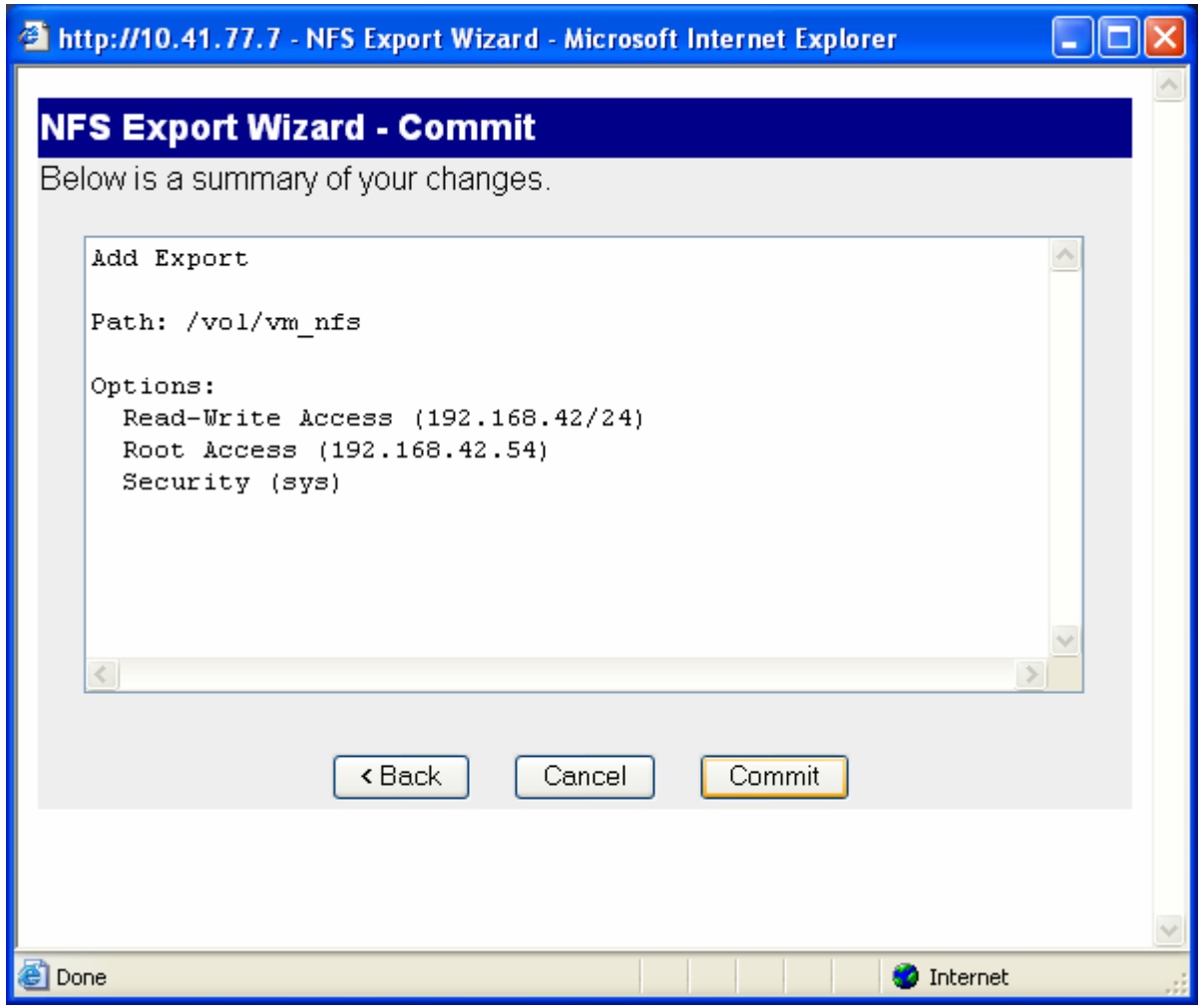
Copyright © 2007 Network Appliance, Inc. All rights reserved.

Export The Volume



Copyright © 2007 Network Appliance, Inc. All rights reserved.

Export The Volume



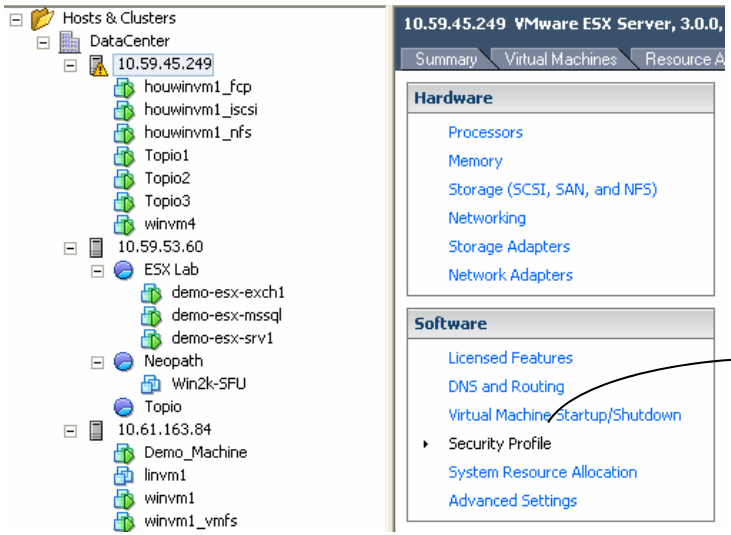
Copyright © 2007 Network Appliance, Inc. All rights reserved.

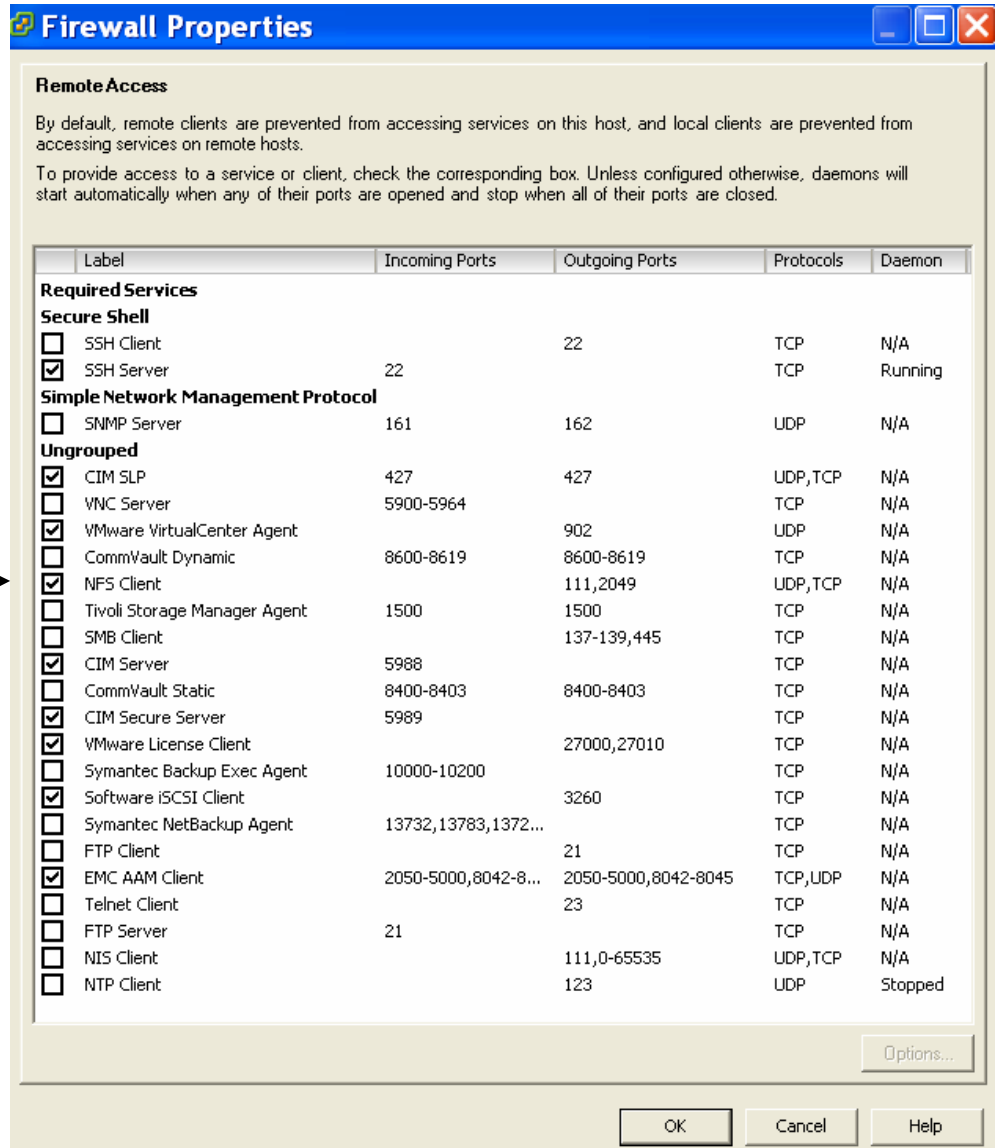
Setting Up ESX

- Enable the NFS client (through security settings)
- Add VMKernel networking
- Increase NFS Max Volumes if necessary
- Add new NFS datastore

Copyright © 2007 Network Appliance, Inc. All rights reserved.

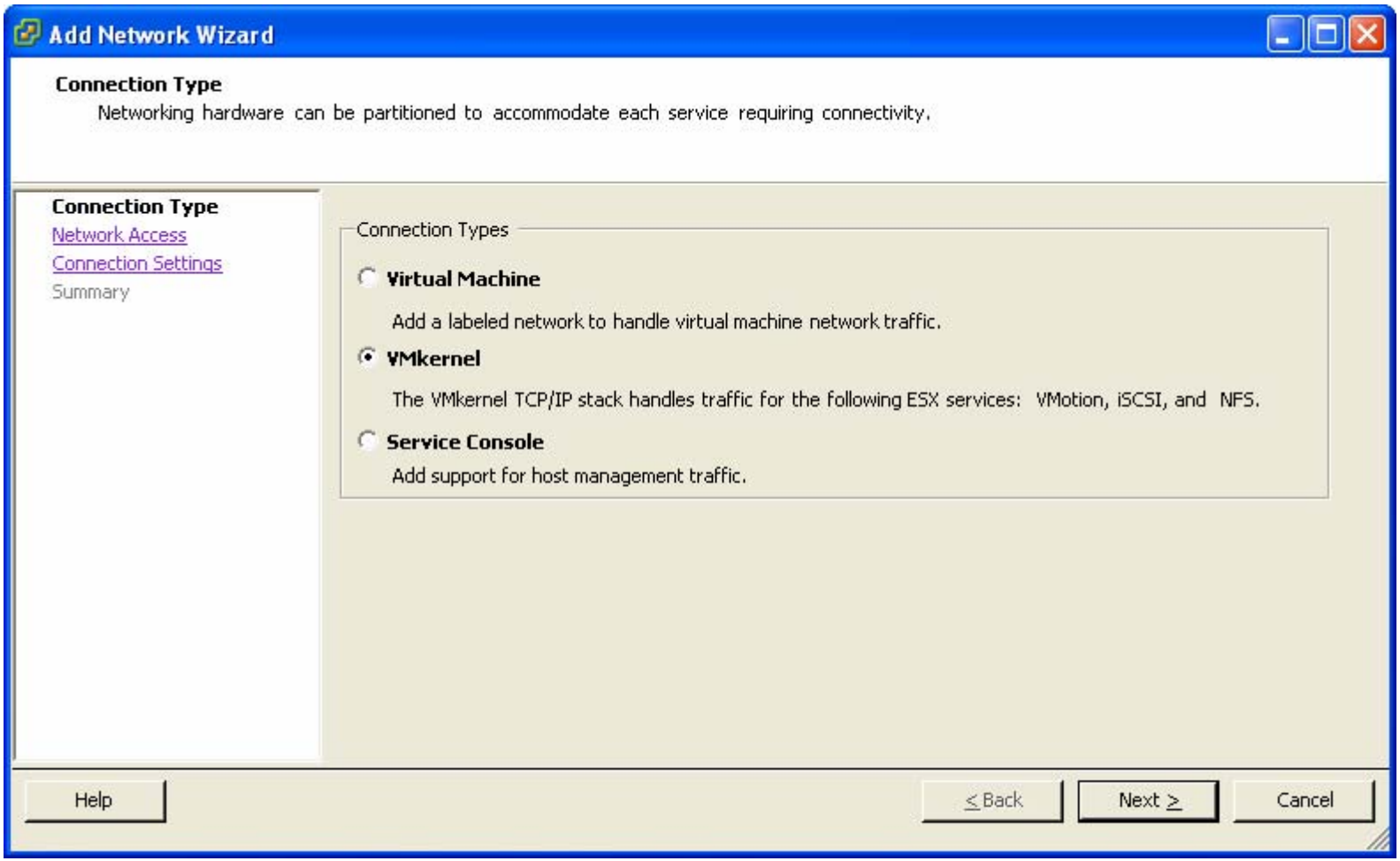
Enabling NFS Client





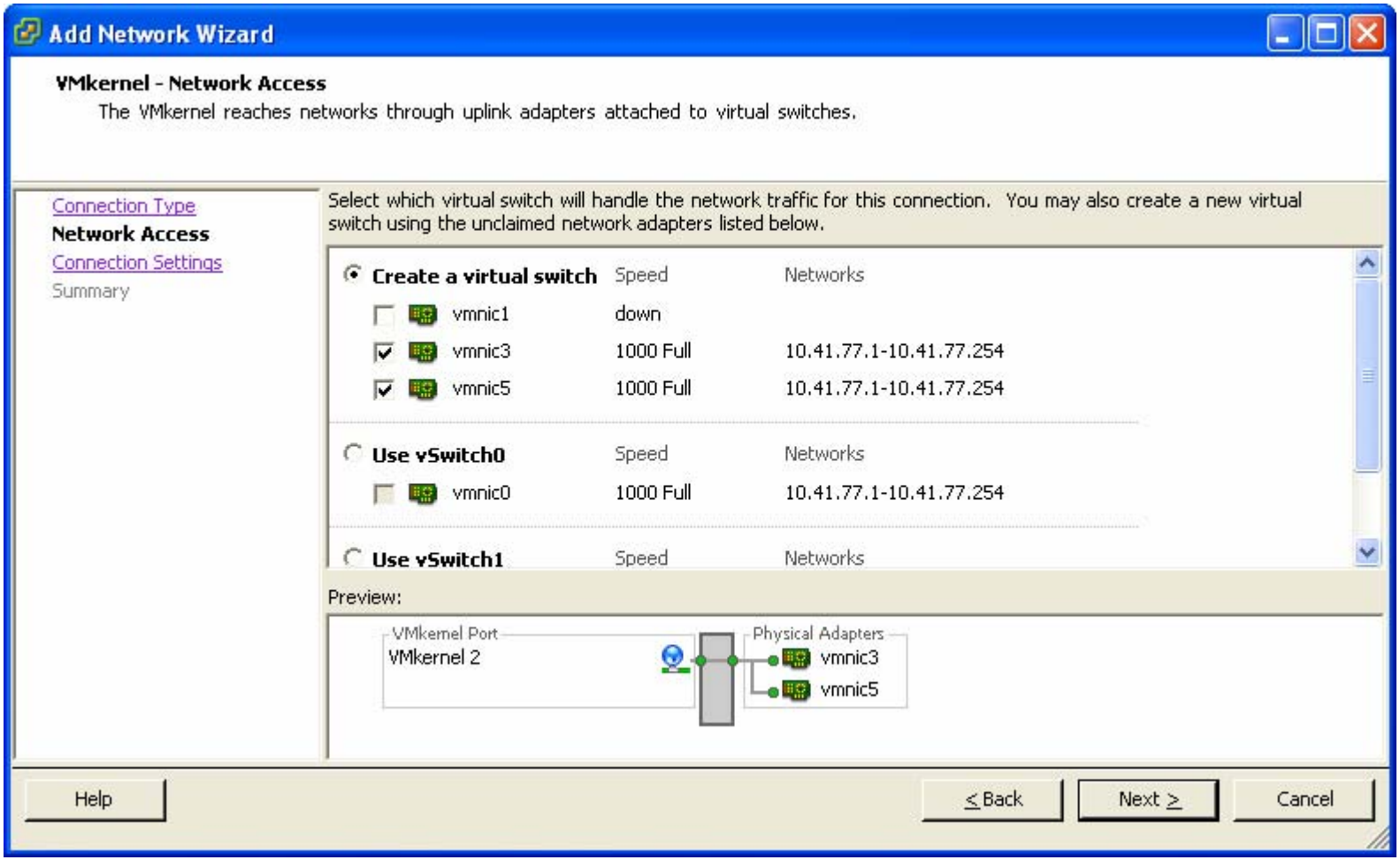
Copyright © 2007 Network Appliance, Inc. All rights reserved.

Add VMKernel Networking



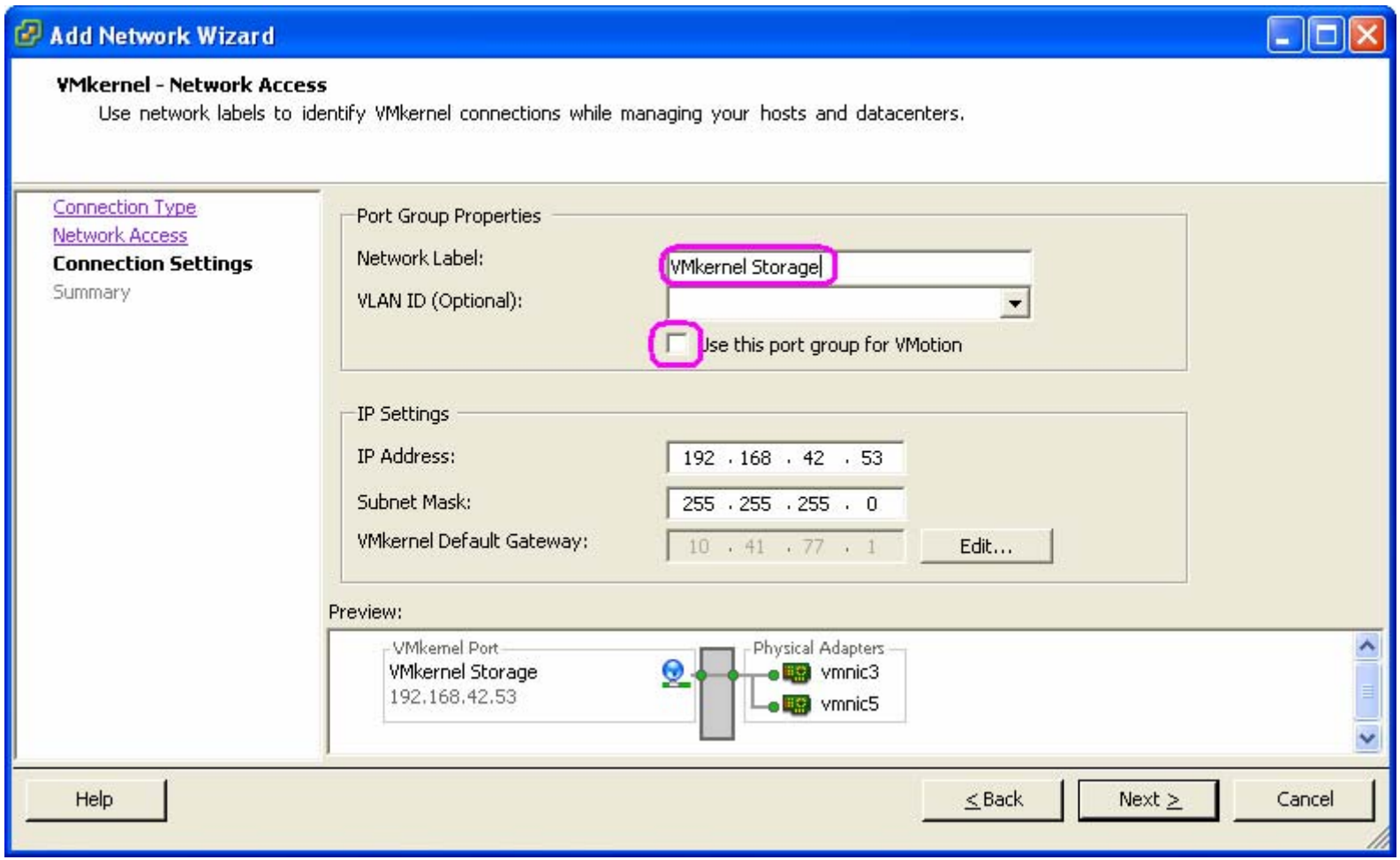
Copyright © 2007 Network Appliance, Inc. All rights reserved.

Add VMKernel Networking



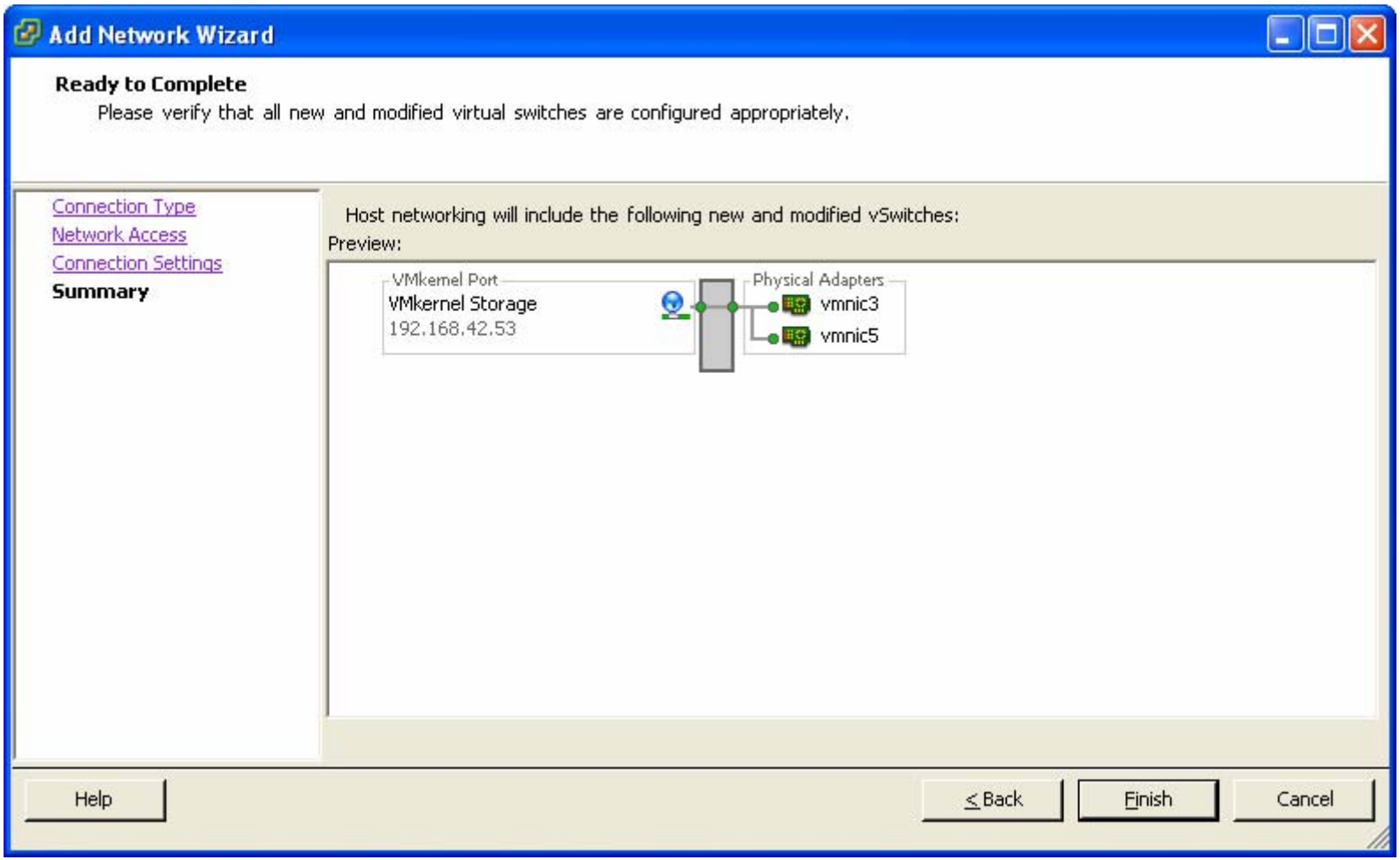
Copyright © 2007 Network Appliance, Inc. All rights reserved.

Add VMKernel Networking



Copyright © 2007 Network Appliance, Inc. All rights reserved.

Add VMKernel Networking

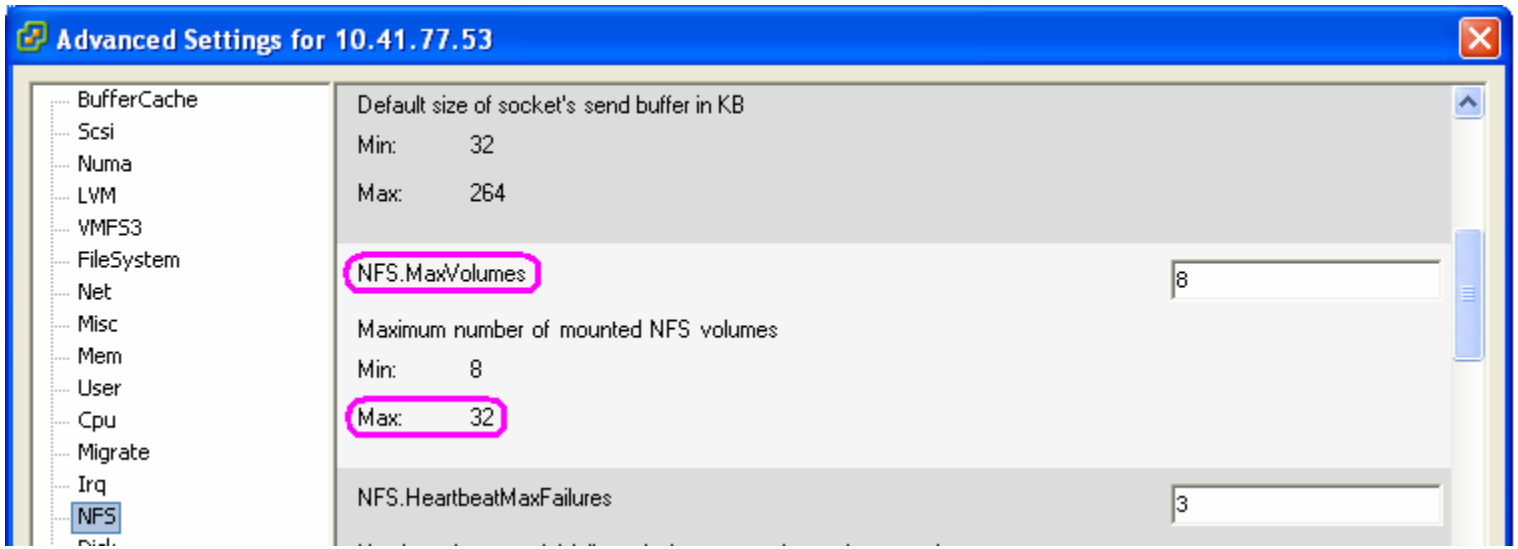


Copyright © 2007 Network Appliance, Inc. All rights reserved.

Increase NFS.MaxVolumes

- **Configuration → Advanced Settings → NFS**
- **Command Line**

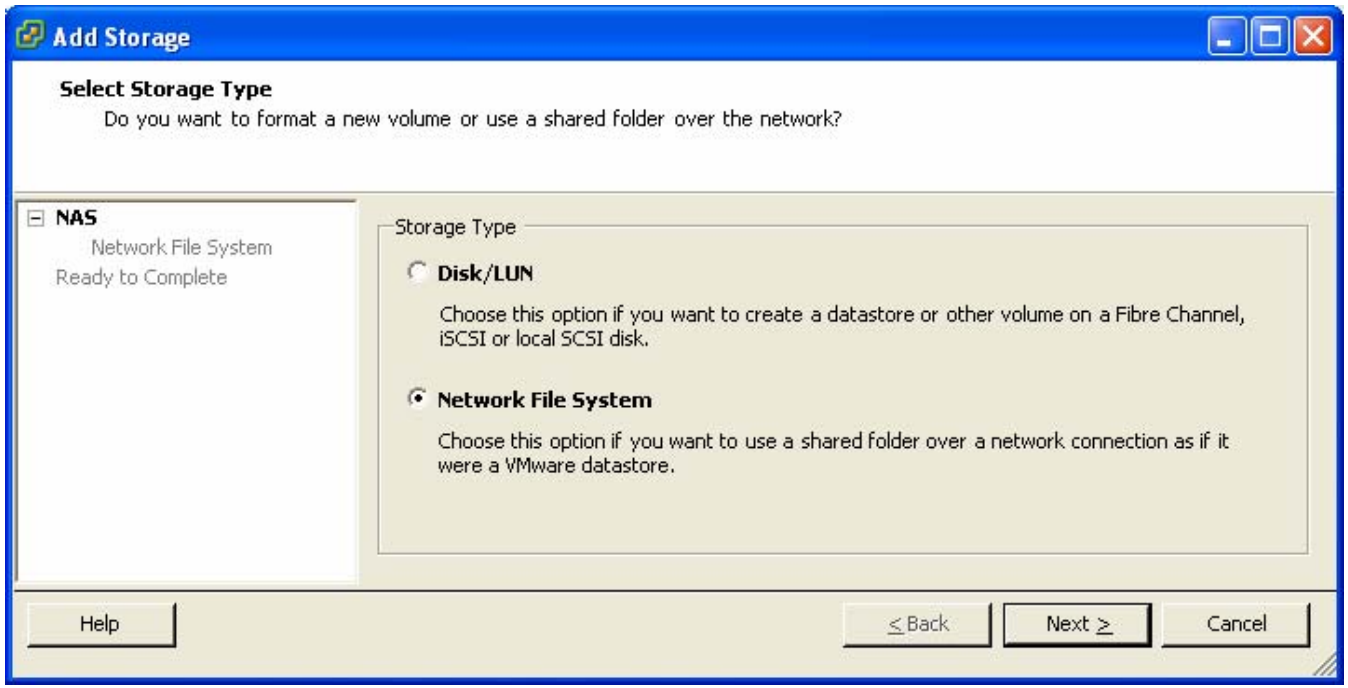
```
# esxcfg-advcfg -g /NFS/MaxVolumes
        Value of MaxVolumes is 8
# esxcfg-advcfg -s 32 /NFS/MaxVolumes
        Value of MaxVolumes is 32
```



Copyright © 2007 Network Appliance, Inc. All rights reserved.

Add New NFS Datastore

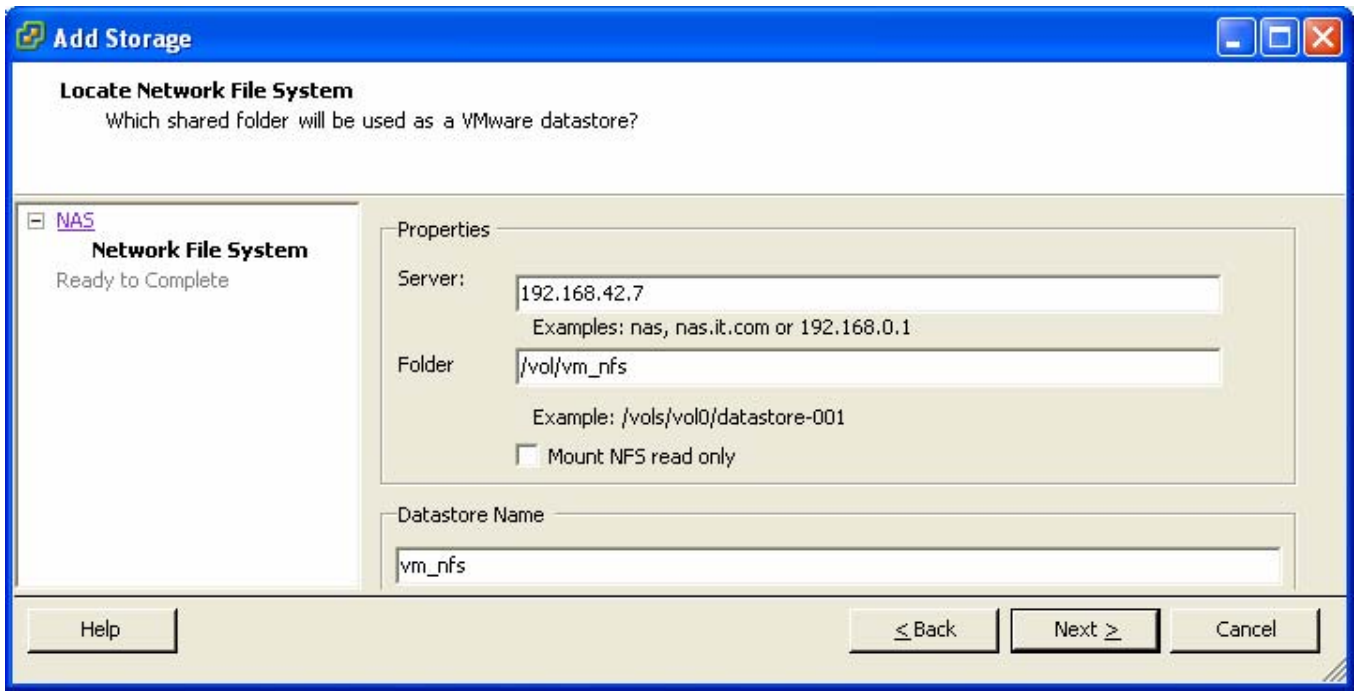
- **Configuration → Storage → Add Storage → NFS**



Copyright © 2007 Network Appliance, Inc. All rights reserved.

Add New NFS Datastore

- **Configuration → Storage → Add Storage → NFS**
- **Datastore Name (akin to a mount point)**
  - **Must be the same name on all ESX hosts in the cluster (VMotion and HA requirement)**



Copyright © 2007 Network Appliance, Inc. All rights reserved.

Backup

- **Snapshot Copies are first defense**
  - **Use VMware snapshots to quiesce, then storage snapshots**
  - **Scripts available to automate**
- **Next, back up off storage**
  - **NDMP (Network Data Management Protocol)**
    - **Back up to tape (and other media) directly from NAS**
  - **Backup proxy server**
    - **VCB does not support NFS**
    - **Same functionality can be provided with Linux backup proxy**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Linux– Loopback Mount NTFS

- **Mount the NFS export on the Linux server.**

```
# mount 10.41.77.76:/vol/vm_nfs /mnt/vm_nfs
```

- **Mount the VMDK as a loopback mount specifying the starting offset and NTFS file system type.**

```
# mount /mnt/vm_nfs/.snapshot/hourly.0/peternfs/peternfs-flat.vmdk
   /mnt/vmdk
   -o ro,loop=/dev/loop2,offset=32768 -t ntfs
```

- **Access NTFS from Linux!**

```
# cd /mnt/vmdk
# ls -l
-r-------- 1 root root         0 Sep 11 2006  AUTOEXEC.BAT
-r-------- 1 root root       210 Dec 18 21:00 boot.ini
-r-------- 1 root root         0 Sep 11 2006  CONFIG.SYS
dr-x------ 1 root root      4096 Dec 18 21:10 Documents and Settings
-r-------- 1 root root     47772 Mar 25 2005  NTDETECT.COM
-r-------- 1 root root    295536 Mar 25 2005  ntldr
-r-------- 1 root root 805306368 Mar 13 16:42 pagefile.sys
dr-x------ 1 root root      4096 Sep 11 2006  Program Files
dr-x------ 1 root root         0 Dec 19 00:35 temp
dr-x------ 1 root root     65536 Mar 13 17:41 WINDOWS
```

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Restore From Snapshot Copies

- **Copy snapshot vmdk back over corrupted vmdk**
  - **Simple and works …**
  - **… but results in I/O and takes time.**
- **Use Instant restore feature (if available)**
  - **SnapRestore (single vmdk or whole datastore) on NetApp**
- **Recover individual files from vmdk**
  - **Connect snapshot VMDK (or clone) to original VM or recovery VM**
  - **Copy missing/corrupt files**

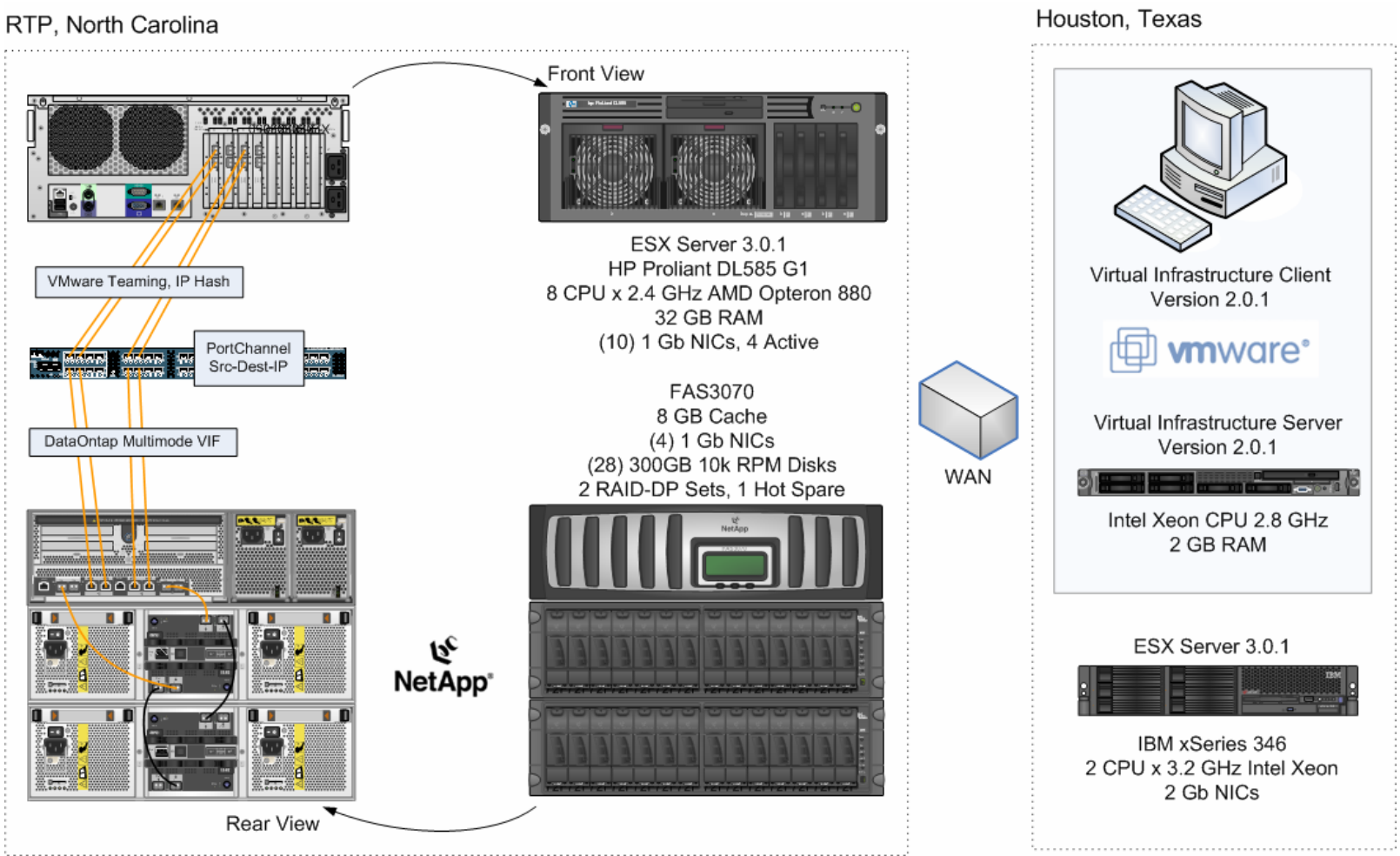
Copyright © 2007 Network Appliance, Inc. All rights reserved.

Beyond Backup → DR

- Use NAS server features for replication, D2D, cloning
- NetApp
  - SnapMirror for DR replication
  - SnapVault for disk to disk backup
  - FlexClone for instant cloning (Smart Copies)
- NAS simplicity over LUNs/VMFS
  - No resignaturing to use datastore snapshot!
    - Just specify different datastore name
      - (You can call it snap-00000001-mynfs if you like!)

Copyright © 2007 Network Appliance, Inc. All rights reserved.

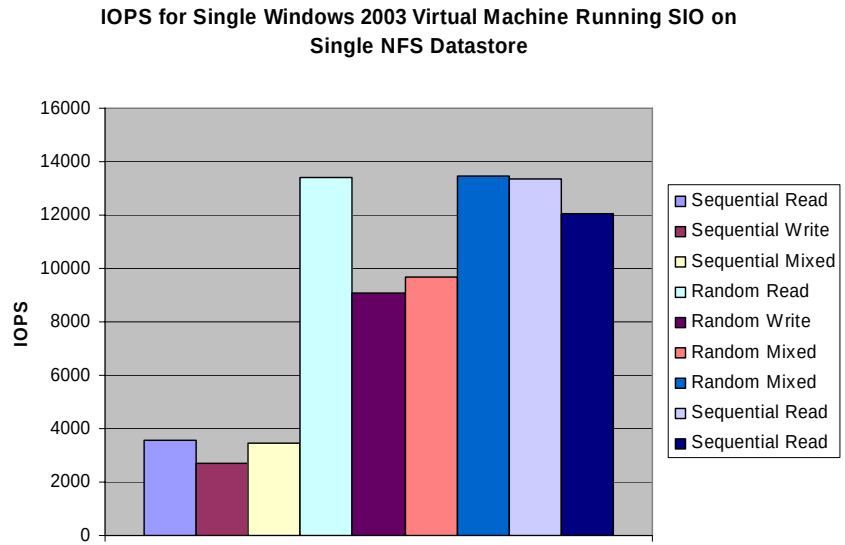
Sample Performance Data: POC Design

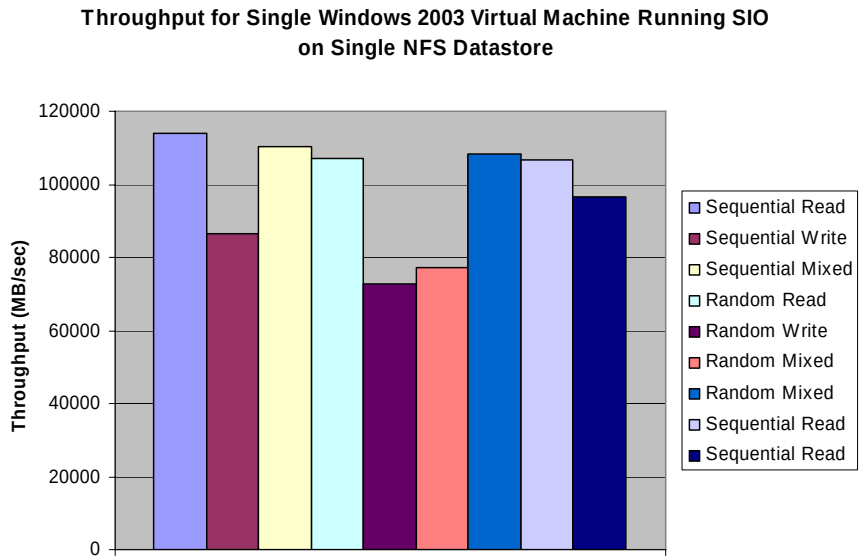


Copyright © 2007 Network Appliance, Inc. All rights reserved.

Sample Performance Data

- **Mixed workload generation for single virtual machine on single NFS datastore**
- **VM Details**
  - **Win 2003 SP1**
  - **4 CPU**
  - **4096 MB mem**
- **Workload Details**
  - **Random: 8K**
  - **Seq: 32K**
  - **16 Streams**

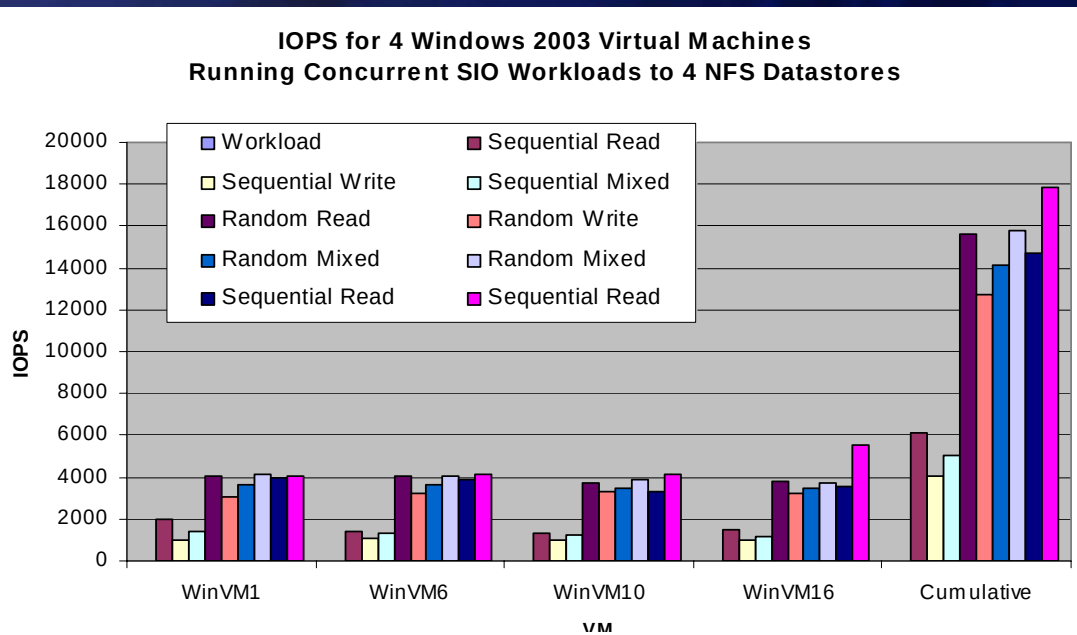


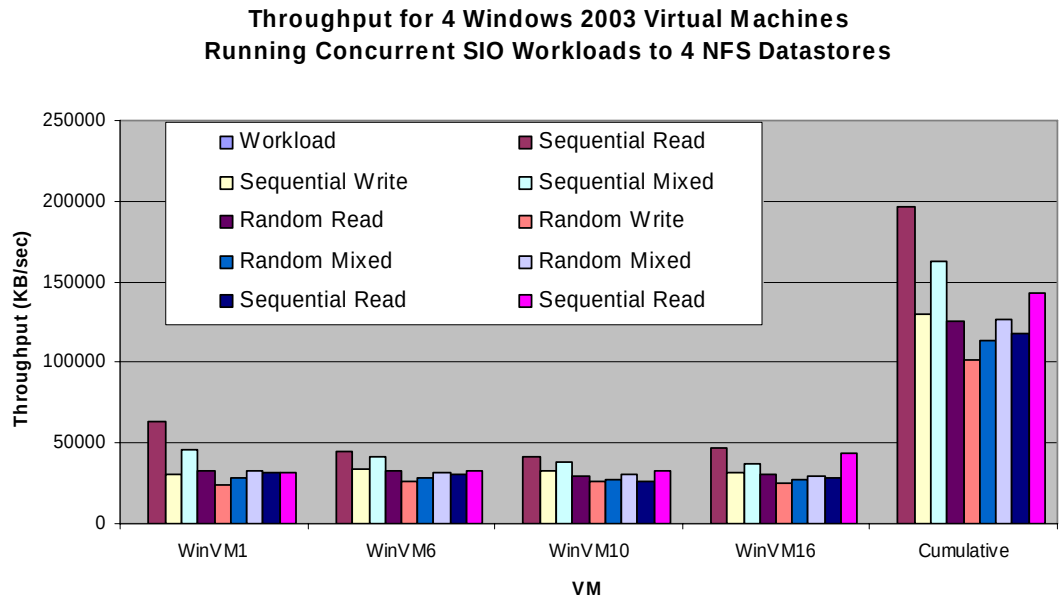


Copyright © 2007 Network Appliance, Inc. All rights reserved.

Sample Performance Data

- **Mixed workload generation for multiple virtual machines residing on multiple NFS datastores.**
- **VM Details**
  - Win 2003 SP1
  - 4 CPU
  - 4096 MB mem
- **Workload Details**
  - Random: 8K
  - Seq: 32K
  - 16 Streams
- **Following two slides depict CPU and network utilization on ESX server during these tests**

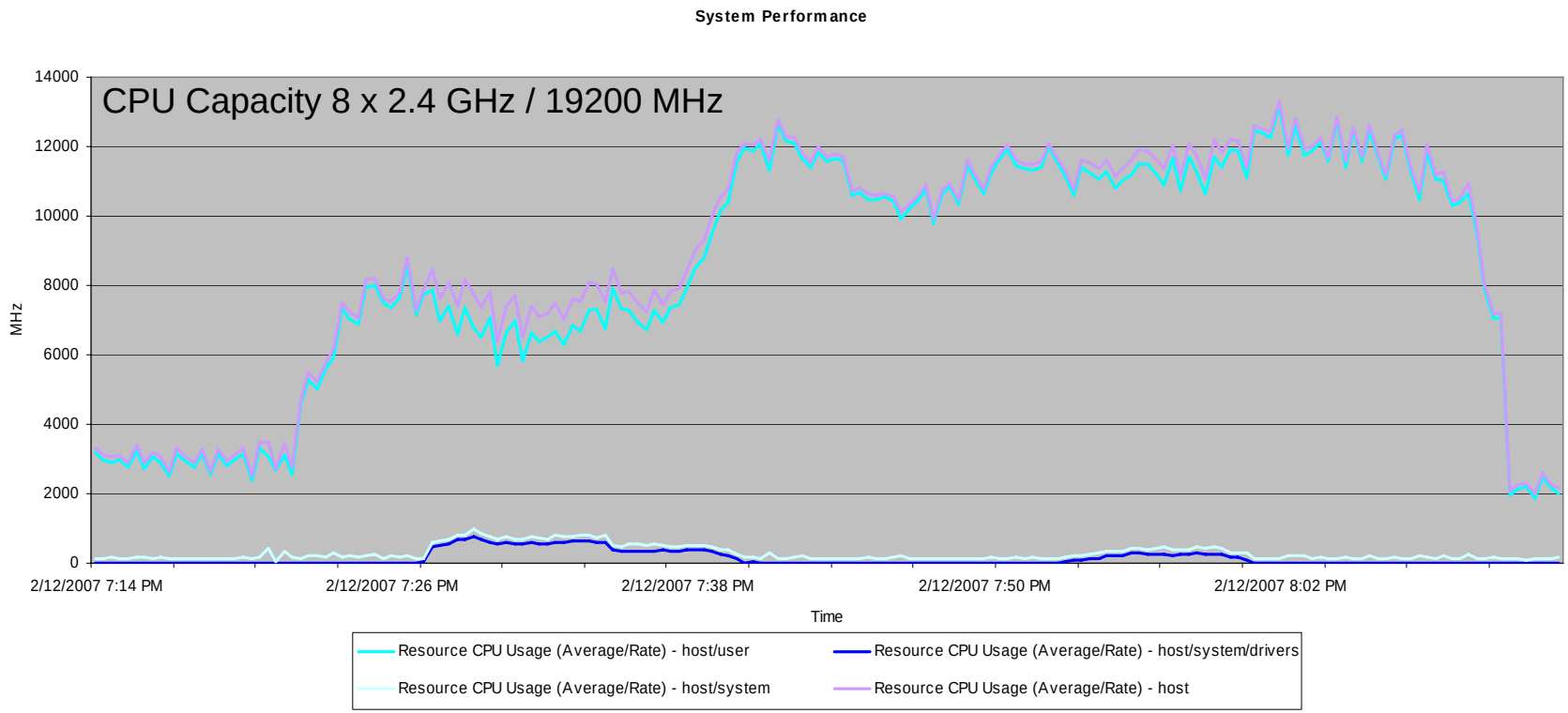




Copyright © 2007 Network Appliance, Inc. All rights reserved.

Comparison Of Protocols For VMware 3.0.1
4 Virtual Machines Running On 4 NFS Datastores*

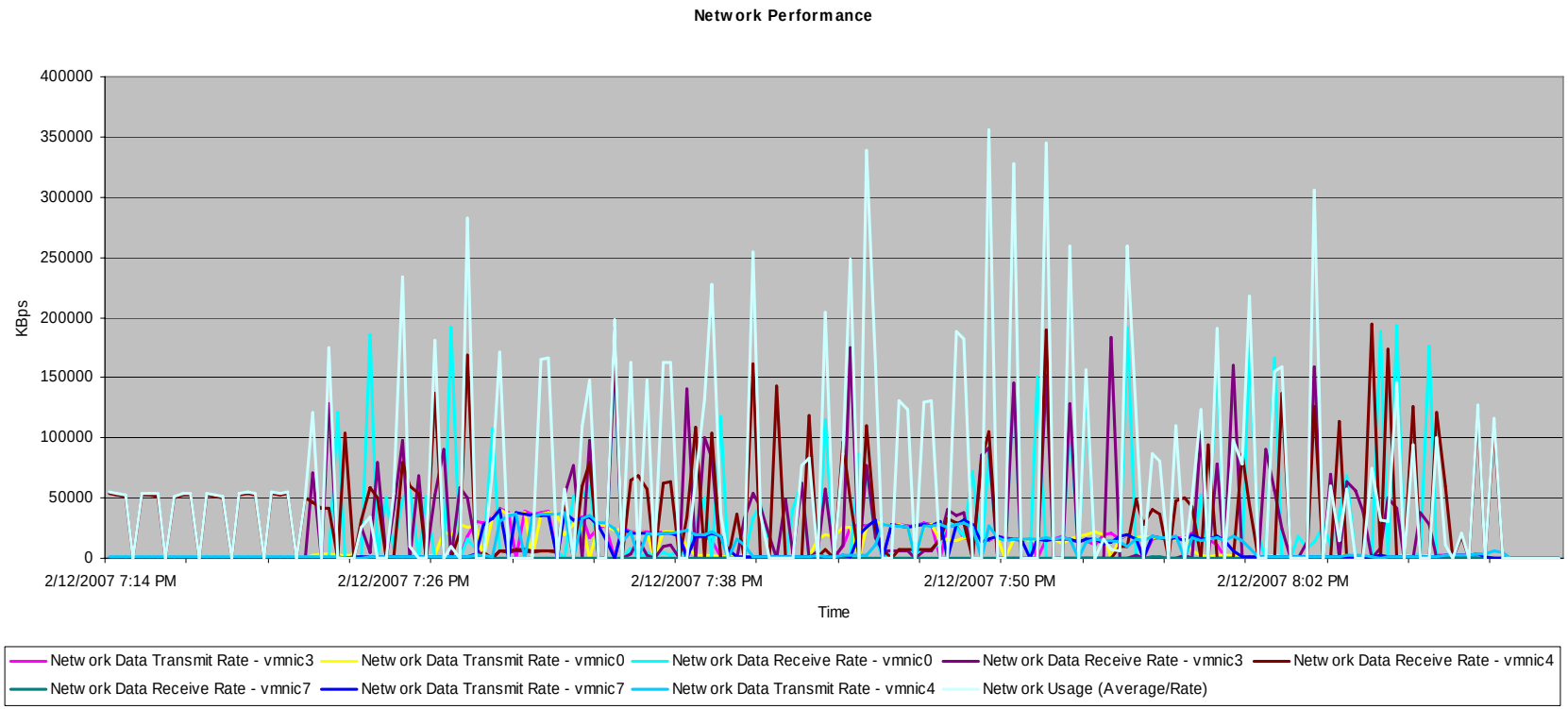
*Single NFS mount point mounted across 4 IP addresses



Copyright © 2007 Network Appliance, Inc. All rights reserved.

Comparison Of Protocols For VMware 3.0.1
4 Virtual Machines Running On 4 NFS Datastores*

*Single NFS mount point mounted across 4 IP addresses



Copyright © 2007 Network Appliance, Inc. All rights reserved.

BEA Implementation

Bud James

BEA Sr. Manager

Data Center Operations Team

Collaboration Team

BEA Environment

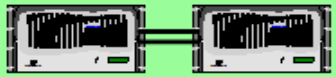
- **Reno Data Center**
  - 13 ESX hosts (growing to 50 in the next 12 months)
  - 48 Virtual Machines (growing to well over 250)
  - 25 Virtual Desktops (VDI) (Also planned for growth)
  - Storage is NetApp FAS940c (Clustered)
  - Storage network is done via dedicated Cisco 6506's (Fiber) with dual ethernet connections from each server to the switches

Copyright © 2007 Network Appliance, Inc. All rights reserved.

NetApp Infrastructure

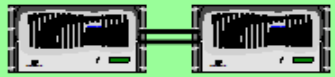
BEA Core Data Center
Implemented NetApp Architecture – May 2007

Production DB



Platform = 960c
Storage = 8 shelves of 144@10K
Usable = Approx. 4TB (nas1) & 4TB (nas2)

Production AP

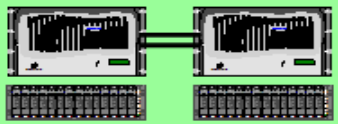


Platform = 940c
Storage = 5 shelves of 144@10K
Usable = Approx. 2TB (nas1) & 3TB (nas2)

ORGSHARE

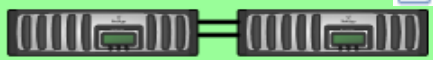
Platform = 270
Storage = 2 shelf of 300@10K
Usable = Approx. 5 TB

Development/Validation Filers



Platform = 940c
Storage = 6 shelves of 300@10K
Usable = Approx. 7TB per head

Bus' Cont' Filers - (MC)



Platform = 3070c
Storage = 12 Shelves of 144@10K
Usable: Approx. 8TB per head

Total NetApp Storage
Currently ~103TB
To grow well over
~250TB

Backup / Nearstore

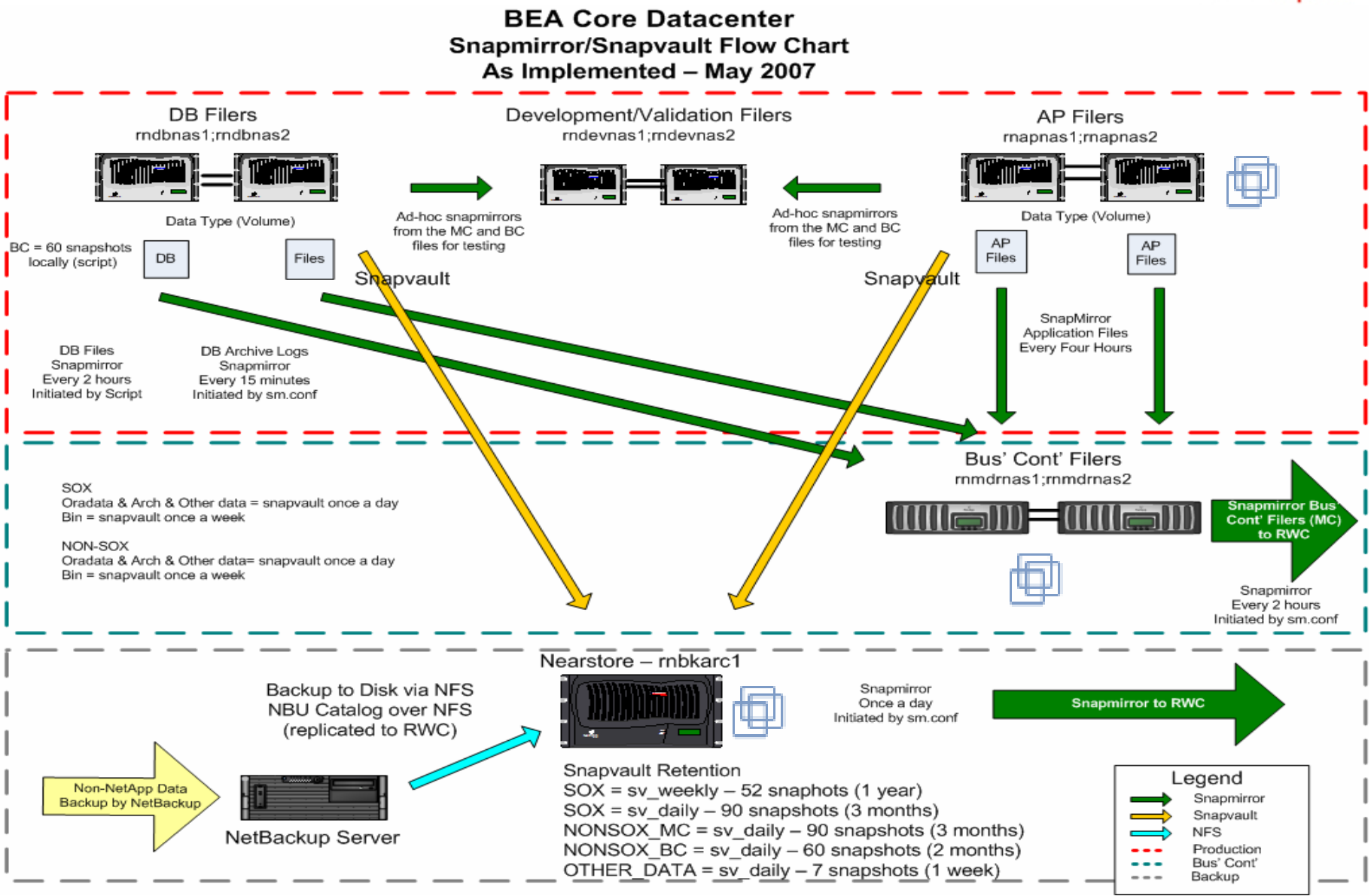
Platform = R200 / 3070
Storage = 320G & 500G Drives
Usable = Approx. 55TB

Storage = Gradually increase storage to 168TB RAW

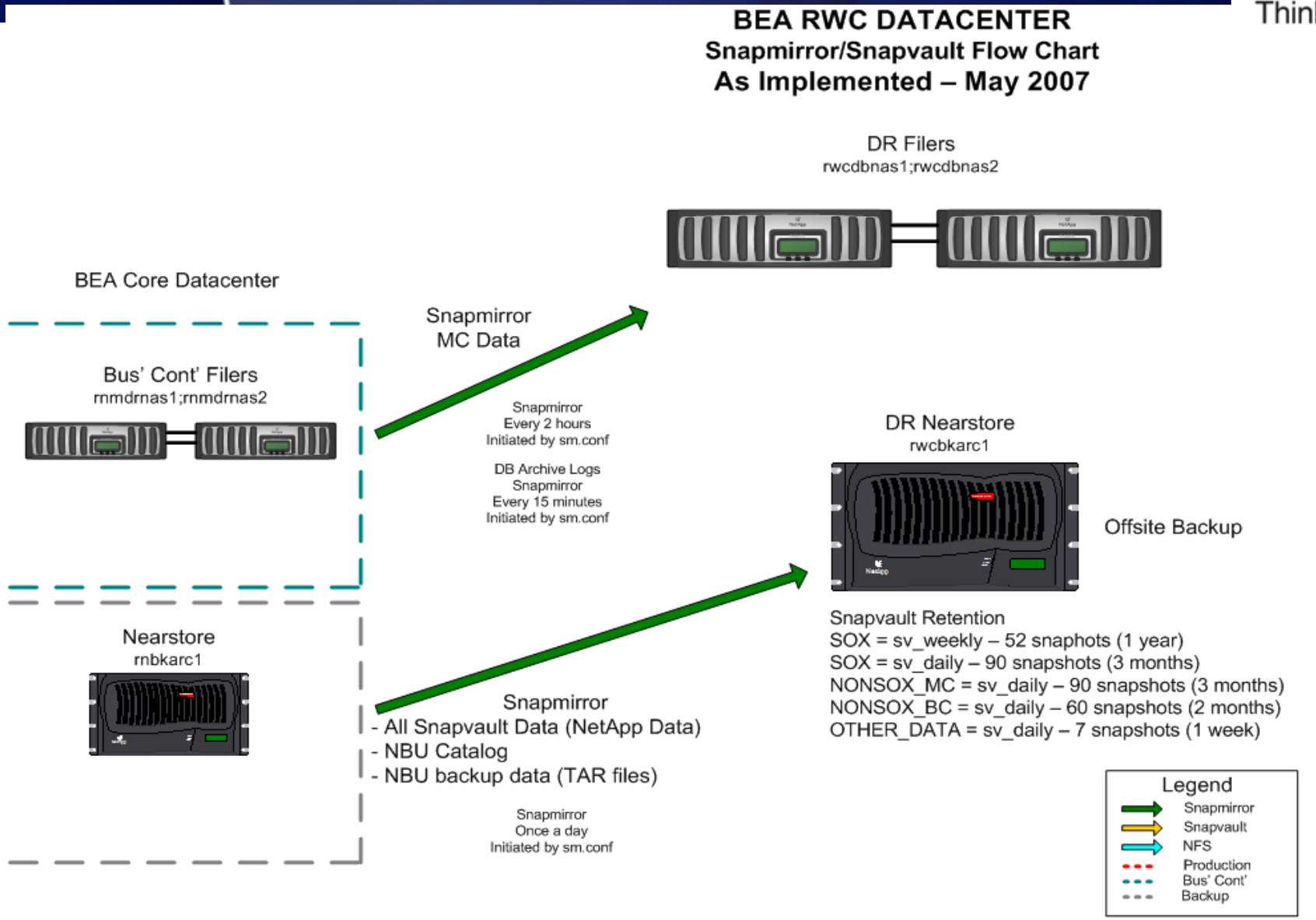
Legend

- New Hardware
- Existing Hardware
- Production
- Bus' Cont'
- Backup
- IT Test / Burn in

How We Back Up

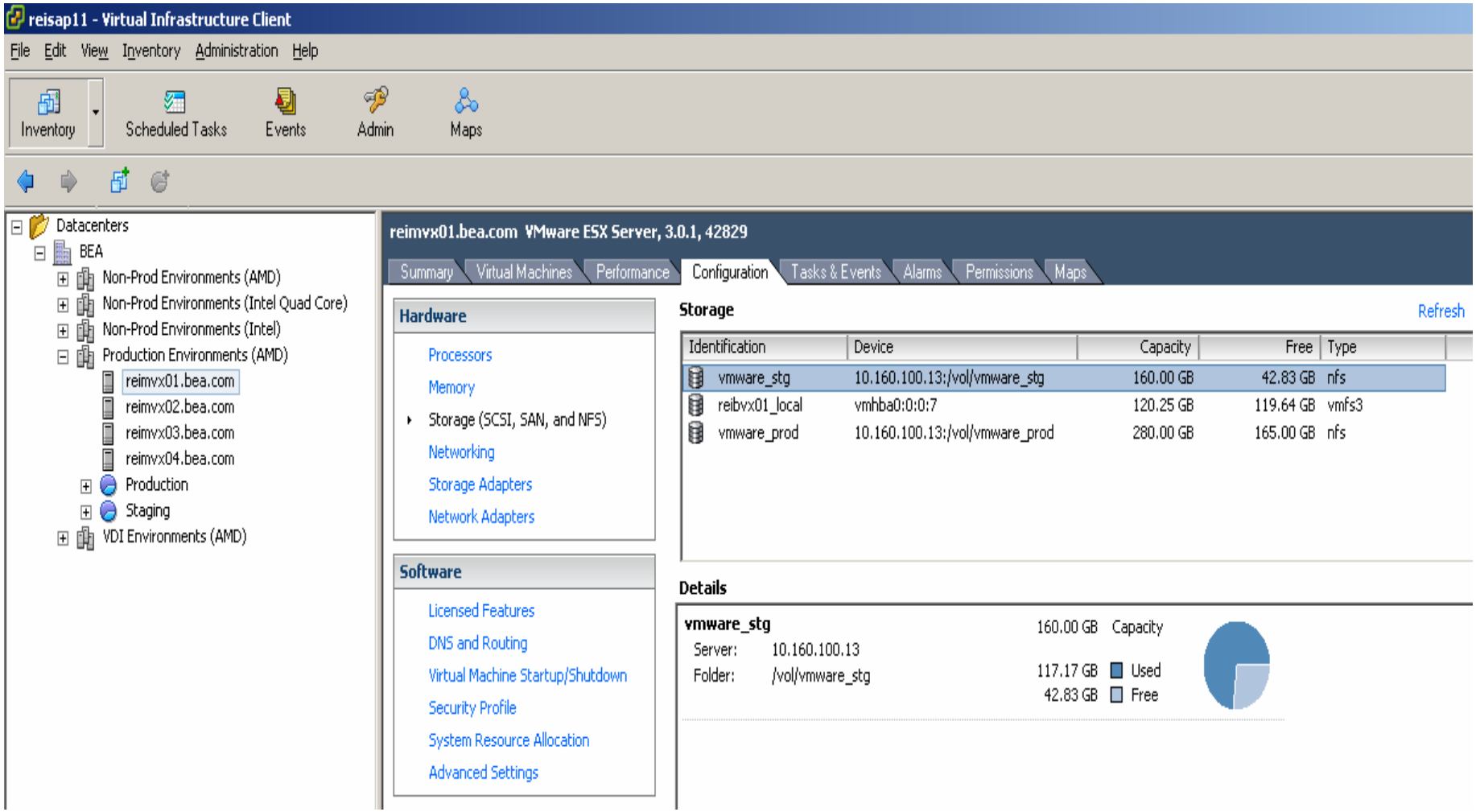


How We Back Up (Cont.)



Copyright © 2007 Network Appliance, Inc. All rights reserved.

NFS Storage Environment



Copyright © 2007 Network Appliance, Inc. All rights reserved.

Lessons Learned / Problems Solved

- **Hardware:**
  - **Make sure your ESX server hosts are certified per the VMware Hardware Compatibility List (HCL)**
  - **Use the same chipset architecture (i.e. all AMD or all Intel) in the cluster for VMotion compatibility**
  - **When using NIC teaming, be sure that physical connections are on separate switches and the standby NIC provides hardware layer redundancy**
  - **Be sure not to over commit resource allocations across VM's**
  - **Size CPU, memory and disk conservatively for the initial rollout and 'tune-down' based upon 'real-world' workloads**
  - **Use SAN/NAS Mass Storage for all VM's**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Lessons Learned / Problems Solved

- **Network:**
  - **Implement VLAN tagging and trunk physical network switches**
  - **Make sure your network switching infrastructure has enough capacity (e.g., uplinks, GigE) given collapsed workloads**

- **Virtual Center:**
  - **Use Virtual Center for virtual machine administration and capacity planning benefits**
  - **Make sure Virtual Center uses a 'mission-critical' RDBMS (SQL, Oracle, etc) during installation since VMware doesn't provide easy porting tools for the embedded MSDE database**
  - **Use Centralized License server for ease of license administration**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Lessons Learned / Problems Solved

- **Training & Support:**
  - **Purchase VMware Platinum Edition for the VMotion feature set (no downtime for scheduled outages, DRS resource scheduling benefits, etc)**
  - **Commit to either direct VMware support or ISV indirect VMware support - - you don't want to have a mixture of both**
  - **Really read the technical whitepapers on the VMware web site**
  - **Use Technical Support for troubleshooting problems**
  - **Send your System Administrators to official VMware training**
  - **Educate business and technical personnel on virtualization benefits to remove organizational obstacles/boundaries**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Why We Love It

- **Quickly take a consistent point in time copy of the ‘quiesced’ guest OS using NetApp Snapshot copies**
  (NOTE: takes just a few seconds for tens of gigabytes of data)
- **Support for both iSCSI & NFS on the same NetApp storage system**
- **Lower Storage Administration & Network support costs using IP-based protocols**
- **Improve Guest OS application recovery point objectives using NetApp SnapMirror for off-site disaster recovery**
- **Mass storage enables the use of VMotion for ‘zero-outage’ scheduled hardware maintenance**
- **Tapeless Back-up Rocks!**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Conclusion

- **NFS is a viable shared storage protocol for VI3 environments**
  - **Flexible**
  - **High-performance**
  - **Reliable**
  - **Scalable**
  - **Simple**
  - **Cost-effective**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

References

- **VMWorld 2006 Storage presentations**
  - http://en.wikipedia.org/wiki/Network_File_System_%28protocol%29
- **O'Reilly *Managing NFS and NIS***
- **Technical Report: Using the Linux® NFS Client with Network Appliance™ Storage**
  - http://www.netapp.com/library/tr/3183.pdf
- **Network Appliance™ and VMware ESX Server 3.0 Storage Best Practices**
  - http://www.netapp.com/library/tr/3428.pdf
- **NetApp stencils are on VisioCafe.Ca**

Copyright © 2007 Network Appliance, Inc. All rights reserved.

Questions?

- **Session ID: IP43**
- **Session Title: Building Virtual Infrastructures with Network Attached Storage**

- **Speaker Name: Peter Learmonth and Kim Weller**
- **Speaker Company: Network Appliance**
- **Special Guest: Bud James, BEA**

- **For more information … visit the NetApp booth or http://www.netapp.com/solutions/infrastructure/server-virtualization/vmware.html**

Copyright © 2007 Network Appliance, Inc. All rights reserved.