White Paper

VMware vSTORAGE APIs FOR ARRAY INTEGRATION WITH EMC VNX SERIES for NAS

Benefits of EMC VNX for File Integration with VMware VAAI

EMC SOLUTIONS GROUP

Abstract

The EMC® VNX™ platform's file service is integrated with VMWare® vSphere™ 5.0's new NAS support in vStorage APIs for Array Integration (VAAI). Based on the VMware vSphere 5.0 Release Candidate (RC) testing, this white paper highlights the benefits of offloading I/O-intensive operations from the VMware ESXi™ hosts to EMC VNX storage arrays. It also details the integration points and the instances when VAAI is not used.

January 2012

EMC²

Part Number h8292.1

Table of contents

Executive summary

VMware® and EMC® partnered to minimize the impact of high I/O virtualization tasks on ESXi® hosts and their networks by offloading these operations to the storage arrays that host the NFS data stores.

Instead of the hypervisor using resources to send large chunks of I/O across the network for common virtualization tasks (such as cloning a virtual machine), with vStorage APIs for Array Integration (VAAI) and EMC VNX, the hypervisor sends only commands to the EMC VNX Data Mover. The Data Mover performs the I/O-intensive operations on behalf of the hypervisor. This saves ESXi host resources and network bandwidth for what are most important—the applications and services that are virtualized. Therefore, by leveraging the new features in VAAI, the network utilization is reduced by 99 percent and the space utilization is 100 percent.

Terminology

Table 1 lists the terminology used in this document.

Table 1. Terminology

Term	Definition
Thick Provision Eager Zeroed	A Thick Provision Eager Zeroed disk has all space allocated and wiped clean of any previous contents on the physical media at the time of creation. Such disks may take longer to create compared to other disk formats.
Thick Provision Lazy Zeroed	A Thick Provision Lazy Zeroed disk has all space allocated at the time of creation. This space may contain stale data on the physical media. Individual blocks within the virtual disk are wiped clean the first time a virtual machine accesses them.
Thin Provision	Space required for a Thin Provision virtual disk is allocated and zeroed on demand as the space is used.

VAAI for NAS

Prior to VMware vSphere™ 5.0 and EMC VNX series build 7.0.35.3, VAAI consisted of a set of APIs communicated through SCSI commands. Since NFS does not support SCSI commands, VAAI did not work in an NFS model. Customers wanted the ability to leverage the offloading capability of VAAI available in SAN environment in a NAS environment. Hence, EMC worked with VMware to create a new way for VNX NFS Data Movers to implement VAAI hardware acceleration.

To achieve VAAI hardware acceleration for NAS, storage vendors must provide a plug-in for ESXi™ to interact with the hypervisor's I/O stack. Allowing NAS vendors to supply their own plug-in enables them to control the communication methods between ESXi and the storage. EMC's approach with its VNX product line is to simply send and receive VAAI communications between ESXi and the Data Movers through previously unused NFS commands.

In vSphere 5.0, VAAI hardware acceleration for NAS has the following VAAI for NAS features:

- Full File Clone — Enables the VNX Data Movers to clone virtual disks.

Figure 1 shows how Full File Clones are offloaded from the ESXi server.

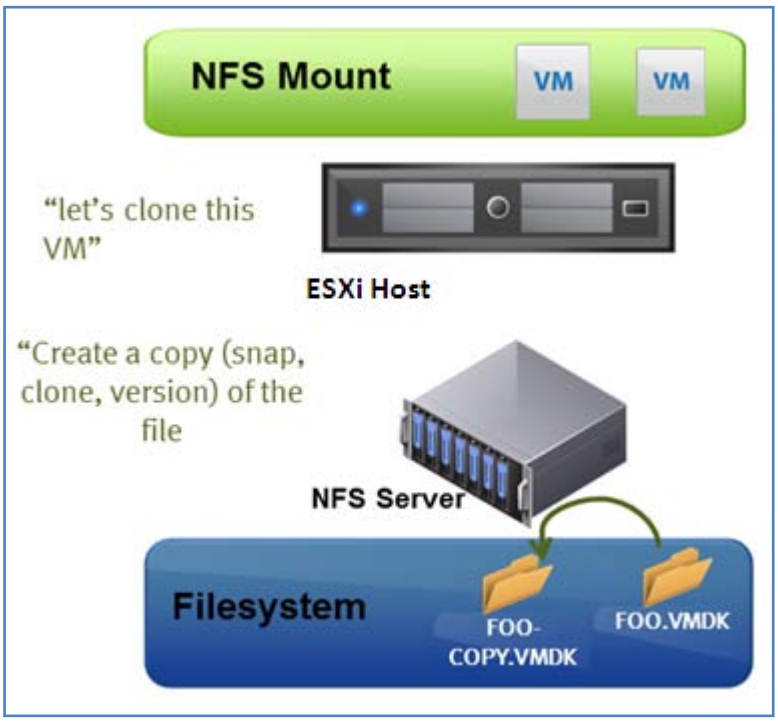

Figure 1. Full File Clones

- Enables vSphere to query space utilization details for files on NFS data stores through the VAAI interface. This includes the size of the file and the amount of space that on the file that is consumed.

Figure 2 shows how accurately the VAAI Extended Statistics feature retrieves the space utilization details from the NAS server.

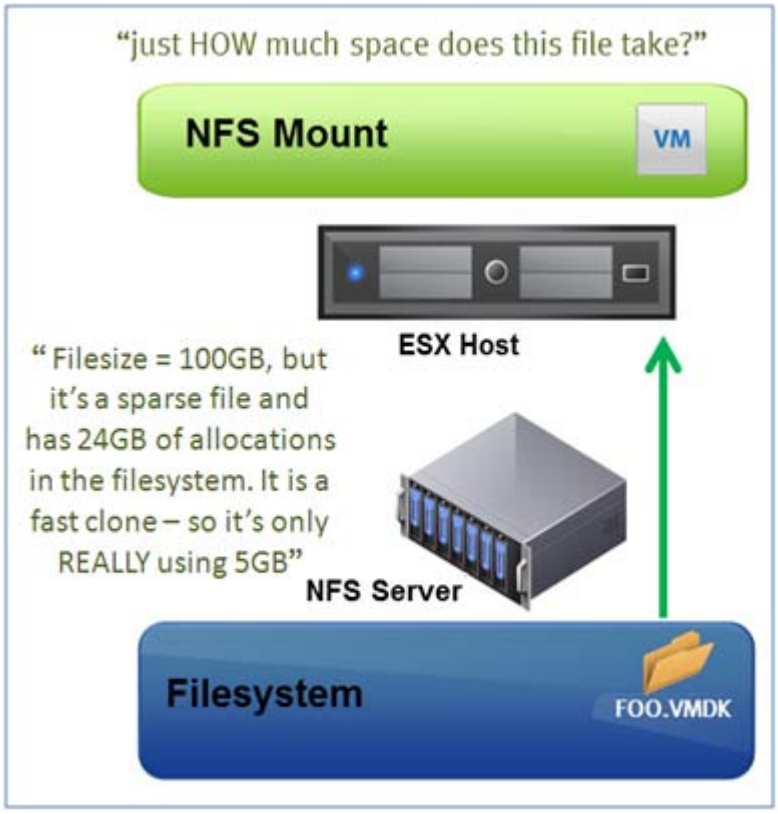

Figure 2. Visual representation of Extended Statistics

Space Reservation—NFS datastores can quickly provision thick virtual disk files with the lazy-zeroed option. Also, the ability to create Thick Provision Eager Zeroed virtual disks is new in vSphere 5.0 for NFS datastores.

VAAI for NAS components

The required hardware and software details for using VAAI for NAS are:

- VMware vSphere 5.0
- The vSphere Storage APIs for Array Integration(VAAI) Plug-in
  - Available for download from the EMC Online Support site and must be installed on each ESXi host.
- EMC VNX for File
  - VNX OE for File version 7.0.35.3 or later.
  - NFSv3-based datastore.

Hardware Acceleration Support Status

To verify the hardware acceleration support of each storage device and datastore through the vSphere Client:

1. Click **Configuration** › **Hardware** › **Storage** to display the **Datastores** or **Devices** view.
2. Review the status for each datastore that appears in the **Hardware Acceleration** column as shown in Figure 3.

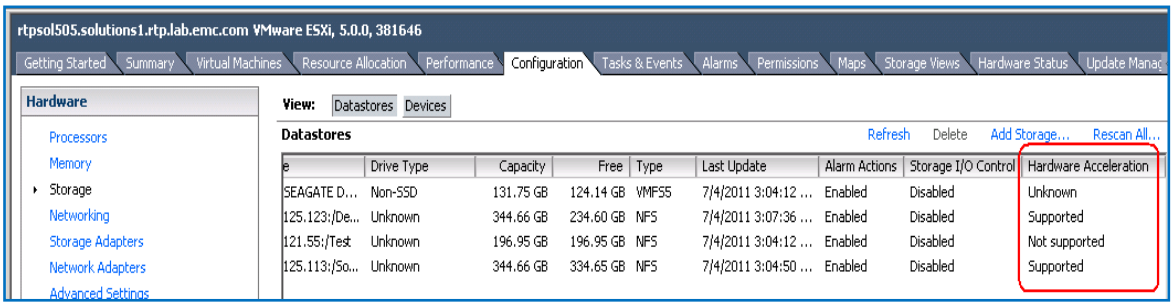

Figure 3. Hardware Acceleration support status

The status values are **Unknown**, **Supported**, and **Not Supported**. The initial value is **Unknown**. The status of the storage that supports NAS acceleration changes to **Supported** after the NAS VAAI plug-in is installed on the ESXi host. Once the storage is identified as **Supported**, the host starts to perform the offload operations automatically.

If the storage device does not support, or partially supports the host operations, the status changes to **Not supported**. Once the storage is identified as **Not supported**, the host reverts to native host-based methods to perform the operations that it cannot offload to the NAS device.

VAAI Statistics from VNX

EMC has provided a new group of statistics starting with the release of VNX OE for File version 7.0.35.3. These statistics provide information about the usage, performance, and characteristics of VAAI operations. The group is called “nfs.v3.vstorage”.

The following example syntax shows how to print new statistics with a CLI command:

```
server_stats server_2 -monitor nfs.v3.vstorage –type accu
```

The output of this command appears as shown in Figure 4 on page 8.

```
[root@ ~]# server_stats server_2 -monitor nfs.v3.vstorage -type accu
server_2      NFS VAAI op        VAAI Op Calls      VAAI Op Total uSecs      VAAI      VAAI Op
Timestamp                                                                  Op Max     Average
                                                                           uSecs      uSec/Op

10:19:17
10:19:32
10:19:47
10:20:02   vaaiOffloadStatus                  2                    0          0           0
           vaaiVxAttrs                        5                    0          0           0
           vaaiReserveSpace                   1                    0          0           0
           vaaiRegister                       7                    0          0           0
10:20:17   vaaiOffloadStatus                  5                    0          0           0
10:20:32   vaaiOffloadStatus                  8                    0          0           0
10:20:47   vaaiOffloadStatus                 11                    0          0           0
10:21:02   vaaiOffloadStatus                 14                    0          0           0
10:21:17   vaaiOffloadStatus                 17                    0          0           0
10:21:32   vaaiOffloadStatus                 20                    0          0           0
10:21:47   vaaiOffloadStatus                 23                    0          0           0
10:22:02   vaaiFullClone                      1                30693      30693       30693
           vaaiOffloadStatus                 25                    0          0           0
           vaaiVxAttrs                       12                    0          0           0
           vaaiRegister                      16                    0          0           0
10:22:17
10:22:32
10:22:47

server_2      NFS VAAI op        VAAI Op Calls      VAAI Op Total uSecs      VAAI      VAAI Op
Summary                                                                    Op Max     Average
                                                                           uSecs      uSec/Op
Minimum    vaaiFullClone                      0                    0          0       30693
           vaaiFastClone                      0                    0          0           -
           vaaiOffloadStatus                  0                    0          0           0
           vaaiOffloadAbort                   0                    0          0           -
           vaaiVxAttrs                        0                    0          0           0
           vaaiReserveSpace                   0                    0          0           0
           vaaiRegister                       0                    0          0           0
Average    vaaiFullClone                      0                 8185       8185       30693
           vaaiFastClone                      0                    0          0           -
           vaaiOffloadStatus                 13                    0          0           0
           vaaiOffloadAbort                   0                    0          0           -
           vaaiVxAttrs                        6                    0          0           0
           vaaiReserveSpace                   1                    0          0           0
           vaaiRegister                       8                    0          0           0
Maximum    vaaiFullClone                      1                30693      30693       30693
           vaaiFastClone                      0                    0          0           -
           vaaiOffloadStatus                 25                    0          0           0
           vaaiOffloadAbort                   0                    0          0           -
           vaaiVxAttrs                       12                    0          0           0
           vaaiReserveSpace                   1                    0          0           0
           vaaiRegister                      16                    0          0           0
```

Figure 4. VAAI Statistics

Physical environment

This section presents the configuration details of the lab environment created to verify the new VAAI for NAS functionality.

Reference architecture

To test the behavior of the VAAI for NAS features introduced in ESXi 5.0, the following network architecture was designed and implemented. The storage layout was kept simple in the test environment. A realworld storage layout may be more complex.

Figure 5 shows the network architecture of the test environment.

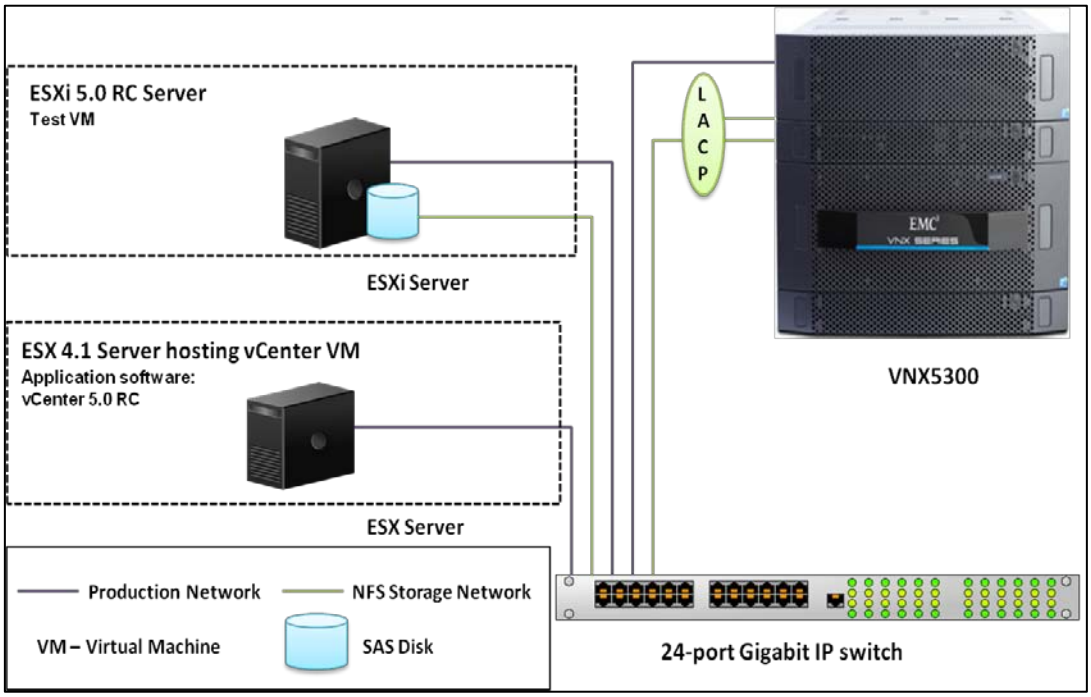

Figure 5. Reference architecture

Hardware resources

Table 2 lists the hardware resources used in this solution.

Table 2. Hardware resources

Hardware	Quantity	Configuration
EMC VNX5300	1	5 DAEs with 300 GB SAS drives
Intel-based rack mount server	1	• Memory: 72 GB of RAM • CPU: Six-core Xeon X5650, 2.67 GHz
Intel-based rack mount server	1	• Memory: 32 GB of RAM • CPU: Quad-core Xeon 5160, 3 GHz

Software resources

Table 3 shows the software used in this environment.

Table 3. Software resources

Software	Configuration
EMC VNX5300	
VNX OE for File	• 7.0.35.3 • NFSv3-based data store
ESX Server	
ESXi	ESXi Server 5.0 RC Build 381646 with the EMC VAAI Plug-in
ESX	vSphere 4.1
vCenter Server	
OS	Microsoft Windows Server 2008 64-bit Enterprise Edition R2
vCenter	vCenter Server 5.0 RC
Exchange Server	
OS	Microsoft Windows Server 2008 64-bit Enterprise Edition
Jetstress	Microsoft Exchange Jetstress 2007–08.02.0060.000

Storage layout

The testing is performed with two file systems of 350 GB in size each mounted as a NFS datastore on the ESXi server. The test virtual machine is located on one of the file systems. Depending on the use case, either one or both the file systems are used.

Figure 6 shows the storage layout used in this test environment.

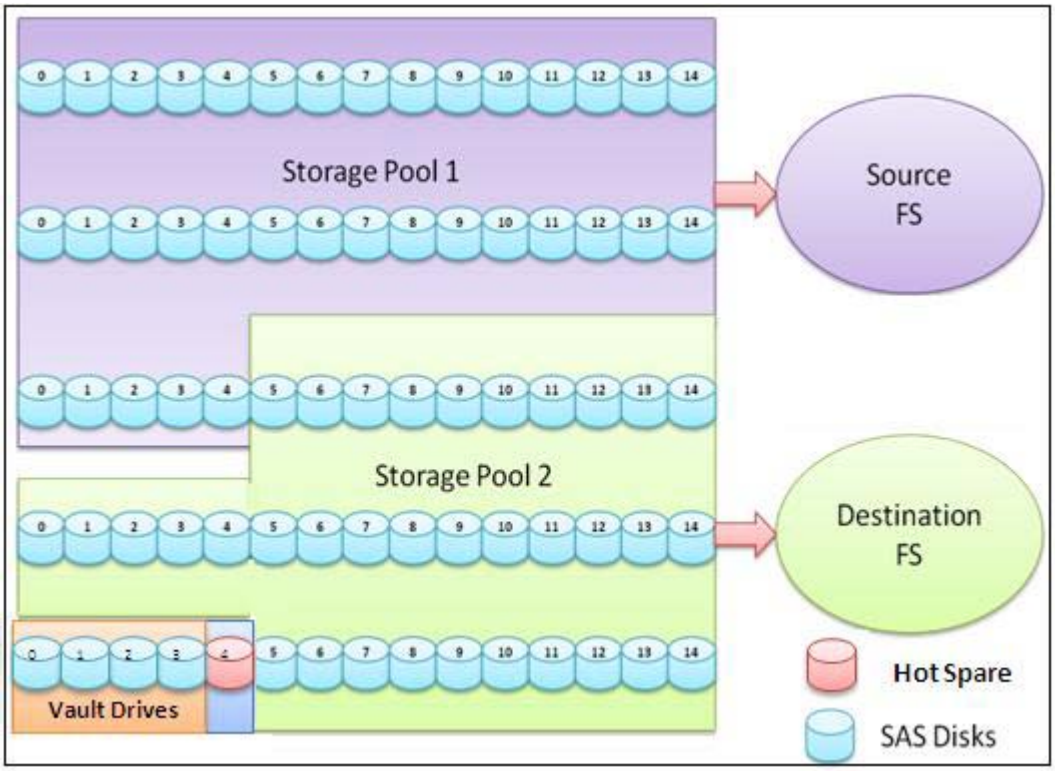

Figure 6. Storage architecture

Use cases and test results

Full File Clone

The benefits of the Full File Clone feature are evident in use cases such as deploying a virtual machine either from a template or cloning a virtual machine. Since the Full File Clone feature offloads these operations to the NAS server, the efficiency of the hypervisor and network infrastructure operation increases. If two datastores are controlled by the same Data Mover, the Full File Clone of a virtual machine uses the Full File Clone Hardware Acceleration feature to clone virtual disks across different datastores of the same Data Mover.

Figure 7 shows how to initiate Full File Clone of a virtual machine from a vCenter Server.

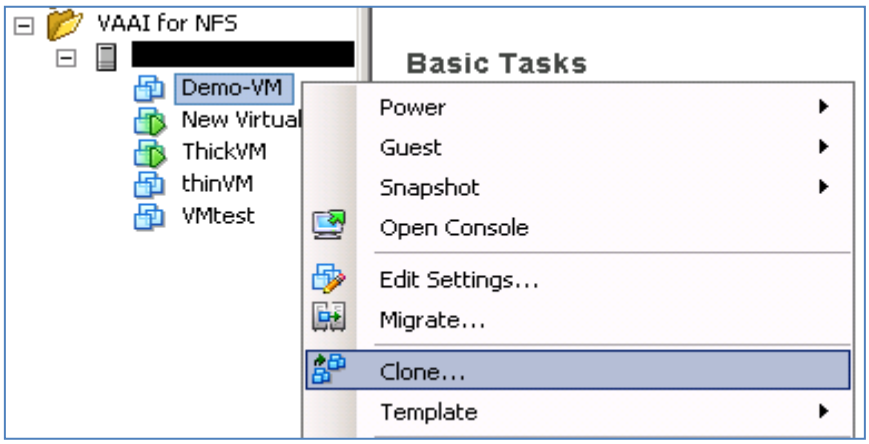

Figure 7. Initiate Full File Clone from vCenter Server

The following example syntax shows how to initiate a Full File Clone at the virtual-disk level using the vmkfstools CLI commands:

```
vmkfstools -i | --clonevirtualdisk <source_disk.vmdk>
           -d --diskformat [zeroedthick|
                thin|
                eagerzeroedthick|
                rdm:<device>|rdmp:<device>|
                2gbsparse]
           -N --avoidnativeclone
<FullClone_name.vmdk>
```

Figure 8 shows how to create a Full File Clone of a virtual disk from the ESXi server's CLI.

```
/vmfs/volumes/1a80df89-eaf30045/Demo-VM # vmkfstools -i Demo-VM.vmdk -d thin FullClone-Demo-VM.vmdk
Destination disk format: Thin
Cloning disk 'Demo-VM.vmdk'...
Clone: 100% done.
/vmfs/volumes/1a80df89-eaf30045/Demo-VM # ls
Demo-VM-ed6833e7.hlog        Demo-VM.vmx                 vmware-1.log
Demo-VM-flat.vmdk            Demo-VM.vmxf                vmware-2.log
Demo-VM.nvram                FullClone-Demo-VM-flat.vmdk vmware-3.log
Demo-VM.vmdk                 FullClone-Demo-VM.vmdk      vmware.log
Demo-VM.vmsd                 vmware-0.log
/vmfs/volumes/1a80df89-eaf30045/Demo-VM # vmkfstools --extendedstat FullClone-Demo-VM.vmdk
Capacity bytes: 42949672960
Used bytes: 13425074176
Unshared bytes: 13425074176
```

Figure 8. Initiate Full File Clone of a virtual disk from CLI

The most impressive part of offloading is the network bandwidth utilization. Figure 9 shows the performance benefit the network bandwidth utilization in various operations.

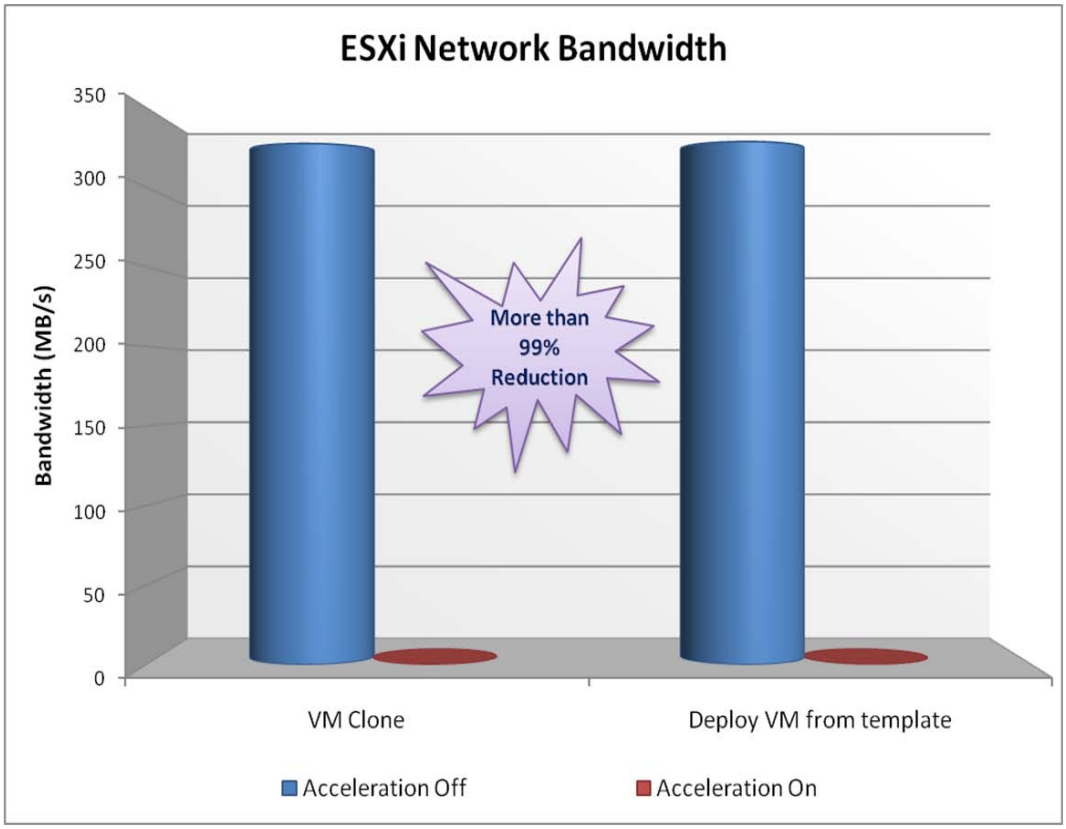

Figure 9. ESXi network utilization

The network utilization is reduced by more than 99 percent with VAAI on the EMC VNX series.

The time taken to complete the virtual machine clone operation was measured. With powered ON virtual machines, there was a lot of variation in time between each clone operation. With powered OFF virtual machines, the clone operation took 19 minutes and 3 seconds without VAAI, and 15 minutes and 8 seconds with VAAI. Hence, VAAI decreased the clone time by 20 percent.

Note Power OFF the source virtual machine while creating a clone of a virtual machine. In addition, ensure that the source virtual machine does not contain any VMware-based snapshots to gain the benefits of the Full File Clone feature. At the time of publication of this document, these are the constraints to implement VAAI for NAS.

Reserve space

NFS servers that do not support VAAI support only Thin Provisioning when creating virtual disks. NFS servers that support VAAI support the additional virtual disk provisioning options such as Thick Provision Eager Zeroed and Thick Provision Lazy Zeroed disks.

Figure 10 shows a VMware environment in which VAAI is not supported. In this example, **Thin Provision** is the only **Disk Provisioning** option available while creating disks with the **Add Hardware** wizard.

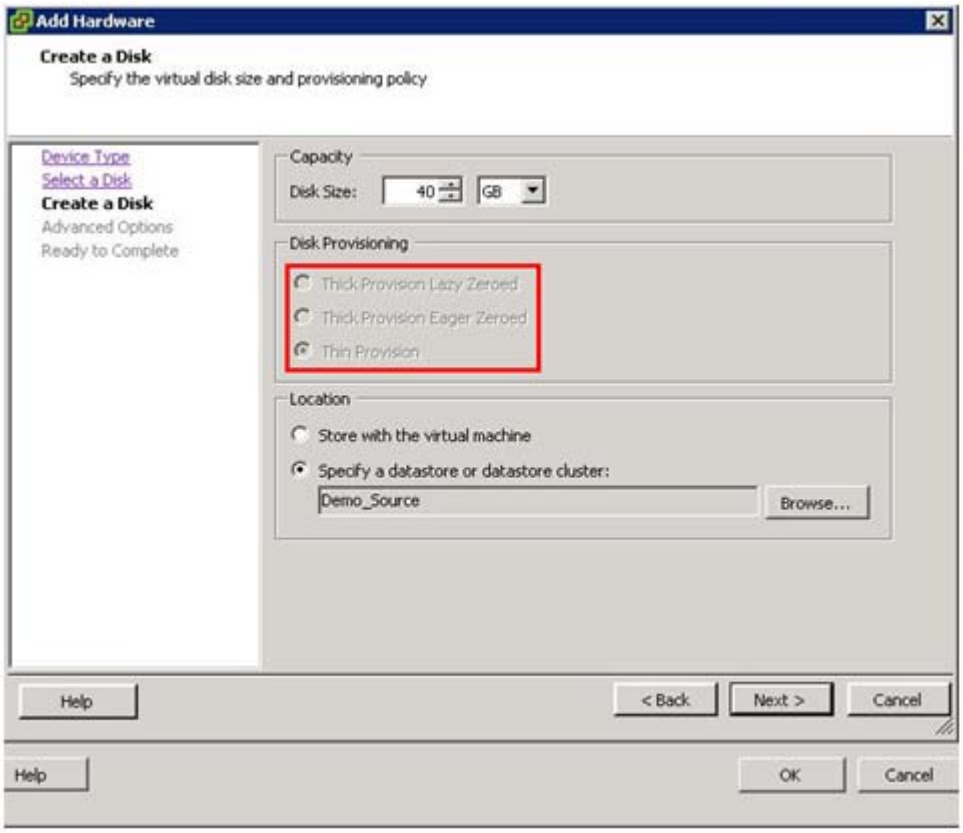

Figure 10. Disk Provisioning options on NFS datastores without VAAI support

If the NFS datastore supports VAAI the option to configure **Thick Provision Lazy Zeroed** and **Thick Provision Eager Zeroed** virtual disks will be enabled as shown in Figure 11 on page 15. The VAAI Space Reservation feature is what enables these additional methods of creating virtual disks.

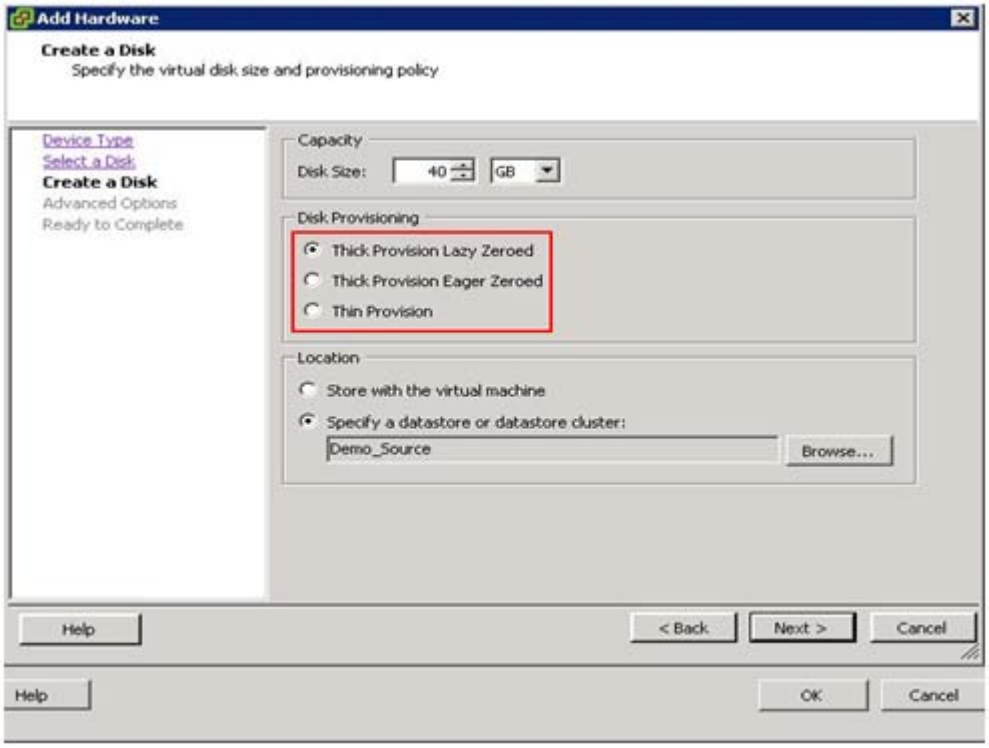

Figure 11. Disk Provisioning options with VAAI

The VAAI Reserve Space operation enables the NFS servers to preallocate the entire size of the virtual disk file and creates a sparse file on the NFS file system. In the case of Thick Provision Eagerly Zeroed disks, the sparse file is immediately filled with zeroes (and becomes dense again) before the VMware operation completes and before the virtual disk is available for use by a virtual machine.

Extended Statistics

This VAAI Extend Statistics feature helps to retrieve the accurate space utilization of a virtual machine by querying the NAS server for file properties that cannot be queried through standard NFS calls. It works in conjunction with the Reserve Space feature to monitor the space utilization within a sparse file and also helps to guarantee adequate space for a virtual machine.

Figure 12 shows the extended stats of a Thin Provision virtual disk. **Capacity bytes** shows the space allocated for the virtual disk, **Used bytes** shows the blocks used for the virtual disk, and **Unshared bytes** shows the actual number of bytes used by the virtual disk.

```
/vmfs/volumes/3affdf56-b845d022/VDI-XP-1 # vmkfstools --extendedstat VDI-XP-1.vmdk
Capacity bytes: 10737418240
Used bytes: 8218484736
Unshared bytes: 8218484736
```

Figure 12. Extended stats of a virtual disk with VAAI

Conclusion

EMC and VMware have effectively partnered to ensure that the VNX product line is fully prepared for the new VAAI functionality for NFS available in vSphere 5.

When customers add VAAI's new NAS features to their environment, consumption of ESXi resources such as CPU, memory, and network utilization are reduced by offloading I/O-intensive tasks to the storage arrays. Key benefits of this solution include:

- Virtual machine cloning and virtual machine deployment from a template are examples of operations that are offloaded to the VNX Data Mover by the Full File Clone feature of VAAI.
- Space Reservation and Extended Statistics features enable administrators to reserve space in the NFS file system for the entire capacity of a Thick Provision Lazy Zeroed virtual disk when provisioning a virtual machine. This prevents the possibility of an out-of-space error during an I/O-write operation.

References

White papers

The following document, located on the EMC Online Support site, provides additional, relevant information. Access to this document is based on the login credentials. If you do not have access to the document, contact your EMC representative:

- *VMware vStorage APIs for Array Integration with EMC VNX Series for SAN*—white paper