CITRIX®

Citrix XenServer and NetApp Storage Best Practices

Citrix Systems, Inc. and NetApp, Inc.

May 2010 | TR-3732 | Rev 2.9

Table of Contents

The Challenge of Today's Enterprise

IT departments face the constant demand to respond more rapidly to changing business priorities, application demands, and user dynamics – all without compromising security or manageability or increasing server count. They must deliver robust data management, business uptime, and complete backup and recovery capabilities. In order to address these challenges, enterprises need to:

- Standardize on a reduced number of hardware configurations – drastically reducing the time required to deploy upgrades and patches

- Provide effective, near-term high-availability for cost-effective protection against physical server and storage failures in an application-agnostic manner

- Adjust allocation of server and storage resources for different application workloads on the fly

- Consider scalability for the largest enterprise applications

- Deploy a tightly unified server and storage virtualization solution that is reliable, not overly complex and leverages all available capabilities

This document presents best practices for using NetApp® storage systems as networked attached storage solutions for Citrix® XenServer™.

Citrix XenServer for Enterprise-Ready Virtualization

Citrix XenServer is a native 64-bit virtualization platform, with the scalability required by Microsoft Exchange Server, Microsoft SQL Server, Citrix XenApp, Citrix XenDesktop and other business-critical applications. The highest host and guest CPU and memory limits available, coupled with fine-grained controls for processor, network, and disk, enable it to deliver optimal quality of service. With Citrix XenServer virtualization, businesses can increase server and storage utilization, reducing costs of equipment, power, cooling, and real estate. By combining servers and storage into resource pools that can be apportioned to the applications with the highest business need, IT operations can be aligned to changing demands and business priorities. With XenMotion™, running virtual machines can be migrated to new servers with no service interruption, allowing essential workloads to get needed resources and enable zero-downtime maintenance. Citrix XenServer products install directly on bare metal servers, requiring no dedicated host operating system. Open command-line (CLI) and programming (API) interfaces make it easy for companies to integrate Citrix XenServer virtualization with existing processes and management tools, rather than requiring rip-and-replace reimplementation. Key benefits and features include:

- Enterprise ready performance and scalability
- Simple deployment and installation
- Flexible shared infrastructure
- On-demand deployment of Windows and Linux virtual machines
- Powerful storage management
- Efficient, secure virtual networking
- Live migration
- XenCenter™ multi-server management, included with product
- Deliver server workloads on demand via streaming

NetApp Storage Solutions for Enterprise-Ready Virtualization

Unified storage solutions from NetApp complement the manageability, utilization and cost-saving benefits of Citrix XenServer. NetApp solutions enable powerful thin provisioning, simplified data management and scalable and consistent I/O performance for all IT applications across NAS, Fibre Channel and iSCSI SAN in a single pool of storage. Key benefits and features include:

- Supports SAN, IP-SAN, or NAS
- Scale non-disruptively to 100's of TB
- Easily installed, configured, managed, and maintained
- Rapid backup and recovery with zero penalty snapshots
- Simple, cost effective replication for Disaster Recovery
- Instant space efficient data clones for provisioning and testing
- Dynamically expand and contract storage volumes as needed
- Data deduplication to reduce capacity requirements

NetApp storage solutions offers these powerful data management and data protection capabilities allowing Citrix XenServer users the ability to lower cost while meeting their capacity, utilization, and performance requirements.

Overview of XenServer storage

Storage repositories (SRs)

The XenServer host accesses containers named Storage Repositories (SRs) in which Virtual Disk Images (VDIs) are stored. A VDI is a disk abstraction which, when attached to a host, appears as a physical disk drive to the virtual machine.

The interface to storage hardware provided on the XenServer host allows VDIs to be supported on a large number of different SR substrate types. VDIs may be files on a local disk, on an NFS share, Logical Volumes within a LUN or a raw LUN itself directly attached to the VM. The SR and VDI abstractions on the host provide for advanced storage features such as sparse provisioning, image snapshots, and fast cloning to be leveraged on storage targets where available.

Each XenServer host can access multiple SRs in parallel of any type. These SRs can be shared between a pool of hosts, or a dedicated repository to a single host. Shared storage must be accessible to all hosts in a resource pool, and is utilized for both non-live and live migration of VMs via XenMotion. When hosting shared Storage Repositories on a NetApp device, there are 4 options; the Citrix StorageLink feature, an NFS file share, an iSCSI LUN or a Fibre Channel LUN.

Virtual Disk Images (VDIs)

There are two fundamental VDI types (Files and LUNs) that can be accessed with a NetApp device as the backend over 4 different SR driver types:

- NetApp managed LUNs: Managed NetApp LUNs are accessible via the StorageLink feature included in Citrix Essentials for XenServer, and are hosted on a NetApp device running a version of Data ONTAP 7.0 or greater. LUNs are allocated on demand via StorageLink and mapped dynamically to the host via the StorageLink service while a VM is active. All the thin provisioning and fast clone capabilities of the device are exposed via StorageLink.

- VHD files. The VHD format can be used to store VDIs in a sparse format. Being sparse, the image file grows proportionally to the number of writes to the disk by the Virtual Machine (VM), so large portions of the disk which are typically unused do not consume unnecessary space. VHD on NFS, iSCSI, or Hardware HBA storage repositories can be shared among all hosts in a pool.

The section entitled 'XenServer Shared Storage Options' discusses each option in more detail.

Managing Storage

There are four XenServer object classes that are used to describe, configure, and manage storage:

- Storage Repositories (SRs) are storage targets containing homogeneous virtual disks (VDIs). SR commands provide operations for creating, destroying, resizing, cloning, connecting and discovering the individual Virtual Disk Images (VDIs) that they contain. A storage repository is a persistent, on-disk data structure. So the act of "creating" a new SR is similar to that of formatting a disk. SRs are long-lived, and may be shared among XenServer hosts or moved between them.

- Physical Block Devices (PBDs) represent the interface between a physical server and an attached SR. PBDs are connector objects that allow a given SR to be mapped to a XenServer host. PBDs store the device configuration fields that are used to connect to and interact with a given storage target. There are several classes of XenServer SRs available for use with NetApp storage:

o NFS. With an NFS SR, a new directory on an existing NetApp NFS share is created for the storage of VDIs in VHD format.

o iSCSI and Fibre Channel. The creation of iSCSI or Fibre Channel (Hardware HBA) SRs involves erasing any existing data on a specified LUN. A LUN will need to be created on the NetApp storage before creating the XenServer SR. Volume management is performed via LVM (Logical Volume Manager), and the underlying VDI storage on an iSCSI or FC SR is VHD.

o Direct StorageLink-NetApp adapter SR. First introduced in XenServer 4.1, this type of SR is created within XenCenter. With this type of SR, FlexVol resources are allocated on the NetApp device and communication is performed via the iSCSI protocol.

o StorageLink Gateway. New with XenServer 5.5, this type of SR is initially created from the StorageLink Manager and visible within XenCenter thereafter. The StorageLink Gateway service runs on a general-purpose Windows server or VM. Once brokered by the StorageLink Gateway, the data path between the virtual machine and storage system can be performed via either iSCSI or Fibre channel. It is important to note that the StorageLink Gateway only serves as a broker and a control path; it is not part of the data path between a XenServer host and the NetApp storage infrastructure.

- Virtual Disk Images (VDIs) are an on-disk representation of a virtual disk provided to a VM. VDIs are the fundamental unit of virtualized storage in XenServer. Similar to SRs, VDIs are persistent, on-disk objects that exist independently of XenServer Hosts.

- Virtual Block Devices (VBDs) are a connector object (similar to the PBD described above) that allows mappings between VDIs and Virtual Machines (VMs). In addition to providing a mechanism to attach (or plug) a VDI into a VM, VBDs allow the fine-tuning of parameters regarding QoS (quality of service), statistics, and the bootability of a given VDI.
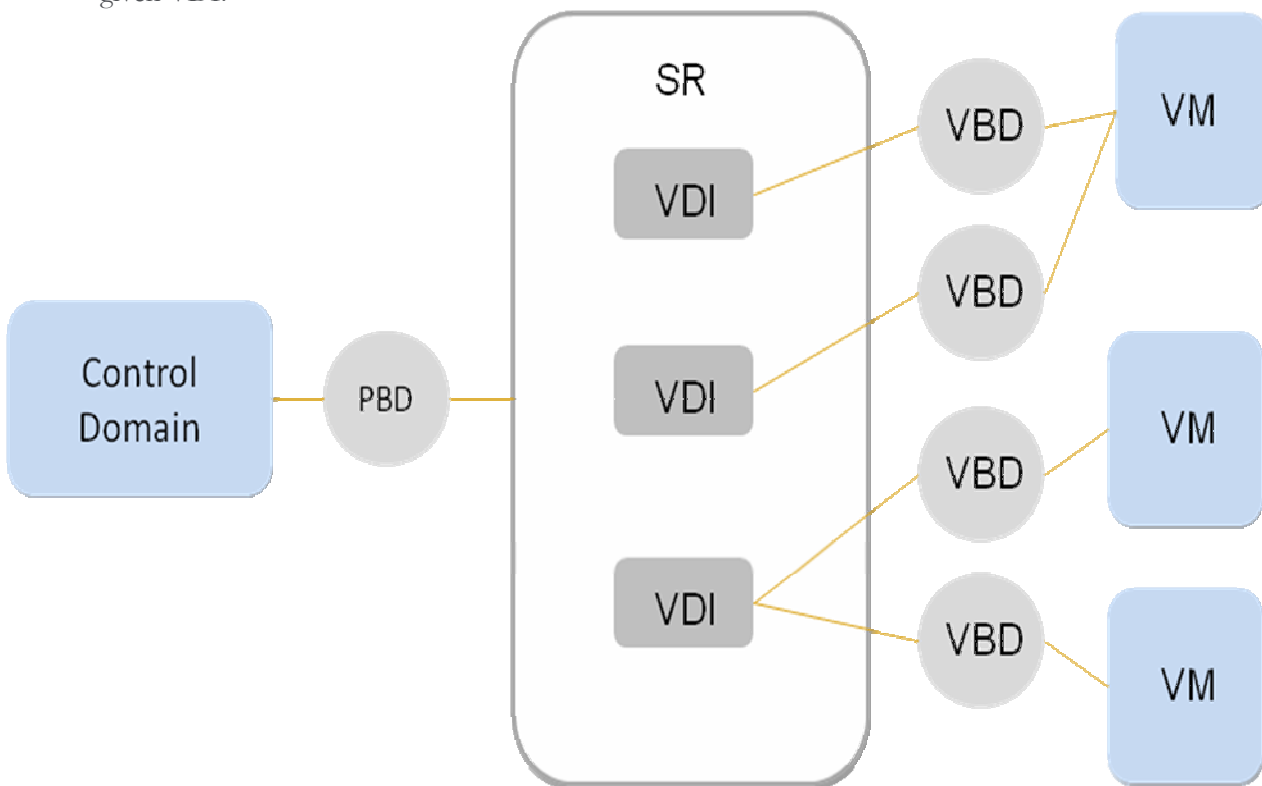

Figure 1 Graphical Overview of Storage Repository and Related Objects

XenServer Shared Storage Options

When using a NetApp device as your networked, backend storage array, it is recommended with XenServer v5.5 onwards to use the StorageLink Gateway SR type. The architecture of the StorageLink Gateway allows XenServer to utilize the capabilities of the NetApp device to provide data efficiency, high performance and ensure compatibility with existing ONTAP device management tools. This allows for:

- Fast provisioning of VDIs
- Fast cloning of VDIs
- Fast Snapshot® of VDIs

To use non-customized storage options with a NetApp device, the following types can also be used:

- Network Attached Storage using NFS
- iSCSI
- Fibre Channel

The following sections give an overview of the above storage types and the benefits associated with them. All shared storage options enable VM agility using XenMotion -- VMs can be started on any XenServer host in a resource pool and migrated between them.

Shared Storage with the StorageLink Gateway

The StorageLink Gateway is a feature offered with Citrix Essentials for XenServer 5.5, Enterprise Edition. It acts as a broker between XenServer and the NetApp storage system. The StorageLink Gateway service runs on a general-purpose Windows server or VM, and this system can typically be shared with other Citrix Essentials for XenServer components such as Workload Balancing and Workflow Studio. Once brokered by the StorageLink Gateway, the data path between the virtual machine and storage system can be performed via either iSCSI or Fibre channel. It is important to note that the StorageLink Gateway only serves as a broker and as part of the control path; it is not part of the data path between a XenServer host and the NetApp storage infrastructure.

Since the adapter exposes LUNs to the XenServer host as VDIs, there is a one-to-one mapping between the Virtual Machine disks in the XenServer environment and the NetApp disk data abstraction. This enables much simpler integration with existing physical host backup and array provisioning tools that already operate at a LUN granularity.
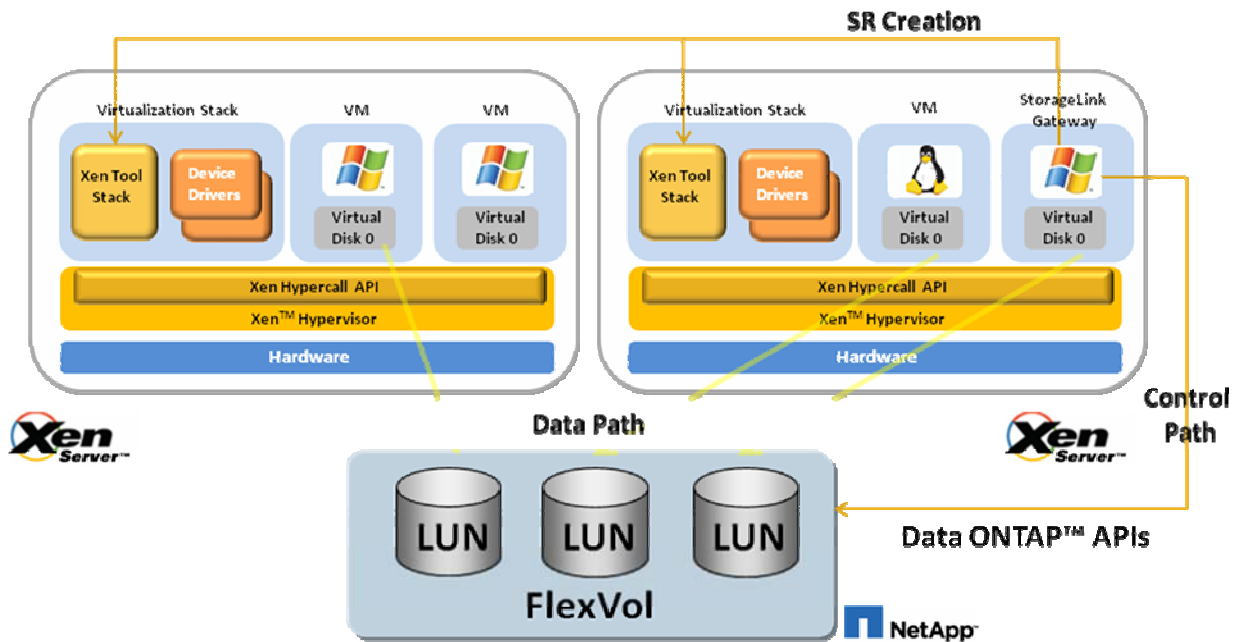
Shared Storage using the direct StorageLink-NetApp adapter

The direct StorageLink adapter for NetApp was first introduced in XenServer 4.1. While it is still available in XenServer 5.5, it is primarily intended to maintain backward compatibility with pre-5.5 deployments. Very small environments and Linux-centric deployments may also find this StorageLink option to be useful. For new XenServer 5.5 deployments, it is recommended to use the StorageLink Gateway.

The following table provides a comparison of the two StorageLink options for NetApp.

	StorageLink Gateway	Direct StorageLink-NetApp Adapter
Deployment	On separate Windows VM or server	Included within XenServer control domain; no additional VMs or servers are required
VM / LUN relationship	LUN per VDI	LUN per VDI
Protocol support	iSCSI and Fibre Channel	iSCSI
Interoperability with Hyper-V	Yes	No
Interoperability with other vendor storage arrays	Yes	Limited

This legacy StorageLink-NetApp adapter for NetApp Data ONTAP uses the Zephyr API (ZAPI) interface to the device to create a group of FlexVol®s which corresponds to a XenServer SR. VDIs are created as virtual LUNs on the device, and attached to XenServer hosts using an iSCSI data path. There is a direct mapping between a VDI and a raw LUN without requiring any additional volume metadata. Thus, at a logical level, the NetApp SR is a managed volume and the VDIs are the LUNs within the volume. Similar to the StorageLink Gateway, the legacy StorageLink adapter exposes LUNs to the XenServer host as VDIs, and there is a one-to-one mapping between the Virtual Machine disks in the XenServer environment and the NetApp disk data abstraction.

For the I/O data path, the NetApp Data ONTAP adapter directly controls the host built-in software initiator and its assigned server iSCSI Qualified Name or IQN to map the data path on demand without requiring user intervention.

Storage can be thinly or fully provisioned depending on the administrative preference for the NetApp array. When thin provisioning is utilized, data de-duping can also be switched on to reclaim common blocks between VM images, thereby conserving even more space. All these configuration options are exposed via the XenServer storage configuration interface and managed by the NetApp Data ONTAP adapter directly.
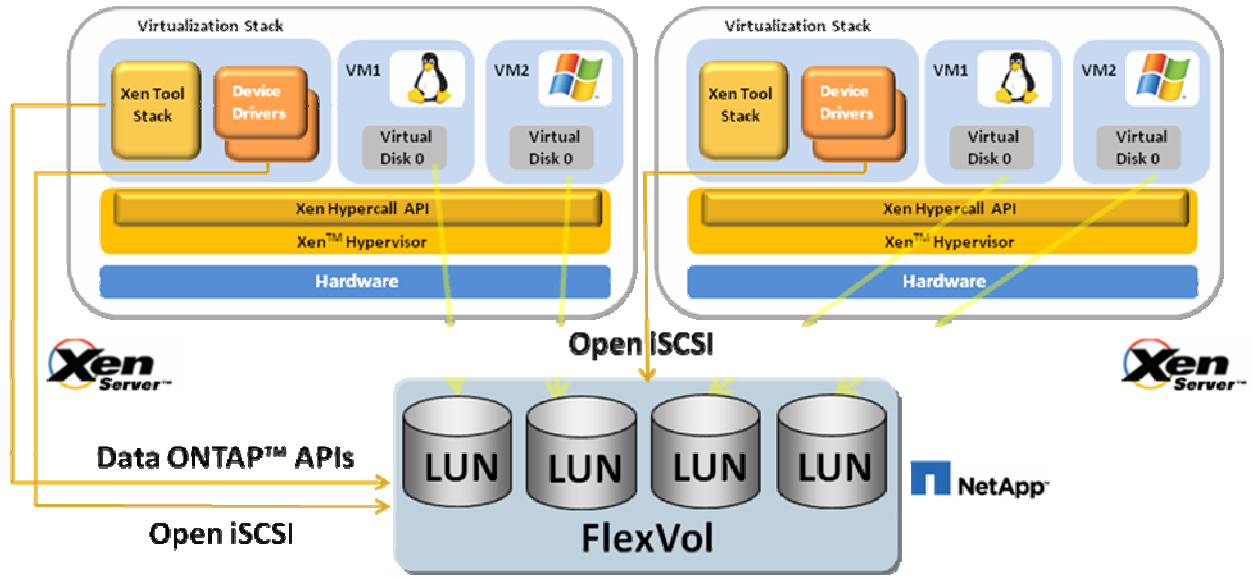
Figure 2 Shared storage using NetApp Data ONTAP

Shared NAS using NFS

XenServer supports shared access to Virtual Disk storage exported as NFS v3 over TCP/IP based on the Microsoft VHD format.

VDIs stored on NFS are *sparse*. The image file grows as the Virtual Machine (VM) writes data into the disk, so unused portions of the virtual disk do not consume space on the array. This is a considerable benefit since VM image files take up only as much space on the NFS device as is required. If a 100-GB VDI is allocated for a new VM and an OS is installed, the VDI file will only reflect the size of the OS data that has been written to the disk.

VHD files may also be *chained*, allowing two VDIs to share common data. In cases where a NFS-based VM is cloned, the resulting VMs will share the common on-disk data at the time of cloning. Each will proceed to make its own changes in an isolated *copy-on-write* version of the VDI. This feature allows NFS-based VMs to be quickly cloned from templates, facilitating very fast provisioning and deployment of new VMs.
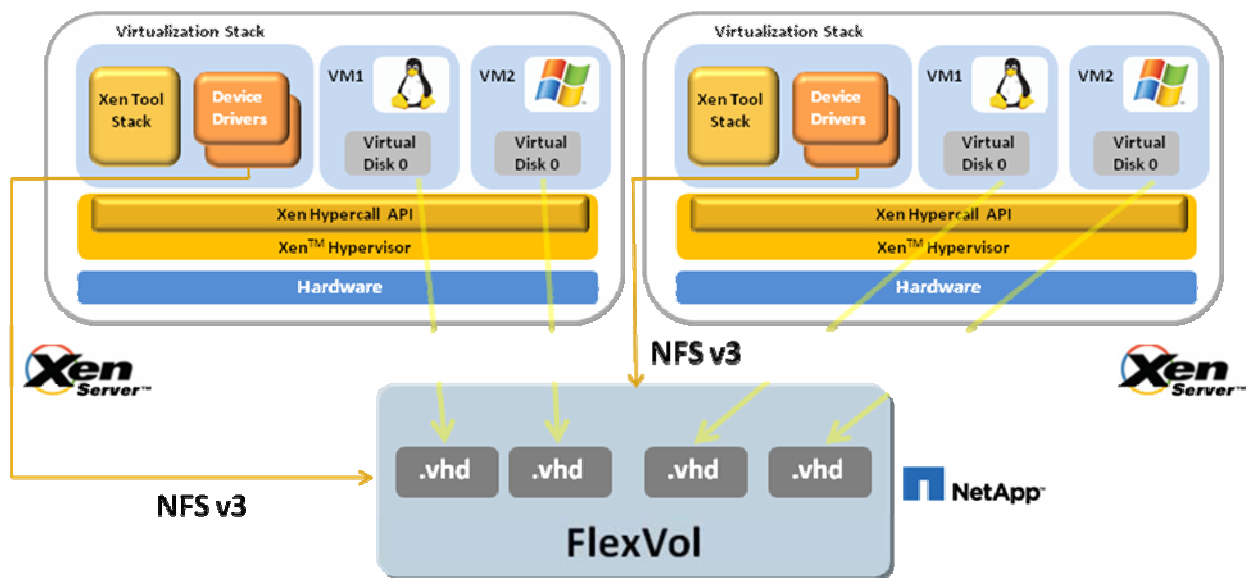
Figure 3 Shared NFS storage

Shared iSCSI Storage

XenServer provides support for shared SRs on iSCSI LUNs. iSCSI is supported using the open-iSCSI software iSCSI initiator or using a supported iSCSI *Host Bus Adapter* (HBA).

Shared iSCSI support is implemented based on XenServer LVHD, a technology that combines the Logical Volume Manager (LVM) and Virtual Hard Disk (VHD) standards. Virtual machine VDIs are stored on an iSCSI LUN created on the NetApp storage system. Shared storage with iSCSI is a good choice for general purpose virtualization deployments, though it may not be as suitable for demanding workloads or deployments.
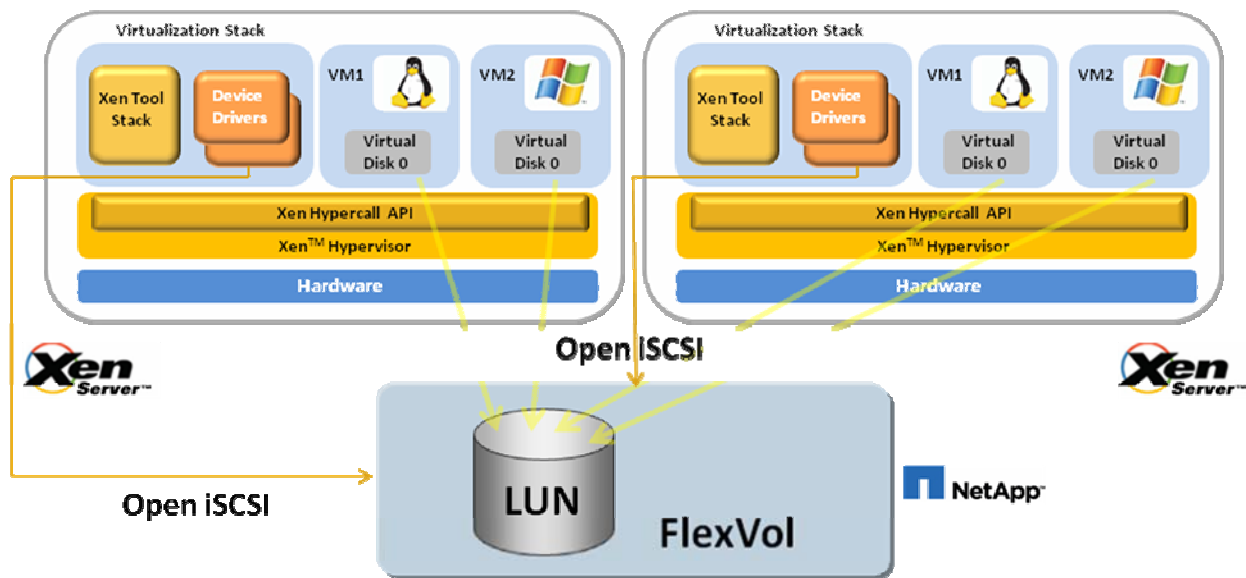
Figure 4 Shared iSCSI storage using Open iSCSI initiator
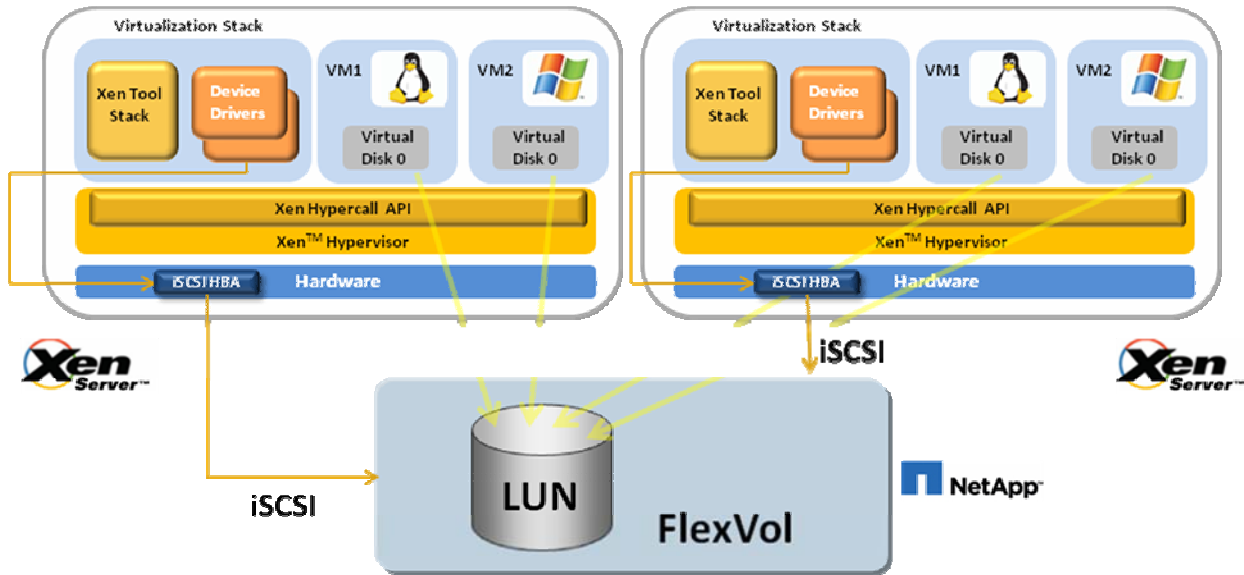
Figure 5 Shared iSCSI storage using iSCSI HBA

Shared Fibre Channel Storage

XenServer hosts support Fibre Channel SANs using an Emulex or QLogic *host bus adapter* (HBA). Logical unit numbers (LUNs) are mapped to the XenServer host as disk devices.

Like iSCSI storage, Fibre Channel storage support is implemented based on the same LVHD technology with the same benefits as iSCSI storage, just utilizing a different data I/O path.
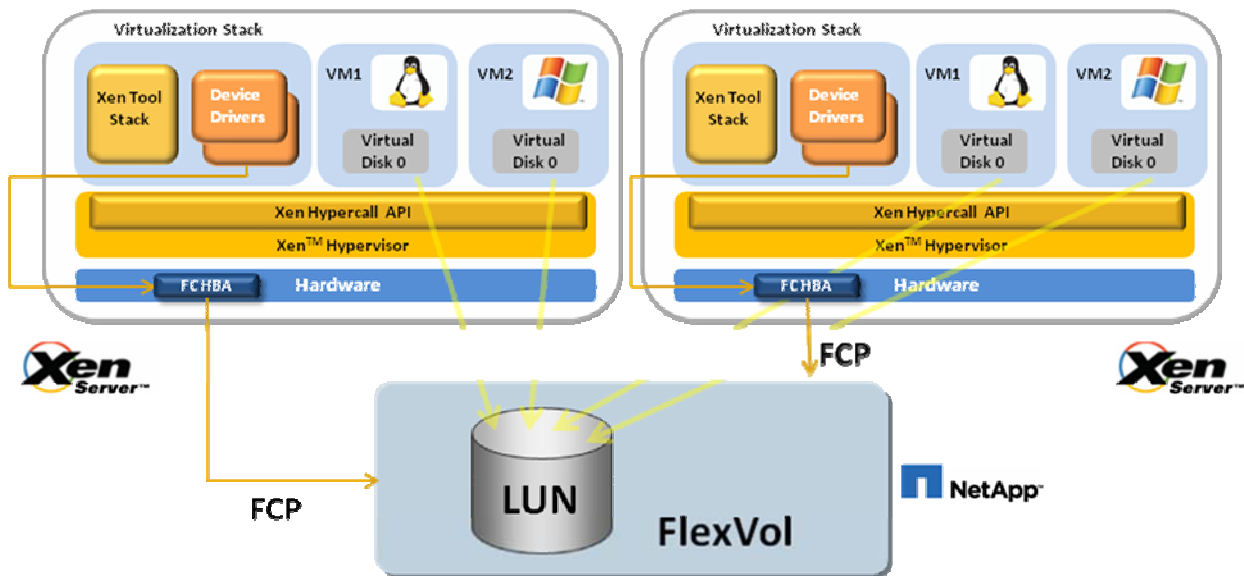
Figure 6 Shared FC storage

Configuration and Setup

NetApp Device Configuration Basics

This section covers the best practices for configuring the NetApp storage box.

Configuring Target Portal Groups for iSCSI connections

Considerations when using Target Portal Groups: (for more information, please consult the Data ONTAP admin guide)

1. Adding or removing a NIC may change the target portal group assignments. Be sure to verify the target portal group settings are correct after adding or removing hardware, especially in Active/Active configurations.

2. When used with multi-connection sessions, the Windows iSCSI software initiator creates a persistent association between the target portal group tag value and the target interfaces. If the tag value changes while an iSCSI session is active, the initiator will be able to recover only one connection for a session. To recover the remaining connections, you must refresh the initiator's target information.

If you want to employ multi-connection sessions to improve performance and reliability, you must use target portal groups to define the interfaces available for each iSCSI session.

Create a target portal group that contains all of the interfaces you want to use for one iSCSI session. However, note that you cannot combine iSCSI hardware-accelerated interfaces with standard iSCSI storage system interfaces in the same target portal group. When you create a target portal group, the specified interfaces are removed from their current groups and added to the new group. Any iSCSI sessions using the specified interfaces are terminated, but the initiator should automatically reconnect. However, initiators that create a persistent association between the IP address and the target portal group will not be able to reconnect.

1. Open console connection to the NetApp device, and run the *iscsi tpgroup* command

   ```
   iscsi tpgroup create [-f] tpgroup_name [-t tag] [interface ...]
   ```

   `-f` forces the new group to be created, even if that terminates an existing session using one of the interfaces being added to the group.

   `tpgroup_name` is the name of the group being created (1 to 32 characters, no spaces or non-printing characters).

   `-t tag` sets the target portal group tag to the specified value. In general you should accept the default tag value. User-specified tags must be in the range 1 to 256.

   `interface ...` is the list of interfaces to include in the group, separated by spaces.

   For example, the following command creates a target portal group named xenstorage_iscsi_group that includes interfaces e0c and e0d.
   ```
   iscsi tpgroup create xenstorage_iscsi_group e0c e0d
   ```

2. Enable target portal group interfaces to accept iSCSI connections
   ```
   iscsi interface enable e0c
   iscsi interface enable e0d
   ```

Configuring a Dynamic Multimode Virtual Interface (VIF)

Throughput can be improved by creating a multimode VIF. With a multimode VIF, all interfaces in the VIF are active and share a single IP address. This logical aggregation of interfaces can be effectively used for NFS connections.

The following prerequisite must be met before creating a multimode VIF:

1. Identify or install a switch that supports link aggregation over multiple port connections in your network, configured according to your switch vendor's instructions.

2. Decide on a case-sensitive name for the VIF that meets the following criteria:

    a. It must begin with a letter.

    b. It must not contain a space.

    c. It must not contain more than 15 characters.

    d. It must not already be in use for a VIF.

3. Decide on the interfaces you want the VIF to consist of.

4. Configure all interfaces that will be included in the VIF to be down using the *ifconfig* command.

To create a dynamic multimode (LACP) VIF for NFS datapaths:

1. Open a console session to the NetApp device, and run the *ifconfig down* command for interfaces that you want to create the vif on.
   ```
   ifconfig down e0c
   ifconfig down e0d
   ```

2. Run the *vif* command to create the new vif
   ```
   vif create [multi|lacp] <vif_name> -b ip [<interface_list>]
   ```

   It is recommended to use the lacp option and use IP address-based load balancing algorithm.
   ```
   vif create lacp nfs_bond -b ip e0c e0d
   ```

3. Run the *ifconfig* command on the newly created vif to assign it an IP address
   ```
   ifconfig nfs_bond 192.168.2.203 netmask 255.255.255.0
   ```

Configuring an Aggregate

An aggregate consists of disk drives; you must specify a number of disks to be added to the new aggregate. Aggregates are the means by which the total IOPs available to all of the physical disks are pooled as a resource. NetApp recommends that whenever possible a small aggregate should be used as the root aggregate. The root aggregate stores the files required for running and providing GUI management tools for the FAS system. The remaining storage should be placed into a small number of large aggregates. On smaller FAS arrays, it may not be practical to have more than a single aggregate, due to the restricted number of disk drives on the system. In these cases, it is acceptable to have only a single aggregate.

1. Open the NetApp FilerView®, and click *Aggregates->Add* to add a new aggregate on the device.
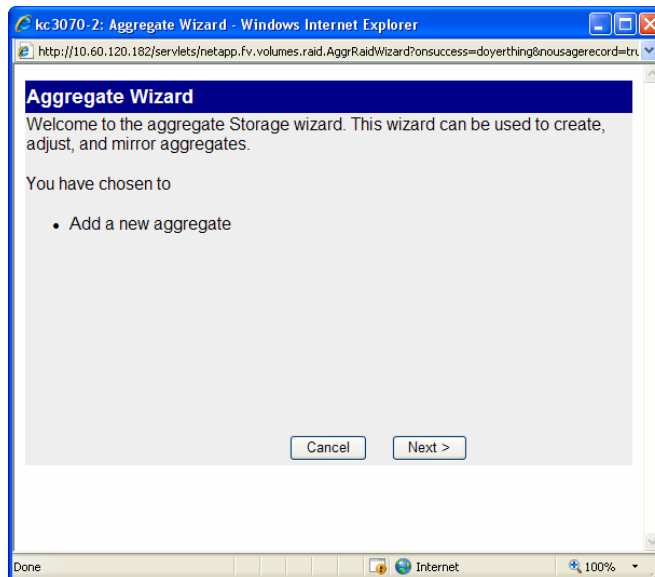
2. Choose an aggregate name that reflects the data center that will be using it, along with "XenServer" in the name. This will also make storage configuration easier when accessed via the XenCenter management interface. Choose the *Double Parity* option if there is an extra disk per RAID group available. This is the recommended RAID-level for the aggregate. NetApp RAID-DP® is an advanced RAID technology that is provided as the default RAID level on all FAS systems. RAID-DP protects against the simultaneous loss of two drives in a single RAID group. It is very economical to deploy; the overhead with default RAID groups is a mere 12.5%. This level of resiliency and storage efficiency makes data residing on RAID-DP safer than data stored on RAID 5 and more cost effective than RAID 10. NetApp recommends using RAID-DP on all RAID groups that store XenServer data.
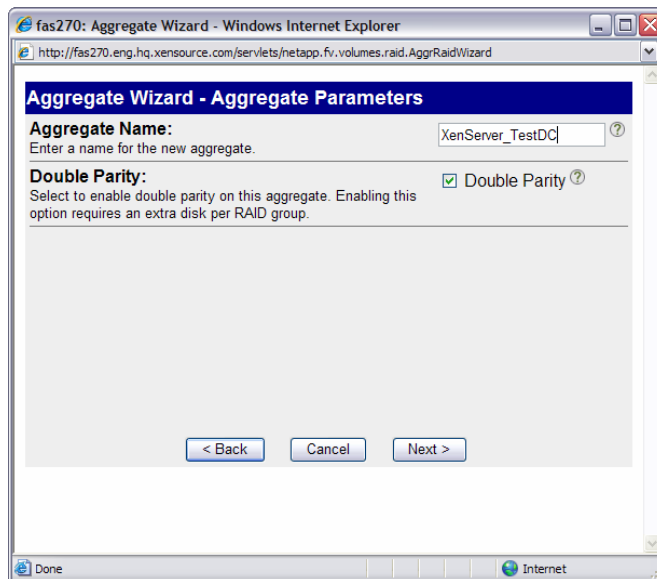
3. Select the number of disks to assign for the *Raid Group Size*. By default the device will default to 16.
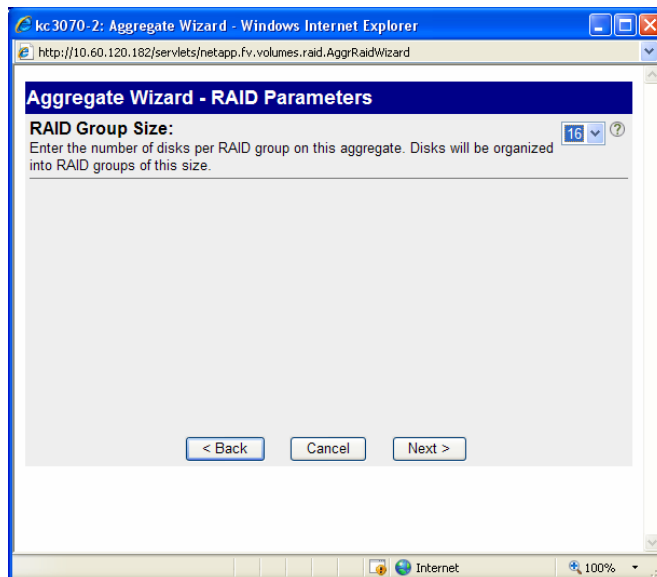
4. Select the disk that will be used for the aggregate. *Automatic* is selected by default for this section.
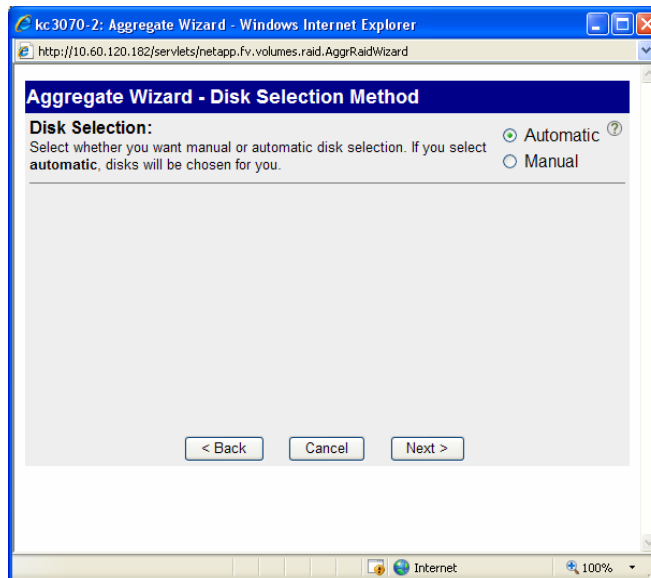
5. Choose the disk size to be used in the aggregate. By default *Any Size* will be selected.
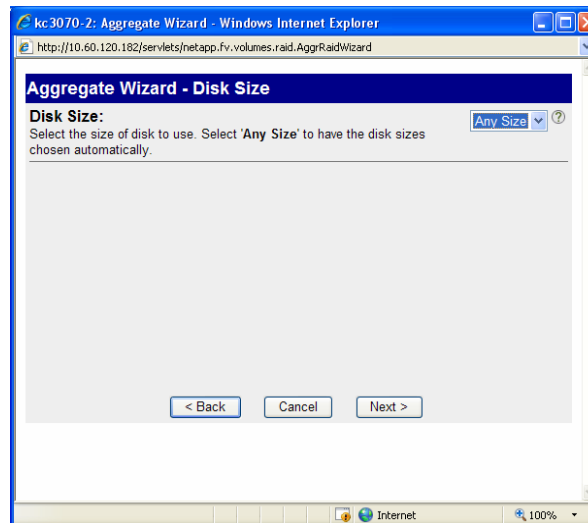
6. Choose the number of disks to be added to the aggregate.
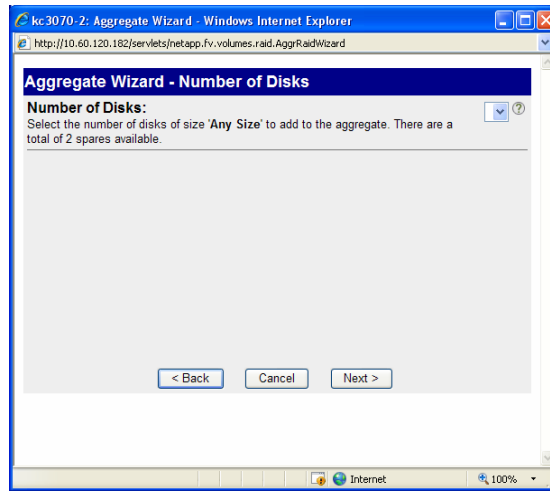
7. Assign at least three disks in order to provision an aggregate. Click *Next* and then *Commit* to create the aggregate.

Thin Provisioning vs Thick Provisioning

Thick provisioning preallocates storage; thin provisioning provides storage on demand. Thin provisioning allows the admin to overprovision storage in order to avoid running out of storage and the associated application downtime when expanding the provisioned storage. Although no system can be run at 100% storage utilization, there are methods of storage virtualization that allow administrators to address and oversubscribe storage in the same manner as with server resources (such as CPU, memory, networking, and so on). This form of storage virtualization is referred to as *thin provisioning*.

The value of thin-provisioned storage is that storage is treated as a shared resource pool and is consumed only as each individual VM requires it. This sharing increases the total utilization rate of storage by eliminating the unused but provisioned areas of storage that are associated with traditional storage. The drawback to thin provisioning and oversubscribing storage is that (without the addition of physical storage) if every VM requires its maximum possible storage at the same time, there will not be enough storage to satisfy the requests present more storage space to the VMs connecting to the SR than is actually available on the SR. It is important to note that there are no space guarantees, and allocation of a LUN does not claim any data blocks in the FlexVol until the VM writes data into it.

The following scenarios benefit from thin provisioning

       a. Quick provisioning
       b. Not realistic to decide how much storage is required up front
       c. Allow storage capacity additions without downtime

NetApp recommends that when you enable NetApp thin provisioning, you also configure storage management policies on the volumes that contain the thin-provisioned LUNs. These policies aid in providing the thin-provisioned LUNs with storage capacity, as they require it. The policies include automatic sizing of a volume, automatic Snapshot deletion, and LUN fractional reserve.

Volume Auto Size is a policy-based space management feature in Data ONTAP that allows a volume to grow in defined increments up to a predefined limit when the volume is nearly full. For Citrix environments, NetApp recommends setting this value to 'on'. Doing so requires setting the maximum volume and increment size options.

0. Log in to NetApp console.
0. Set Volume Auto Size Policy

```
Vol autosize <vol-name> [-m <size>[k|m|g|t]][-i <size>[k|m|g|t]] on.
```

Snapshot Auto Delete is a policy-based space-management feature that automatically deletes the oldest

Snapshot copies on a volume when that volume is nearly full. NetApp recommends setting this value to delete Snapshot copies at 5% of available space. In addition, you should set the volume option to have the system attempt to grow the volume before deleting Snapshot copies.

2. Log in to NetApp console.
2. Set Snapshot Auto Delete Policy
2. snap autodelete <vol-name> commitment try trigger

```
snap autodelete <vol-name> commitment try triggervolume target_free_space 5
delete_order oldest_first.
```

LUN Fractional Reserve is a policy that is required when you use NetApp Snapshot copies on volumes that contain XenServer LUNs. This policy defines the amount of additional space reserved to guarantee LUN writes if a volume becomes 100% full. For XenServer environments where Volume Auto Size and Snapshot Auto delete are in use, NetApp recommends setting this value to 0%. Otherwise, leave this setting at its default of 100%. To enable this option, follow these steps.

3. Log in to NetApp console.
3. Set Volume Snapshot Fractional Reserve:

```
vol options <vol-name> fractional_reserve 0.
```

Using FAS De-Dupe

FAS De-Dupe reclaims redundant disk space by dividing newly-stored data objects into small blocks, each block containing a digital signature, which is compared to all other signatures in the data volume. If an exact block match exists, the duplicate block is discarded and the disk space reclaimed.

Deduplication is enabled on a per flexible volume basis. It can be enabled on any number of flexible volumes in a storage system. It can be run one of four different ways:

- Scheduled on specific days and at specific times
- Manually, by using the command line
- Automatically, when 20% new data has been written to the volume
- Automatically on the destination volume, when used with SnapVault®

Only one deduplication process can run on a flexible volume at a time.

Up to eight deduplication processes can run concurrently on eight volumes within the same NetApp storage system.

Beginning with Data ONTAP 7.3.1, deduplication checkpoint restart allows a deduplication process that was interrupted to continue from the last checkpoint. Prior to Data ONTAP 7.3.1, an interrupted deduplication process would result in a restart of the entire deduplication process. If the system is restarted while deduplication is in process, when the system is once again online, the deduplication process automatically restarts from the last checkpoint.

Below are some results on the testing FAS De-Dupe1

- Supported by all storage data access types; iSCSI, FCP & NFS
- >70% with virtualization environment
- Saves up to 95% for full backups; 25% to 55% for most data sets.

If FAS De-Dupe is going to be enabled for the FlexVol® from XenCenter or StorageLink, then note that the volume size should match the maximum supported FAS De-Dupe limit for the device.

The maximum flexible volume size limitation for deduplication varies based on the platform (this number depends primarily on the amount of system memory). When this limit is reached, writes to the volume fail just as they would with any other volume after it is full.

This could be important to consider if the flexible volumes are ever moved to a different platform with a smaller maximum flexible volume size. The current limits are listed in the table below:

Data ONTAP 7.2.X (Starting with 7.2.5.1) and Data ONTAP 7.3.0							
FAS2020	FAS3020 N5200 FAS2050	FAS3050 N5500	FAS3040 FAS3140 N5300	R200	FAS3070 N5600 FAS3160	FAS6030 FAS6040 N7600 FAS3170	FAS6070 FAS6080 N7800
0.5TB	1TB	2TB	3TB	4TB	6TB	10TB	16TB
Data ONTAP 7.3.1 and later							
FAS2020	FAS3020 N5200 FAS2050	FAS3050 FAS2040 (7.3.2 only)2 N5500	FAS3040 FAS3140 N5300	R200	FAS3070 N5600 FAS3160	FAS6030 FAS6040 N7600 FAS3170	FAS6070 FAS6080 N7800
1TB	2TB	3TB	4TB	4TB	16TB	16TB	16TB

The maximum shared data limit per volume for deduplication is 16TB, regardless of the platform type. Once this limit is reached, there is no more deduplication of data in the volume, but writes to the volume continue to work successfully until the volume is completely full.

---

1 The testing was done in NetApp's Kilo Client Center as part of a Virtual Desktop Infrastructure Proof of Concept. The backend filer used was a FAS 3050 clustered, and XenServer running on IBM Blade Center H.

2 The FAS 2040 supports deduplication for Data ONTAP 7.3.2 only. Prior versions are not supported.

XenServer Shared Storage Configuration Basics

This section covers the best practices for configuring the various available storage options (Data ONTAP, NFS, iSCSI, FP) with a NetApp storage box. The recommended storage configuration is to utilize the StorageLink Gateway SR type since it provides the most optimized performance as well as maximum flexibility in configuration and protocol selection.

Storage Networking

1. Configure physical network infrastructure so that different traffic is on different subnets.
2. Configure a new network to make the bond and/or VLAN available
3. Create a management interface to use the new network

The next few sections will cover how multipathing and NIC bonding can be used for storage networks. The above three steps will be enumerated in each section.

Storage Multipathing

XenServer 5.0 introduced Active/Active multipathing for iSCSI and FC protocols for I/O datapaths. It is recommended to use the following diagram as a guide to configure multipathing (**D**ynamic **M**ulti**P**athing or DMP)
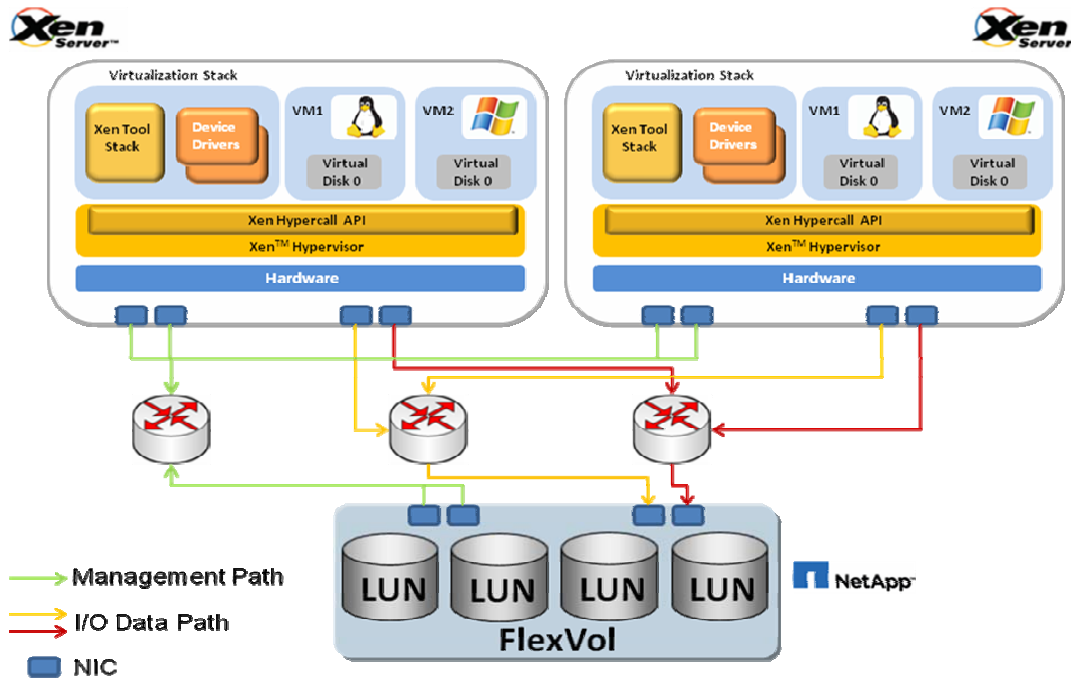
Figure 7 Graphical representation of DMP configuration

Dynamic multipathing uses a round-robin mode load balancing algorithm, so both routes will have active traffic on them during normal operations. Multipathing can be enabled via XenCenter or on the command line. However, before attempting to enable multipathing, verify that multiple targets are available on your storage server.

XenServer 5.5 supports ALUA, asymmetric logical unit access. ALUA is a relatively new multipathing technology for asymmetric arrays. NetApp is ALUA compliant. Note ALUA is only available with fiber channel.

Dedicate NICs for software iSCSI storage traffic

1. From *XenCenter*, right click on the pool (or standalone server) in the *Resources* tab. Choose the *Management Interfaces…* option
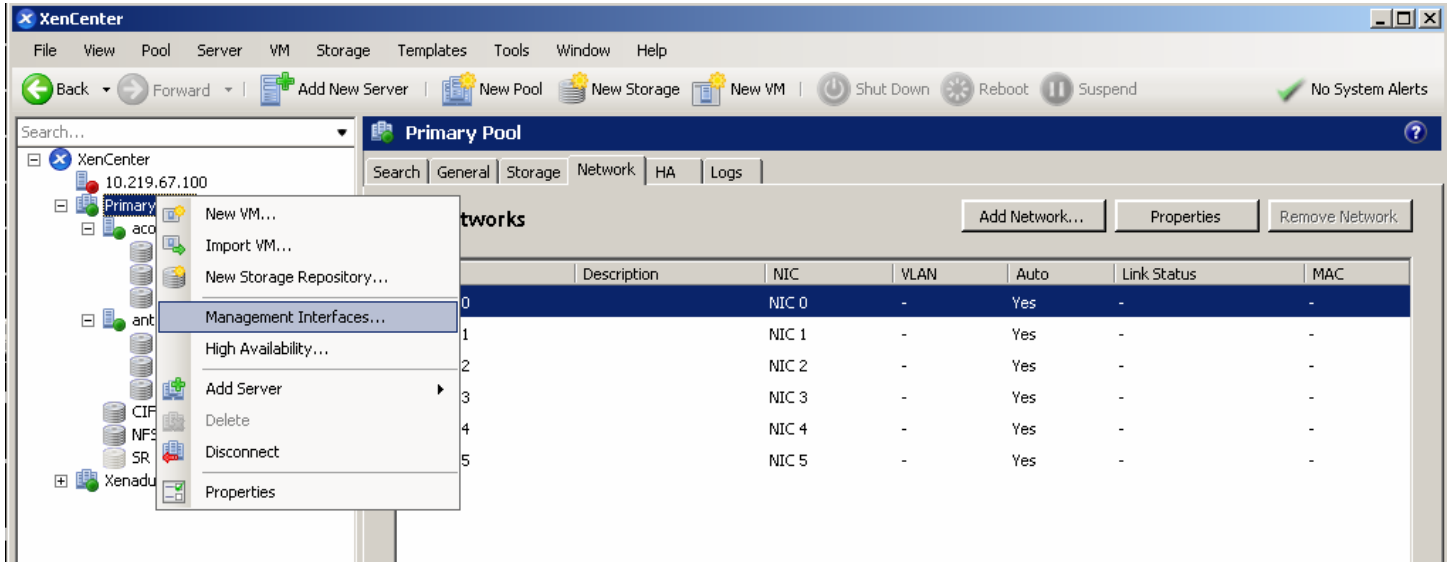
2. On the *Management interfaces* wizard, click on the *New Interface* button
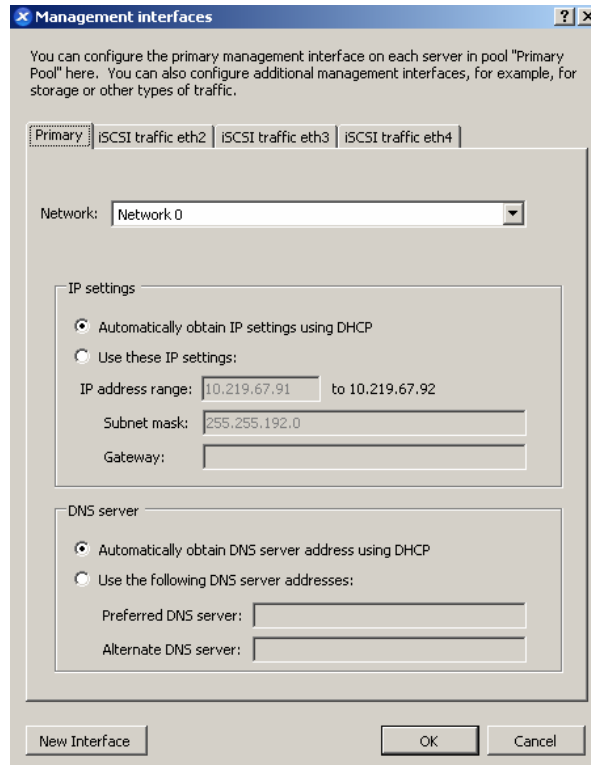
3. Give the new interface a recognizable name, and select the *Network* you want the dedicated interface on.
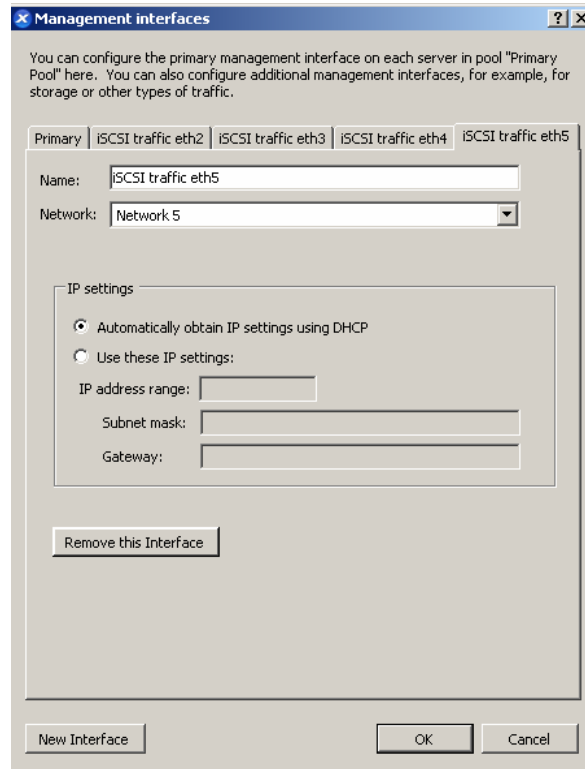
4. Click on the *Use these IP settings:* radio button and enter a starting IP address for the NICs in the *Network*.
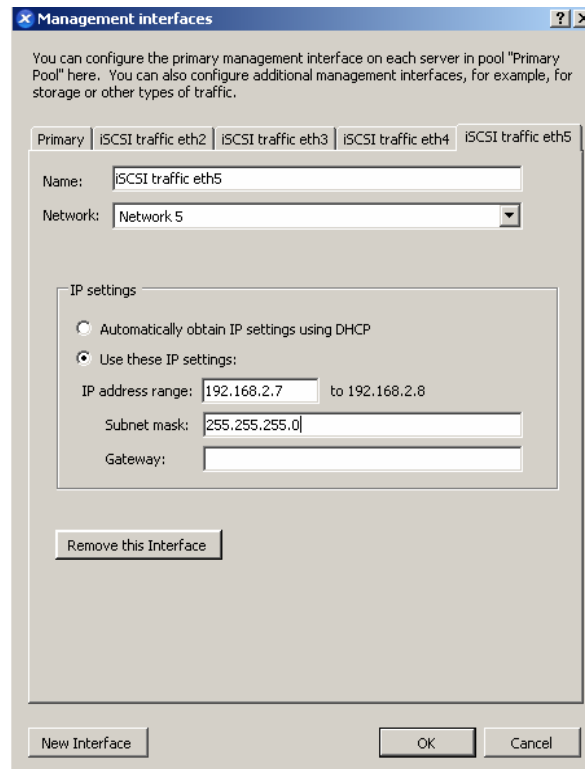
5. Repeat the above steps for each NIC dedicated to storage, and click *OK*

6. Open a console session to the XenServer pool master, and run the *iscsiadm* command with *discovery* and *sendtargets* option.

    ```
    [root@antisana-v5 ~]# iscsiadm –m discovery --type sendtargets --portal
    192.168.2.200
    192.168.2.200:3260,1 iqn.1992-08.com.netapp:sn.101173424
    192.168.3.200:3260,1 iqn.1992-08.com.netapp:sn.101173424
    ```

    The command returns multiple targets on the NetApp device showing that the backend storage device is configured properly.

Configuring Multipathing for iSCSI

0. Modify /etc/*multipath-enabled.conf* file to include the following in the *devices* section.

    ```
    device
        {
        vendor                     "NETAPP"
        product                    "LUN"
        path_grouping_policy       multibus
        getuid_callout             "/sbin/scsi_id –g –u –s /block/%n"
        prio_callout               "/sbin/mpath_prio_ontap /dev/%n"
        features                   "1 queue_if_no_path"
        path_checker               directio
        failback                   immediate
        }
    ```

Configuring Multipathing for Fiber Channel

0. Open a console session to the NetApp device, set igroup to ALUA.

    ```
    igroup set <initiator_group> alua yes
    igroup show –v
    ```

0. Modify */etc/multipath-enabled.conf* file to include the following in the *devices* section
    ```
    defaults
    {
          user_friendly_names yes
          max_fds max
          queue_without_daemon no
    }
    devices
    {
           device{
    vendor "NETAPP"
    product "LUN"
    getuid_callout "/sbin/scsi_id –g –u –s /block/%n"
    prio_callout "/sbin/mpath_prio_alua /dev/%n"
    features "1 queue_if_no_path"
    hardware_handler "0"
    path_grouping_policy group_by_prio
    failback immediate
    path_checker directio
    flush_on_last_del yes
             }
    }
    ```

Enable Multipathing at XenCenter

1. From *XenCenter*, right click the server in the pool from the *Resources* pane, and select the option *Enter Maintenance Mode…*
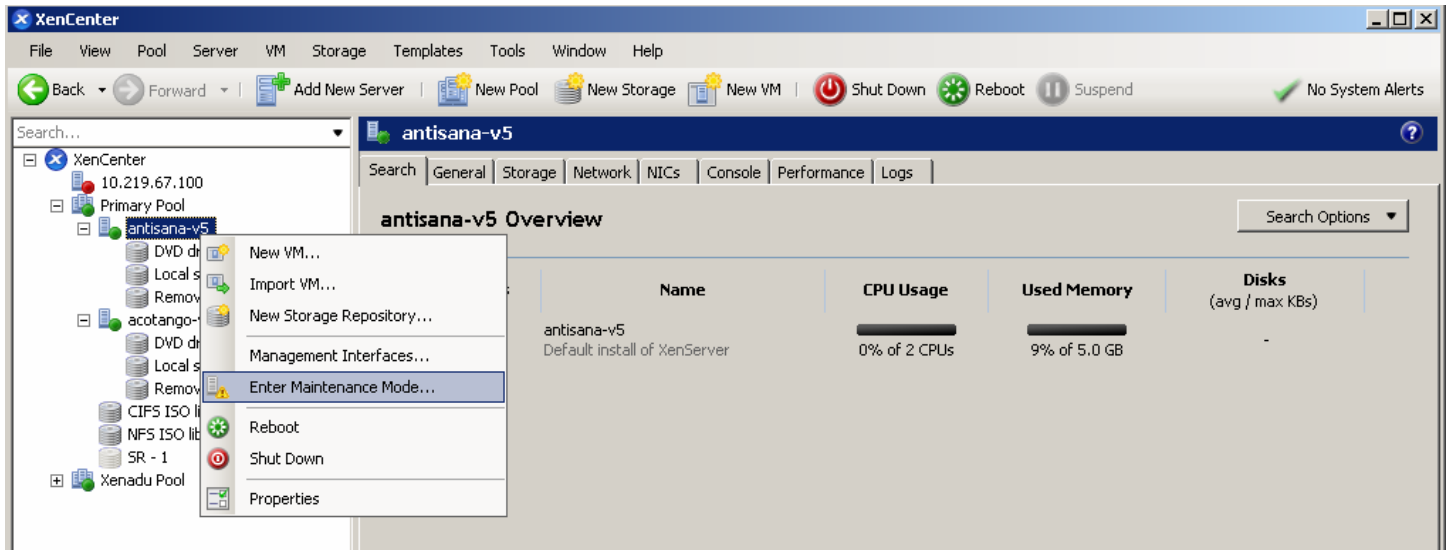
Select *Enter maintenance mode* from the dialog box. Choose any other server node in the pool as the temporary master.
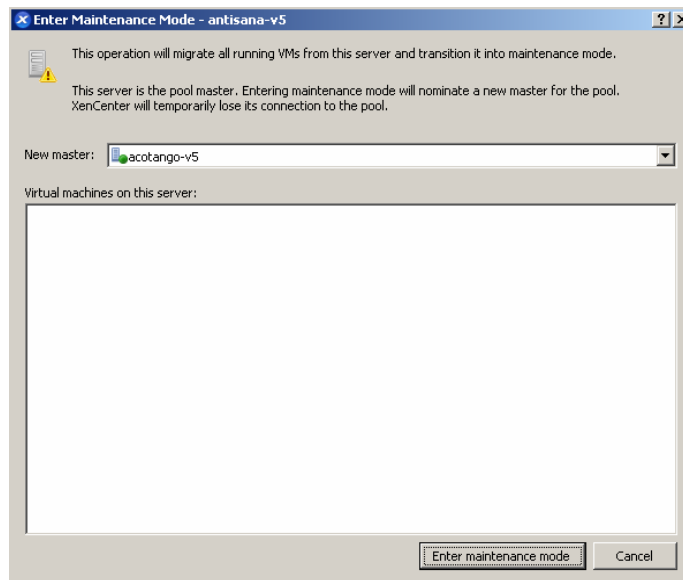
There is a short delay while *XenCenter* migrates any active virtual machines and unplugs the existing storage; if the server is a pool master, it will be disconnected and may disappear from the *Resources* pane temporarily while a new pool master is assigned.

2. When connection is restored, click on the server and then the *General* tab.
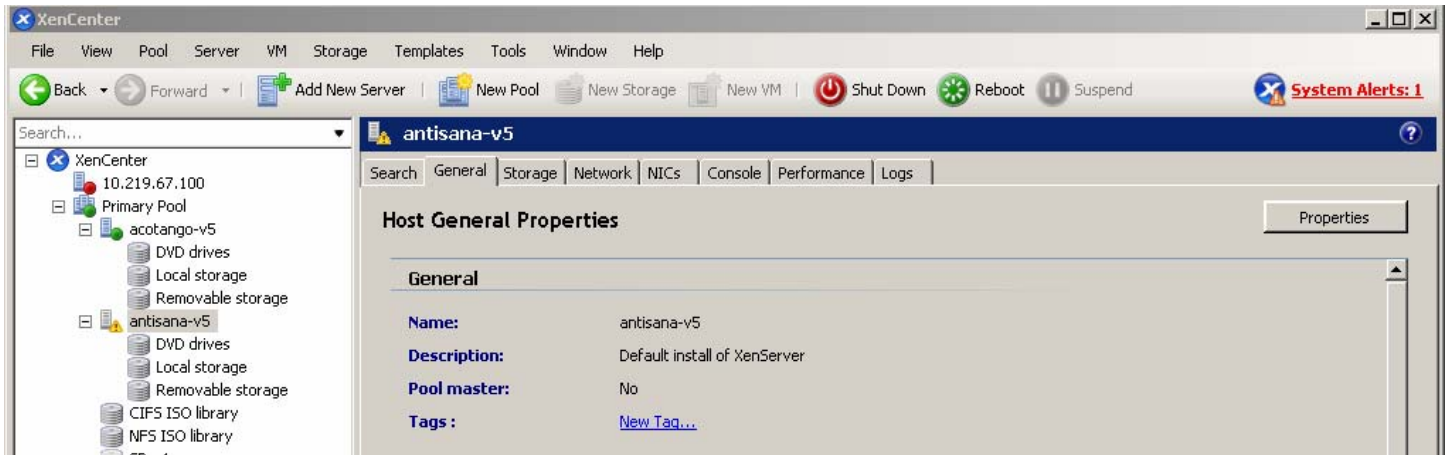
Click the *Properties* button and then click on the *Multipathing* tab. Select the *Enable multipathing on this server* check box, and click *OK*.
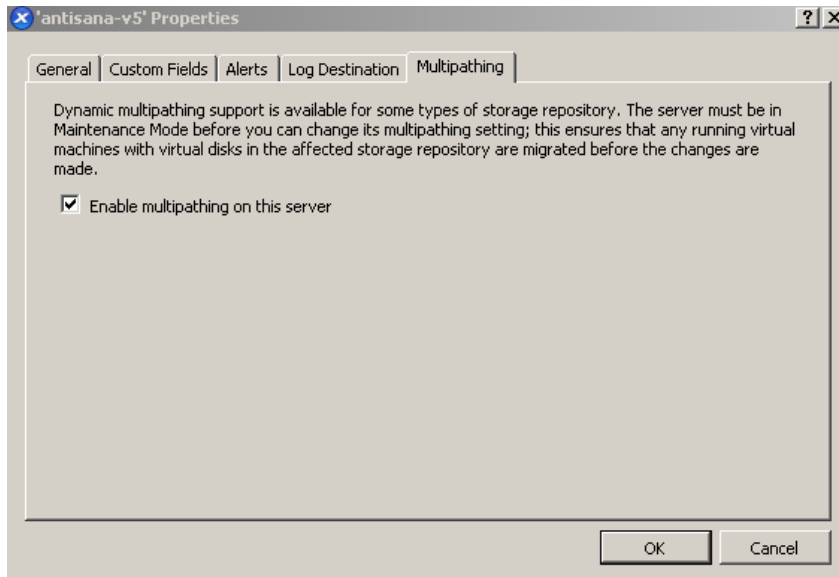
There is a short delay while *XenCenter* saves the new storage configuration.

3. Take the server back out of Maintenance mode by right clicking on the server in the *Resources* pane and click *Exit Maintenance Mode.*
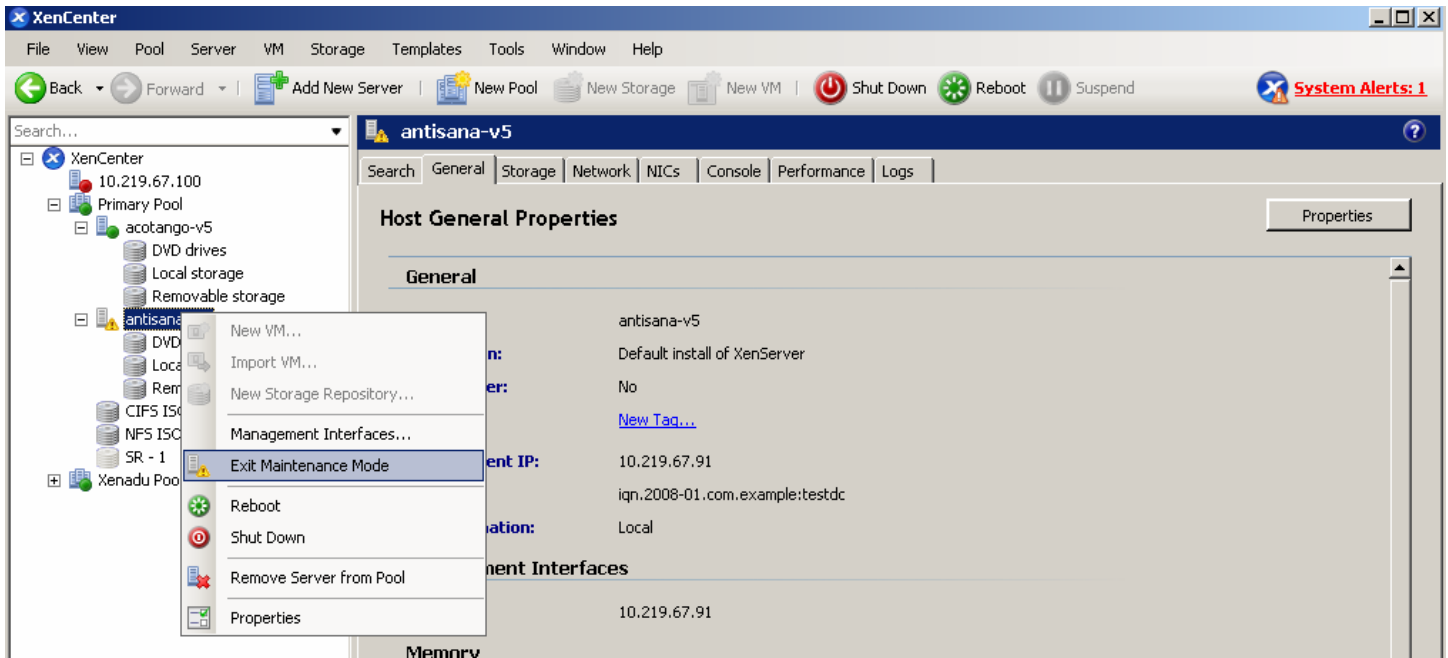
3. Repeat the above steps for all servers in the pool that multipathing needs to be enabled on.

NIC Bonding

NIC bonds can improve XenServer host resiliency by using two physical NICs as if they were one. If one NIC within the bond fails the host's network traffic will automatically be routed over the second NIC. NIC bonds work in an active/active mode, with traffic balanced between the bonded NICs.

As such, XenServer host NICs should be bonded for NFS traffic to the NetApp device. It is recommended to use the following diagram as a guide to configure NIC bonding.
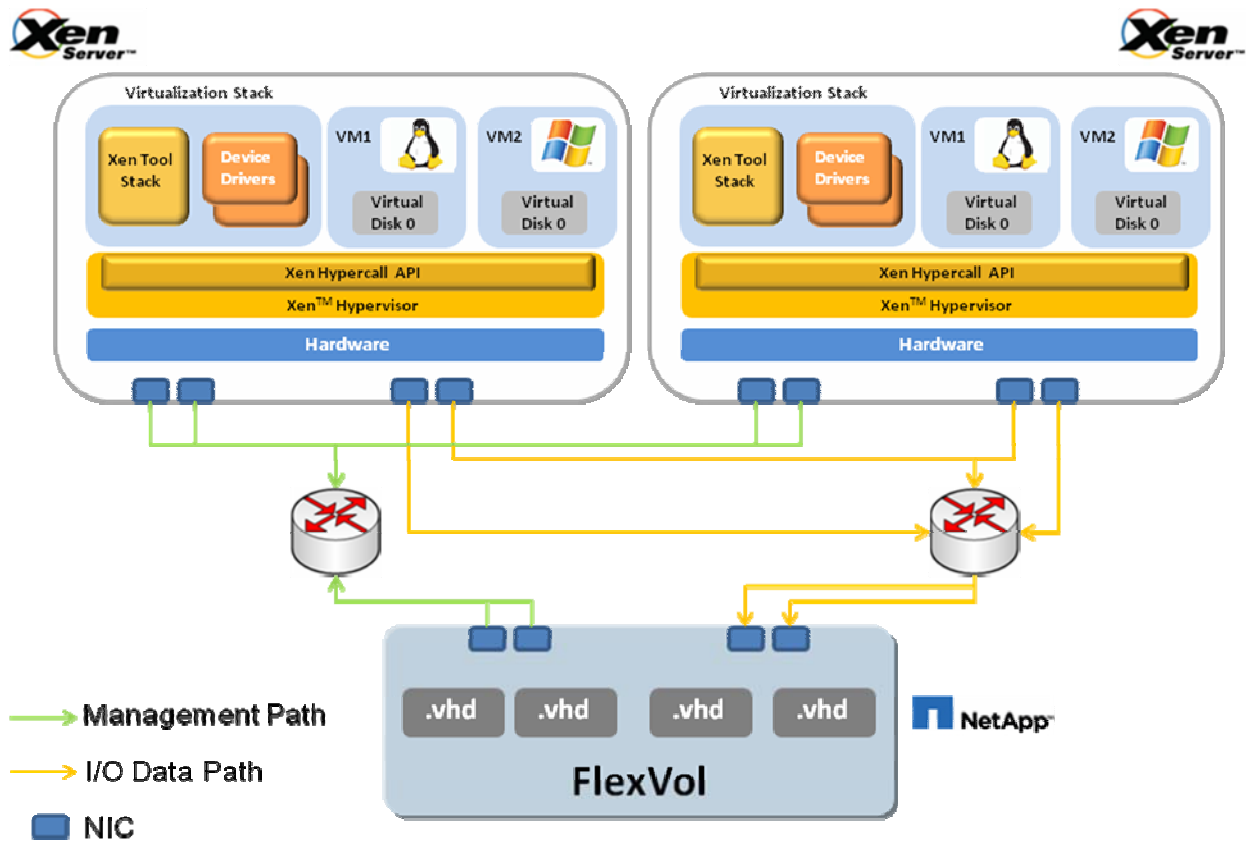
Figure 8 Graphical representation of NIC bonding configuration

1. From XenCenter, click on the pool in the *Resources* pane (or server if this is a standalone server), and click on the *Network* tab.
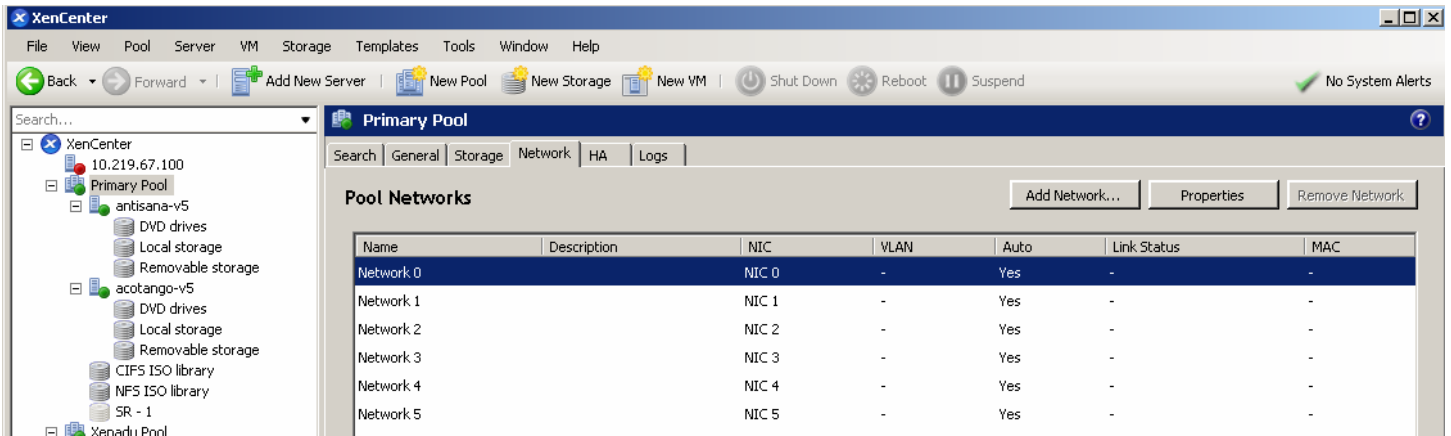
2. Click on the *Add Network…* button. Select the *Bonded Network* radio button, and click *Next>*.
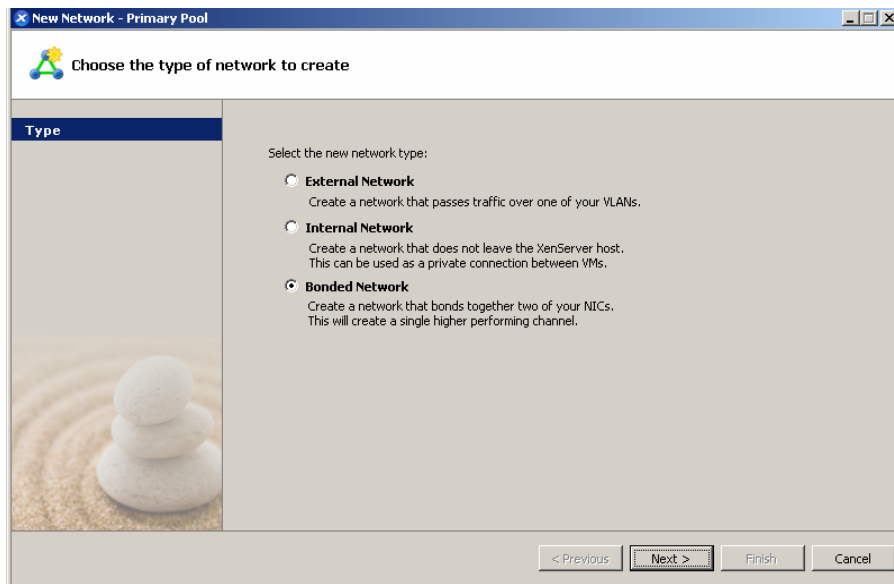
3. Add the NICs that need to be bonded from the *Available NICs* pane. Click *Finish*.
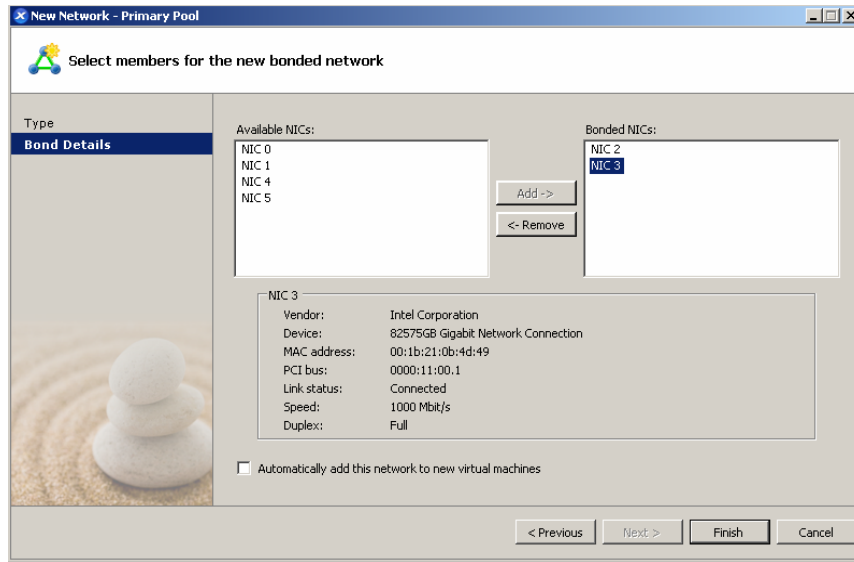
4. This will create a new network named *Bond 2 + 3*.
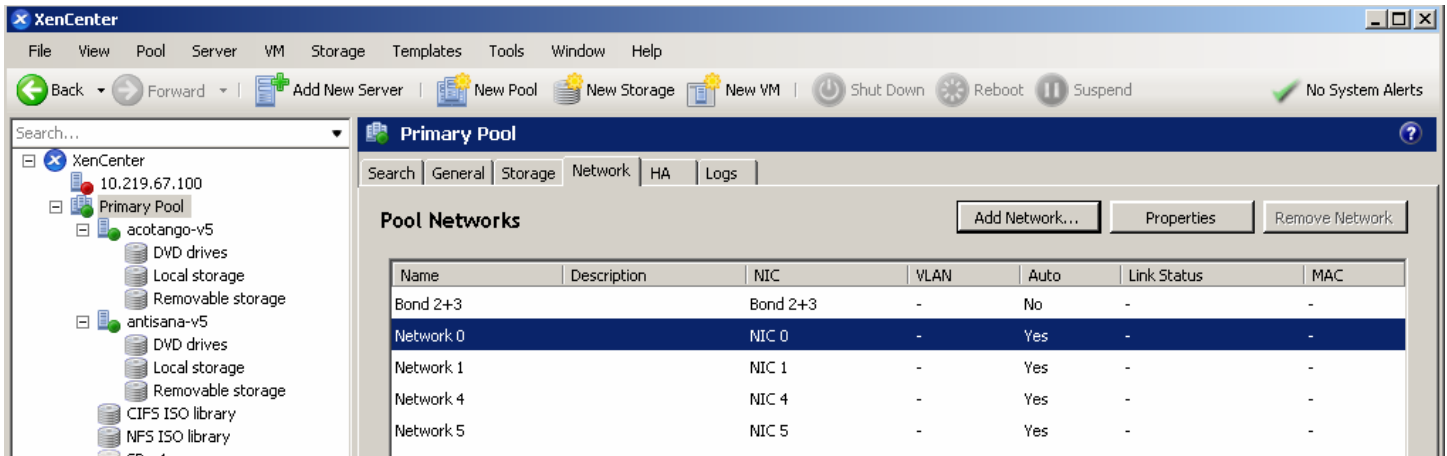
5. Select *Bond 2 + 3,* and click the *Properties* button to give the new network a more meaningful name and description.
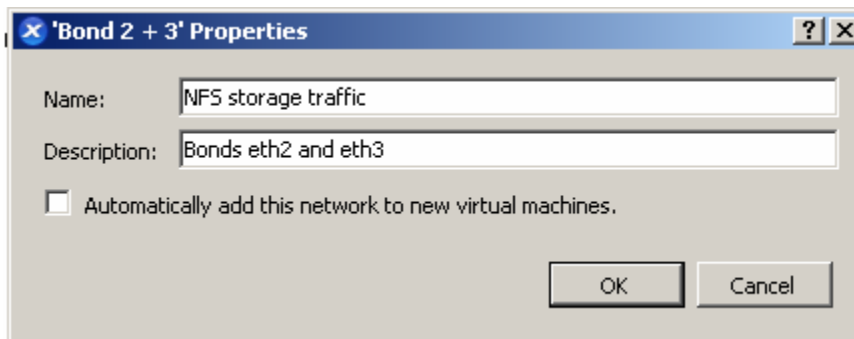
6. Right click on the pool (or server if in standalone mode) in the *Resources* pane, and click on *Management Interfaces…*
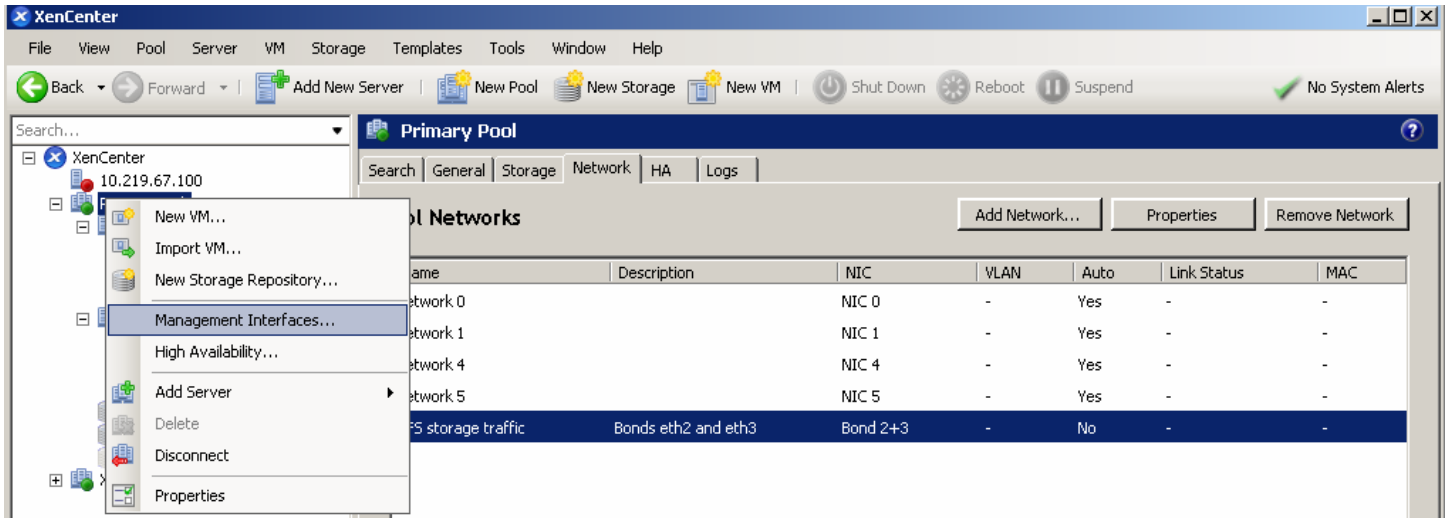
7. Click on *New Interfaces* in the *Management interfaces* wizard.
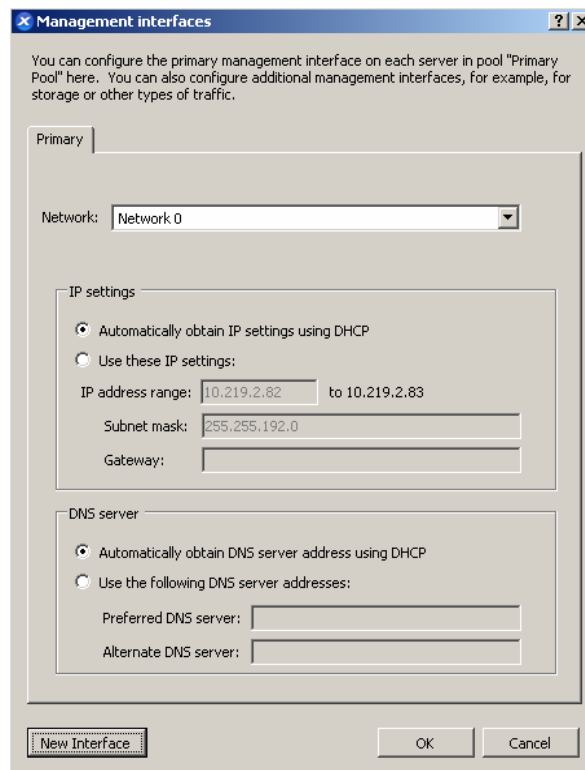
8. Give the new interface a name, and choose the *Network* created in previous steps from the drop down. Click *OK*.
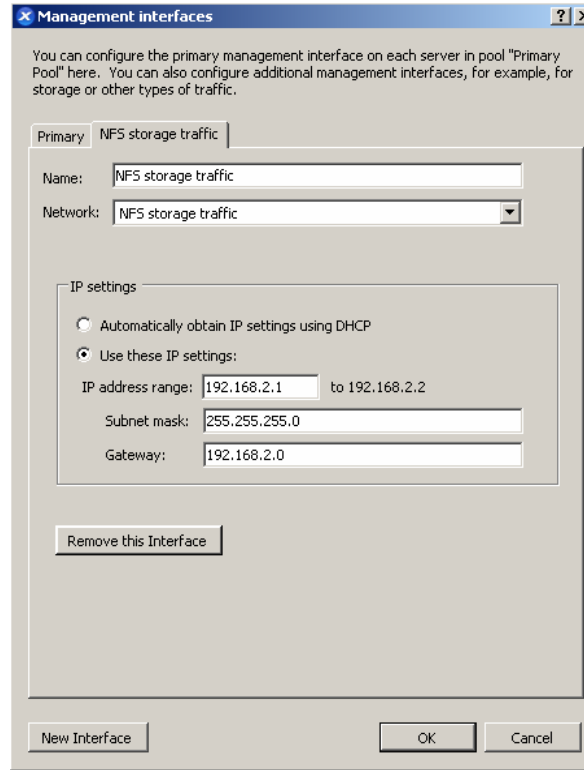
Shared Storage using StorageLink Gateway

Use of the StorageLink Gateway for shared storage is the recommended approach for deployments of XenServer with NetApp. This type of storage repository is available with Citrix Essentials for XenServer, Enterprise Edition, and requires the installation of the StorageLink Gateway service on a Windows Server virtual machine or physical server. In addition to XenServer, StorageLink Gateway also supports Hyper-V hosts and provides storage interoperability between the two platforms. As mentioned before, the StorageLink Gateway only serves as a broker, and is not part of the data path between a XenServer host and the NetApp storage infrastructure.

If installing the StorageLink Gateway on a XenServer virtual machine, the virtual machine disks will need to be connected via a standard NFS, iSCSI, or Hardware HBA storage repository type. For highly available configurations, it is recommended that StorageLink be implemented in an isolated resource pool and using an iSCSI or Hardware HBA storage repository type (ideally the same SR used for the XenServer HA heartbeat disk). The StorageLink Gateway services do not need to be available while VMs are running, however many operations such as VM start require the service for a short period of time.

Once the StorageLink Gateway Service and Manager (management console) have been installed on a virtual machine or physical server, the following process can be followed to create the storage repository.

1. Open the StorageLink Manager console.

2. Click "Connect to the StorageLink Gateway" in the center pane, and enter the credentials entered during installation.

3. Add the Hypervisor hosts. Use the name or IP address of the XenServer resource pool master and the XenServer root credentials. (Note: if you receive a networking error during this step, you may need to temporarily remove NIC bonds from your XenServer host(s). The bonds can be recreated once the hosts have been added to the StorageLink Manager.)
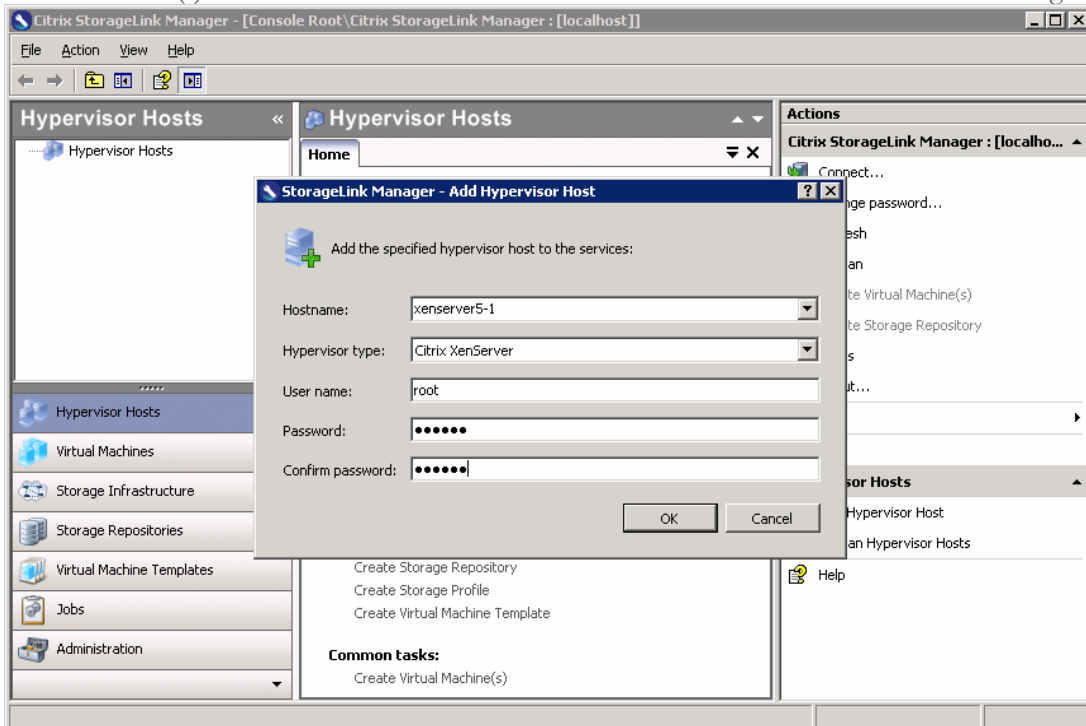
0. Once all of the XenServer hosts have been enumerated, your screen should look something like this:
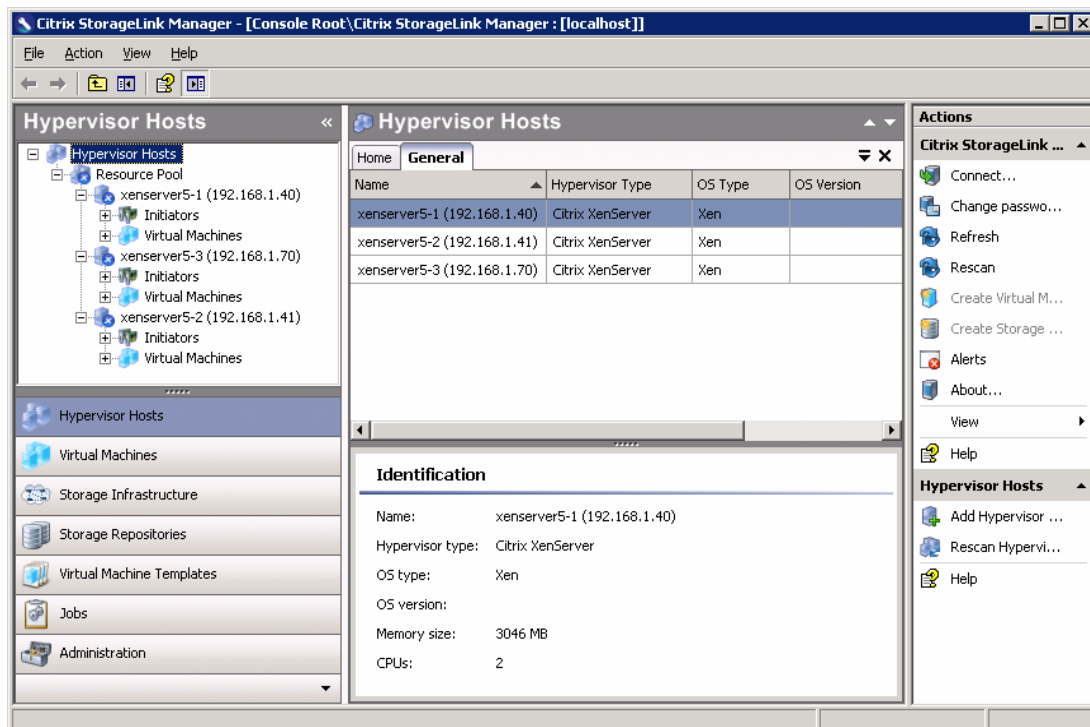
5. Next, add the NetApp storage system. Select the "Storage Infrastructure" tab and then "Add Storage System."
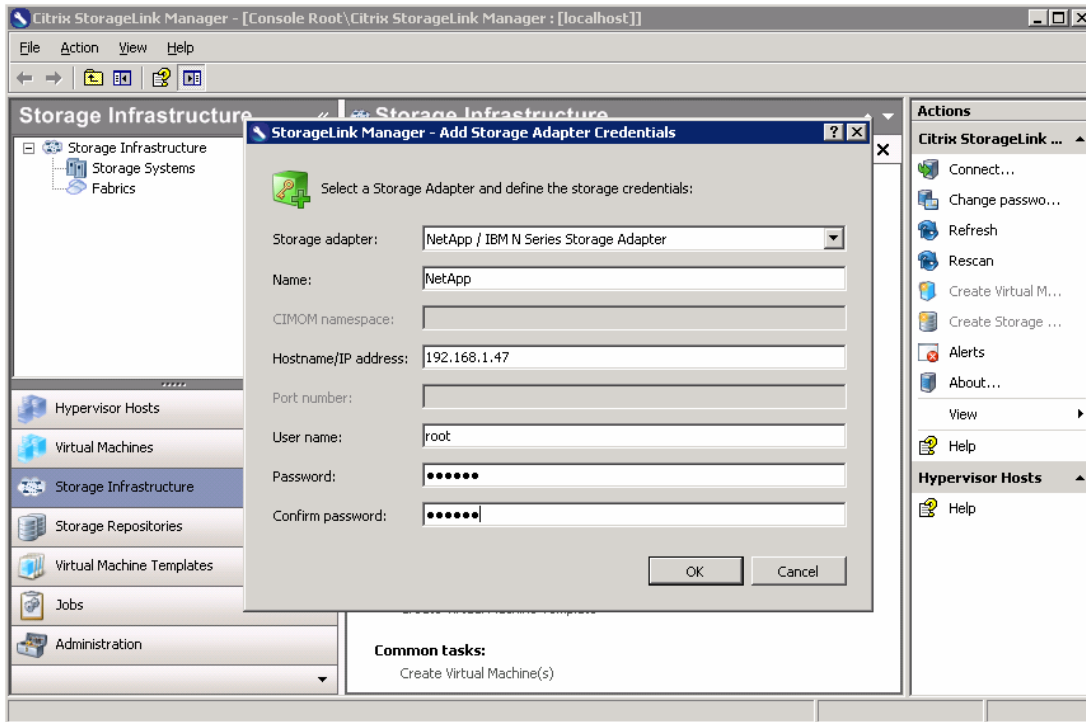
6. After the storage system has been discovered, your screen should look similar to this. You will see the any aggregates on the NetApp storage within the StorageLink Manager console (in this case aggr0 and aggr1)
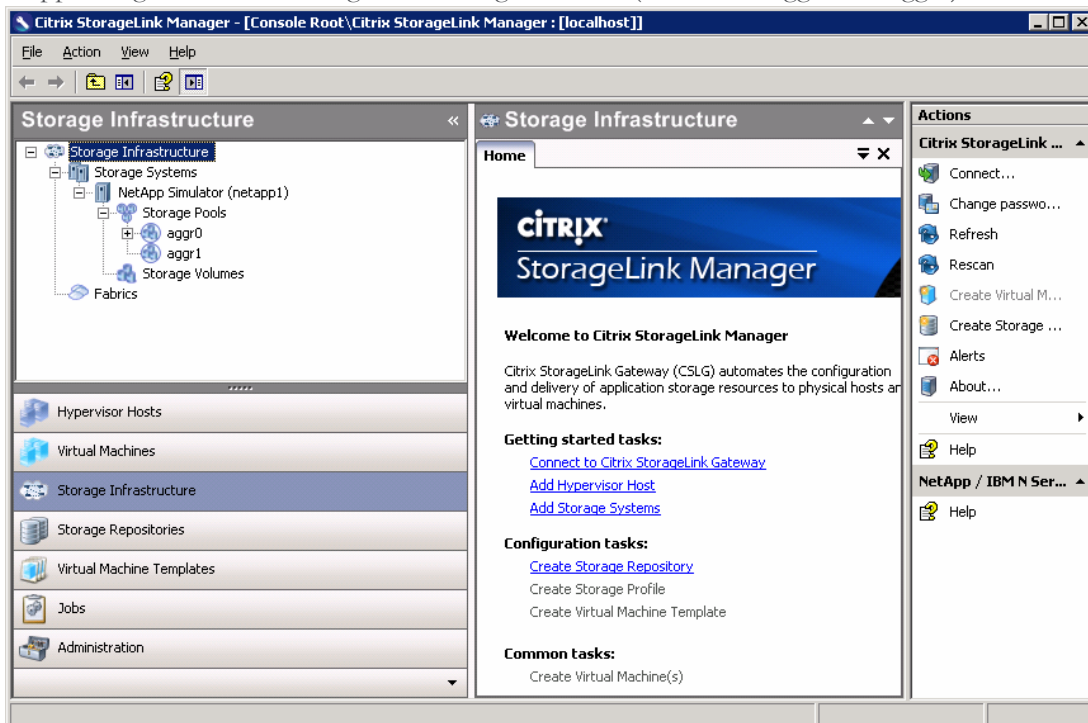
7. Create the Storage Repository. Click on the "Create Storage Repository" link in the middle pane.
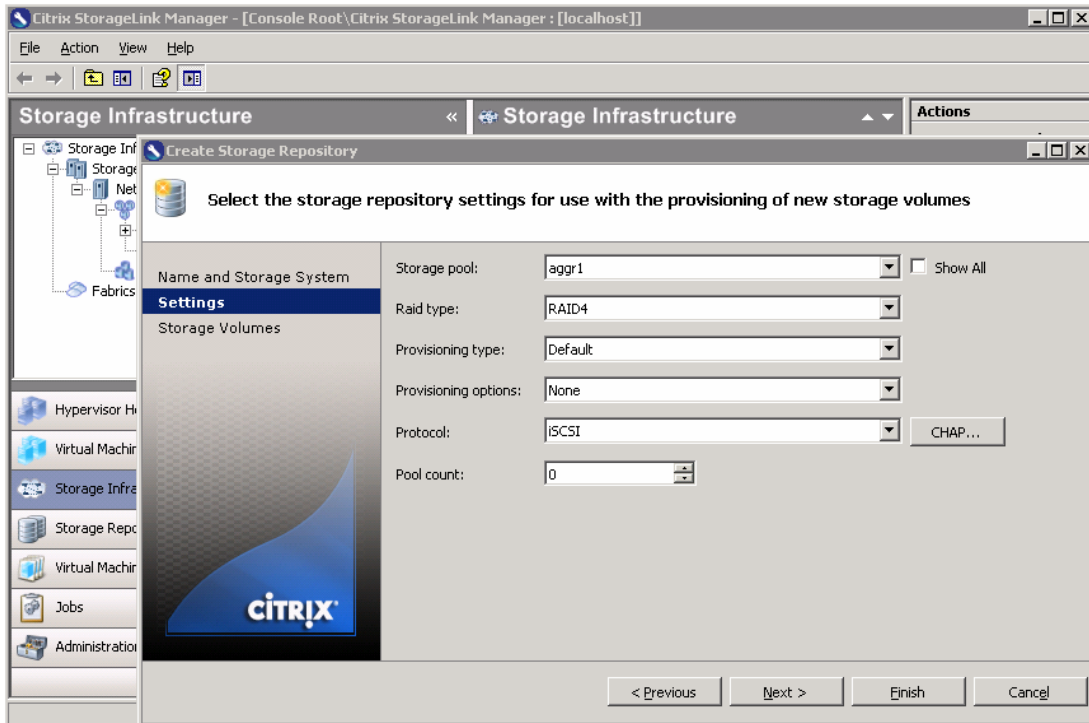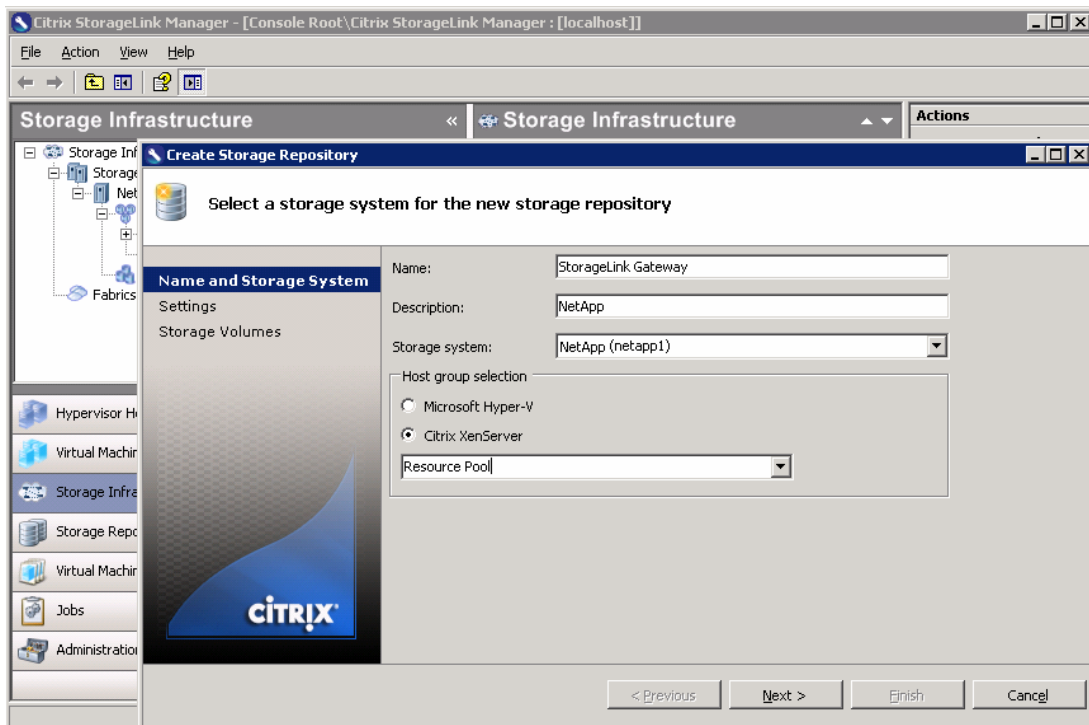
1. After clicking "Finish" an SR named "StorageLink Gateway" (or whatever name you specified) will be visible within both the StorageLink Manager as well as XenCenter, similar to what is shown below.
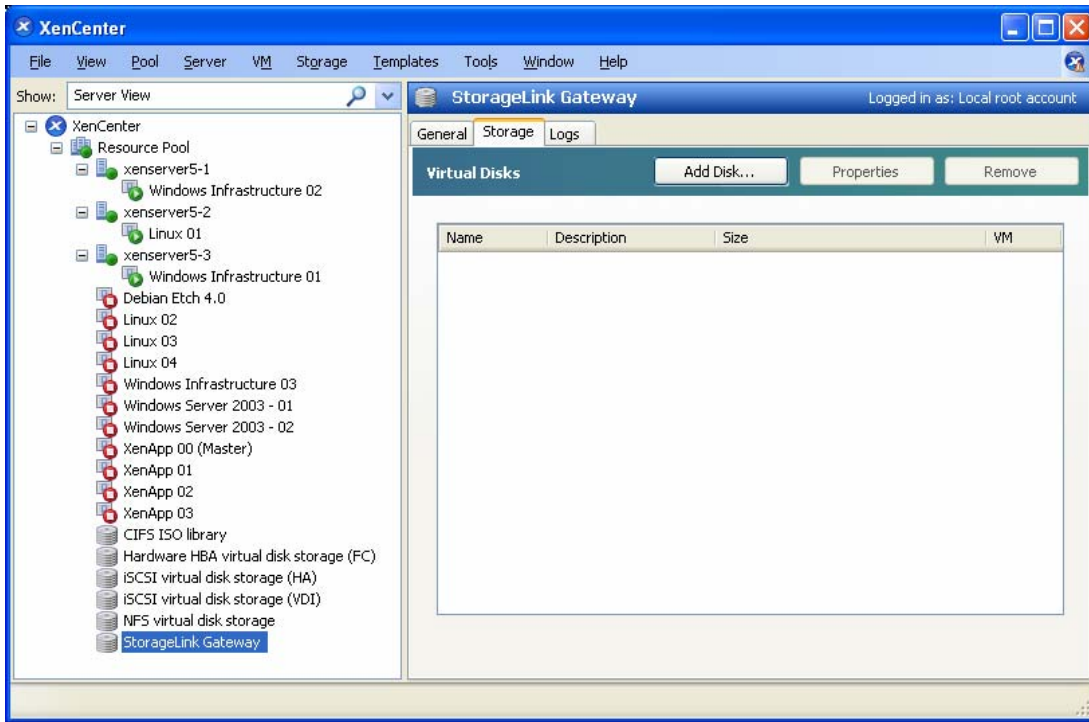
1. In addition, a new volume will be created on the NetApp aggregate selected in step 7. The name of this new volume will start with the name of the SR.

1. At this point, the StorageLink Manger console will not be required for regular administration activities. The StorageLink Gateway service does need to be running for many VM operations within XenCenter.

2. When new VMs are created on the storage repository, the StorageLink Gateway will automatically provision LUNs on the NetApp storage system. For example, after using XenCenter to create a new Debian VM with two disks (one 4 GB and another 512 MB), the following two LUNs will be visible in NetApp FilerView:
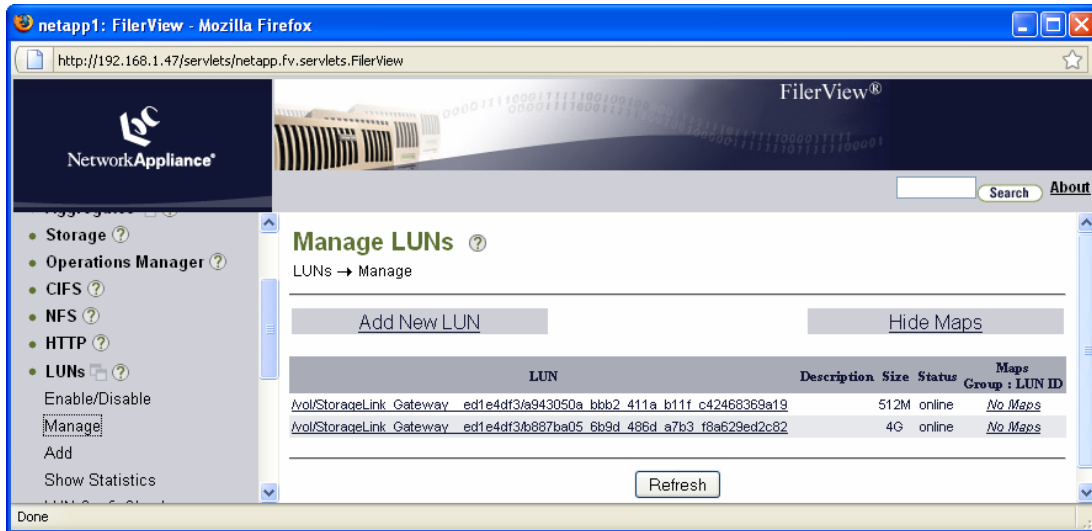
2. If a virtual machine snapshot is performed from XenCenter, additional LUNs will be visible in FilerView with a "snapXX" appended to the name, similar to below:
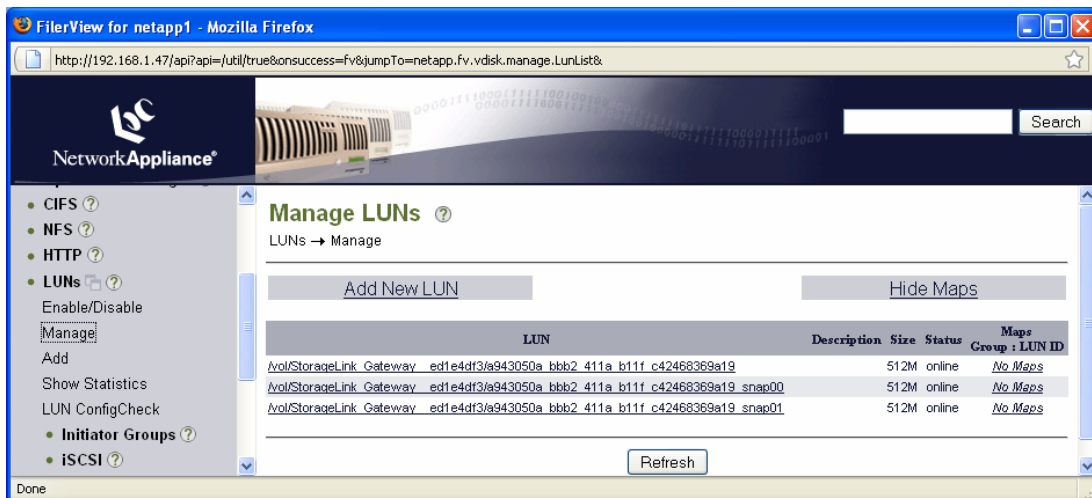
Shared Storage using the direct StorageLink Adapter for NetApp

The direct StorageLink adapter for NetApp was first introduced in XenServer 4.1. While it is still available in XenServer 5.5, it is primarily intended to maintain backward compatibility with pre-5.5 deployments. For new XenServer 5.5 deployments, it is recommended to use the StorageLink Gateway.

The direct StorageLink-NetApp storage repository uses NetApp Data ONTAP for its control path, with in-band configuration and management of the data path via the host software initiator. The only backend NetApp storage configuration required is to create an aggregate that will house the FlexVols used by this storage repository type. See Appendix B for information for security considerations to allow XenServer administrators to have root access to the device to be able to provision FlexVols and generate Snapshots on the device.

1. For using NetApp Data ONTAP to create the storage repository, in XenCenter, choose *New Storage*. Select *NetApp*.
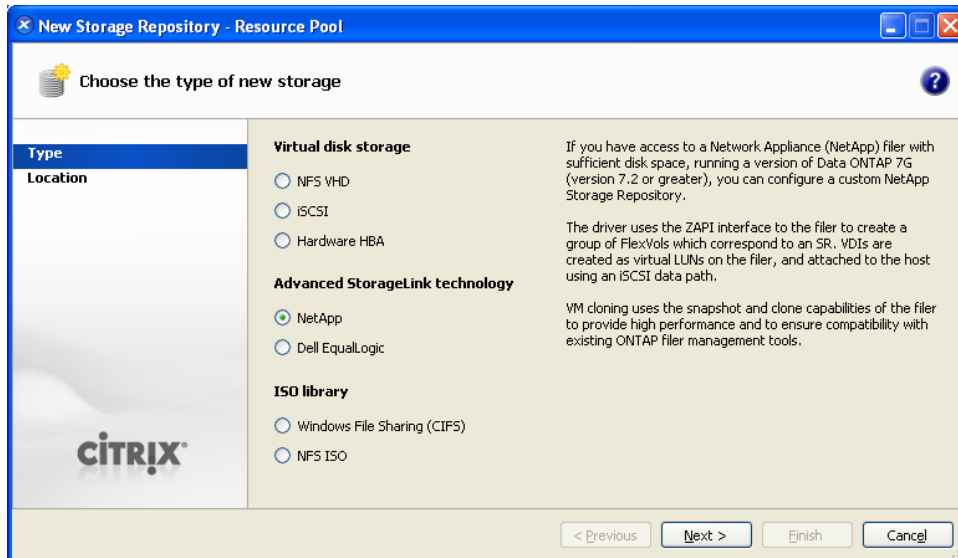
2. Provide the name of the device (or its IP address), and authentication credentials. If CHAP is required, then select *Use CHAP* and provide the username and password. CHAP is only required if there is a security requirement between the NetApp storage and the XenServer.
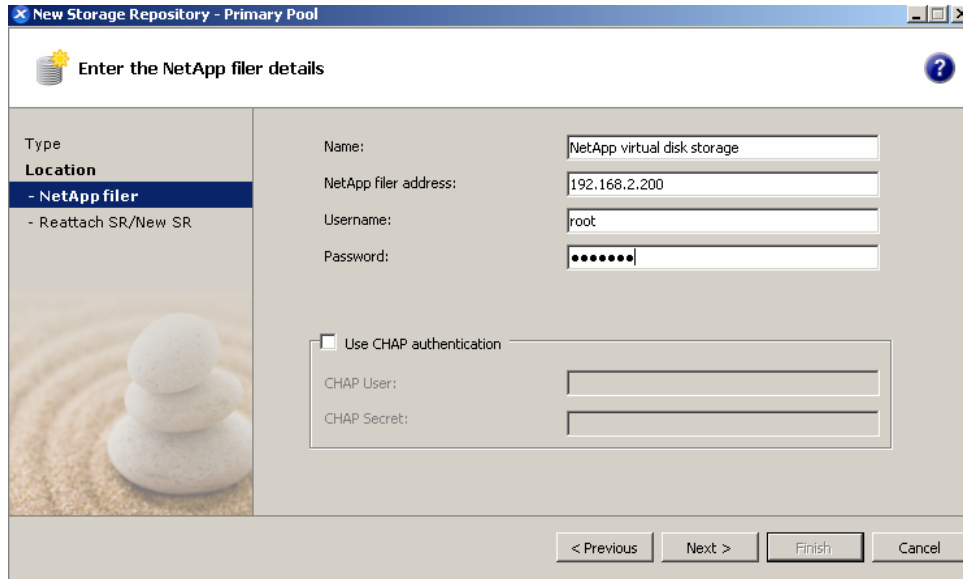
Note, the device manages CHAP authentication credentials per host IQN. This must be managed and configured directly on the device if required.

3. The NetApp Data ONTAP adapter will poll the device for existing aggregates. Choose the aggregate to create the storage repository in. If thin provisioning is required, then check the box entitled *Use NetApp thin provisioning*. With thin provisioning, it is also possible to make enable FAS de-dupe on the backend.

Thin provisioning is a very useful space conserving mechanism for Virtual Machine disk storage, since many Virtual Machines are likely to significantly under-utilize all the virtual disk space allocated. Furthermore, FAS de-dupe can be very effective where many VMs contain the same, or similar, Operating System, and there is likely to be a significant amount of duplication of data across disks. Selecting *Use NetApp thin provisioning* with *Enable FAS deduplication* can significantly reduce the amount of space required on disk. Note, however that there are no space guarantees when operating in thin provisioning mode, so it is quite feasible to over-provision the amount of allocated space. If an over-provisioned aggregate runs out of space, a LUN will be forcibly set offline when any data writes are received which may cause a Virtual Machine to crash. Management of space usage can be greatly improved by utilizing the NetApp Data ONTAP alert mechanisms. Alternatively, Virtual Machine disk space can be guaranteed by disabling thin provisioning.

Deduplication is enabled on a volume, and the amount of data deduplication realized is based on the commonality of the data stored in a deduplication-enabled volume. For the largest storage savings, NetApp recommends grouping similar operating systems and similar applications into one volume. For deduplication best practices, including scheduling and performance considerations, see TR 3505 NetApp FAS Dedupe: Data Deduplication Deployment and Implementation Guide.

When we consider how many volumes we create, we need to consider the total VM number, backup strategy and space saving. The more same OS VMs in one volume, the more deduplication storage saving. NetApp recommends grouping VMs with similar backup requirement in other words VMs with the same snapshot schedule in one volume.

For example, a 1000 virtual desktops environment, we can put the 250 VMs in one volume and have 4 volumes to host the 1000 VMs. And use SnapMirror to backup gold image volume to disaster recovery site storage. Another example, we can use one volume to host 100 VMs if all the VMs have same snapshot backup requirement.
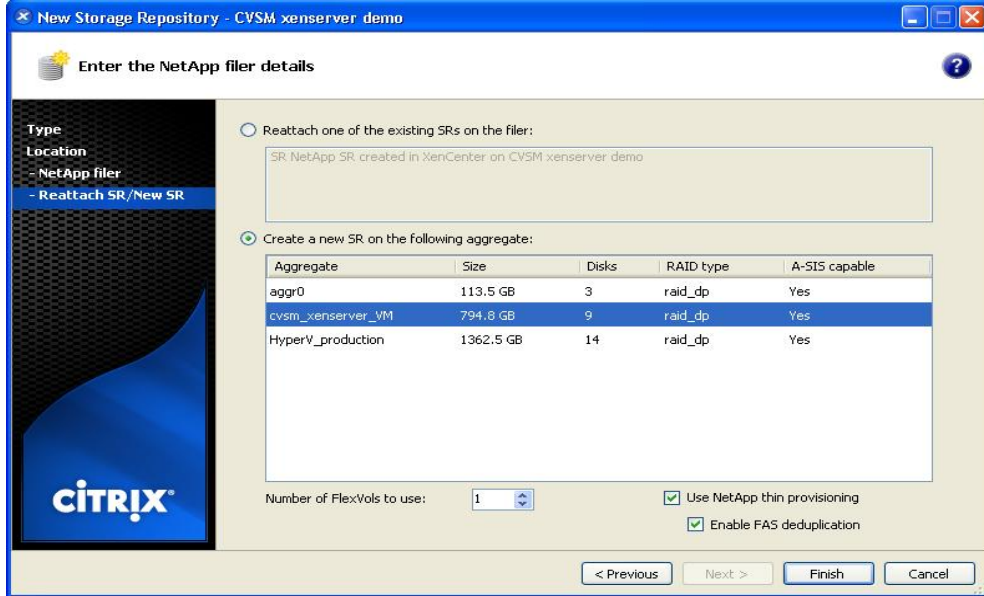
4. Click *Finish* for the storage repository to be created using Data ONTAP for the control path.

NOTE: If multipathing is enabled, the NetApp SR in XenCenter will show multipathing as *Not Active*. The reason is that the NetApp SR type is a LUN-per-VDI model, and no LUNs or VDIs exist when the SR is created.
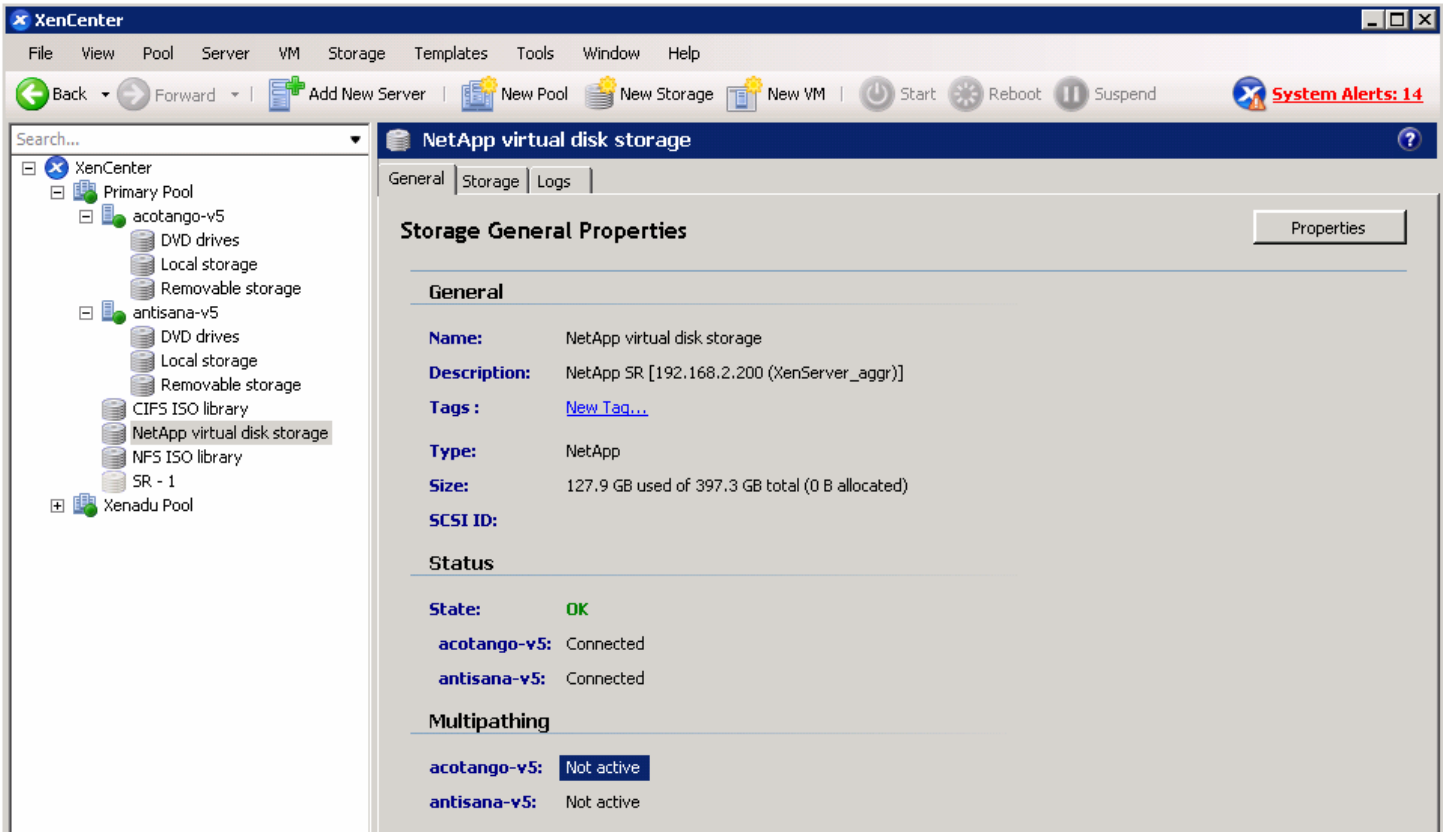
However, once a VM has been created with VDI's attached and is started up, XenCenter will show multipathing as *Active*. Henceforth, the SR will remain in multipath mode.
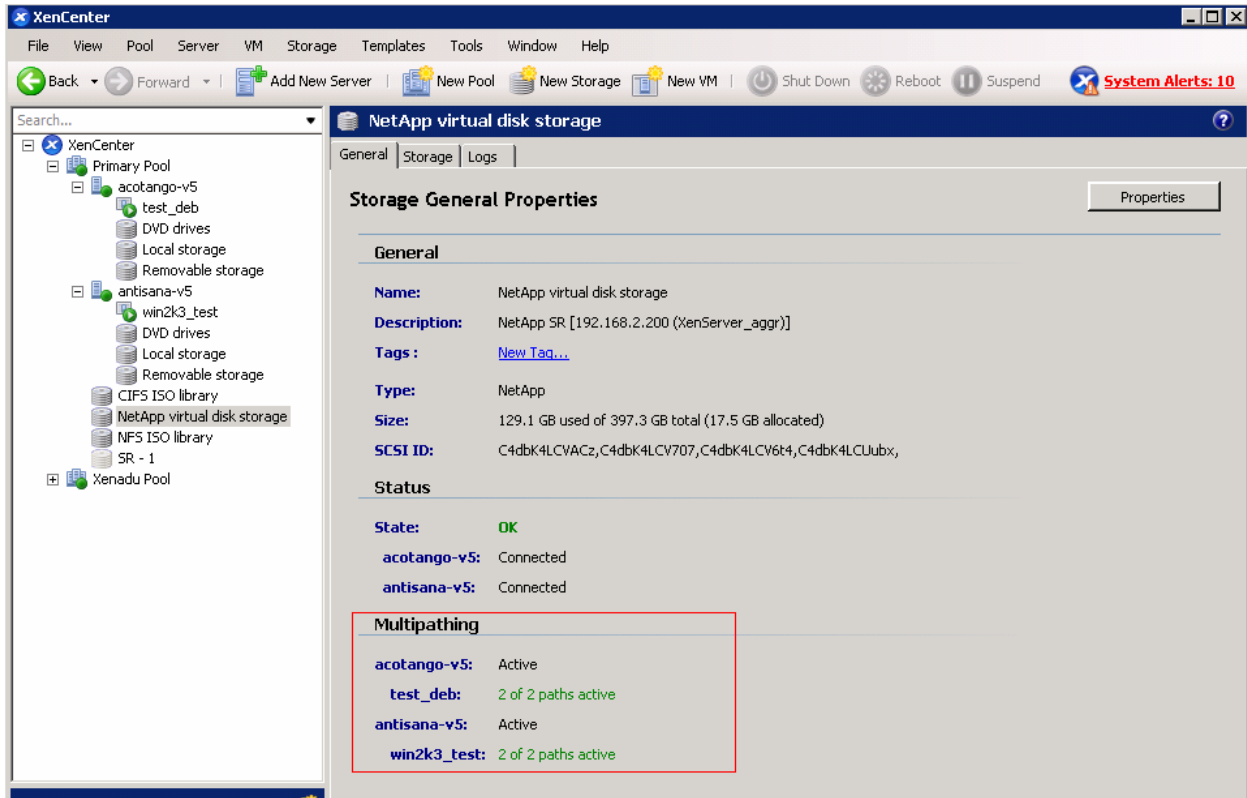
Identifying XenServer objects on the NetApp Device

The NetApp SR is mapped to a FlexVol in the NetApp device. Each VDI is represented as a LUN within a FlexVol, with VDI's of the same VM being within a single FlexVol.

When the NetApp SR is created in Direct StorageLink-NetApp Adapter, the FlexVols created on the NetApp device follow a naming pattern:

XenStorage_*UUID*_FV*n*

where *UUID* is the UUID of the NetApp SR in XenServer, and *n* is the FlexVol number

The LUN that are created within the FlexVol use as their name the UUID of the VDI that they represent.

When the NetApp SR is create in StorageLink gateway, the FlexVol created on the NetApp device follows a naming pattern: <CustomizedName>_UUID.

To view the FLexVols that correspond to the NetApp SR

1. In XenServer console, get the UUID of the NetApp SR using the *xe sr-list* command.

2. On the NetApp FilerView, click on *Volumes->Manage.*
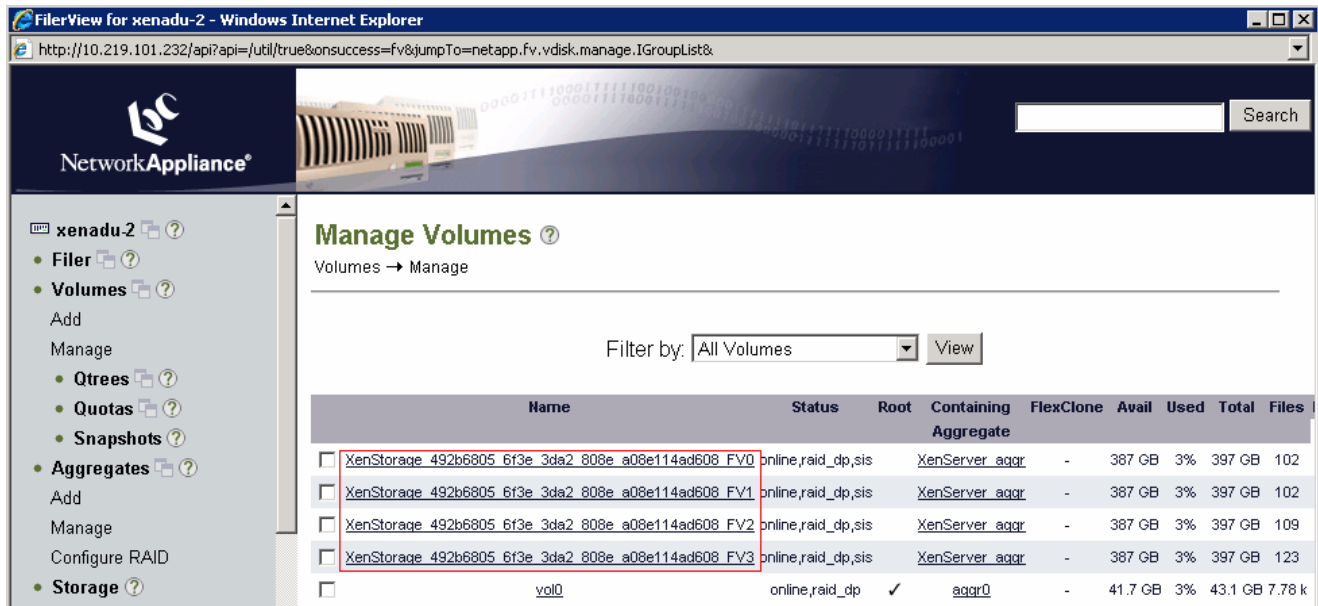
In the FilerView above, *XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV0* to *XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV3* are the FlexVols that make up the NetApp SR which has UUID 492b6805-6f3e-3da2-808e-a08e114ad608.

On the NetApp device, the FlexVol can also be displayed via ONTAP CLI command.

```
vol status
```

To view the LUNs associated with the VDIs in the NetApp SR

1. In XenServer console, get the UUID of VDIs created on the NetApp SR using the *xe vdi-list* command.

2. In the NetApp FilerView, click on *LUNs->Manage.*

In the above example, /vol/XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV3/b138d821-6c2d-4c88-b631-bb8202fc6465 is the LUN that represents the VDI with UUID b138d821-6c2d-4c88-b631-bb8202fc6465.

On the NetApp device, the LUNs can also be displayed via ONTAP CLI command.

        lun show

```
xenadu-2> lun show
        /vol/XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV2/33f18309-33ef-413b-bdf3-5055ce5710c8      25
0m (262144000)    (r/w, online)
        /vol/XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV3/111264a8-616b-46ae-9475-71c33056fef1
8g (8589934592)    (r/w, online)
        /vol/XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV3/88dddbdc-1568-4cb6-bfcb-1114312e73b5
8g (8589934592)    (r/w, online)
        /vol/XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV3/b138d821-6c2d-4c88-b631-bb8202fc6465
8g (8589934592)    (r/w, online, mapped)
```

Configuration Shared NAS using NFS

To use the NetApp storage box as a shared NAS storage option using NFS, it is recommended that a separate volume be created for VDI storage. To do so:

1. Open the FilerView for the storage box by pointing a browser to http://device/na_admin.

2. Click on *Volumes*, and then *Add* to open the *Volume Wizard*.

3. Click *Next* and select the volume type as *Flexible*.
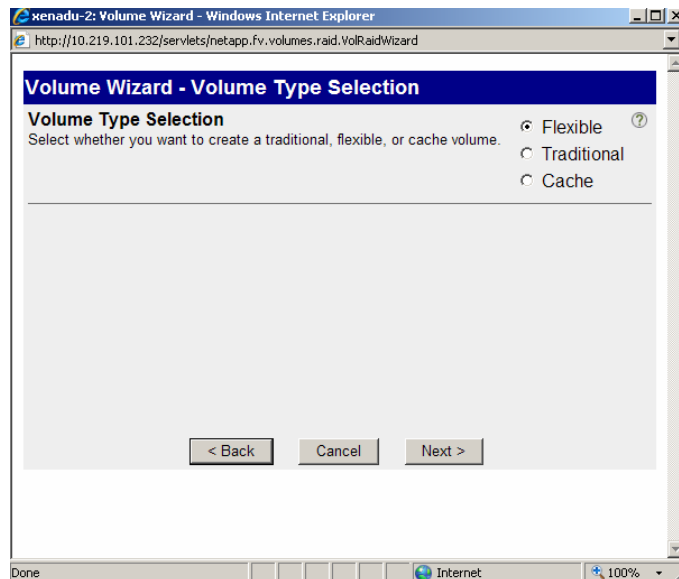
4. It is recommended to give a name that the NetApp storage server's automatic support system can identify as specific to XenServer storage, for example, a name such as "XenServer_NFS_Storage" would be appropriate.
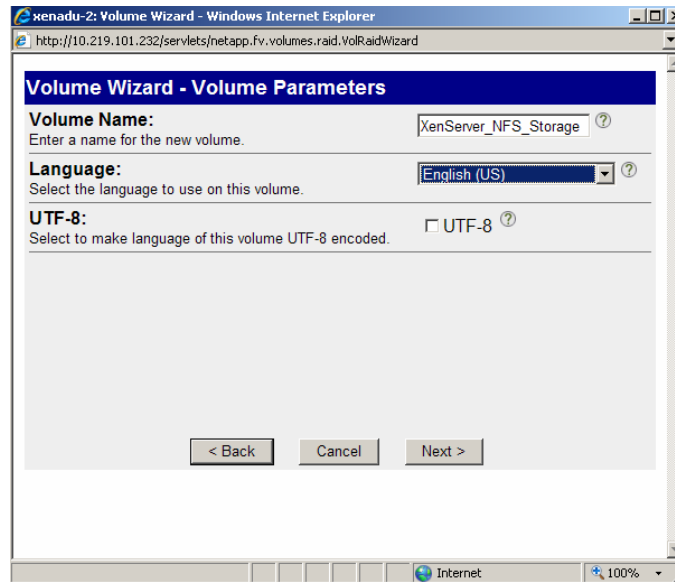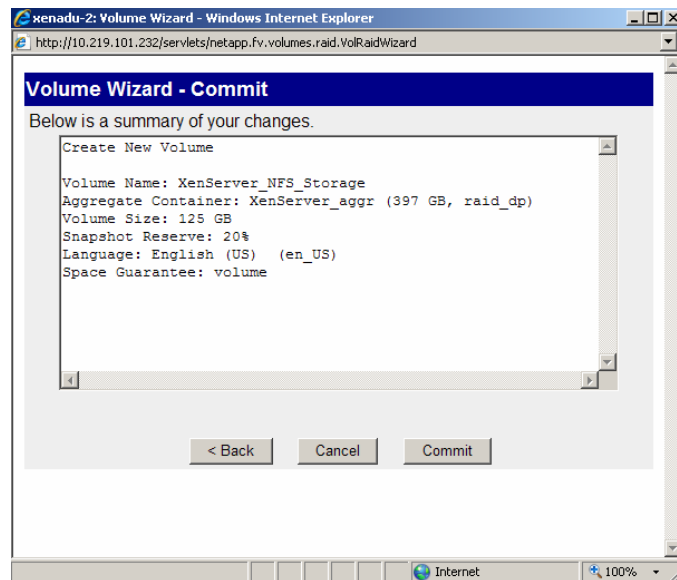

5. Click *Next* to choose the appropriate aggregate. For storage saving, you can select "None" for Space Guarantee. Click *Next*.
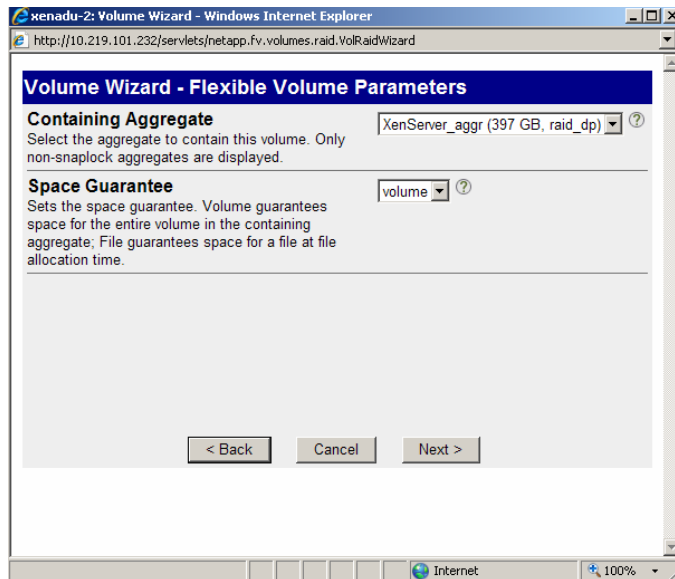
6. Set the size required for the NFS SR. If snapshot is not required, you can set 0% snapshot reserve to save space.
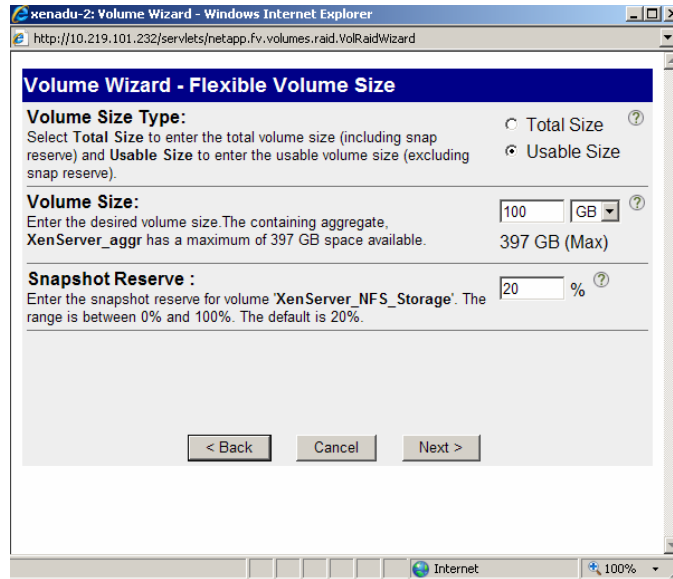

7. Click *Next*, and then *Commit* to create the FlexVol.

8. To create an NFS export on the above volume, select *NFS->Manage Exports*. Click on the above created volume to modify the options associated with the export. For the XenServer host to be able to create SRs on the exported NFS target, the host's IPAddress or subnet mask needs to be granted *Root Access*. Select *Root Access* from the export wizard.
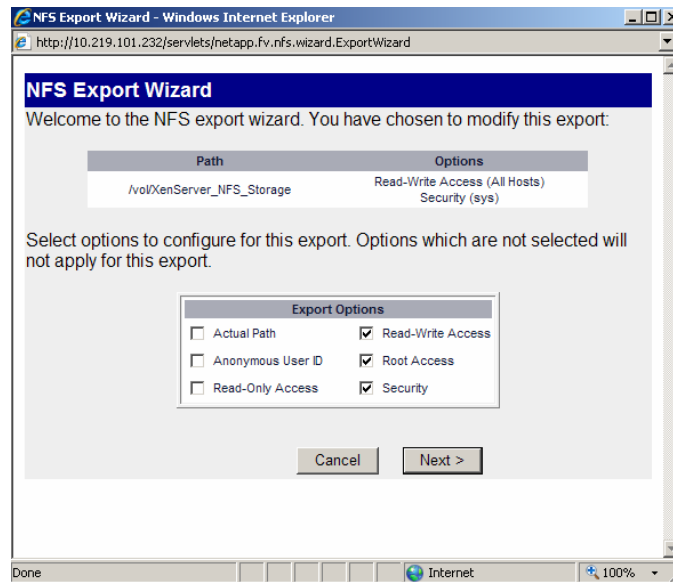
9. For the *Root Hosts* page, click *Add….* If all hosts on a particular subnet should be given access to the NFS storage repository, then enter the subnet. Else, enter the individual host names (or IP addresses) separated by a comma. In the case of subnet, the format is *dotted_ip/num_bits*, where *dotted_ip* is the IP subnet and *num_bits* is the number of leading 1 bits on the subnet mask. For the example in the screen shot below, the XenServer host is in the 192.168.2.0 subnet – so the entry is 192.168.2.0/8s.
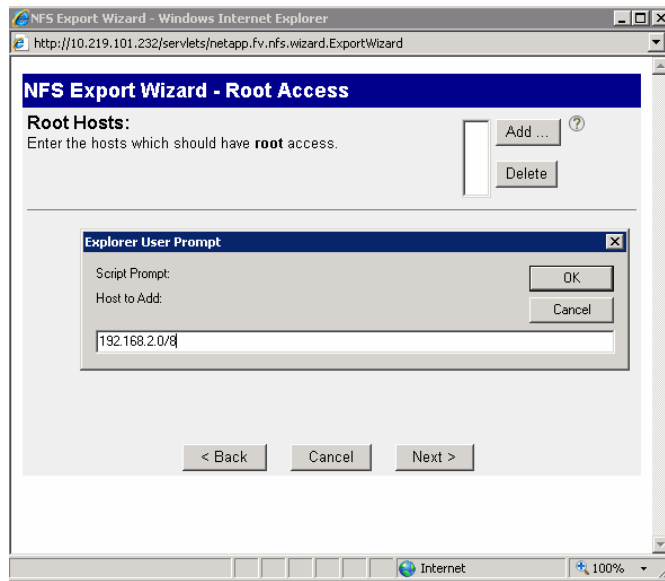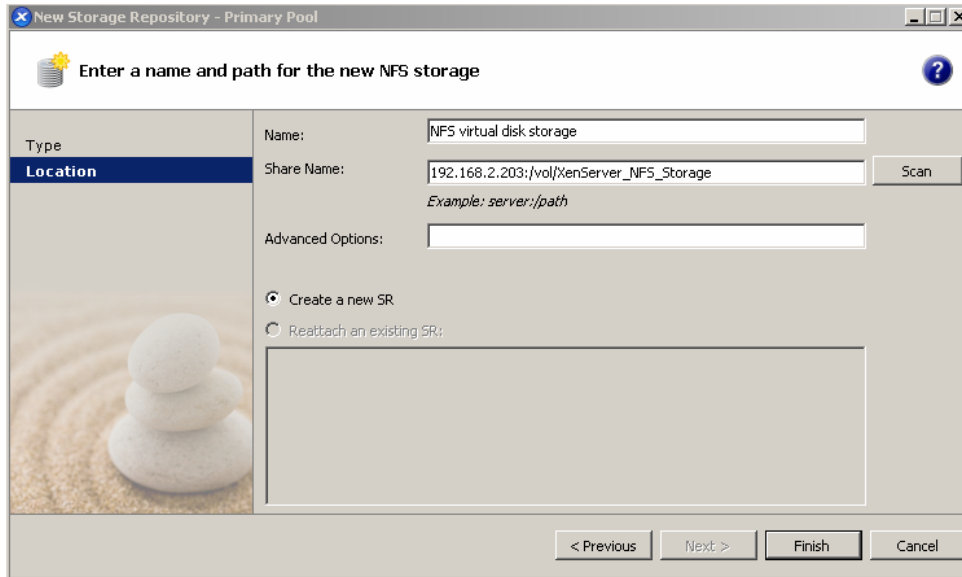

In the following wizard panes, set the *Security* to *Unix Style*, and *Commit* the changes in the end.

10. In XenCenter, connect to the XenServer host, and choose *New Storage.* Select *NFS*.
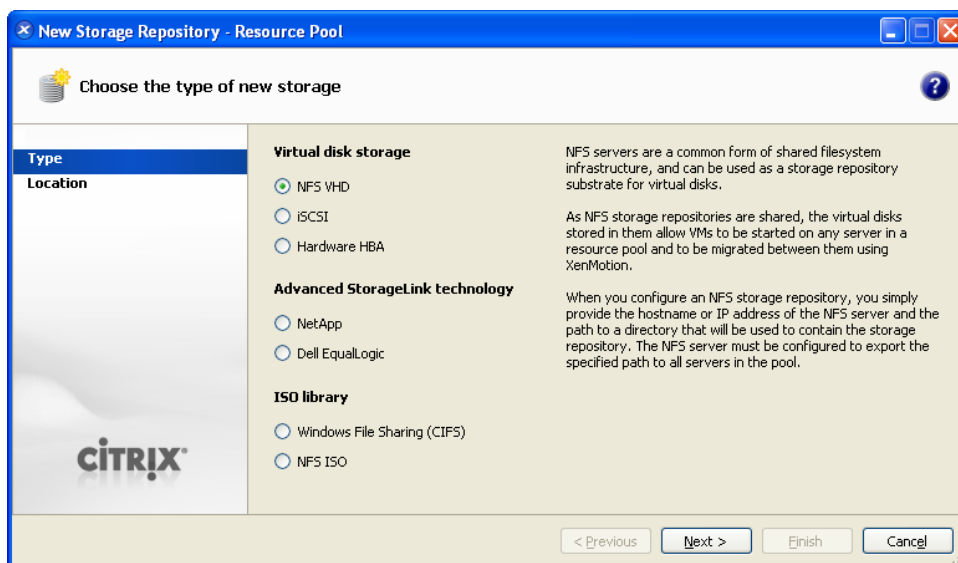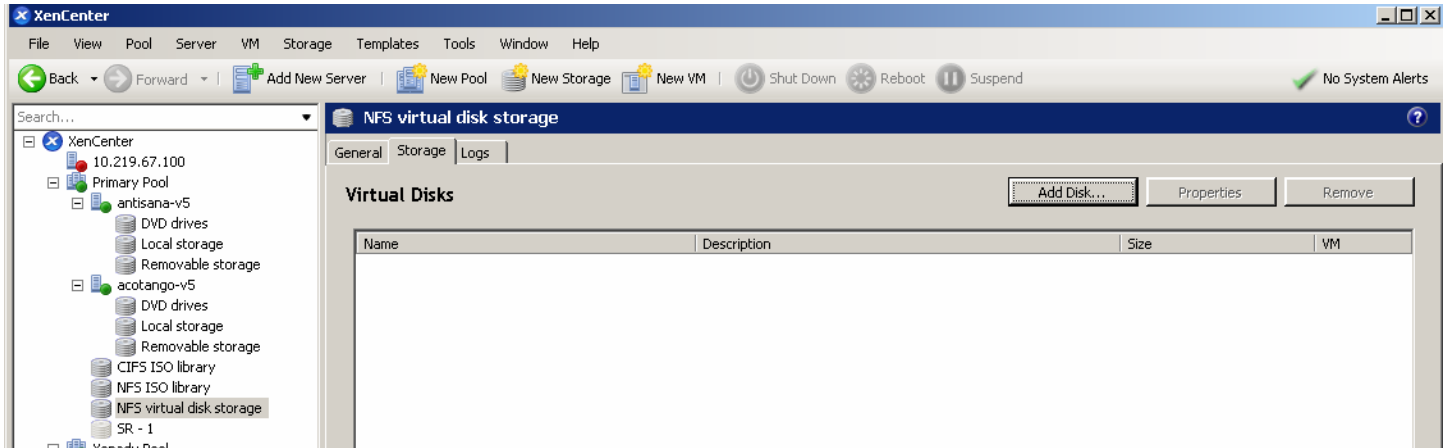
11. Give the NFS share a name, and set the path to the NFS export point from the device and click *Finish*.

By default, the VDI that is created in the NFS SR is thin-provisioned.

Creating a Fully Provisioned VDI in the NFS SR

1. In XenCenter, click on the newly created NFS SR in the *Resources* pane, and click the *Storage* tab.

2. Click on *Add Disk…* button, and enter details for the size of VDI you want. Make sure the newly created NFS SR is highlighted, and click *Add*.
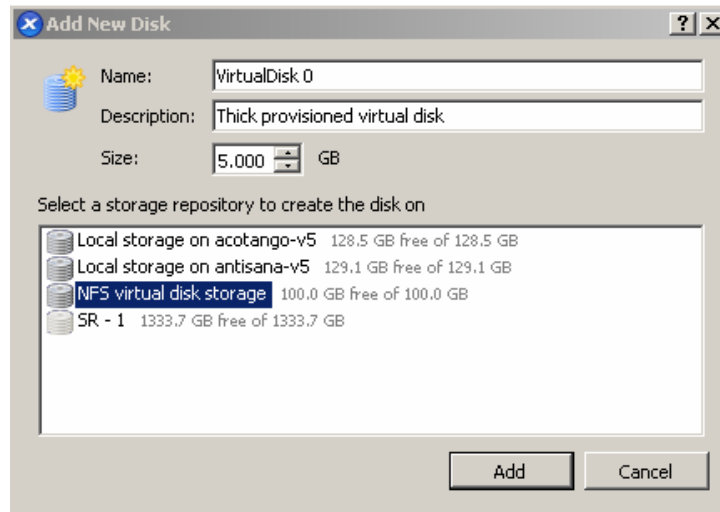
3. From a XenServer console, find the mount point for the newly created NFS SR using the *mount* command.

```
[root@antisana-v5 ~]# mount
/dev/sda1 on / type ext3 (rw)
none on /proc type proc (rw)
none on /sys type sysfs (rw)
none on /dev/pts type devpts (rw)
none on /dev/shm type tmpfs (rw)
none on /proc/sys/fs/binfmt_misc type binfmt_misc (rw)
sunrpc on /var/lib/nfs/rpc_pipefs type rpc_pipefs (rw)
192.168.2.203:/vol/XenServer_NFS_Storage/ee26e6a4-42c8-2c81-10ef-77372d39890e on
 /var/run/sr-mount/ee26e6a4-42c8-2c81-10ef-77372d39890e type nfs
(rw,soft,timeo=66,retrans=1,tcp,addr=192.168.2.203)
```

In the example above, */var/run/sr-mount/ee26e6a4-42c8-2c81-10ef-77372d39890e* is the mount point.

4. Change directory to the mount point, and run the *vhd-util* command to write zeroes into the .vhd file that represents the newly created VDI that needs to be fully provisioned

```
vhd-util
```

usage: vhd-util COMMAND [OPTIONS]
COMMAND := { create | snapshot | query | read | set | repair | resize | fill | coalesce | modify | scan | check }
In the example VDI above, the command is
[root@antisana-v5 ee26e6a4-42c8-2c81-10ef-77372d39890e]# vhd-util fill -n 8f78b8a5-d560-4004-b6e7-caec8a6773a8.vhd
You can now attach the VDI to a virtual machine as a fully provisioned virtual disk.

Configuring iSCSI Storage

To set up the NetApp storage box to be used for an iSCSI SR from XenServer, follow steps 1-7 from the previous section to create a FlexVol.   In step 4, give the FlexVol a name that can easily be identified by the device's autosupport program as a volume being used by a XenServer host, for example XenServer_iSCSI_Storage used in illustrations below. Keep the *Snapshot Reserve* to 20% if Snapshot will be done from the backend – XenServer does not support VDI snapshots for VDIs in an iSCSI SR. Otherwise, bring it down to 0% if there are concerns for space usage.

Once the FlexVol has been created, the initiator groups (igroups) need to be created. These igroups are used to provide a mechanism of host access to a LUN via LUN masking. It is recommended that a single igroup be used for a data center. To simplify the management of igroups, we suggest that the igroup name contain "XenServer", the protocol type (*iSCSI* in this case), and the data center name. This also helps the XenServer administrator to choose the correct igroup LUNs for the pool of hosts in the data center. The igroup itself contains iSCSI qualified names or IQNs that match the IQN of each XenServer host in the pool.

The LUN wizard from the NetApp FilerView can be used to create LUN that will be used by the iSCSI SR.

1. Select *Wizards*, and then *LUN Wizard* from the FilerView.

2. Enter the path for the LUN for the above created volume as /vol/XenServer_iSCSI_Storage/{LUN Name}. Set the *Protocol Type* to *Xen*.

3. Click *Next>*.
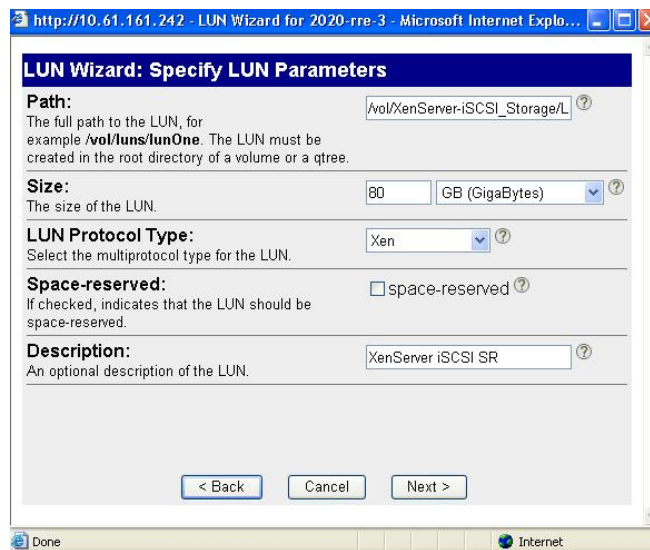
4. To enable LUN masking, the LUN must be mapped to the igroup. Click *Add Group* to either create a new igroup or assign an already existing igroup. Select *Create new initiator group* and continue.
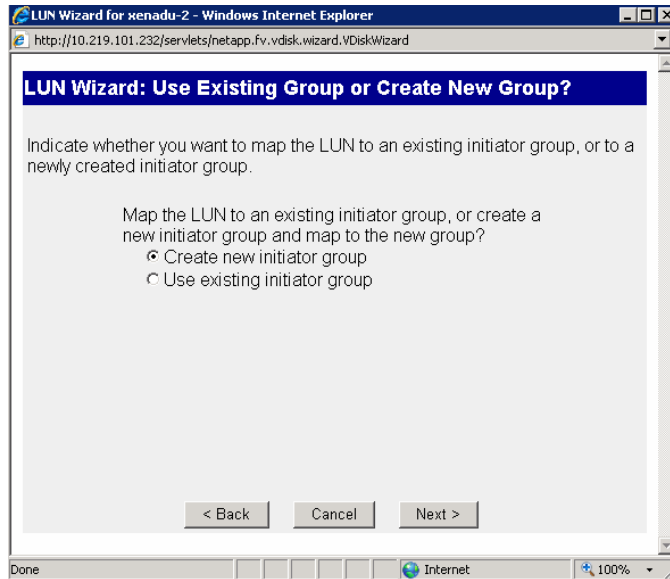

5. Give a name that uniquely identifies the igroup to be used by XenServer hosts in the data center. Select the initiator group type as *iSCSI*, and the operating system as *Xen*.
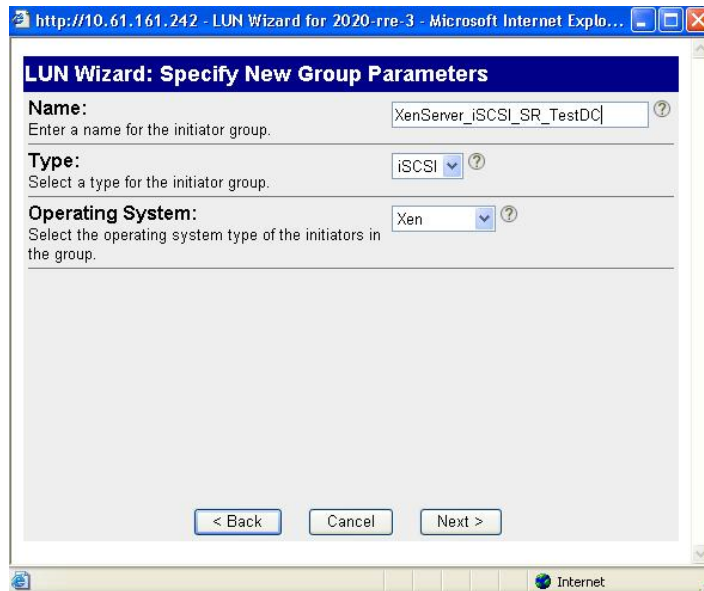
6. The next steps need IQNs from the XenServer host. The IQN of a XenServer host can be seen from the *General* tab for the host in XenCenter (in the screen shot below, the IQN for XenServerTest host is "iqn.2008-01.com.example:44b80d2a"). To change the IQN to a more recognizable name, click on the *Properties* button at the top right of the *General* tab, and modify the *iSCSI IQN* field (see the screen shot below where IQN is changed to iqn.2008-01.com.example:testdc).
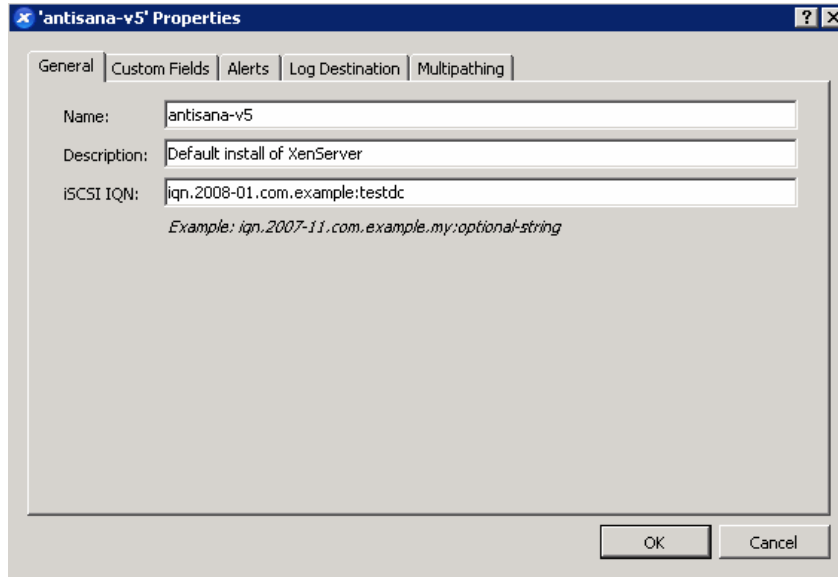

7. Once you have set the appropriate IQN on the XenServer host, it needs to be added to the igroup in the LUN wizard from the NetApp FilerView. Click *Add Initiator*, and enter the IQN from the XenServer host here, or chose from known list of iSCSI node names if they already exist.
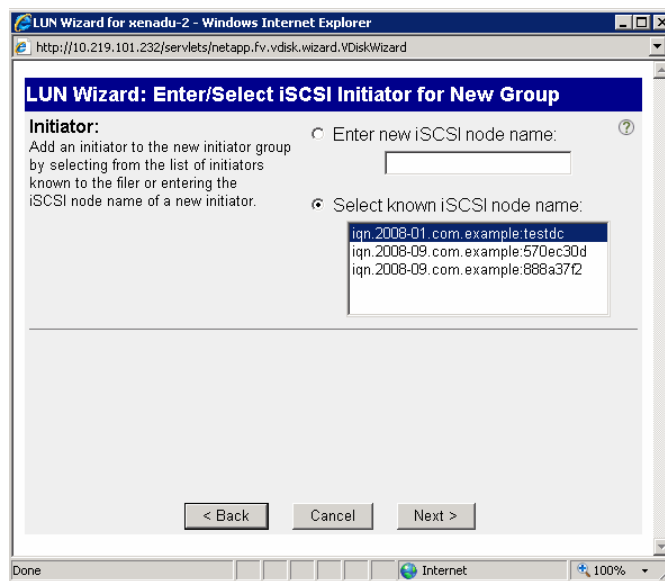
8. Repeat step 6 so that IQNs from all XenServer hosts in the pool are added to the igroup.

9. Continuing on, the LUN Wizard will ask for a LUN ID that the newly created LUN should be mapped to in the igroup. This can be left blank so that lowest available LUN ID used.
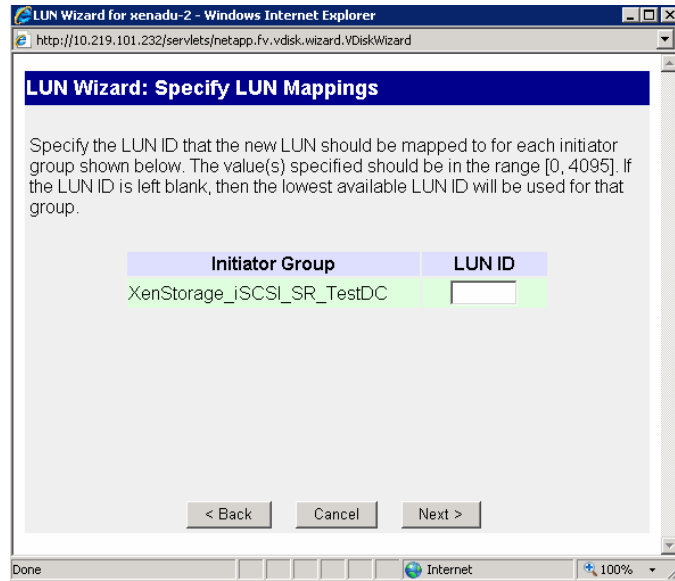

10. Commit the changes from the wizard, and switch to the XenCenter GUI to start creating an iSCSI based storage repository on the newly created LUN.

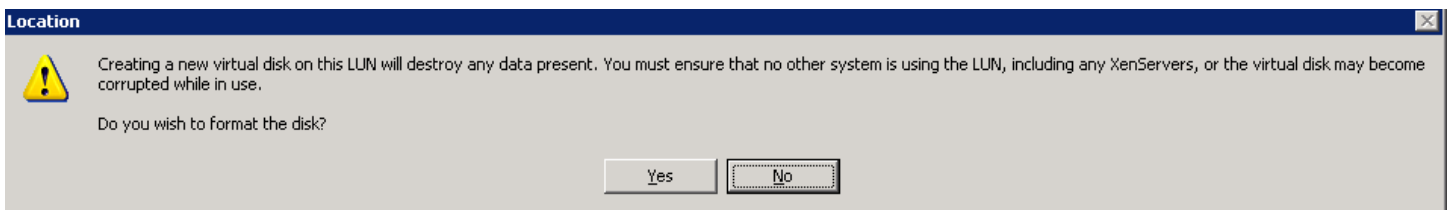## Creating an iSCSI SR Using the iSCSI Software Initiator

1. To create the storage repository, in XenCenter, choose *New Storage*. Select *iSCSI*.
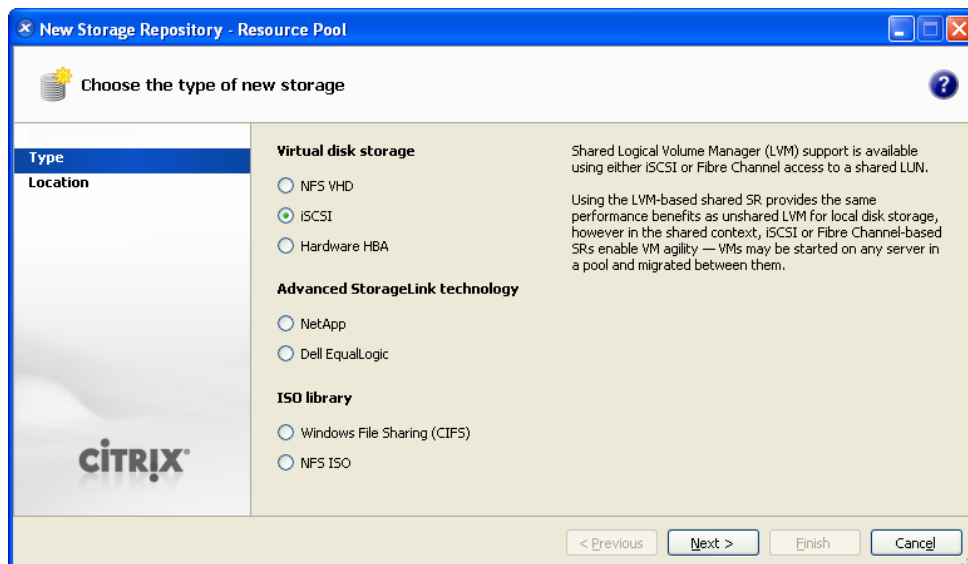
2. Enter the *Target Host* as the hostname or IP address of the NetApp device in which the LUN was set up in steps 1-9. If *CHAP Authentication* was set up for iSCSI security on the device, then enter the *CHAP User* and *CHAP Secret*. It is recommended to have the same CHAP username/password for initiators in the same igroup (as is the case with a pool of XenServer hosts connecting to the same igroup). Most of the time, customers typically do not enable security for iSCSI unless there is a security requirement. If you are required to enable security for the iSCSI connection, then we recommend that you utilize the CHAP option. Click on *Discover IQNs* to get the list of *Target IQN*s on the NetApp device. Select the relevant *Target IQN* and click on *Discover LUNs* to get a list of LUNs associated with the mapped igroup and LUNs in it. From the initial steps, the LUN that was created should show up in this list. Select the LUN, and click *Finish*.
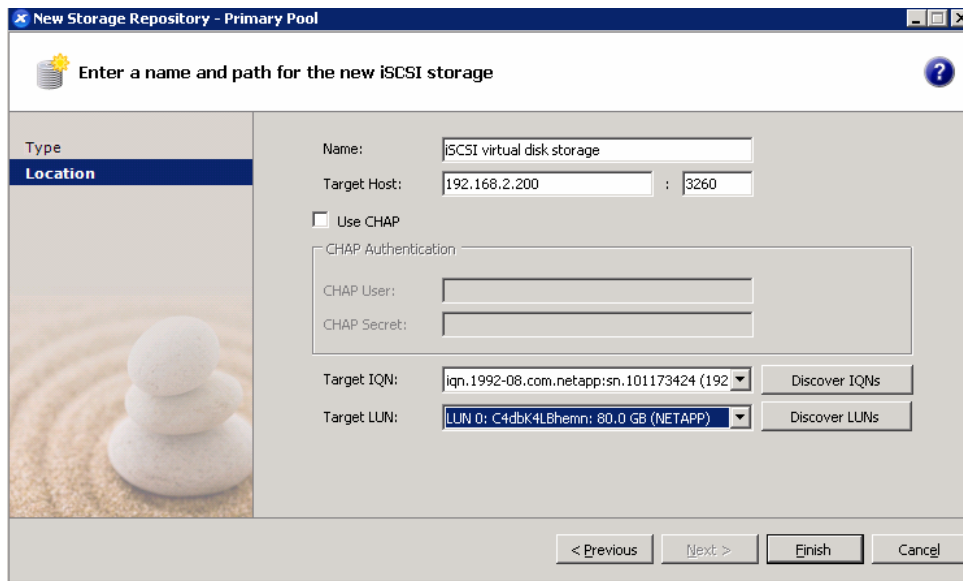
3. The new LUN will be overlaid with LVM, and XenCenter will ask the LUN to be formatted as such. Click *Yes* on the pop-up for the LUN to be formatted with LVM.

Creating an iSCSI SR Using an iSCSI HBA

For using an iSCSI HBA to create the iSCSI SR, the CLI from the control domain needs to be used. Depending on what HBA is being used, the initiator IQN for the HBA needs to be configured. Given the type of HBA used, the documentation for that HBA should be consulted to configure the IQN.

Once the IQN has been configured for the HBA, use the NetApp FilerView to create a new LUN as in steps 1-9 at the beginning of this section that covers LUN creation. However, instead of using the XenServer's IQN, specify the IQN of the various ports of the HBA. Do this for every XenServer host in the pool.

Two HBA CLI's are included in the XenServer host to configure the HBA:

> Emulex: /usr/sbin/hbaanyware
>
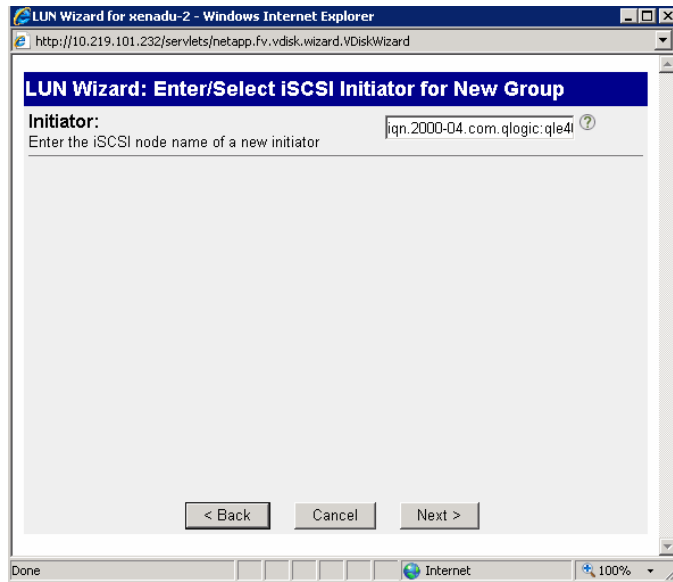> QLogic iSCSI: /opt/QLogic_Corporation/SANsurferiCLI/iscli

For the purposes of an example, this guide illustrates how the QLogic iSCSI HBA CLI *iscli* can be used.

1. Get the IQN of the ports of the iSCSI HBA from the output of *iscli* command, and add them as initiators to the initiator group being used in the LUN creation wizard.

2. After creating the LUN, set the IP address for the HBA. In the control domain, use the *iscli* CLI to do so. Choose option 4, then option 2 to enter *Port Network Settings Menu*. Enter option 4 to *Select HBA Port* and then option 2 to *Configure IP Settings*.

3. Now add a persistent target to the HBA. The target iSCSI IQN can be retrieved from the NetApp FilerView by clicking *LUNs->Initiator Groups->iSCSI->Manage Names*.
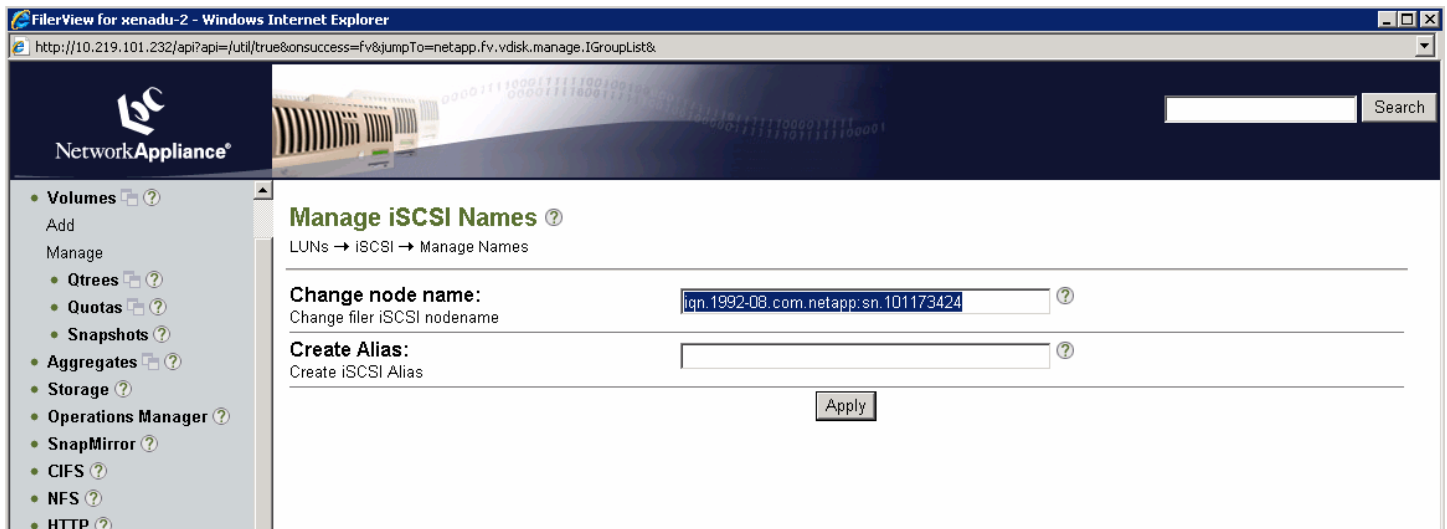

```
/opt/QLogic_Corporation/SANsurferiCLI/iscli –pa 0 <iSCSI device IP Address>
-INAME <target iSCSI IQN>
```
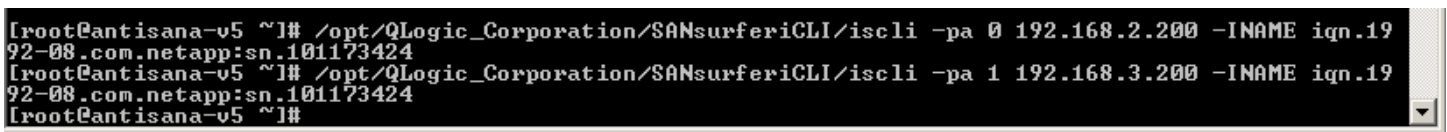

Note that the above command was run for 2 ports of the iSCSI HBA, each port connecting to a different subnet (as indicated in the multipathing configuration at the beginning of the XenServer Storage Configuration section). This command will assign the port the specified iSCSI IP address and specific target IQN.

4. Use the *xe sr-probe* command to force a scan of HBAs installed on the system and detect the new LUN zoned to the host. It will return the list of properties for each LUN found. One of these properties will be *<path>* which is the global device path of the HBA LUN. Specify the host-uuid of the system from where the *xe sr-probe* command is run.

```
xe sr-probe type=lvmohba host-uuid=<UUID of host>
```


To validate that the device path is for the newly created LUN on the device, match the serial number from the *<serial>* field of the *xe sr-probe* output with the serial number of the LUN in FilerView. Note that there are 2 paths to the LUN indicating that multipathing is active.

To determine the LUN serial number from the NetApp FilerView, click *LUNs->Manage* and click on the newly created LUN. Then note the *Serial Number* field.

5. Repeat the above steps for all hosts in the pool.

6. To create the iSCSI SR over HBA on this LUN, in XenCenter click *New Storage* and select *Hardware HBA* option. Click *Next*.
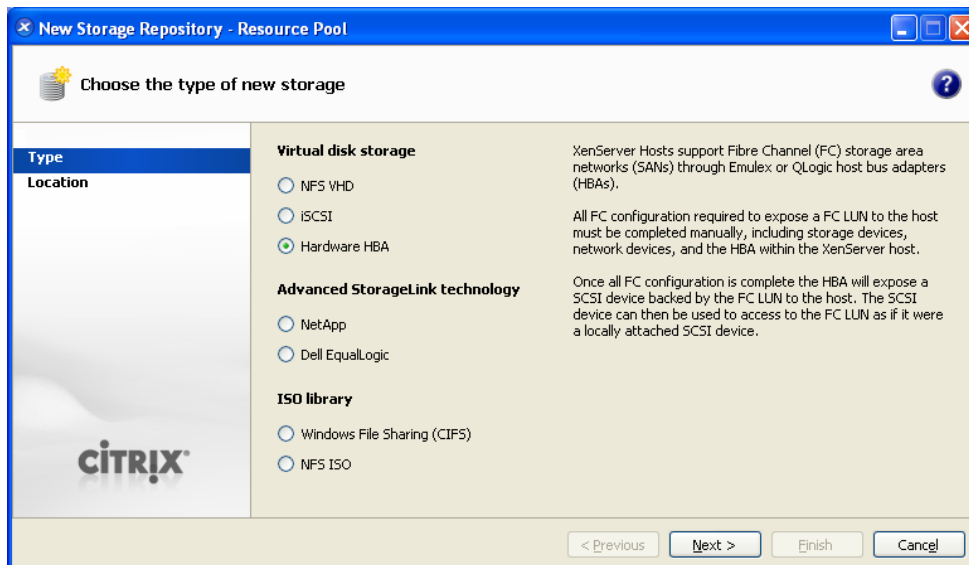
There is short delay while XenServer probes for available LUNs.

7. Select the appropriate LUN. Give the SR an appropriate *Name* and click *Finish*.
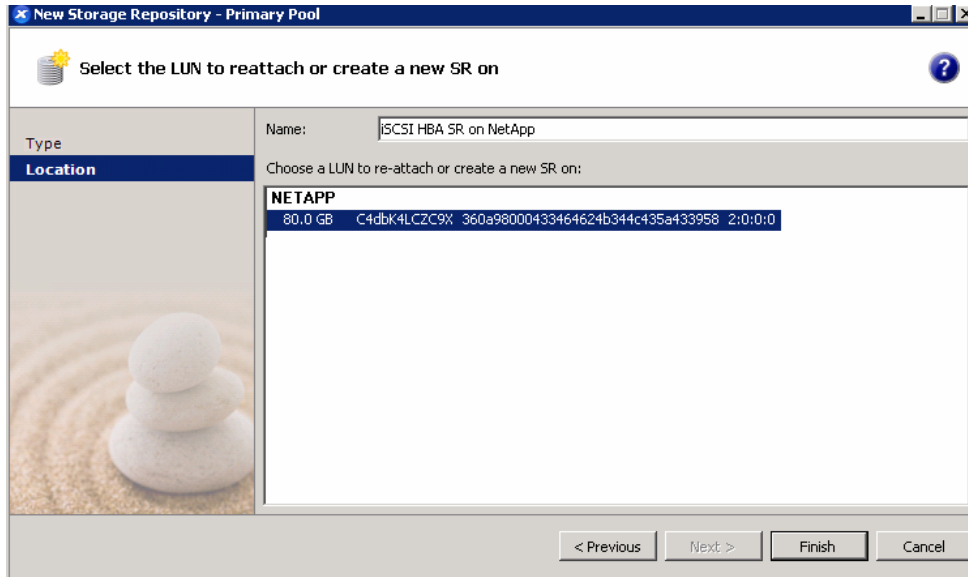
8. A warning is displayed that the LUN will be formatted and any data present will be destroyed. Click *Yes* to format the disk.
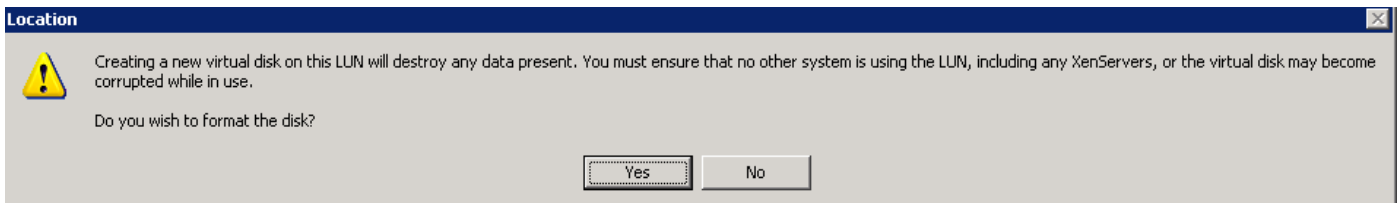
Note that multipathing is active by clicking on the newly created SR in the *Resources* pane, and then the *General* tab.
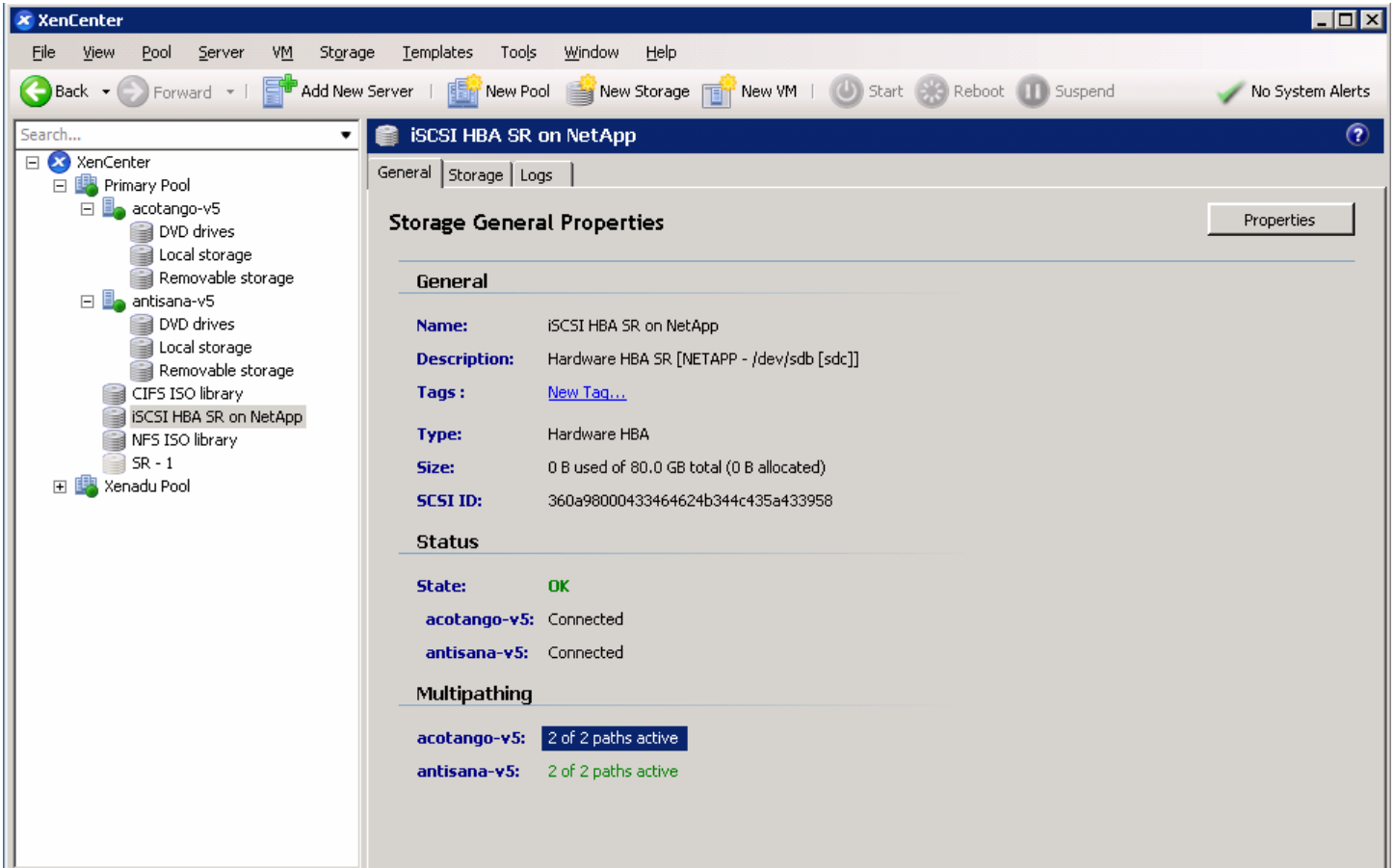
Configuring Fibre Channel Storage

To set up the NetApp storage box to be used for a Fibre Channel (FC) SR from XenServer, follow steps 1-7 from the previous section to create a FlexVol.   In step 4, give the FlexVol a name that can easily be identified by the device's autosupport program as a volume being used by a XenServer host, for example XenServer_FCP_Storage used in illustrations below. Keep the *Snapshot Reserve* to 20% if Snapshot will be done from the backend – XenServer does not support VDI snapshots for VDIs in a FCP SR. Otherwise, bring it down to 0% if there are concerns for space usage.

Once the FlexVol has been created, the initiator groups (igroups) need to be created. These igroups are used to provide a mechanism of host access to a LUN via LUN masking. It is recommended that a single igroup be used for a data center. To simplify the management of igroups, it is suggested that the igroup name contain "XenServer", the protocol type (*FCP* in this case), and the data center name. This also helps the XenServer administrator to choose the correct igroup LUNs for the pool of hosts in the data center. The igroup itself contains World Wide Port Names (WWPNs) that match the WWPN for the HBA ports on the XenServer host.

The LUN wizard from the NetApp FilerView can be used to create LUN that will be used by the FC HBA SR.

1. Select *Wizards*, and then *LUN Wizard* from the FilerView.

2. Enter the path for the LUN for the above created volume as /vol/XenServer_FCP_Storage/{LUN Name}. Set the *Protocol Type* to Xen. Click *Next*.

3. To enable LUN masking, an igroup needs to be assigned to the LUN being created.   Click Add Group to either create a new igroup or assign an already existing igroup. Select Create new initiator group and continue. Give a name that uniquely identif*ies t*he igroup to be used by XenServer hosts in *the data* center. Se*lect the init*ator group type as *FCP*, and the *operating system a*s *Linux*.
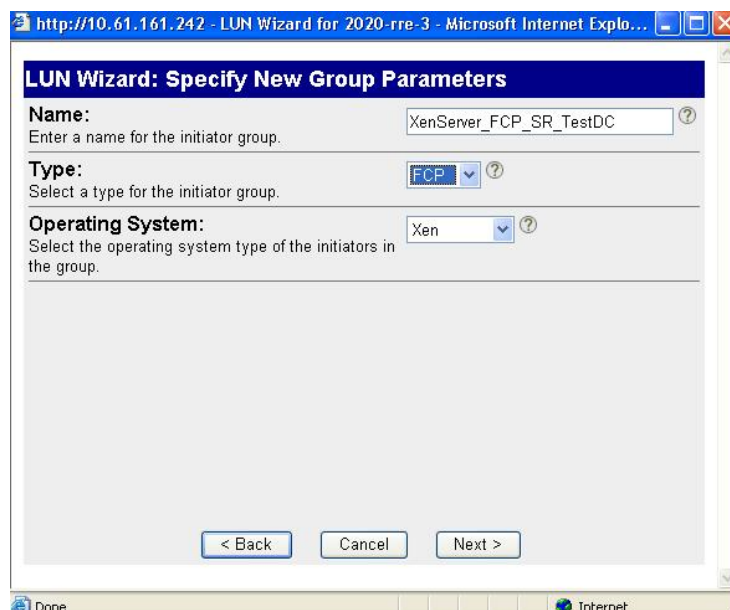
4. The next steps need WWPN for *the H*BA from the XenServer host. Since differe*nt FC ven*dors have s*pecific and d*ifferent configuratio*n* requirements, *it is recommen*de*d that th*e documentation for the specific HBA be consulted for configuration settings.

This guide will assume a QLogic 2342 HBA, and as such use the /opt/QLogic_Corporation/SANsurferCLI/scli to get configuration information. Run /opt/QLogic_Corporation/SANsurferCLI/scli in the control domain, and enter *5* from the main menu. The screen shot below shows the WWPN for the ports highlighted..

5. Once you have set the appropriate WWPN from the XenServer host, it needs to be added to the igroup in the LUN wizard from the NetApp FilerView. Click *Add Initiator*, and enter the WWPN for the FC HBA on the XenServer host here. Note that the WWPN in the wizard needs to be in xx:xx:xx:xx:xx:xx:xx:xx format where 'x' is a hexadecimal digit.
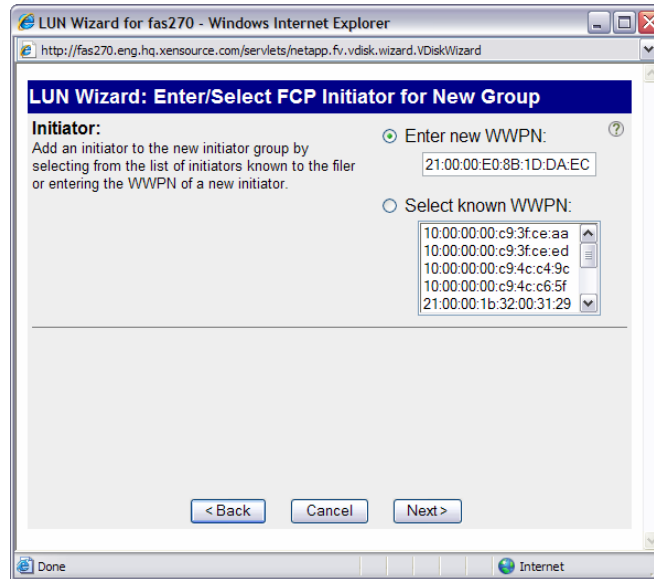
6. Continuing on, the LUN Wizard will ask for a LUN ID that the newly created LUN should be mapped to in the igroup. This can be left blank so that lowest available LUN ID used.

7. Commit the changes from the wizard.

8. The newly created LUN now needs to be zoned in to the XenServer host and will appear as a SCSI device. For this, use the *xe sr-probe* command similar to the usage as when creating an iSCSI HBA SR. The command to force a scan of HBAs installed on the system and detect the new LUN zoned to the host. It will return the list of properties for each LUN found. One of these properties will be *<path>* which is the global device path of the HBA LUN. Specify the host-uuid of the system from where the *xe sr-probe* command is run.

```
xe sr-probe type=<SR type> host-uuid=<UUID of host>
```
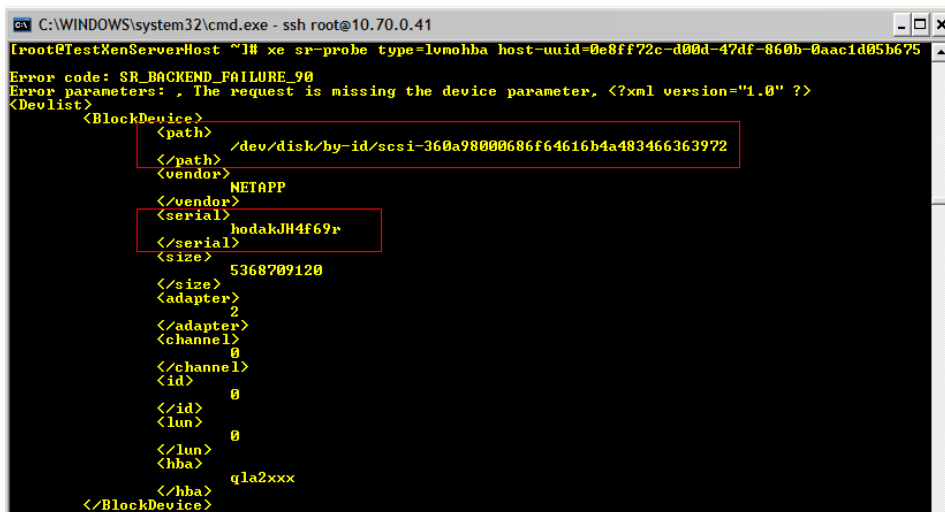
9. To validate that the device path is for the newly created zoned in LUN on the device, match the serial number from the *<serial>* field of the *xe sr-probe* output with the serial number of the LUN in FilerView. To determine the LUN serial number from the NetApp FilerView, click *LUNs->Manage* and click on the newly created LUN. Then look at the *Serial Number* field.

10. Repeat the above steps for all hosts in the pool.

11. To create the FC SR over HBA on this LUN, follow the same steps used to create the iSCSI SR over HBA.

Fixing Misaligned Windows Guest Operating System

Disks use geometry to identify themselves and their characteristics to the upper layer operating system. The upper layer operating system uses the disk geometry information to calculate the size of the disk and partition the disk into predetermined addressable blocks. Just as with physical disks, logical disks (LUNs) report disk geometry to the host so that it can calculate space and partition the LUN into addressable blocks.

NetApp uses 4KB blocks (4 x 1,024 = 4,096 bytes) as its basic storage building block. Writes can consume no less than a single 4KB block and can consume many 4KB blocks depending on the size of the write operation. Files that are smaller than 4KB are actually stored in the inode that contains their metadata. When a LUN is created in Data ONTAP, a certain amount of space is carved out of the disk in the form of 4KB blocks. These are then reported to the host OS with a specified geometry, depending on the LUN type selected. The host OS then takes the presented LUN and partitions it appropriately.

The problem of misaligned LUN I/O occurs when the partitioning scheme used by the host OS does not match the block boundaries inside the LUN.

Note: This problem is not peculiar to NetApp. All SAN vendors have the potential to experience misalignment issues. Misalignment issues can be caused because the partition size of the blocks in the host OS use something other than 4KB blocks. Misalignment issues can be caused if the host OS imposes an offset that has not been compensated for. NetApp compensates for offsets by identifying the OS and then adjusting the offset. This is done during LUN creation when the user enters the LUN type.

For a deeper discussion of disk geometry, consider attending the NGS SAN Fundamentals class or reading ntapcs7976: "FCP SCSI Geometry FAQs."
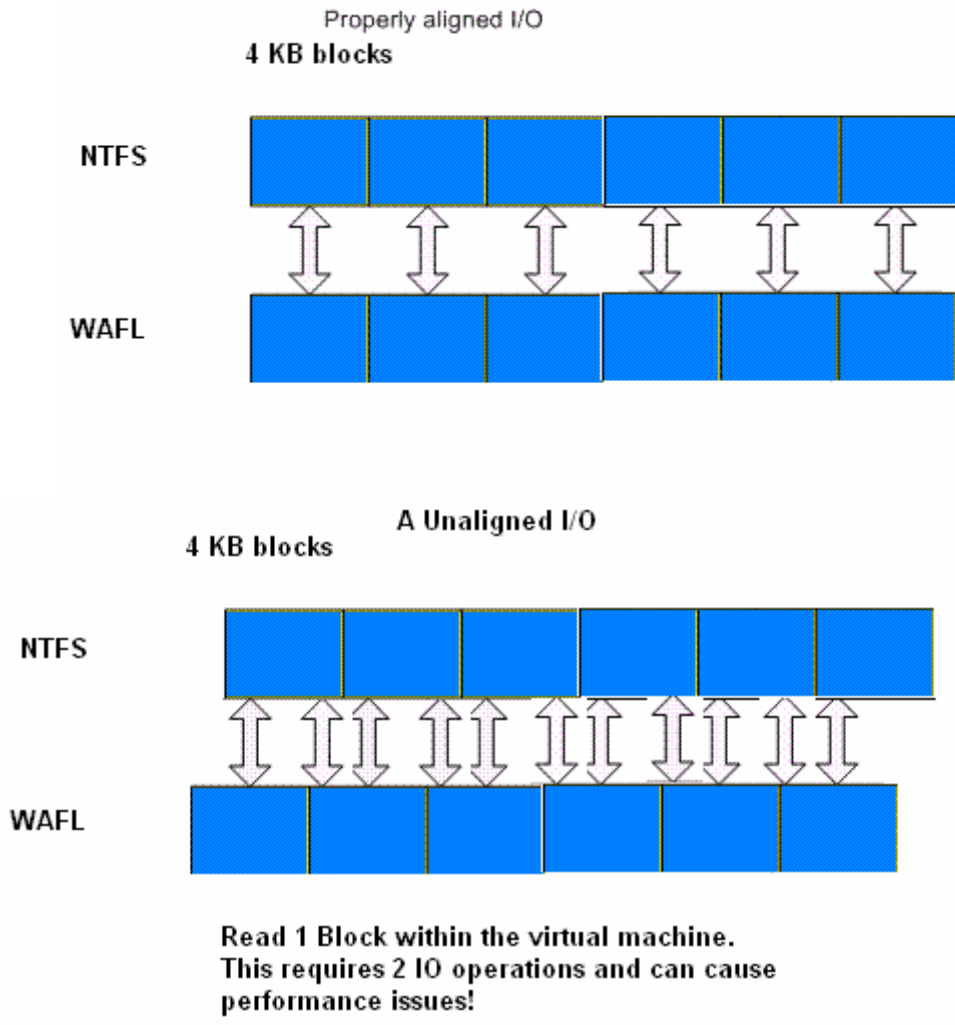
Figure 9 Properly aligned and unaligned I/O.

When aligning the partitions of virtual disks for use with NetApp FAS systems, the starting partition offset must be divisible by 4,096. The recommended starting offset value for Windows 2000, 2003, and XP operating systems is 32,768. Windows 2008 and Vista default at 1,048,576 and do not require any adjustments.

To verify this value, we need to run msinfo32.exe from the Windows command console. And you will typically find that the VM is running with a default starting offset value of 32,256. To run msinfor32, you select start > All Programs > Accessories > System Tools > System Information. Notice the partition starting offset is 32,256 bytes in Figure 10. This indicates disk misalignment.
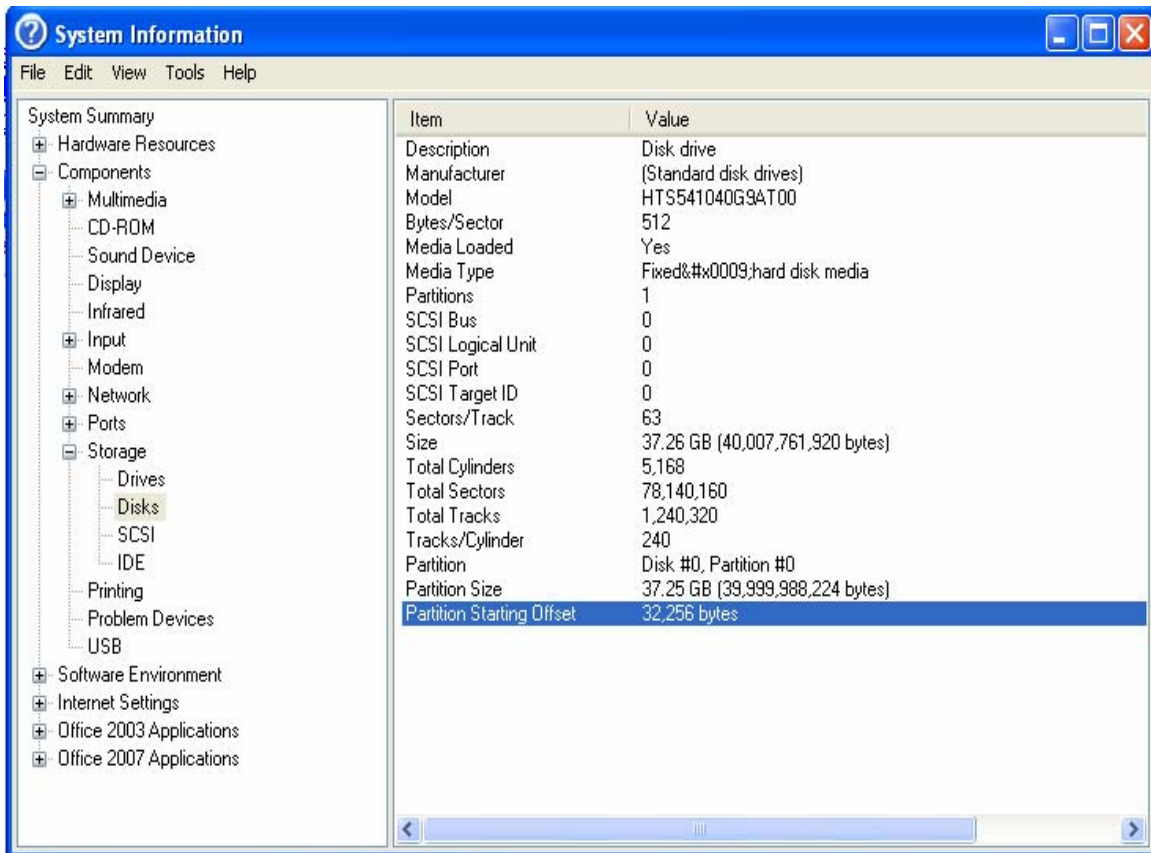
Figure 10 Using system information to identify the starting partition offset.

Resolution

Virtual disks can be formatted with the correct offset at the time of creation by simply booting the VM before installing an operating system and manually setting the partition offset. For Windows guest operating systems, consider using the Windows Preinstall Environment boot CD or alternative tools such as Bart's PE CD. To set up the starting offset, follow these steps:

1.	Boot the VM with the WinPE CD.
2.	Select Start > Run and enter DISKPART..
3.	Enter Select Disk0.
4.	Enter Create Partition Primary Align=32.
5.	Reboot the VM with WinPE CD.
6.	Install the operating system as normal.


```
C:\>diskpart

Microsoft DiskPart version 5.1.3565

Copyright (C) 1999-2003 Microsoft Corporation.
On computer: VSTEWART01-LXP

DISKPART> select disk 0

Disk 0 is now the selected disk.

DISKPART> create partition primary align=32_
```
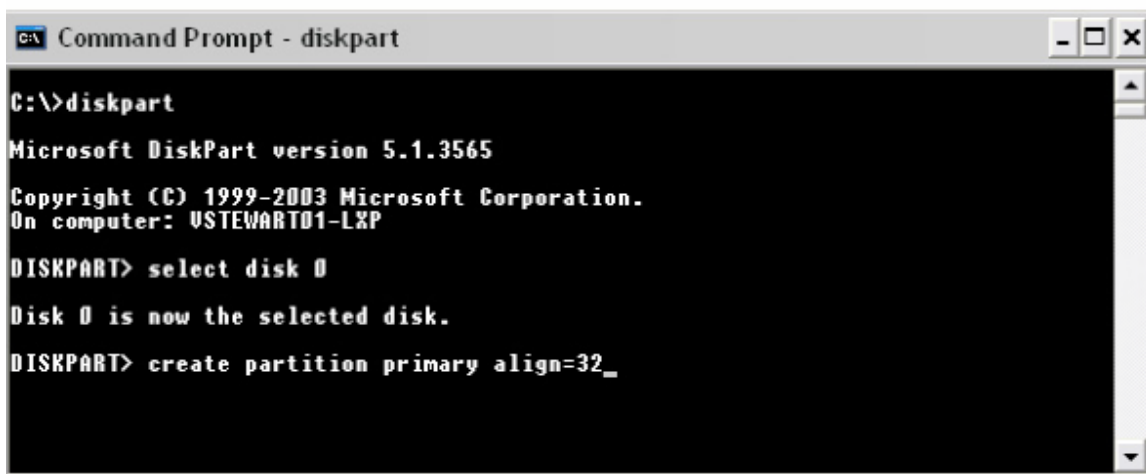
Figure 11 Running diskpart to set a proper starting partition offset.

Before running *diskpart.exe* for Windows guests using the NFS SR, run the v*hd-util* to fully provision the VHD VDI, as detailed in the section describing the configuration of the NFS SR.

For more information on misalignment, read Best Practice for File System Alignment in Virtual Environment.

Backup and Recovery

There are various elements for backup and recovery:

- Snapshot the virtual machines time to time,

- Backup VM metadata

- Backup the snapshots and metadata, and

- Recover data/virtual machines from snapshots and metadata when needed.

Snapshot for Virtual Machines on NetApp Data ONTAP and NFS Storage Repositories

Creating VM snapshots for those VMs whose VDIs are resident on the NetApp Data ONTAP SR utilizes the NetApp device Snapshot technology directly by invoking the Snapshot at the device level. This results in minimal resource usage on the XenServer host in terms of CPU and memory during the snapshot process.

The snapshot process for VMs using the standard NFS, iSCSI, and Hardware HBA SRs, however, do not invoke any NetApp device capability. It uses the VHD capability for its VDIs to allow chaining for the original and snapshot VDI to share common data. The original VDI proceeds to make its own changes in an isolated copy-on-write version, with the snapshot VDI being Read Only.

XenServer 5 provides a convenient snapshot mechanism that can take a snapshot of a VM's storage and metadata at a given time. Where necessary, IO is temporarily halted while the snapshot is being taken to ensure that a self consistent disk image can be captured.

Snapshot operations result in a snapshot VM that is similar to a template. The VM snapshot contains all the storage information and VM configuration, including attached VIFs, allowing them to be exported and restored for backup purposes.

The snapshot operation is a 2 step process:

- Capturing metadata as a template.

- Issuing a VM snapshot.

Two types of VM snapshots are supported: regular and quiesced.

Regular Snapshots

Regular snapshots are crash consistent and can be performed on all VM and Storage Repository types, including Linux VMs.

1. Within XenCenter, select the VM in the left-hand pane and then the "Snapshot" tab. VMs snapshots can be created for both running and powered down VMs.
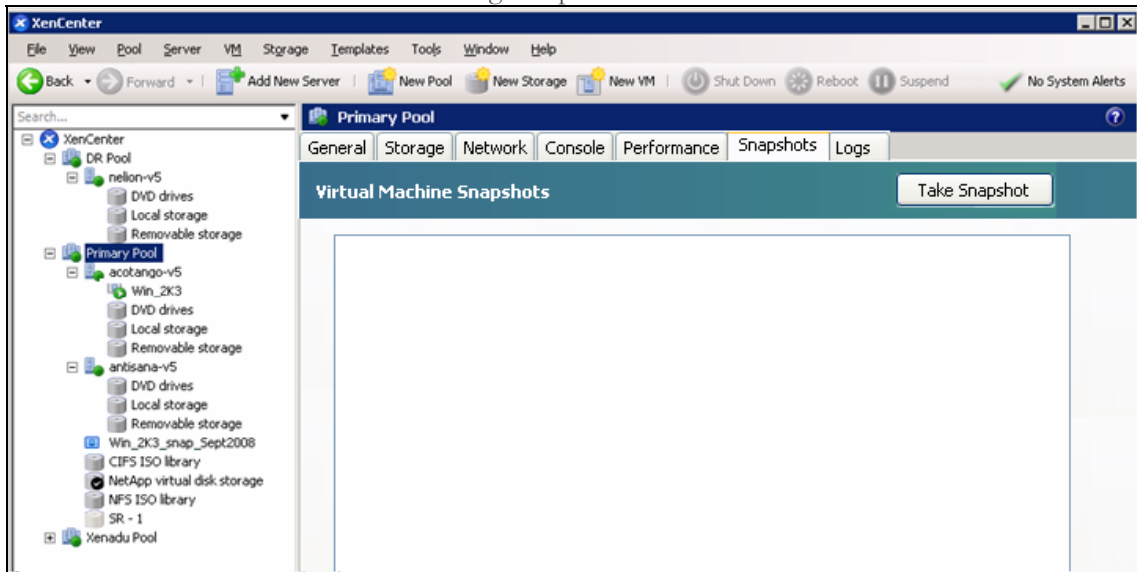
2. After clicking "Take Snapshot" and providing a name, the snapshot will appear in XenCenter
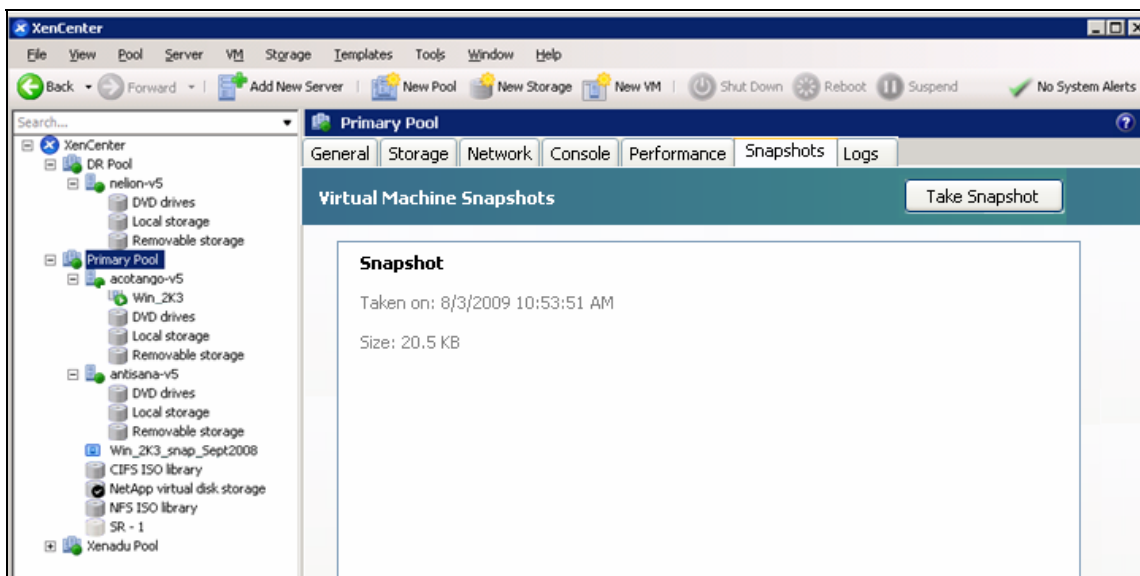
3. Snapshots can be used to restore a VM. Right click the snapshot and select "Create new VM from Snapshot." Note that the MAC address(es) will need to be manually changed to make the new VM identical to the original. Snapshots can also be used to create templates, either for creating new VMs or for backup.
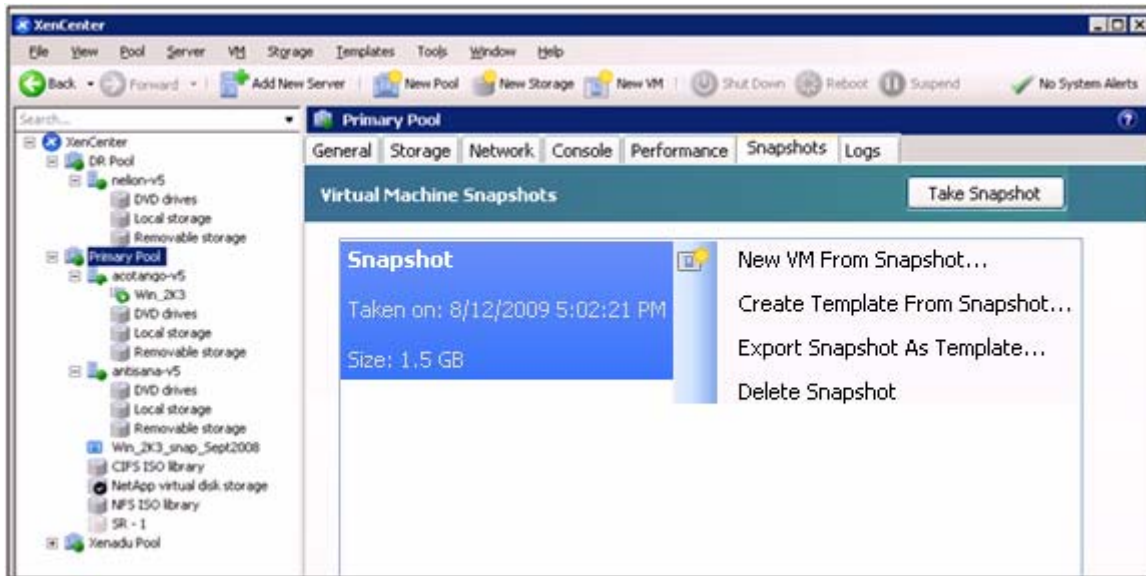
Quiesced Snapshots

Quiesced snapshots are a special case that take advantage of the Windows Volume Snapshot Service (VSS) for services that support it, so that a supported application (for example Microsoft Exchange or SQLServer) can flush data to disk and prepare for the snapshot before it is taken.

Quiesced snapshots are therefore safer to restore, but can have a greater performance impact on a system while they are being taken. They may also fail under load so more than one attempt to take the snapshot may be required.

It is essential to install the Xen VSS provider in the Windows guest in order to support VSS. This is done via the *install-XenProvider.cmd* script provided with the Windows PV drivers.

1. In the Windows VM that needs to be snapshot in quiesced mode, open a command window and change directory to where the Citrix XenTools are installed (this is by default in %ProgramFiles%\Citrix\XenTools. Run the *install-Provider.cmd*.

2. Open a console session to the XenServer master and run the *xe vm-snapshot-with-quiesce* command to snapshot the VM.

```
xe vm-snapshot-with-quiesce vm=<vm_name> new-name-label=<vm_snapshot_name>
```


3. Once a quiesced snapshot has been created, it will appear in the VM's "snapshot" tab within XenCenter.

Backing up VM Metadata Information on NetApp Device

1. Open an *xsconsole* session to the XenServer master from XenCenter (if in a shell prompt, type *xsconsole* and press *<Enter>*).
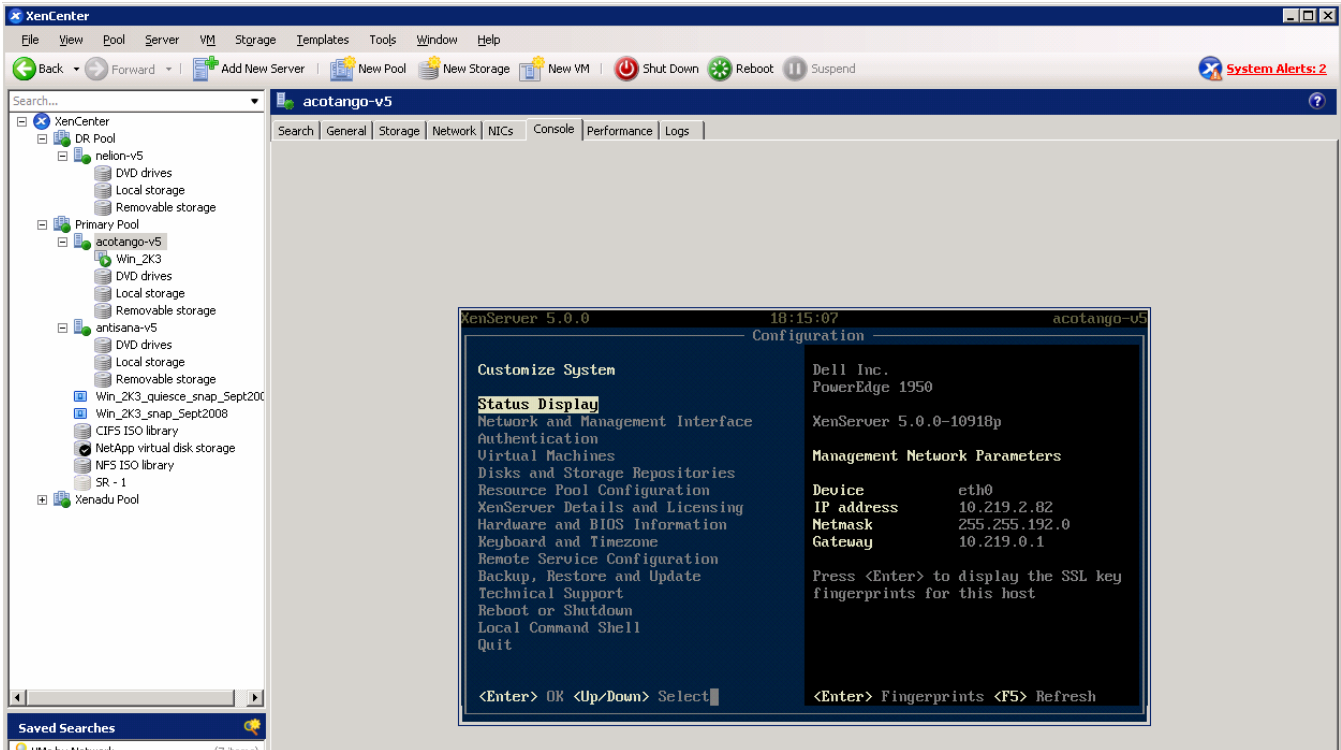
2. Scroll to the *Backup, Restore and Update* option and hit *Enter*. Choose the *Backup Virtual Machine Metadata* option and press *<Enter>*.
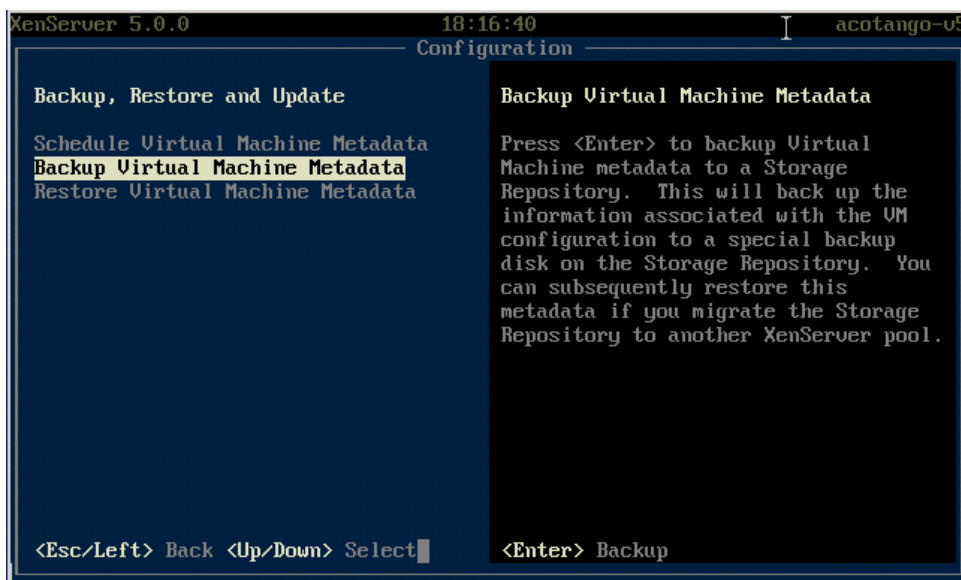
3. Enter the login credentials for the XenServer host if prompted.

4. Select the NetApp SR, and press *<Enter>*.
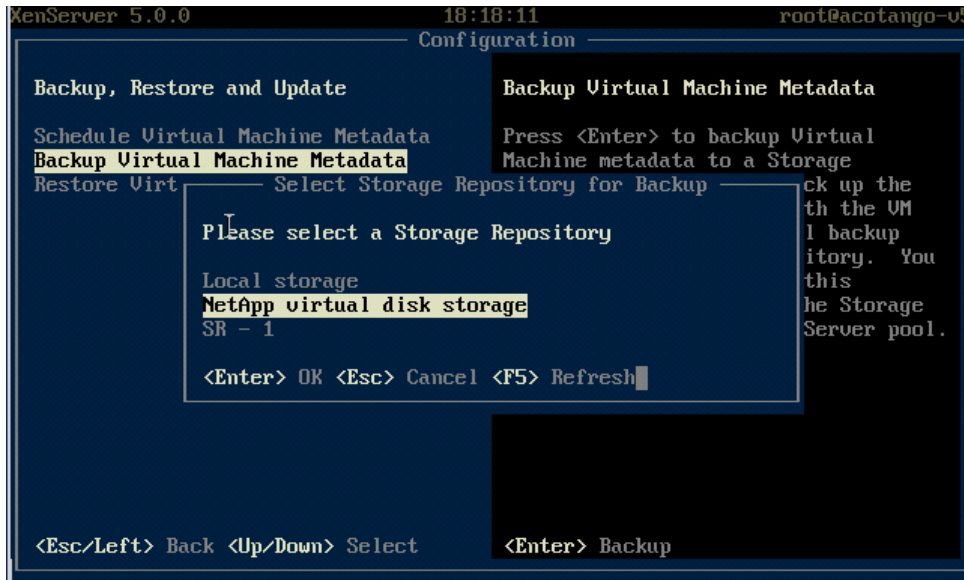
There will be a short delay while the VM metadata is backed up to the NetApp device

5. A success dialog box shows up indicating that the backup of the metadata was successful
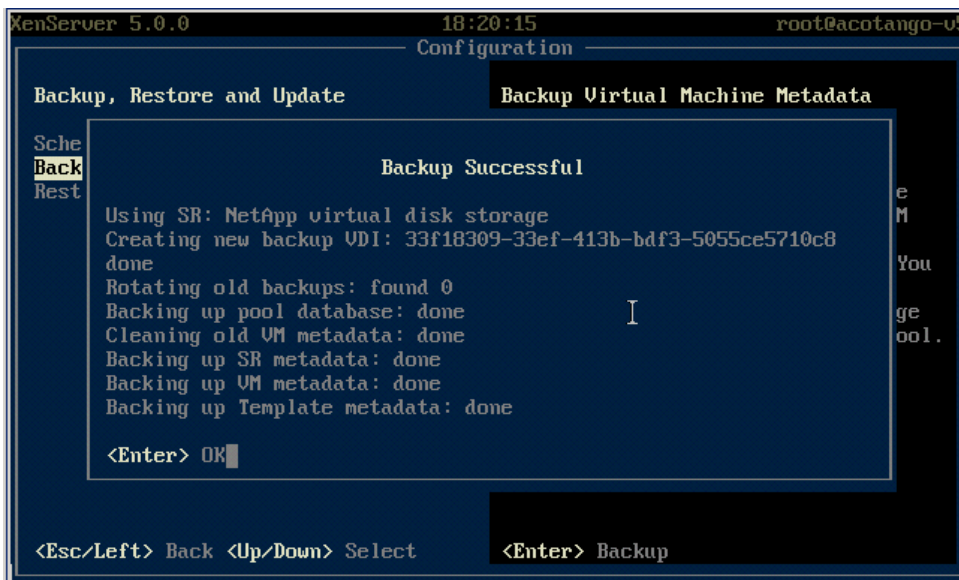
Backing up Storage Repositories on NetApp Device

Utilize NetApp's SnapMirror® technology to backup the FlexVols that make up the SR. For more information on use of SnapMirror, please refer to the SnapMirror Administration guide at http://now.netapp.com/NOW/knowledge/docs/ontap/rel73/pdfs/ontap/onlinebk.pdf.

Some considerations when using SnapMirror:

1. For SnapMirror volume replication, the destination storage system must use a version of Data ONTAP that is the same as or later than that of the SnapMirror source storage system. For volume SnapMirror to support replication for the purpose of disaster recovery, both the source and destination storage systems must use the same version of Data ONTAP.

2. For SnapMirror volume replication, the capacity of the destination volume must be greater than or equal to the capacity of the source volume

3. To optimize performance, stagger Snapshot copy update schedules so that SnapMirror activity does not begin or end at the exact minute a *snap sched* command operation attempts to create a Snapshot copy. If the SnapMirror feature is scheduled to perform Snapshot management at the same time as a snap sched activity, then the Snapshot management operations scheduled using the *snap sched* command might fail with syslog messages: *Skipping creation of hourly snapshot* and *Snapshot already exists*.

4. For optimum SnapMirror volume replication performance, ensure that the SnapMirror source volume and destination volume contain disks of the same size.
5. To achieve maximum space savings on the destination volume, scan the entire file system to recreate the deduplication metadata for the destination volume. Use the *sis start* -*s* command to do so. **Note:** The destination volume is accessible for read-write operations when the deduplication scan is in progress. The *sis* start command, without the -*s* option, has the potential for space savings on the destination volume is reduced because only the new data written to the volume will be scanned for deduplication.
6. For Firewall setting, allow a range of TCP ports from 10565 to 10569.

It is recommended to use the below diagram configuration as a guide to setup replication between the primary and DR site.
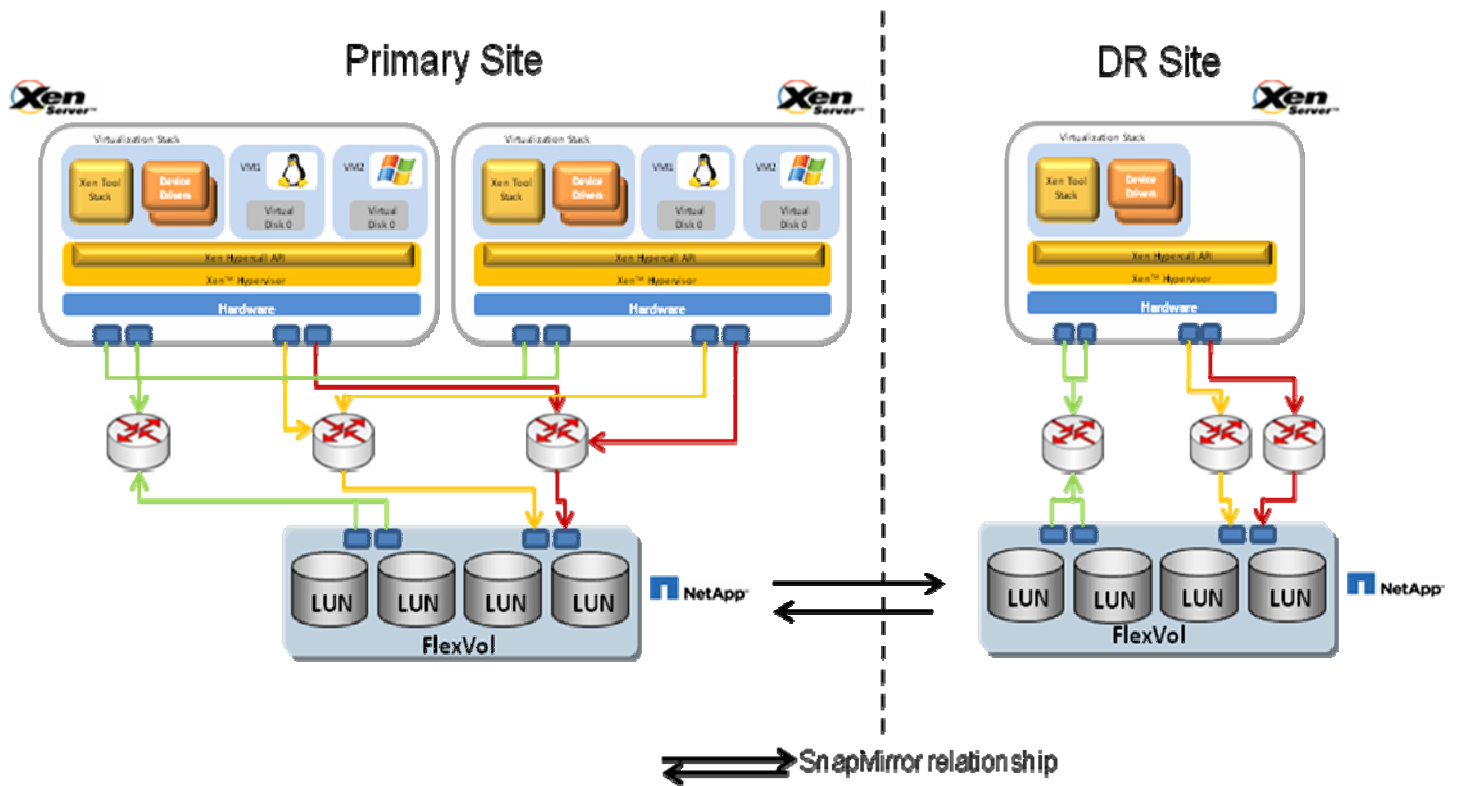
Figure 12 Graphical representation of replication setup

Creating the SnapMirror Relationship

(For purposes of illustration, the NetApp SR type is used for the backup and recovery steps. The same steps can be used for other SR types given the FlexVol)

1. Determine the SR UUID from XenServer host console (as indicated in previous sections). For NFS/iSCSI/LVMoHBA SR on the NetApp device, the FlexVol is the name determined when creating the FlexVol from the FilerView FlexVol creation wizard.

2. Open a console session to the NetApp device on the primary site. Convert the FlexVols in the NetApp SR to use ucode using the *vol options* command

   ```
   vol options <vol-name> <option-name> <option-val>
   ```

   ```
   xenadu-2> vol options XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV0 create_ucode on
   xenadu-2> vol options XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV0 convert_ucode on
   ```

3. Repeat the above step for all FlexVols that make up the NetApp SR

4. Ensure SnapMirror is turned on using the *options* snapmirror.enable command

   ```
   xenadu-2> options snapmirror.enable
   snapmirror.enable           on
   ```

   If not enabled, turn snapmirror on using the same command with *on* option

   ```
   options snapmirror.enable on
   ```

5. Authorize DR site host to replicate the primary site storage using the *options snapmirror.access* command

   ```
   options snapmirror.access host=<IP address or DNS name of DR site device>
   ```

   ```
   xenadu-2> options snapmirror.access host=10.219.101.234
   ```

6. On the primary site NetApp SR, disable any system Snapshots that may be on by default. In FilerView, click *Volumes->Snapshots->Configure* and select the FlexVols that make up the NetApp SR. Click *Unselect All*, and set the *Number of Scheduled snapshots to Keep* to *0*. Ensure that *Scheduled Snapshots* is unselected as well. Click *Apply*
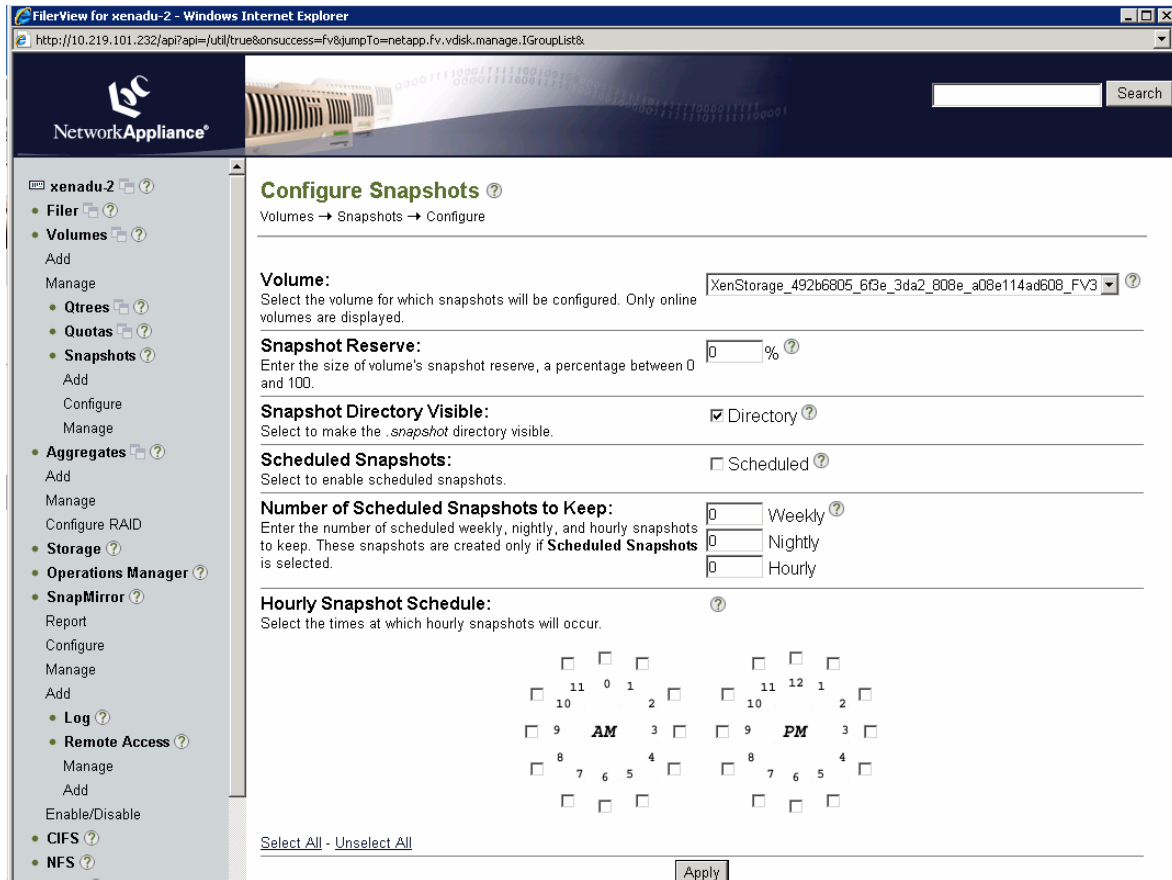
7. Open the NetApp FilerView for the NetApp device at the DR site, and create an aggregate that is larger than the aggregate on the primary site that has the NetApp SR FlexVols. The aggregate can be created using steps outlined in the NetApp device configuration section.

8. Create FlexVols within the aggregate created in the previous step, such that the FlexVols have the same name as the FlexVols that make up the NetApp SR at the primary site. FlexVols can be created using the steps outlined earlier in the document.   Ensure that the space guarantee for the FlexVol is set to *none* and the size if at least 5% greater than the FlexVol in the NetApp SR at the primary site. Also set the *Snapshot Reserve* to 0%.
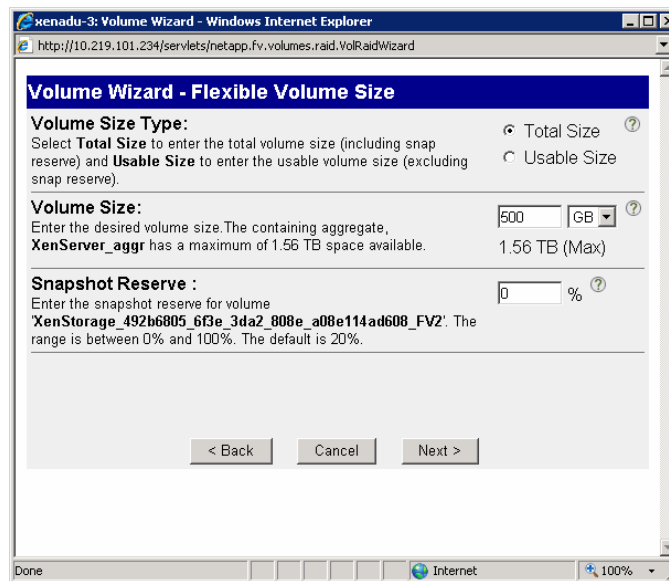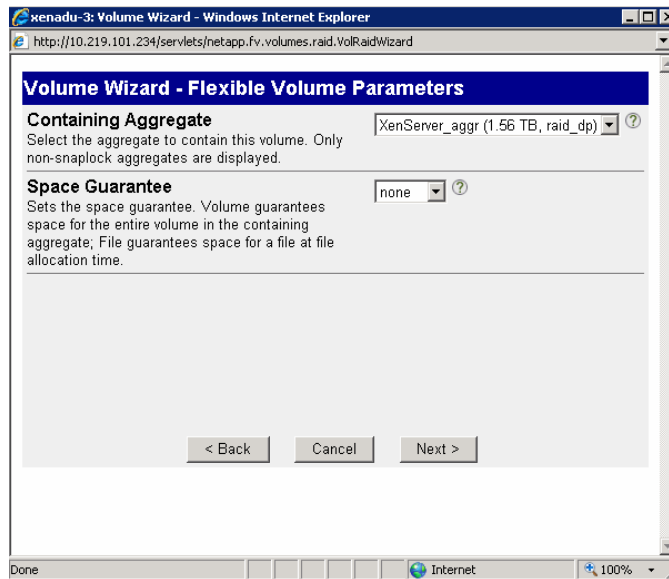
9. Set the ucode options on the FlexVols.

10. Turn off Snapshot on the FlexVols using the *vol options* command.

```
vol options <flexvol name> nosnap on
```

11. If Data Deduplication is turned on at the primary site, turn on dedupe on the DR stie FlexVols as well using the *sis on* command.

```
sis on <path>
```

12. Restrict the FlexVols by clicking on *Volumes->Manage*, selecting the FlexVols and then clicking the *Restrict* button.
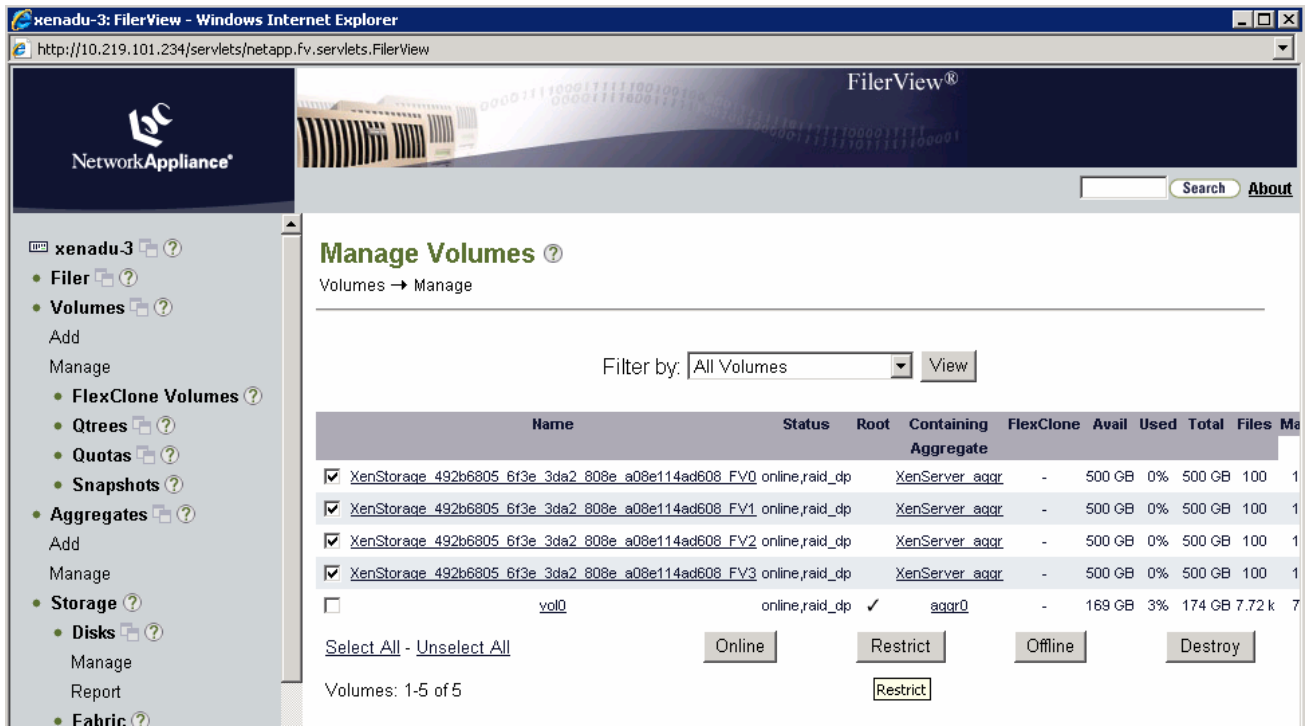
13. Give access to the primary site NetApp device to the DR site NetApp device and vice-versa. From FilerView, click *SnapMirror->Remote Access->Add* and add the DNS name or IP address of the filer will request access.
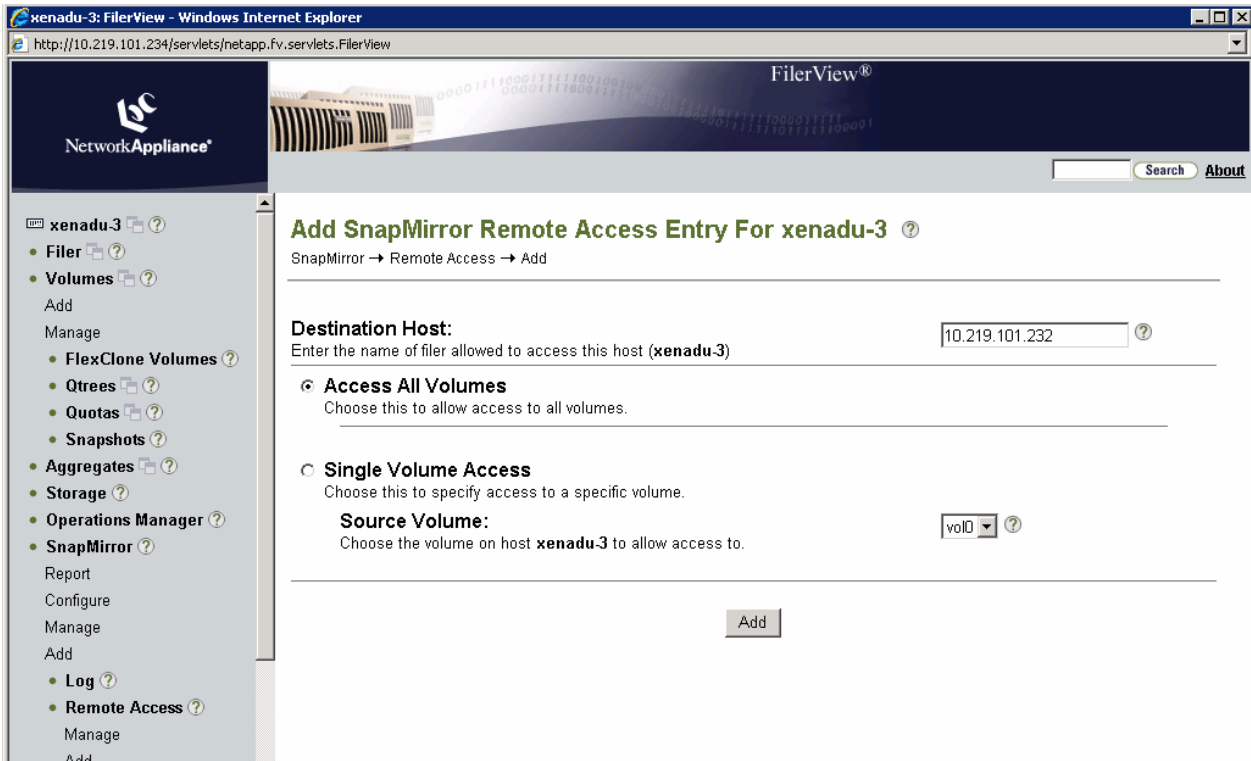
14. Set up SnapMirror from FilerView by clicking *SnapMirror->Add*.

15. Select the first FlexVol and click *Next*.
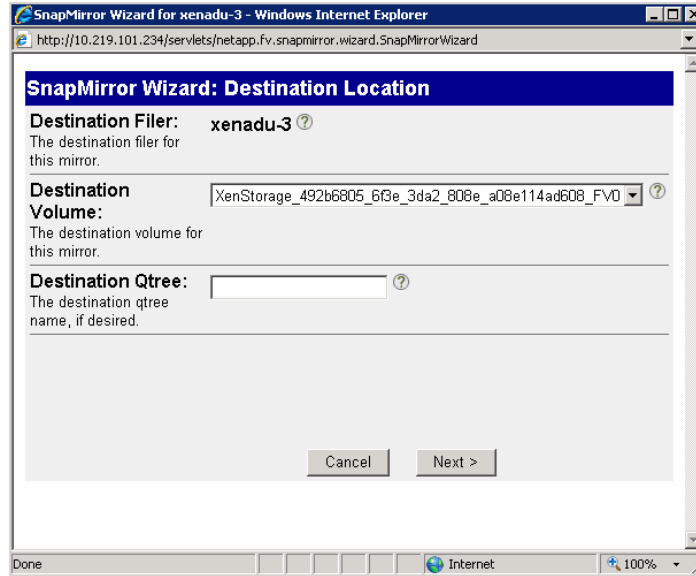
16. Enter the primary site NetApp device name or IP address and FlexVol for the first FlexVol that needs to be replicated.
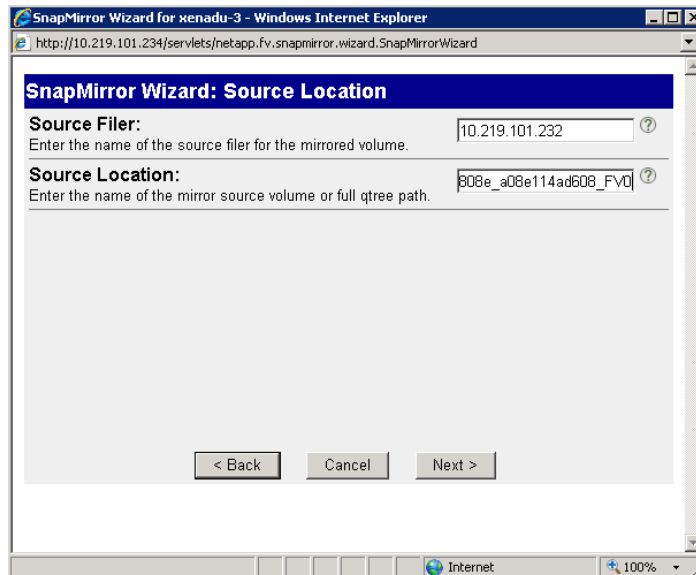
17. Leave the *Restart Mode* and *Maximum Transfer Rate* as the default. The data transfer limit may be changed depending on the network and storage configuration at the deployment site.

   a. Calculate a TCP window size that works well with the network, using the following formula:

   Window Size = (Round Trip Delay) × (Desired Rate)

   b. Adjust the TCP window size by entering the following command

   ```
   options snapmirror.window_size rate
   ```

   rate = the desired TCP window size.

18. Set the schedule for SnapMirror to occur as determined by the business needs. Having a small window will affect disk I/O on both the primary and DR sites, depending on the rate of change of the data blocks.

19. Commit the changes to create the new SnapMirror.

10.219.101.232:XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV1	xenadu-3:XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV1	idle	unknown	00:00:00	Modify View Delete Advanced
10.219.101.232:XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV2	xenadu-3:XenStorage_492b6805_6f3e_3da2_808e_a08e114ad608_FV2	idle	unknown	00:00:00	Modify View Delete Advanced

20. Click on *SnapMirror->Manage* and click *Advanced* for the SnapMirror just created.
21. Click on *Initialize* to setup the initial base line transfer.
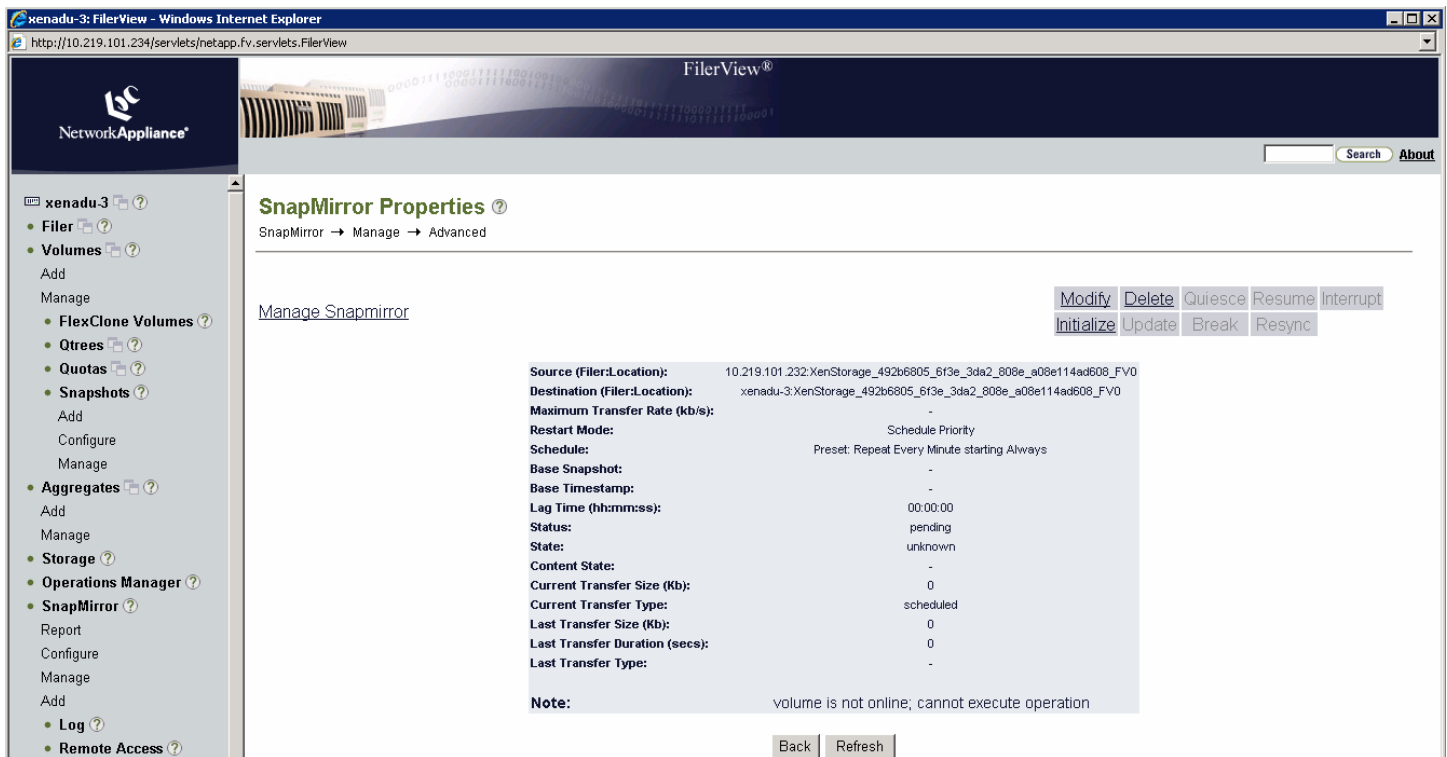
22. Once all FlexVols have been initialized, the SnapMirror *State* should show *snapmirrored*. To verify, in FilerView, click on *SnapMirror->Manage*.

Configuring XenServer at DR site at time of disaster

The steps to restore a FlexVol that has been backed up using SnapMirror are:

1. Quiesce the SnapMirror relationship.

2. Break the SnapMirror relationship.

3. Once the SnapMirror relationship has been broken, the mirrored FlexVol is attached to the backup XenServer host at the DR site.

Quiesce the SnapMirror relationship

1. In FilerView, click on *SnapMirror->Manage*. Click *Advanced* for the FlexVol that needs to be quiesced.
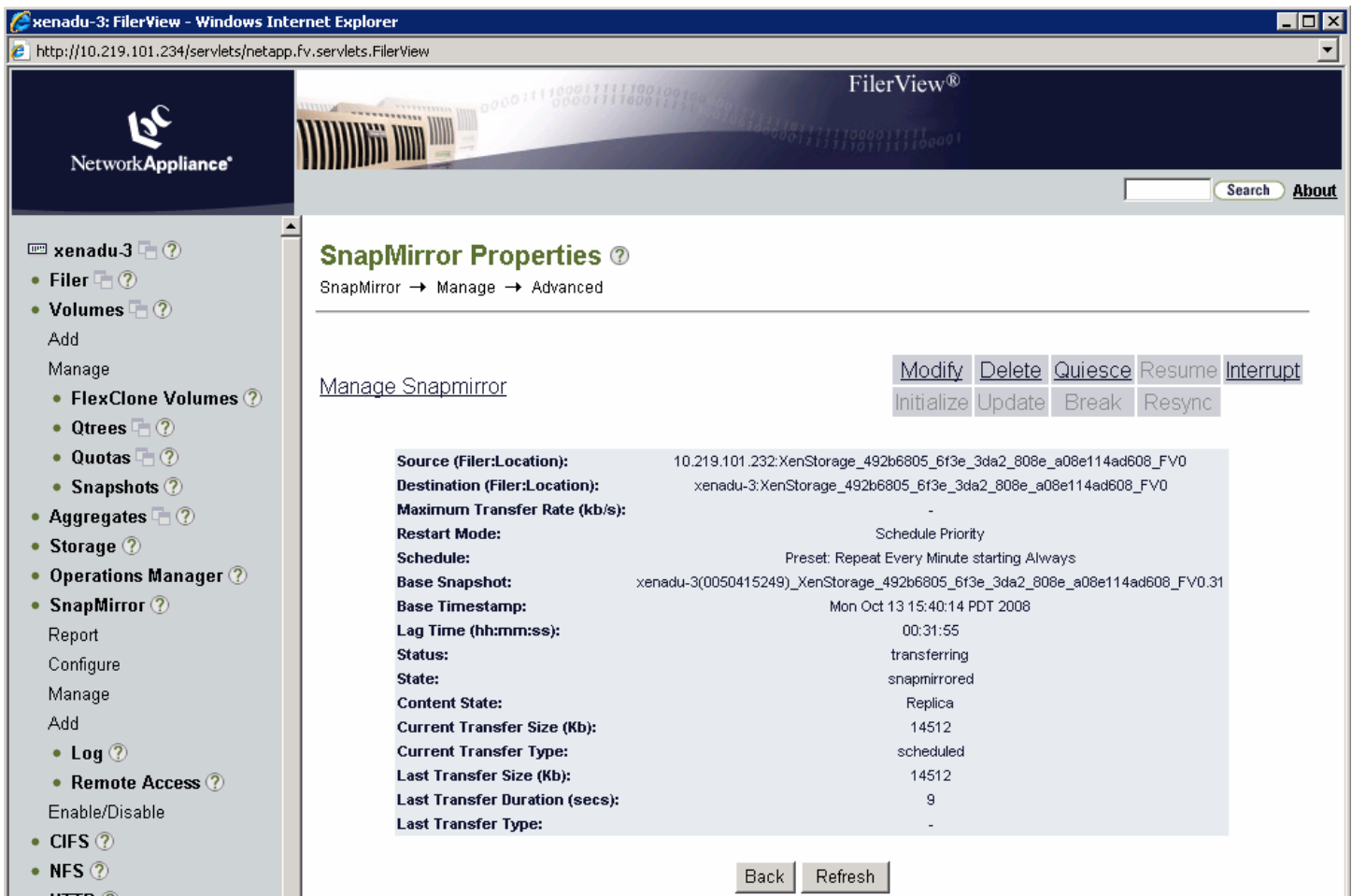
2. Click *Quiesce* and then *OK* in the confirmation dialog box.

Break the SnapMirror Relationship

1. Once the mirror has quiesced, click *Break* and *OK* in the confirmation dialog box. The state of the mirror should show *broken-off*.
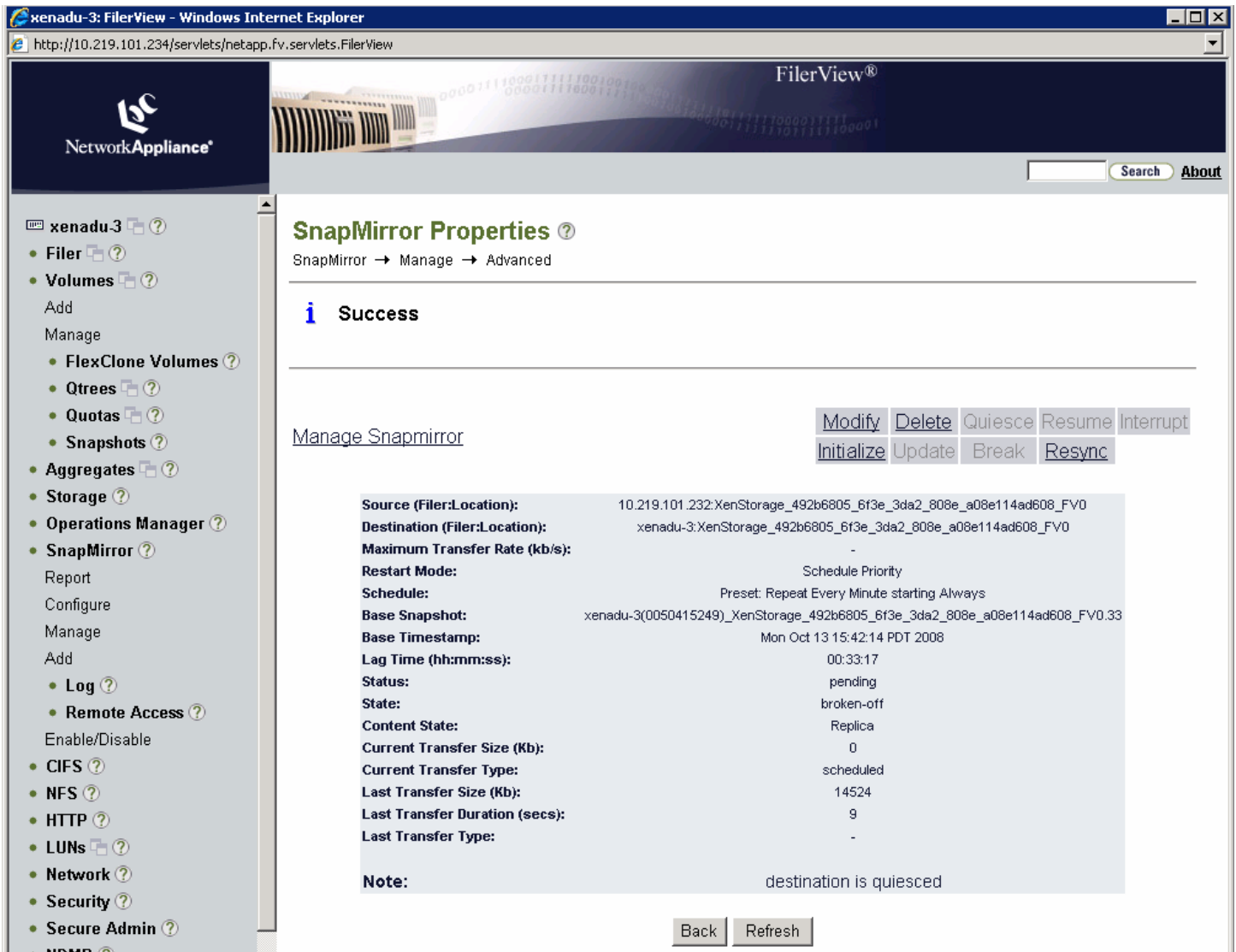
Repeat the quiesce and break-off process for each FlexVol that makes up the mirror'ed NetApp SR..

Attaching the mirror'ed SR to XenServer host at DR site

1. From XenCenter, connect to the XenServer host at the DR site..

2. Click on *New Storage*, and choose the *NetApp* option.

3. Enter the name/IP address of the NetApp device at the DR site that has the mirror'ed FlexVols that make up the NetApp SR.

4. Click on *Reattch on the existing SRs on the filer* and choose the SR that was mirror'ed. Click *Finish*.
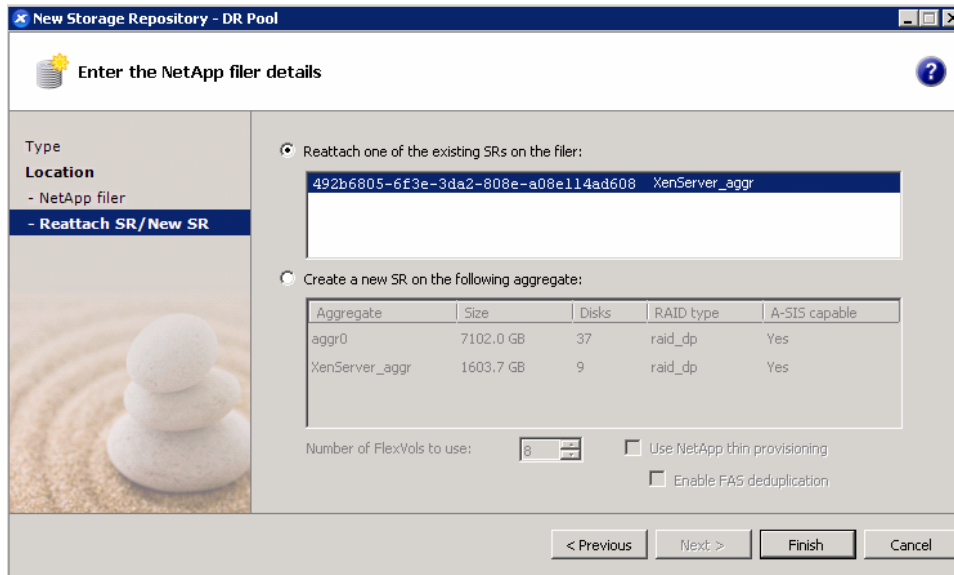
5. Click *Yes* to confirm attaching the SR..

6. Open the *xsconsole* to the XenServer host at the DR site. Scroll to *Backup, Restore and Update* and press *<Enter>*.
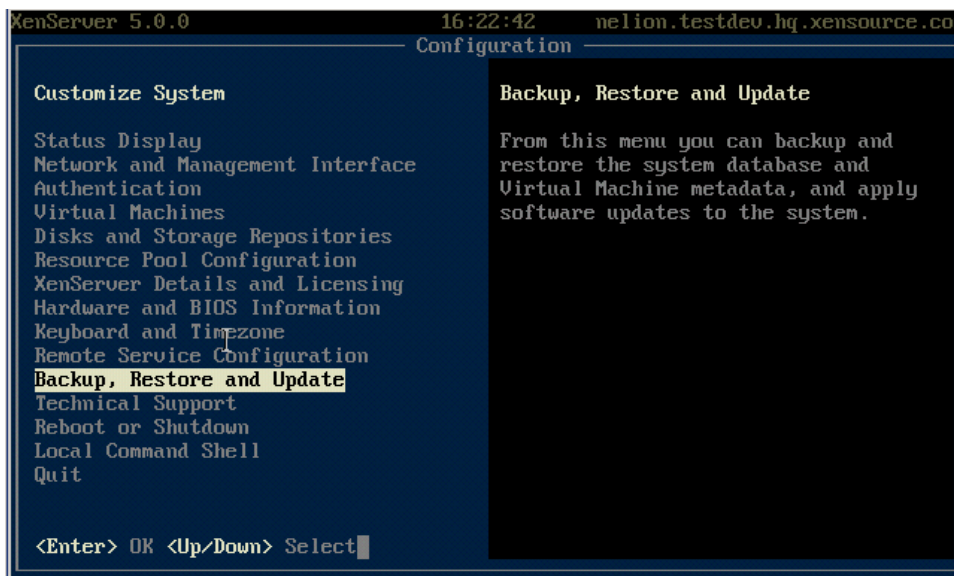
7. Scroll to *Restore Virtual Machine Metadata* and press *<Enter>*.
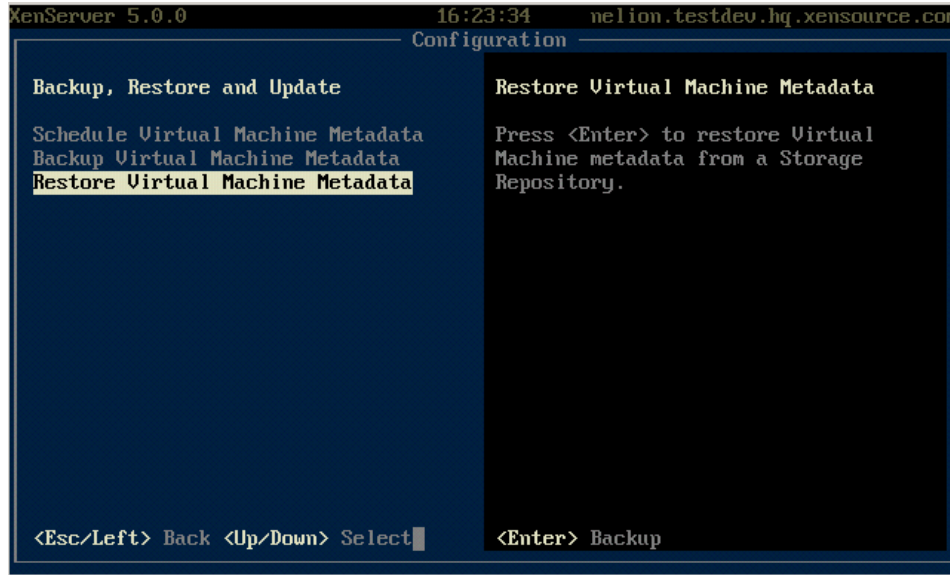
8. If prompted, provide XenServer login credentials to perform the metadata restoration.

9. Select the attached NetApp SR from the previous steps and press *<Enter>*.   There is a slight delay while XenServer looks for the backup VDI.

10. Once the backup VDI is found, select the appropriate metadata to restore from and press *<Enter>*.
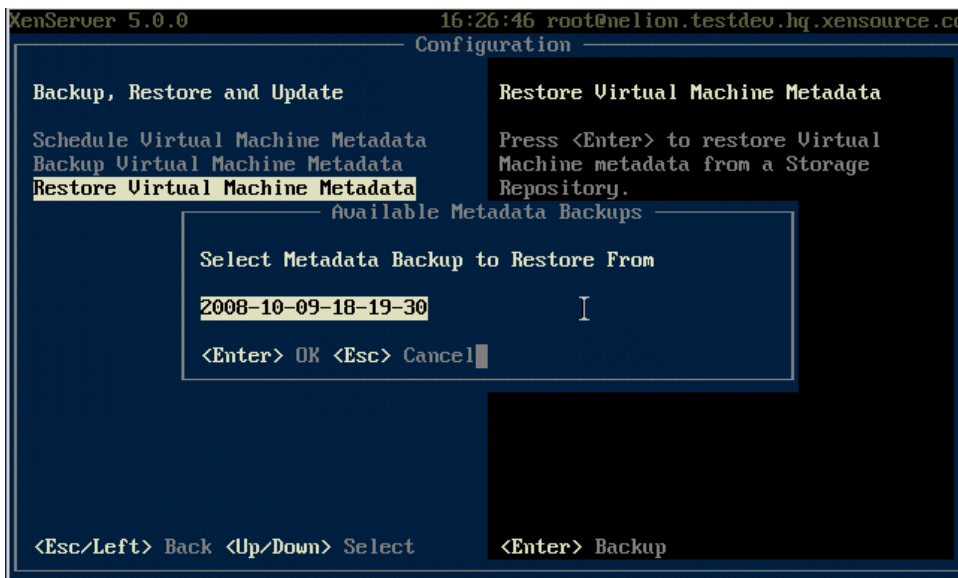
11. Select the *Only VMs on this SR* option to import all the VM metadata into the XenServer host at the DR site.
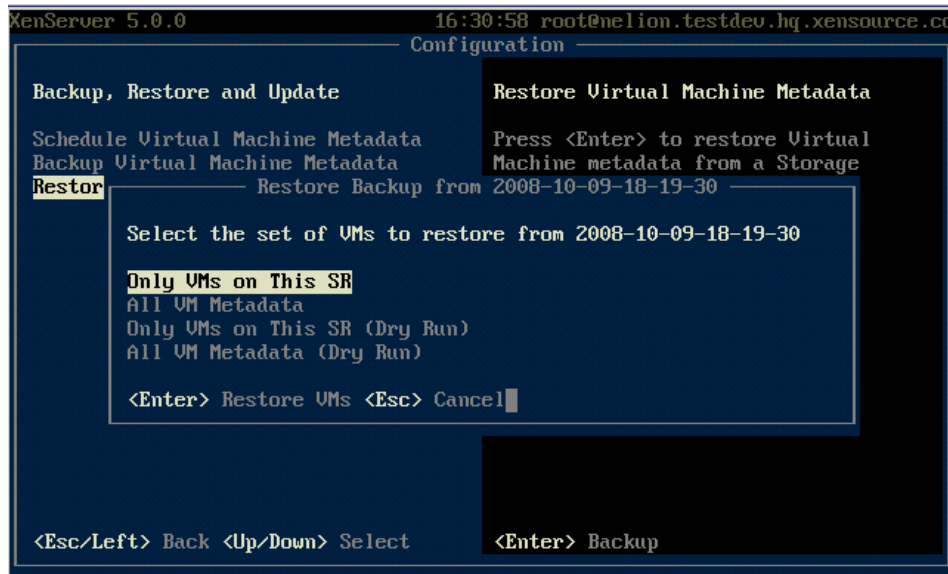
There is a slight delay while all metadata is imported.

Restoring SnapMirror relationship to bring primary site back up

Once the primary site has recovered from the disaster, use the below steps to restore information back to primary site.

To restore the SnapMirror relationship:

1. On primary site, destroy all data by re-initializing the primary NetApp device

2. Set up a SnapMirror from the DR NetApp device to the primary site NetApp device:

    a. Initialize a base line transfer

    b. Quiesce the SnapMirror relationship

    c. Break the SnapMirror relationship

    d. Delete the SnapMirror configuration

    This restores all the updated data from the DR site back to primary site.

3. Delete the SnapMirror configuration on the DR site NetApp device

4. Repeat the steps in creating the SnapMirror relationship to set up replication between the primary site and DR site.

Summary

While other virtualization products consume valuable server resources to run proprietary management tools to manage storage services, Citrix XenServer enables each hardware component to do what it does best: NetApp storage systems can deliver optimal storage features and performance, and servers can direct their power to business-critical application workloads.

Above all, by incorporating the best practices for configuration, allocation and administration into the NetApp Data ONTAP Adapter, Citrix XenServer makes optimal management of storage in virtualized environment seamless – allowing organizations to unify virtualization management for increased agility and reduced management costs.

Appendix A

Script to perform Snapshot and Recovery of VMs

The script below is an example of how a virtual machine may be snapshot for backup and recovery.

Usage:

Copy/paste the script below to a file with .py extension. Copy the .py flie to your XenServer host, and execute it.

Sample run output:

python snapshot_restore.py

```
        Citrix XenServer VM Snapshot/Restore Utility
        Copyright (c) 2008 Citrix Systems, Inc.

VMs available:

        1. Win_2K3

Do you want to snapshot|snapshot with quiesce (Windows guests only)|recover?
(input "s" or "sq" or "r")
```

```python
#!/usr/bin/env python
# Copyright (c) 2008 Citrix Systems, Inc.
#
# Permission to use, copy, modify, and distribute this software for any
# purpose with or without fee is hereby granted, provided that the above
# copyright notice and this permission notice appear in all copies.
#
# THE SOFTWARE IS PROVIDED "AS IS" AND THE AUTHOR DISCLAIMS ALL WARRANTIES
# WITH REGARD TO THIS SOFTWARE INCLUDING ALL IMPLIED WARRANTIES OF
# MERCHANTABILITY AND FITNESS. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR
# ANY SPECIAL, DIRECT, INDIRECT, OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES
# WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN
# ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF
# OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

import sys, time
import XenAPI

def snapshot_vm(session, vm_rec, snap_choice):
        epoch = time.strftime("%Y%m%d-%H%M%S", time.localtime())
        vm_snap_name = "%s - %s" % (vm_rec['name_label'], epoch)
        print "\nNow creating snapshot for VM %s..." % vm_rec['name_label']
        if snap_choice == 'sq':
                vm_snap = session.xenapi.VM.snapshot_with_quiesce(\
                  session.xenapi.VM.get_by_uuid(vm_rec['uuid']), \
                  vm_snap_name)
        else:
                vm_snap = session.xenapi.VM.snapshot(\
                  session.xenapi.VM.get_by_uuid(vm_rec['uuid']), \
                  vm_snap_name)
        print "VM snapshot created as : %s" % vm_snap_name

def recover_vm(session, vm_rec):
        snaps = vm_rec['snapshots']
        if len(snaps) == 0:
```

```
            print "No snapshots available for VM %s" % vm_rec['name_label']
        else:
            snap_num = 0
            snap_dict = {}
            print "List of snapshots available for VM %s:\n" % vm_rec['name_label']
            for snap in snaps:
                snap_rec = session.xenapi.VM.get_record(snap)
                snap_num += 1
                snap_dict[snap_num] = snap
                print "\t%i. %s" % (snap_num, snap_rec['name_label'])
            if (snap_num == 1):
                snap_num_choice = 1
                raw_input("\nPress Enter to restore from this snapshot")
            else:
                snap_num_choice = 0
                while (snap_num_choice == 0):
                    snap_num_choice_s = raw_input ('\nPlease choose a snapshot to
Restore to [1 - %i]: ' % snap_num)
                    if (snap_num_choice_s == ""):
                        print "\nInvalid entry.  Please try again...\n"
                        continue
                    snap_num_choice = int(snap_num_choice_s)
                    if (snap_num_choice > snap_num):
                        print ("\nInvalid entry.  Please try again...\n")
                        snap_num_choice = 0
                        continue
            snap_choice = snap_dict[snap_num_choice]
            snap_choice_rec = session.xenapi.VM.get_record(snap_choice)

            # Steps to restore
            # 1. Force shutdown VM
            # 2. Destroy VBDs
            # 3. Clone VDI's of snapshot
            # 4. Sync state (VBD, VIF, VCPU, Memory, etc)
            # 5. Destroy subsequent snapshots

            print "\nCleanup VM..."
            vm_name = vm_rec['name_label']
            vm_power_state = vm_rec['power_state']
            if not(vm_power_state == 'Halted'):
                print "Halting VM..."

        session.xenapi.VM.hard_shutdown(session.xenapi.VM.get_by_uuid(vm_rec['uuid']))

            vbds = vm_rec['VBDs']
            for vbd in vbds:
                vbd_rec = session.xenapi.VBD.get_record(vbd)
                if vbd_rec['type'] == 'Disk':
                    vdi = vbd_rec['VDI']
                    session.xenapi.VDI.destroy(vdi)
            vifs = vm_rec['VIFs']
            for vif in vifs:
                session.xenapi.VIF.destroy(vif)

            print "Restoring VM..."

            # Restore VIFs
            snap_vifs = snap_choice_rec['VIFs']
            for snap_vif in snap_vifs:
                snap_vif_rec = session.xenapi.VIF.get_record(snap_vif)
                restore_vif = { 'device' : snap_vif_rec['device'],
                                'network' : snap_vif_rec['network'],
```

```
                                'VM' : session.xenapi.VM.get_by_uuid(vm_rec['uuid']),
                                'MAC' : snap_vif_rec['MAC'],
                                'MTU' : snap_vif_rec['MTU'],
                                "other_config" : snap_vif_rec['other_config'],
                                "qos_algorithm_type" : snap_vif_rec['qos_algorithm_type'],
                                "qos_algorithm_params" :
snap_vif_rec['qos_algorithm_params'] }
                    session.xenapi.VIF.create(restore_vif)

            # Restore VBDs
            snap_vbds = snap_choice_rec['VBDs']
            for snap_vbd in snap_vbds:
                    snap_vbd_rec = session.xenapi.VBD.get_record(snap_vbd)
                    if snap_vbd_rec['type'] == 'Disk':
                            snap_vdi = snap_vbd_rec['VDI']
                            restore_vdi = session.xenapi.VDI.clone(snap_vdi, {})
                            restore_vbd = { 'VM' :
session.xenapi.VM.get_by_uuid(vm_rec['uuid']),
                                        'VDI' : restore_vdi,
                                        'userdevice' : snap_vbd_rec['userdevice'],
                                        'bootable' : snap_vbd_rec['bootable'],
                                        'mode' : snap_vbd_rec['mode'],
                                        'type' : snap_vbd_rec['type'],
                                        'unpluggable' : snap_vbd_rec['unpluggable'],
                                        'empty' : snap_vbd_rec['empty'],
                                        'other_config' : snap_vbd_rec['other_config'],
                                        'qos_algorithm_type' :
snap_vbd_rec['qos_algorithm_type'],
                                        'qos_algorithm_params' :
snap_vbd_rec['qos_algorithm_params'] }
                            session.xenapi.VBD.create(restore_vbd)

            # Restore config information
            vm_rec['platform'] = snap_choice_rec['platform']
            vm_rec['VCPUs_max'] = snap_choice_rec['VCPUs_max']
            vm_rec['recommendations'] = snap_choice_rec['recommendations']
            vm_rec['user_version'] = snap_choice_rec['user_version']
            vm_rec['HVM_shadow_multiplier'] = snap_choice_rec['HVM_shadow_multiplier']
            vm_rec['memory_dynamic_max'] = snap_choice_rec['memory_dynamic_max']
            vm_rec['HVM_boot_policy'] = snap_choice_rec['HVM_boot_policy']
            vm_rec['PV_legacy_args'] = snap_choice_rec['PV_legacy_args']
            vm_rec['actions_after_shutdown'] = snap_choice_rec['actions_after_shutdown']
            vm_rec['xenstore_data'] = snap_choice_rec['xenstore_data']
            vm_rec['actions_after_crash'] = snap_choice_rec['actions_after_crash']
            vm_rec['PCI_bus'] = snap_choice_rec['PCI_bus']
            vm_rec['tags'] = snap_choice_rec['tags']
            vm_rec['other_config'] = snap_choice_rec['other_config']
            vm_rec['VCPUs_at_startup'] = snap_choice_rec['VCPUs_at_startup']
            vm_rec['actions_after_reboot'] = snap_choice_rec['actions_after_reboot']
            vm_rec['PV_bootloader'] = snap_choice_rec['PV_bootloader']
            vm_rec['resident_on'] = snap_choice_rec['resident_on']
            vm_rec['HVM_boot_params'] = snap_choice_rec['HVM_boot_params']
            vm_rec['blocked_operations'] = snap_choice_rec['blocked_operations']
            vm_rec['PV_args'] = snap_choice_rec['PV_args']
            vm_rec['affinity'] = snap_choice_rec['affinity']
            vm_rec['PV_ramdisk'] = snap_choice_rec['PV_ramdisk']
            vm_rec['memory_static_min'] = snap_choice_rec['memory_static_min']
            vm_rec['PV_bootloader_args'] = snap_choice_rec['PV_bootloader_args']
            vm_rec['PV_kernel'] = snap_choice_rec['PV_kernel']
            vm_rec['memory_dynamic_min'] = snap_choice_rec['memory_dynamic_min']

            if not(vm_power_state == 'Halted'):
```

```
                        print "Powering up VM..."
                        session.xenapi.VM.start(session.xenapi.VM.get_by_uuid(vm_rec['uuid']),
False, False)

                print "Destroying orphaned snapshots..."
                for snap in snaps:
                        snap_rec = session.xenapi.VM.get_record(snap)
                        if snap_rec['snapshot_time'] > snap_choice_rec['snapshot_time']:
                                vbds = snap_rec['VBDs']
                                for vbd in vbds:
                                        vbd_rec = session.xenapi.VBD.get_record(vbd)
                                        if vbd_rec['type'] == 'Disk':
                                                vdi = vbd_rec['VDI']
                                                session.xenapi.VDI.destroy(vdi)
                                session.xenapi.VM.destroy(snap)

def snapshot_recover_vm(session, vm_rec):
        while (True):
                choice = raw_input("Do you want to snapshot|snapshot with quiesce (Windows
guests only)|recover?\n(input \"s\" or \"sq\" or \"r\") ")
                if (choice == 's') or (choice == 'sq'):
                        snapshot_vm(session, vm_rec, choice)
                        break
                elif (choice == 'r'):
                        recover_vm(session, vm_rec)
                        break
                else:
                        print "\nInvalid choice.  Please try again...\n"
                        continue

def main(session):
                vm_records = session.xenapi.VM.get_all_records()
                vm_num = 0
                vm_dict = {}
                print "VMs available:\n"
                for vm_record in vm_records:
                    vm_rec = vm_records[vm_record]
                  if not(vm_rec['is_a_template']) and not(vm_rec['is_control_domain']):
                        vm_num += 1
                        vm_dict[vm_num] = vm_rec
                        print "\t%i. %s" % (vm_num, vm_rec['name_label'])
                 print ""
                if (vm_num == 0):
                  print ("No VMs available for snapshot/recover")
                  return
                elif (vm_num == 1):
                  vm_num_choice = 1
                else:
                  vm_num_choice = 0
                  while (vm_num_choice == 0):
                        vm_num_choice_s = raw_input('Please choose VM to snapshot or recover [1 -
%i]: ' % vm_num)
                        if (vm_num_choice_s == ""):
                                print "\nInvalid entry.  Please try again...\n"
                                continue
                         vm_num_choice = int(vm_num_choice_s)
                        if (vm_num_choice > vm_num):
                                print "\nInvalid entry.  Please try again...\n"
                                vm_num_choice = 0
                                continue

                vm_rec_choice = vm_dict[vm_num_choice]
```

```
                snapshot_recover_vm(session, vm_rec_choice)

if __name__ == "__main__":
    print "\n\tCitrix XenServer VM Snapshot/Restore Utility"
    print "\tCopyright (c) 2008 Citrix Systems, Inc.\n"

    session = XenAPI.Session("http://_var_xapi_xapi", transport=XenAPI.UDSTransport())
    session.xenapi.login_with_password("root", "xensource")

    try:
        main(session)
    except Exception, e:
        print str(e)
        raw_input ("\n\n--Press Enter to exit--")
        raise
    raw_input ("\n\n--Press Enter to exit--")
```

Appendix B

Security Considerations

To allow XenServer administrators to login into the device with root privileges and create the FlexVols for the NetApp Storage Repository, use the following guidelines.

Equating Windows 'Domain\Administrator' privileges to UNIX 'root' in a multi-protocol environment
1. To equate Windows 'Domain\Administrator' privileges to UNIX 'root' in a multi-protocol environment, on the device enter:

```
device> options wafl.nt_admin_priv_map_to_root on
```

2. Authorizing a Unix User Account to Login As Root on the Device

Data ONTAP uses the /etc/usermap.cfg file to map user names. In its simplest form, each /etc/usermap.cfg entry contains a pair of names: the Windows name and the UNIX name. Data ONTAP can translate the Windows name to the UNIX name or vice versa.

When a connection is started, if the /etc/usermap.cfg file is missing, a default file is created. It contains commented-out sample map entries that are useful for improving security. When Data ONTAP receives a connection request from a user, it searches the /etc/usermap.cfg file to see whether an entry matches the user's Windows domain name and user name. If an entry is found, Data ONTAP uses the UNIX name specified in the entry to look up the UID and GID from the UNIX password database. If the UNIX name is a null string, Data ONTAP denies access to the user.

If an entry is not found, Data ONTAP converts the Windows name to lowercase and considers the UNIX name to be the same as the Windows name. Data ONTAP uses this UNIX name to look up the UID and GID from the UNIX password database. Data ONTAP http://www.citrix.com/English/partners/partner.asp?partnerID=950197scans the file sequentially. It uses the first matching entry for mapping.

For information about character coding of the /etc/usermap.cfg file, see the information about the contents of the /etc directory in the Storage Management Guide.

Specify each entry using the following format:

```
[IP_qualifier:] Windows_name [direction] [IP_qualifier:] UNIX_name
```

where
IP_qualifier field is an IP address that qualifies the user name by narrowing the match.
Windows_name field consists of a Windows domain name, which is optional, and a Windows user name.
Direction field indicates the direction of the mapping.
UNIX_name field is a UNIX name in the UNIX password database.

You can embed comments in the file by beginning the comment lines with #. Comments at the end of an entry are also allowed if preceded by #. Blank lines are ignored.

The way in which Data ONTAP interprets a domain name in the /etc/usermap.cfg file that contains a dot depends on whether storage system is in a Windows NT domain or a Windows Active Directory domain. Follow some guidelines to keep entries simple and easy to understand, and add several entries to the /etc/usermap.cfg file to prevent unauthorized users from accessing the storage system.

Appendix C

Enhancements in ONTAP 7.3.1

Enhancements in ONTAP 7.3.1 affect the ability to do a LUN clone split when using NetApp's ZAPI. ONTAP has two ways (styles) to do the LUN clone split. Beginning with ONTAP 7.3.1, it will default to the new style. Citrix XenServer's NetApp Adapter will attempt to use the old style LUN clone split.

You will see the following error message if you are affected by this issue:

"Error parameters: General IO error [opterr=creating flexvol snapshot:Snapshot operation not allowed during LUN clone split. (13023)]"

If you are affected, the issue can be verified and fixed with the following commands.

```
> options lun.clone_restore
> options lun.clone_restore {off | on}
```

Setting to 'off' will mimic the old behavior; the system defaults to 'on'.

References

XenServer 5.5 Administration Guide: http://support.citrix.com/article/CTX124024

XenServer 5.5 Virtual Machine Installation Guide: http://support.citrix.com/article/CTX124026

Citrix StorageLink User Guide: http://support.citrix.com/article/CTX120791

Data ONTAP® 7.3 System Administration Guide: http://now.netapp.com/NOW/knowledge/docs/ontap/rel73/pdfs/ontap/sysadmin.pdf

Data ONTAP® 7.3 Data Protection Online Backup and Recovery Guide: http://now.netapp.com/NOW/knowledge/docs/ontap/rel73/pdfs/ontap/onlinebk.pdf

Citrix Worldwide

Worldwide headquarters
Citrix Systems, Inc.
851 West Cypress Creek Road
Fort Lauderdale, FL 33309
USA
T +1 800 393 1888
T +1 954 267 3000

Regional headquarters

Americas
Citrix Silicon Valley
4988 Great America Parkway
Santa Clara, CA 95054
USA
T +1 408 790 8000

Europe
Citrix Systems International GmbH
Rheinweg 9
8200 Schaffhausen
Switzerland
T +41 52 635 7700

Asia Pacific
Citrix Systems Hong Kong Ltd.
Suite 3201, 32nd Floor
One International Finance Centre
1 Harbour View Street
Central
Hong Kong
T +852 2100 5000

Citrix Online division
5385 Hollister Avenue
Santa Barbara, CA 93111
USA
T +1 805 690 6400

www.citrix.com

About Citrix Systems, Inc.
Citrix Systems, Inc. (NASDAQ: CTXS) is the global leader in access infrastructure and the most trusted name in secure access. More than 200,000 organizations around the world use the Citrix Access Platform to provide secure, well-managed access to business information wherever it lives —on demand. Citrix customers include 100 percent of the Fortune 100 companies, 99 percent of the Fortune 500, and 97 percent of the Fortune Global 500, in addition to hundreds of thousands of smaller businesses and individuals. Based in Fort Lauderdale, Florida, Citrix has offices in 22 countries and approximately 6,200 channel and alliance partners in more than 100 countries. Annual revenue in 2007 was $1.4 billion.

About NetApp, Inc.

NetApp creates innovative storage and data management solutions that accelerate business breakthroughs and deliver outstanding cost efficiency. Discover our passion for helping companies around the world go further, faster at www.netapp.com