White Paper

VMware vSTORAGE APIs FOR ARRAY INTEGRATION WITH EMC VNX SERIES FOR SAN

Benefits of EMC VNX for Block Integration with VMware VAAI

EMC SOLUTIONS GROUP

Abstract

This white paper highlights the benefits of EMC® VNX™ integration with VMware® vStorage API for Array Integration (VAAI) for VNX Block storage protocols—Fibre Channel (FC), iSCSI, and FCoE. With the advanced storage capabilities of EMC, the VAAI features enable certain I/O-intensive operations to be offloaded from the VMware ESXi® host to the storage array enhancing performance and reducing the load on the ESXi host.

September 2011

EMC²

Copyright © 2011 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is.” EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

VMware, ESXi, vSphere, and vMotion are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions.

All trademarks used herein are the property of their respective owners.

Part Number H8293

Table of contents

Executive Summary

EMC® and VMware® partnered to provide intelligent solutions to minimize the impact of high I/O virtualization tasks on ESXi® hosts and their networks by offloading these operations to the storage array that hosts the VMFS datastores. Instead of the hypervisor using its resources to send large chunks of I/O across its networks for common virtualization tasks (such as cloning a virtual machine), with vStorage APIs for Array Integration (VAAI) and the EMC VNX™ platform, VMware vSphere™ only has to send commands to the EMC VNX platform to perform the I/O-intensive operations on behalf of vSphere. This saves ESXi host resources and network bandwidth for what are most important—the applications and services that are virtualized.

VAAI

VMware vStorage APIs for Array Integration (VAAI) enables very tight integration between the EMC VNX platform and VMware vSphere5.0.This integration reduces the load on the hypervisor from storage-related tasks to free resources for other operations.VAAI is a set of APIs and SCSI commands that offload certain I/O-intensive functions from the ESXi host to the VNX platform for more efficient performance.

VAAI was first introduced with vSphere 4.1 to enable the offload capabilities support for the following three features:

- Full Copy or Hardware-Assisted Move
- Block Zero or Hardware-Assisted Zeroing
- Hardware-Assisted Locking or Atomic Test and Set (ATS)

vSphere 5.0 introduced an additional feature, Thin Provisioning. In addition, EMC continues to improve the implementation of the offloaded features, and further extend the tight integration between the VNX platform and VMware vSphere.

The following sections explain the features in detail.

Thin Provisioning

vSphere 5.0 introduces multiple VAAI enhancements for environments that use array-based thin provisioning capabilities. The two new enhancements of VAAI Thin Provisioning are:

- Dead Space Reclamation
- Out-of-space conditions

Dead Space Reclamation

Historically, when virtual machines were migrated from a datastore, or when virtual disks were deleted, the blocks that were used by the virtual machines prior to the migration were still reported as “in use” by the array. This means that the usage statistics from the storage array might have been misleading, and expensive disk space may have been wasted.

EMC VNX integration with vSphere 5.0 mitigates these problems and offers the ability to reclaim blocks on a thin-provisioned LUN on the array when a virtual disk is deleted or migrated to a different datastore.

Benefit

With this new VAAI feature, the storage device is communicated that the blocks are no longer used. This leads to more accurate reporting of disk space consumption, and enables the reclamation of the unused blocks on the thin LUN.

Out-of-space conditions

An out-of-space condition is a significant problem in array-based, thin-provisioned environments. Storage oversubscription in thin-provisioned environments leads to catastrophic scenarios when an out-of-space condition is encountered.

EMC VNX integration with vSphere 5.0 mitigates these problems and simplifies storage management. If a thin-provisioned datastore reaches 100 percent, only the virtual machines that require extra blocks of storage space are paused, while virtual machines on the datastore that do not need additional space continue to run.

Benefit

The VAAI out-of-space condition in array-based thin provisioning temporarily pauses a virtual machine when disk space is exhausted. Administrators can allocate additional space to the datastore, or migrate an existing virtual machine without causing the virtual machine to fail.

Theory of operation

For thin-provisioned LUNs, vSphere 5.0 uses the SCSI UNMAP command to immediately free physical space on a LUN when a virtual disk is deleted, migrated to a different datastore, or when a snapshot is deleted.

Full Copy

This feature enables the storage arrays to make full copies of data within the array without the need for the VMware ESXi server to read and write the data.

Effective usage

The following scenarios make effective use of the Full Copy Feature:

- Clone a virtual machine
- Performa Storage vMotion
- Deploy virtual machines from a template

Benefits

On the EMC VNX platform, copy processing is faster. The server workload and I/O load between the server and storage are reduced.

Theory of operation

Without VAAI, the ESXi server reads (SCSI Read) every block from the VNX platform, and then writes (SCSI Write) the blocks to a new location. Therefore, server resources are consumed by transmitting large amounts of data between the ESXi server and the VNX platform.

With VAAI, the ESXi server sends a single SCSI (Extended Copy) command for a set of contiguous blocks to instruct the storage array to copy the blocks from one location to another. The command across the network is small, and the actual work is performed on the storage array. This minimizes data transmission, and speeds up copy processing. Figure 1 on page 8 shows a graphical representation of how the Full Copy feature offloads the copying of blocks with Extended Copy from ESXi to the VNX platform.

Figure 1. Full Copy feature

Block Zero

This feature enables storage arrays to zero out a large number of blocks to speed up virtual machine provisioning.

Effective usage

The following scenarios make effective use of the Block Zero feature:

- Create Thick Provision Eager Zeroed virtual disks
- Write data to an unused area of a Thick Provision Lazy Zeroed virtual disk

These two virtual disk formats zero out virtual disks in different ways, and therefore, benefit differently from the Block Zero feature:

- Thick Provision Eager Zeroed virtual disks are zeroed out when created, and are not usable until the process is completed. These disks are primarily used with virtual machines that are configured for VMware Fault Tolerance (FT).
- Thick Provision Lazy Zeroed virtual disks can be used immediately after they are created. Their blocks are zeroed on the first access.

Benefits

With Block Zero, the process of writing zeros is offloaded to the storage array. Redundant and repetitive write commands are eliminated to reduce the server load and the I/O load between the server and storage. This results in faster capacity allocation.

Theory of operation

Without VAAI, zeroing disk blocks sends redundant and repetitive write commands from the ESXi host to the storage array to explicitly write zeroes to each block. The host waits for each request until the zeroing task is complete. This process is time-consuming and resource-intensive.

With VAAI, ESXi uses the SCSI Write Same command to instruct the storage device to write the same data to a number of blocks. Instead of having the host wait for the operation to complete, the storage array returns to the requesting service as though the process of writing zeros has been completed. Internally, the VNX platform finishes zeroing out the blocks. Figure 2 on page 9 shows a graphical representation of how

the Block Zero feature offloads the process of writing zeros with the Write Same request from the ESXi host to the VNX platform.

Figure 2. Block Zero feature

Hardware-Assisted Locking

Hardware-Assisted Locking provides an alternate method to protect the metadata for VMFS cluster file systems and improve the scalability of large ESXi servers sharing a VMFS datastore. ATS allows locking at the block level of a logical unit (LU) instead of locking the whole LUN.

Effective usage

The following scenarios make effective use of the Hardware-Assisted Locking feature:

- Create a VMFS datastore
- Expand a VMFS datastore onto additional extents
- Power on a virtual machine
- Acquire a lock on a file
- Create or delete a file
- Create a template
- Deploy a virtual machine from a template
- Create a new virtual machine
- Migrate a virtual machine with vMotion
- Grow a file (for example, a snapshot file or a thin- provisioned virtual disk.)

Benefits

Hardware-assisted locking provides a much more efficient method to avoid retries for getting a lock when many ESXi servers are sharing the same datastore. It offloads the lock mechanism to the array, and then the array performs the lock at a very granular level. This permits significant scalability without compromising the integrity of the VMFS-shared storage pool metadata when a datastore is shared on a VMware cluster.

Theory of operation

Before VAAI, VMware had implemented locking structures within the VMFS datastores that were used to prevent any virtual machine from being run on, or modified by more than one ESXi server at a time. The initial implementation of mutual exclusion for updates to these locking structures was built on the use of the SCSI RESERVE and RELEASE commands. This protocol claims the sole access to an entire logical unit for the reserving host until it issues a subsequent release. Under the protection of a SCSI RESERVE command, a server node can update metadata records on the device to reflect the usage of the device without the risk of interference from any other host that also claims the same portion of the device. This approach, shown in Figure 3, significantly impacts the overall cluster performance because all other access to any portion of the device is prevented while SCSI RESERVE is in effect. As the size of the ESXi clusters and the frequency of modification of the virtual machines grow, the performance degradation from the use of SCSI RESERVE and RELEASE commands is unacceptable.

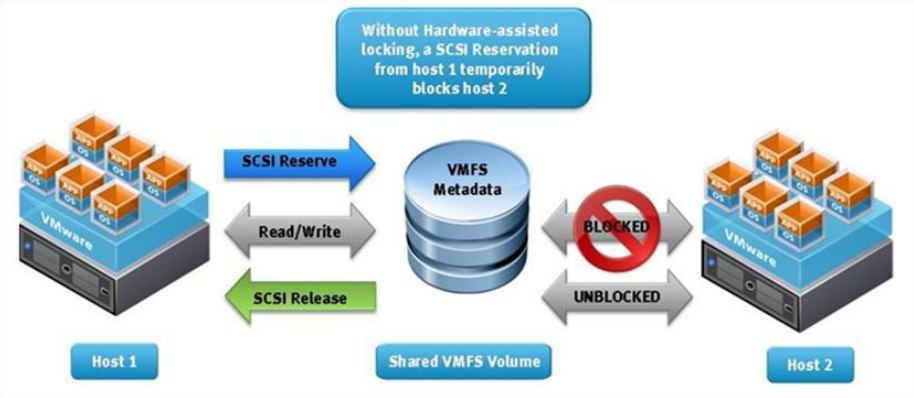

Figure 3. VMFS locking without VAAI

With VAAI, Hardware-Assisted Locking provides a more granular method to protect the VMFS metadata than was available with SCSI reservations. Hardware-Assisted Locking leverages a storage array ATS capability to enable a fine-grained block-level locking mechanism as shown in Figure 4 on page 11. First, Hardware-Assisted Locking replaces the sequence of RESERVE, READ, WRITE, and RELEASE SCSI commands with a single SCSI COMPARE AND WRITE (CAW) request for an atomic read-modify-write operation, based on the presumed availability of the target lock. Second, this new request only requires exclusion of other accesses to the target locked block, not the entire VMFS volume containing the lock. This locking metadata update operation is used by VMware when the state of a virtual machine changes. This may be a result of the virtual machine being powered ON or OFF, having its configuration modified, or being migrated from one ESXi server host to another with vMotion or Dynamic Resource Scheduling (DRS).

Figure 4. Hardware-Assisted Locking with VAAI

Hardware acceleration support status

The hardware acceleration support of each storage device and datastore can be verified in the vSphere Client. Navigate to **Configuration›Hardware›Storage›View: Datastores Devices**. The list of datastores appears as shown in Figure 5. The **Hardware Acceleration** column shows the status for each datastore or device.

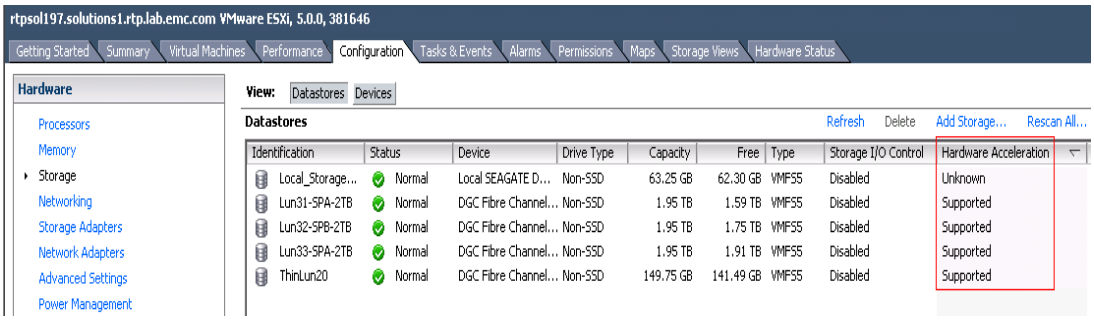

Figure 5. Hardware acceleration support status

Table 1 shows the possible hardware acceleration status values.

Table 1. Hardware acceleration status values

Status value	Description
Supported	The storage devices support VAAI
Not Supported	The storage devices do not support VAAI for FC datastores
Unknown	Local datastores

Physical Environment

This section presents the configuration details of the test environment created to verify the VAAI for Block functionality.

Reference architecture

Figure 6 shows the network architecture that was designed and implemented to test the behavior of VAAI Block features available on ESXi 5.0. A simple storage layout was used in the test environment. The realworld storage layout may be more complex.

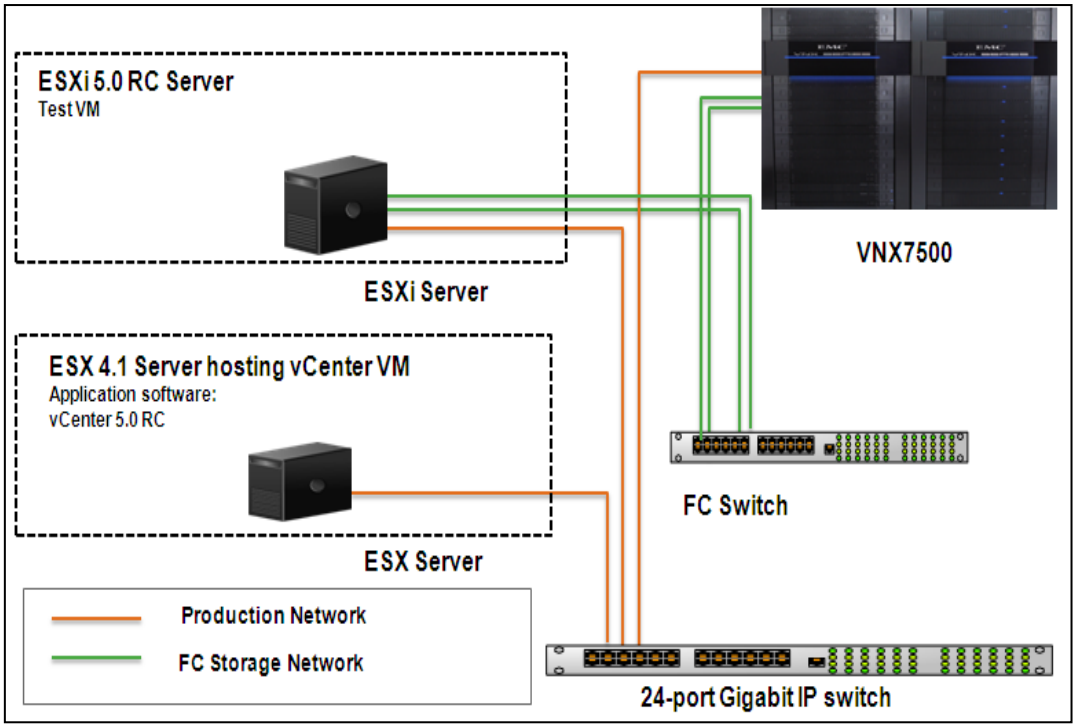

Figure 6. Reference architecture of the test environment

Hardware resources

Table 2 lists the hardware resources used in this solution.

Table 2. Hardware resources

Hardware	Quantity	Configuration
EMC VNX7500™	1	2 DAEs with 600 GB SAS drives
Intel-based rackmount server	10	Memory: 72 GB of RAM CPU: Quad-core Xeon X5550, 2.67 GHz

Software resources

Table 3 lists the software resources used in this solution.

Table 3. Software resources

Software	Configuration
EMC VNX7500	
VNX OE for File	7.0.35.x
VNX OE for Block	05.31.000.5.502
ESX server	
ESXi	ESXi 5.0 Build 441354
ESX	vSphere 4.1
vCenter Server	
OS	Microsoft Windows Server 2008 64-bit Enterprise Edition R2
vCenter	vCenter Server 5.0 Build 441357

Storage layout

The testing was performed on two VMFS datastores provisioned over FC. The test virtual machine was located on one of the VMFS datastore, and it was cloned or migrated to another datastore.

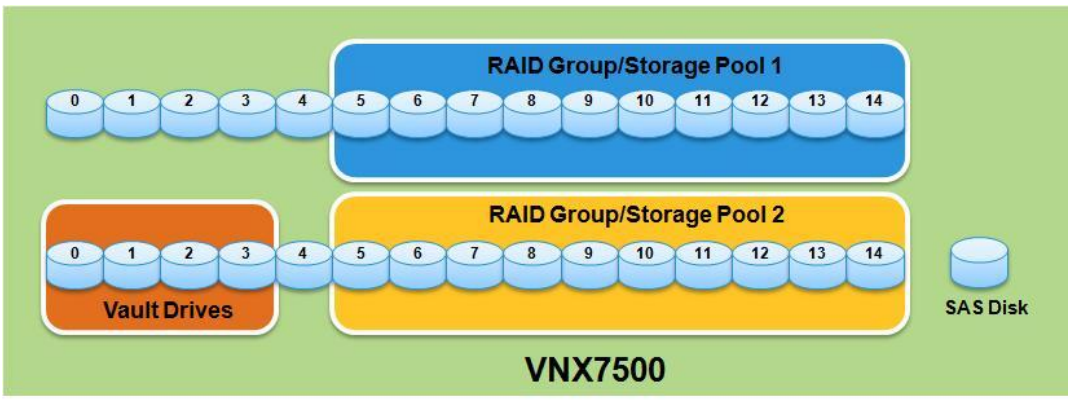

Figure 7. Storage layout used in the test environment

Use Cases and Test Results

Full Copy

Verification steps

The following steps were performed to verify the Full Copy feature:

1. Create a virtual machine with a 150 GB Thick Provision Lazy Zeroed virtual disk on a VMFS datastore. The actual storage consumption on the datastore was 58 GB.
2. Migrate or clone the virtual machine from one RAID group to another RAID group.
3. Measure the time required to migrate or clone the virtual machine.
4. Compute the network traffic on the ESXi host.
5. Repeat the tests with VAAI OFF and VAAI ON.

Table 4 shows the time taken to perform Storage vMotion and virtual machine clone operations with Full Copy VAAI ON and OFF.

Table 4. Time taken for Full Copy

Full Copy use case	VAAI OFF	VAAI ON
Storage vMotion	4 minutes 26 seconds	3 minutes 45 seconds
Virtual machine clone	4 minutes 38 seconds	3 minutes 38 seconds

Key findings

The following key findings are based on the testing of the Full Copy feature:

- Storage vMotion is15 percent faster and virtual machine cloning is 21 percent faster using VAAI with EMC VNX.
- There was a drastic decrease in the network traffic from the ESXi host to the VNX storage with VAAI ON. Figure 8 on page 15 shows that 8,000 copy commands are reduced to 115 commands to complete the Storage vMotion or virtual machine clone operation. This is the result of offloading 98 percent of the disk commands from the ESXi host to the EMC VNX platform using VAAI.

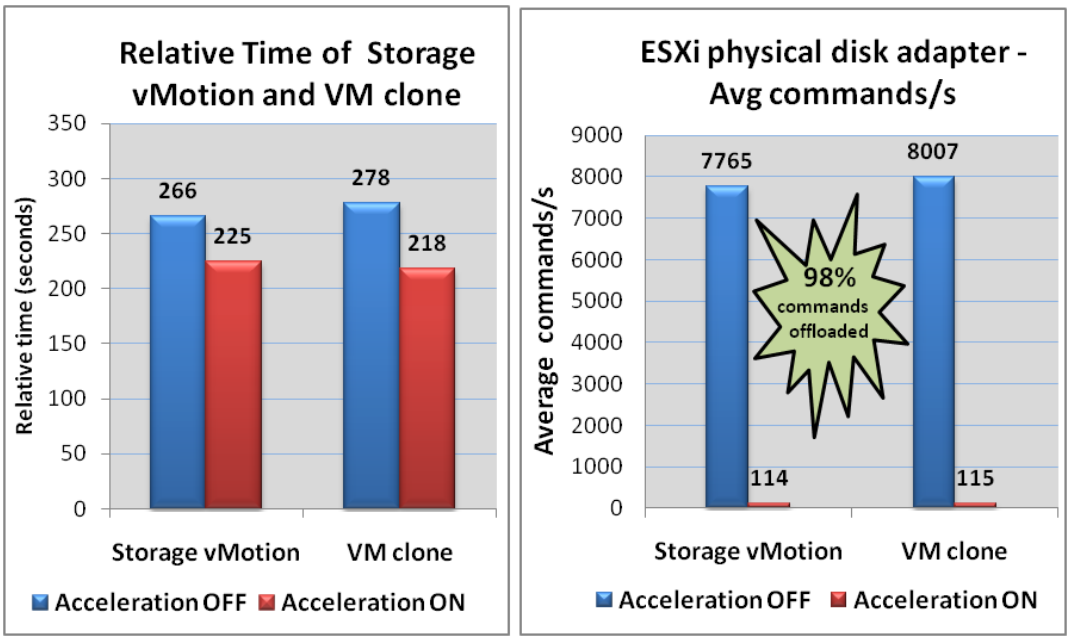

Figure 8. Storage vMotion and virtual machine clone with VAAI (OFF and ON)

Block Zero

Verification steps

The following steps were performed to verify the Block Zero feature:

1. Measure the time taken to create a 100 GB Thick Provision Eager Zeroed virtual disk on a virtual machine.
2. Compute the network traffic on the ESXi host.
3. Repeat the tests with VAAI OFF and VAAI ON.

Table 5 shows the time taken to create a Thick Provision Eager Zeroed virtual disk on both a thick pool LUN and a traditional LUN with VAAI ON and OFF.

Table 5. Time taken for Block Zero

Block Zero use case	VAAI OFF	VAAI ON
Thick pool LUN	4 minutes 33 seconds	2 minutes
Traditional LUN	4 minutes 34 seconds	1 minute 23 seconds

Thick pool LUN — A LUN that provides storage through a pool is called a Thick pool LUN.

Traditional LUN — A LUN that provides storage directly from the VNX OE for Block is called traditional LUN. This is any LUN that is not a pool LUN.

Key findings

The following key findings are based on the testing of the Block Zero feature:

1. Creation of a Thick Provision Eager Zeroed virtual disk with VAAI enabled is 56 percent faster with a VNX thick pool LUN, and 70 percent faster with a traditional LUN.
2. There was a drastic reduction in network traffic on the ESXi host with VAAI enabled. Figure 9 shows that 75 percent of write commands are offloaded from the ESXi host to the EMC VNX platforms with VAAI.

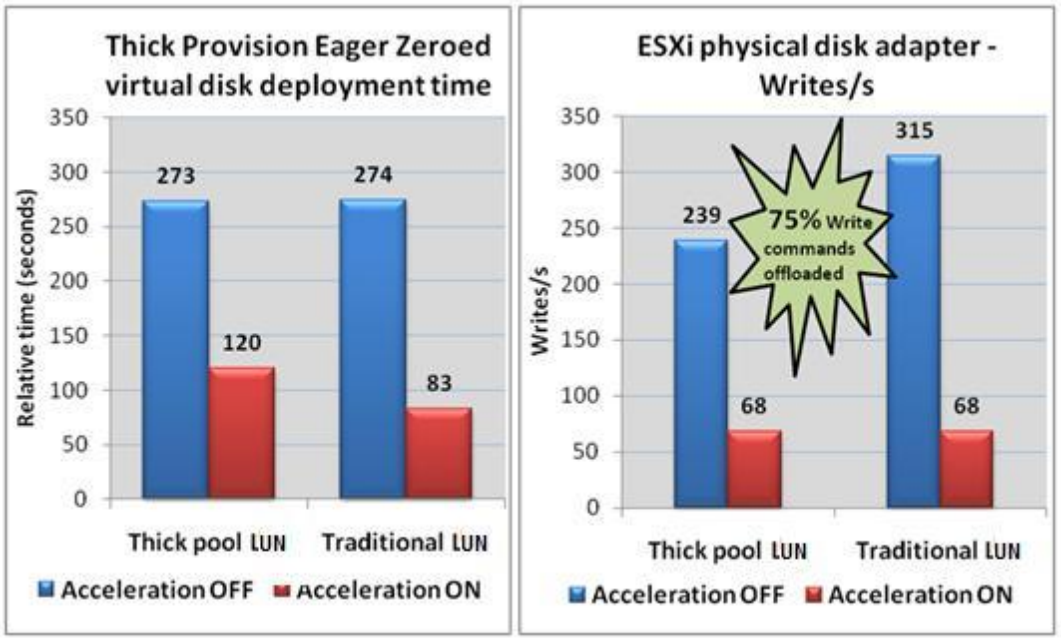

Figure 9. Thick Provision Eager Zeroed virtual disk creation with VAAI (OFF and ON)

Hardware-Assisted Locking

This feature delivers improved locking controls on a VMware VMFS datastore, enabling a greater number of virtual machines per datastore and supporting larger ESXi clusters while maintaining a high level of performance. This feature increases the performance of simultaneous metadata operations on a shared VMFS datastore during common tasks such as creating and deleting of virtual disks, simultaneously powering many virtual machines (ON or OFF), and snapshot operations.

Most users will not face the LUN-level locking issue because the EMC VNX series already handles these operations efficiently. Hardware-Assisted Locking is beneficial in extreme situations, and EMC recommends enabling the feature to improve performance when an extreme situation arises.

An extreme use case was required to demonstrate the performance benefits of the Hardware-Assisted Locking feature.

Verification steps

The following steps were performed to verify the Hardware-Assisted locking feature:

1. Create a virtual machine on ESXi and generate I/O load with IOmeter on a shared VMFS datastore.
2. Use several other ESXi hosts to perform continuous virtual machine power ON or OFF operations simultaneously on a shared VMFS datastore.
3. Repeat the tests with VAAI OFF and VAAI ON.

Key findings

The testing showed that with VAAI Hardware-Acceleration Locking ON, the datastore was able to service 70 percent more IOPS than it was able to service without VAAI.VAAI enabled the VMFS datastore to service a significantly higher number of IOPS.

Thin Provisioning

Verification steps

The following steps were performed to verify the Dead Space Reclamation capability of the Thin Provisioning feature:

1. Create a thin LUN on the EMC VNX platform.
2. Create two VMFS datastores on the thin LUN.
3. Create a virtual machine with a 100 GB virtual disk on one VMFS datastore, and provision data on it.
4. Calculate the consumed space on the thin LUN before and after the migration of the virtual disk with VAAI ON.

```
[root@rtpsol20 ~]# navicli -h spa lun -list -l 99 | grep Capacity | grep GBs
User Capacity (GBs):  1000.000
Consumed Capacity (GBs):  106.207
[root@rtpsol20 ~]# navicli -h spa lun -list -l 99 | grep Capacity | grep GBs
User Capacity (GBs):  1000.000
Consumed Capacity (GBs):  5.010
```

Figure 10. Consumed space on a thin LUN before and after migration

The portion highlighted in green in Figure 10 shows the **User Capacity** and **Consumed Capacity** of a thin LUN before the migration of the virtual disk to another datastore. Because the virtual machine has a fully provisioned100 GB virtual disk, it shows a **Consumed Capacity** of 106 GB.

The portion highlighted in red in Figure 10 shows the capacity that is available after the migration of the virtual disk to another datastore. The **Consumed Capacity** shows the reclamation of all 100 GB on the thin LUN.

Key finding

The Thin Provisioning feature enables 100 percent reclamation of the dead space on a thin LUN, and avoids the waste of expensive disk space.

Conclusion

The tight integration of the EMC VNX platform with VMware vSphere5.0 and VAAI provides significant benefits and optimal performance for customers to build and maintain scalable, efficient virtual environments. The combination of EMC VNX and VMware VAAI achieves high-speed processing, reduced server load, and reduced I/O load.

The key benefits of the VAAI features are:

- The Full Copy feature speeds up the Storage vMotion or virtual machine clone operations and greatly reduces ESXi network traffic by offloading the operations to the VNX platform instead of sending the traffic through the ESXi hosts.
- The Block Zero feature speeds up the deployment of Thick Provision Eager Zeroed virtual disks by offloading the redundant and repetitive zeroing of large numbers of blocks to the VNX platform, to free ESXi host resources for other tasks.
- The Hardware-Assisted Locking feature provides a much more efficient means to avoid retries for getting a lock when many ESXi servers are sharing the same datastore. It enables the offloading of the lock mechanism to the VNX array, which does the locking at a very granular level. This provides an alternative method to protect the metadata of VMware VMFS cluster file systems and thereby improves the scalability of large ESXi servers sharing a VMFS datastore.
- The Dead Space Reclamation capability of the Thin Provisioning feature enables the reclamation of blocks from a thin-provisioned LUN on the VNX platforms. This provides the ability to overcome out-of-space conditions by temporarily pausing the virtual machine when disk space is exhausted. With this feature, the administrator can allocate additional space to the datastore, or migrate an existing virtual machine, without causing the virtual machine to fail.

References

EMC documentation

The following documents, located on EMC Powerlink®, provide additional and relevant information:

- *Using VMware vStorage APIs for Array Integration with EMC Symmetrix VMAX* — White paper
- *VMware vStorage APIs for Array Integration with EMC VNX Series for NAS* — White paper
- *Using EMC VNX Storage with VMware vSphere* — TechBook

VMware documentation

The following VMware document, located on the VMware website, also provides useful information:

- What's New in VMware vSphere™ 5.0 — Storage Technical Marketing Documentation