VMware VAAI Integration

VMware vSphere 5.0 VAAI primitive integration and performance validation with Dell Compellent Storage Center 6.0

Document revision

Date	Revision	Comments
3/29/2012	A	Initial Publication
7/30/2012	B	Update Publication

Contents

Table of Contents

Preface

Audience

The audience for this document is VMware system administrators who are responsible for the setup and maintenance of ESX/ESXi hosts and associated storage. Readers should have a working knowledge of VMware vSphere 5.0 and the Dell Compellent Storage Center.

Purpose

This document provides an overview of VMware's VAAI API to offload command to the storage array. VMware VAAI primitives were created to offload VMware vSphere storage tasks to the storage array. Performance gains garnered by the Copy Offload and Block Zeroing VAAI primitives will vary depending on workload of the storage array.

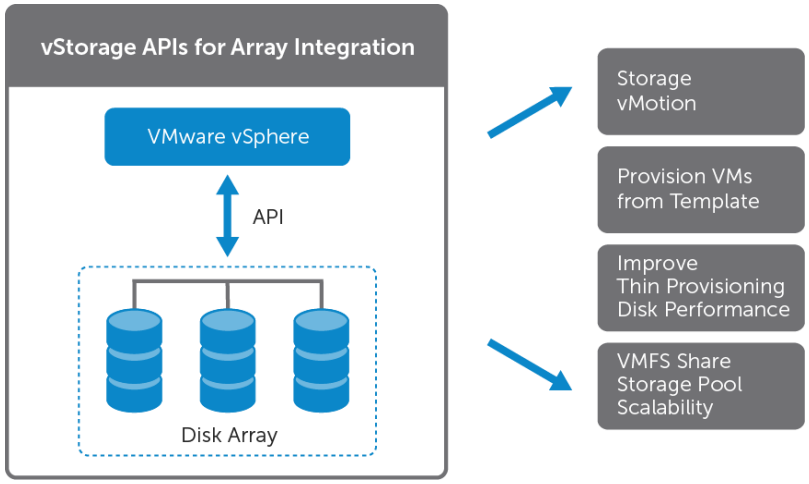
Customer support

Dell Compellent provides live support 1-866-EZSTORE (866.397.8673), 24 hours a day, 7 days a week, 365 days a year. For additional support, email Dell Compellent at support@compellent.com. Dell Compellent responds to emails during normal business hours.

Introduction

Introduction to VMware VAAI

VMware vStorage APIs for Array Integration offload specific storage operations to the Compellent Storage Center 6.0 for unparalleled performance and efficiencies. With VMware vStorage APIs for Array Integration, VMware vSphere hosts can perform key operations faster and consume less CPU, memory and storage network bandwidth. The four VMware vStorage APIs for Array Integration in VMware vSphere 5.0 are Full Copy, Block Zeroing, Hardware Assisted Locking, and Array-based Thin Provisioning.

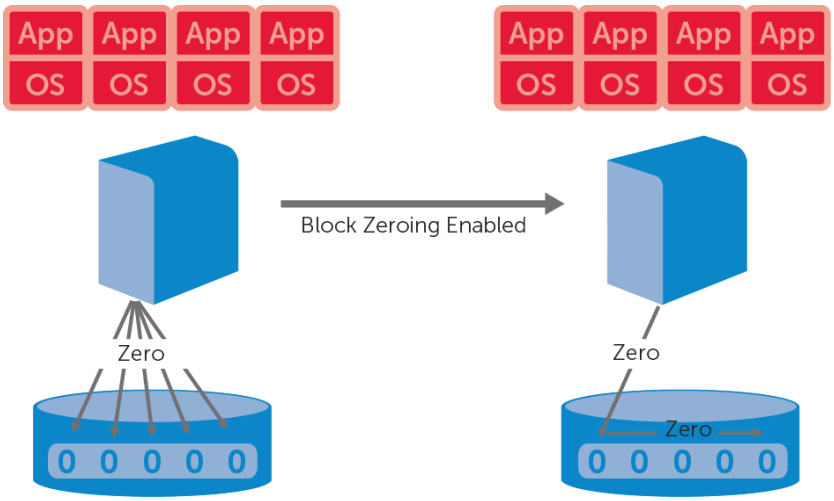


Block Zero

Fault Tolerant or Microsoft Cluster Service virtual machines require VMDKs of the type Thick Provision Eager-Zeroed . These differ from standard Thick Provision Lazy Zeroed or Thin Provision VMDKs in that the VMDK's blocks are zeroed out at the time that the VMDK is created. For large disks this can take a significant amount of time as each zero is written from the vSphere host to the array, and an acknowledgment of each write is sent back from the array to the vSphere host.

Utilizing the Block Zeroing primitive, the vSphere host off loads the task of zeroing out the blocks during the Virtual Machine provisioning to the Dell Compellent Storage Center array, permitting the vSphere host to return to its task of creating the Fault Tolerant Virtual Machine, all while the array completes the task of zeroing the disk. With Block Zeroing offloaded to the Dell Compellent Storage Center array, Fault Tolerant or Microsoft Cluster Service Virtual Machines can be created significantly faster.

Tier 1 applications have suggested that Thick Provision Eager Zeroed VMDK's are prefered because of the emination of the on first write zeroing of blocks. Dell Compellent has issued a guide which compares the different types of VMDK's and performance assosciated with each. See the appendix for a link to the Dell Compellent document.

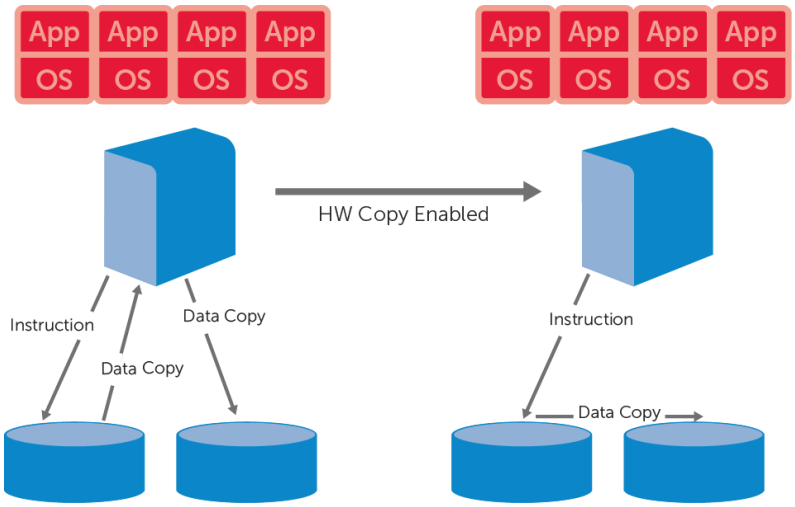


Copy Offload

A common day to day task in any IT environment is the deployment of servers in support of new business applications. Virtualization of servers changed this task from a labor intensive task of racking a server and installing the OS, to a couple of mouse clicks that deploys a Virtual Machine from a preconfigured template.

While this change has resulted in significant savings of time spent on task, there was still a significant amount of time spent waiting while the virtual machine is copied and deployed.

Traditionally, prior to Copy Offload, the process of deploying a Virtual Machine involved reading its data from the array, across the storage network to the vSphere host, and then written back across the storage network to the array. This placed a non-production workload on both the storage network and the vSphere host, in addition to the production workload of the running environment. Now, with the benefit of the Copy Offload primitive, vSphere hosts can offload this task to the Dell Compellent Storage Center array where it can be completed in a much more efficiently manner with a significant reduction in the workload on the vSphere host and on the storage network.

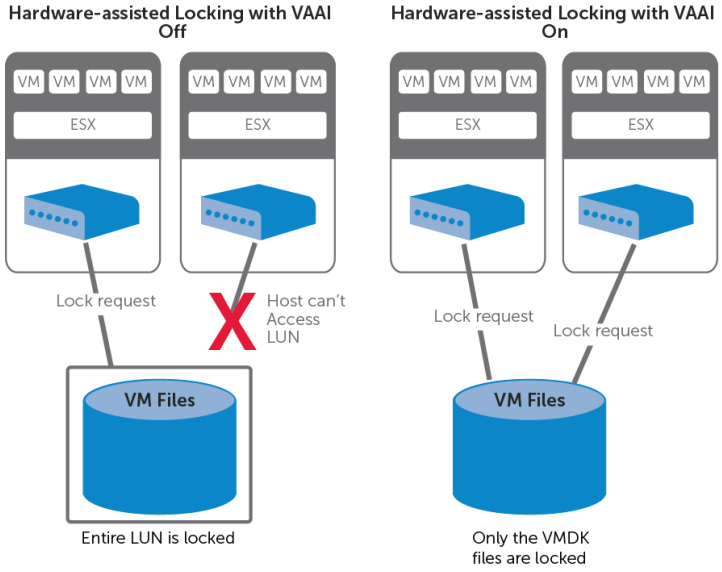


The benefits of Copy Offload do not end with deploying virtual machines from templates, but also extend to Virtual Machine tasks such as Storage vMotion, as well as Cloning of Virtual Machines.

Hardware-Assisted Locking

The Hardware-Assisted Locking primitive is a more granular locking method for protecting the VMFS metadata than the SCSI-2 reservations used by previous versions. Formerly whenever a Virtual Machine was powered on, powered off, grew a thin provisioned virtual disk, or was vMotioned to another vSphere host, a SCSI-2 reservation lock would be issued by the vSphere host to the datastore's underlying volume. This prevented other vSphere hosts from issuing a similar SCSI-2 reservation to service a similar request to change metadata. While SCSI-2 reservations are short lived, the impact can easily be observed when powering on a large number of virtual machines simultaneously, as would be typical in a Virtual Desktop Infrastructure environment.

The Hardware-Assisted Locking primitive resolves the contentention by working with the Dell Compellent array to lock only the necessary blocks rather than the entire volume. This enables other vSphere hosts to perform similar locking operations against that same volume concurrently.

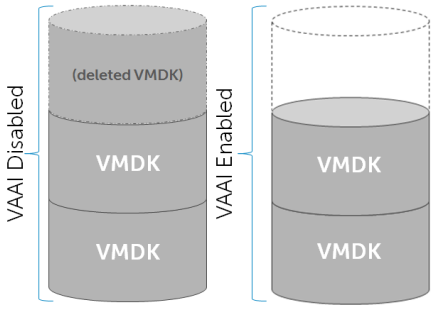


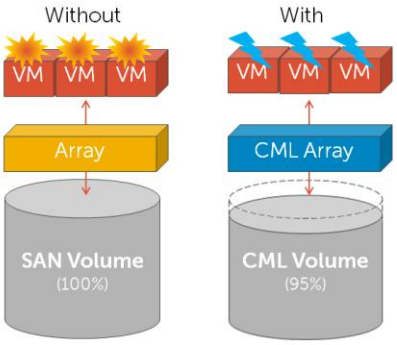
Thin Provisioning (STUN, UNMAP)

With the launch of VMware vSphere 5.0 a new primitive was introduced called the VMware APIs for Array Intergration Thin Provisioning. This new primitive was created to alliviate inefficencies with running VMware Virtual Machines on a thinly provisioned storage array. The Thin Provisioning feature is actually two separate primitive rolled into one single primitive. (See Appendix for link explaining Enabling/Disabling Thin Provisioning) The VMware API's for Thin Provisioning are certified and availible in Dell Compellent Storage Center 6.1+ code releases.

UNMAP

A new VAAI primitive (using the SCSI UNMAP command) allows an ESXi to tell the storage array that space that was occupied by a VM (whether it be deleted or migrated to another datastore) can be reclaimed. This allows an array to correctly report space consumption of a Thin Provisioned datastore, and allows customer to correctly monitor and correctly forecast new storage requirements.





STUN

If a Thin Provisioned datastore reaches 95%, only those VMs which require extra blocks of storage space will be paused, while VMs on the datastore that do not need additional space continue to run. This prevents VM's from crashing and allows the customer to allocate more storage to the Thinly Provisioned array. A warning is also raised when a thinly provisioned datastore reaches 75% utilization.

Introduction to Dell Compellent Storage Center 6.0

Storage Center 6.0 for provides a new foundation for the future of Compellent's enterprise storage, delivering the enhanced efficiency and agility SAN administrators expect from a migration to 64-bit while extending the virtualized storage capabilities of Compellent's Storage Center. Storage Center 6.0 also expands on Dell and VMware's long-standing integration and engineering collaboration. As customers accelerate their move to dynamic data centers and cloud computing environments.

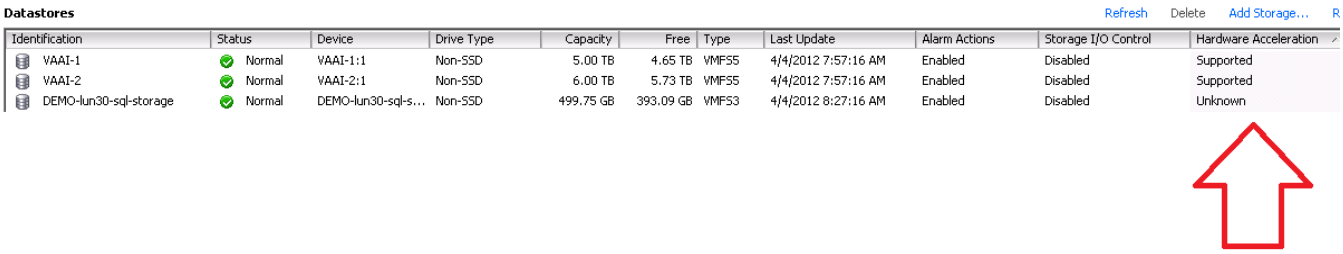
VAAI and VMware vSphere 4.1

The VAAI primitives Block Zero, Copy Offload and Hardware Assisted Locking were all available with release of VMware vSphere 4.1 but did require a special VAAI plug-in developed by each storage vendor. A Dell Compellent VAAI plugin for vSphere 4.1 will be made available on Dell Compellent Knowledge Center coniciding with the release of Dell Storage Center 6.0. The vSphere VAAI 4.1 Plugin for Dell Compellent Storage Center will be a free download.

VAAI primitives in vSphere 5.0 *do not* require any plugin's to utilize the VAAI primitives.

The VMware VAAI primitives are not easily distingusishable as to which primitives are enabled and which primitives the storage vendor supports. The Datastores view on a vSphere host shows a column for Hardware

Acceleration. If Supported is show for a Datastore, all VAAI primitives are availible. If Unknown is shown, one or more but not all of the primitives are supported. If Not Supported is shown none of the VAAI primitives are availible.



Datastores

Refresh Delete Add Storage... R

Identification	Status	Device	Drive Type	Capacity	Free	Type	Last Update	Alarm Actions	Storage I/O Control	Hardware Acceleration
VAAI-1	Normal	VAAI-1:1	Non-SSD	5.00 TB	4.65 TB	VMFS5	4/4/2012 7:57:16 AM	Enabled	Disabled	Supported
VAAI-2	Normal	VAAI-2:1	Non-SSD	6.00 TB	5.73 TB	VMFS5	4/4/2012 7:57:16 AM	Enabled	Disabled	Supported
DEMO-lun30-sql-storage	Normal	DEMO-lun30-sql-s...	Non-SSD	499.75 GB	393.09 GB	VMFS3	4/4/2012 8:27:16 AM	Enabled	Disabled	Unknown

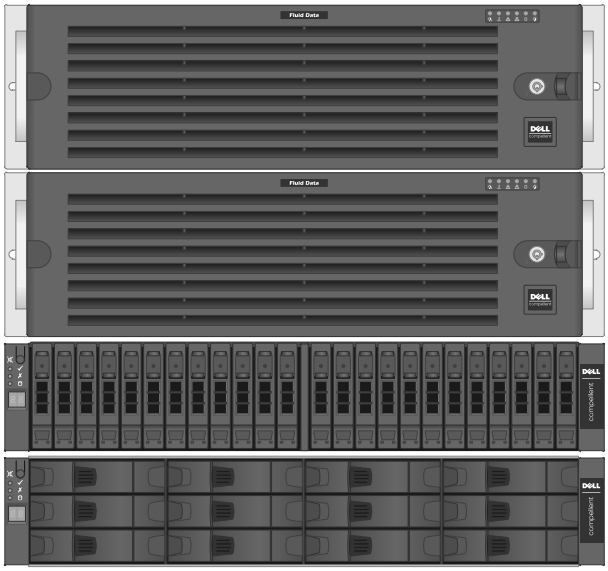
VMware VAAI Test Lab Validation

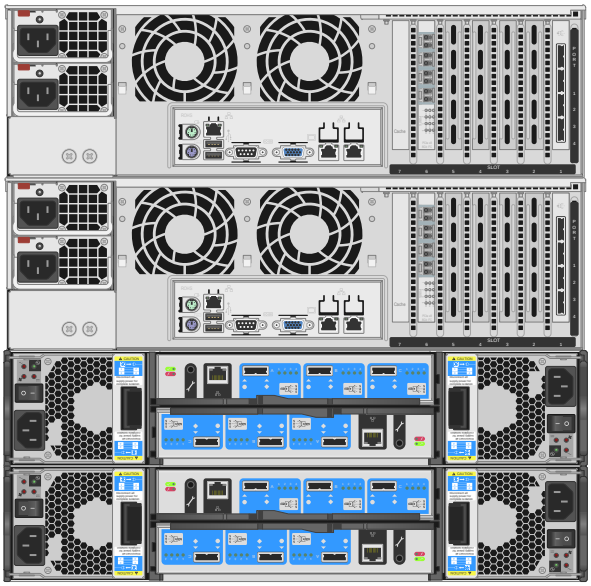
A test lab consisting of VMware vSphere 5.0 hosts and Dell Compellent Storage Center 6.0.5 was configured to validate and guage the efficiencies gained by a particular VMware VAAI primitive. The VMware VAAI primitives primary function is to offload common storage tasks to the array, freeing up vSphere host cycles of CPU, memory, storage bandwith, and network bandwith. Depending on the arrays workload, specific vSphere tasks could see a dramatic improvment in time to complete the task assigned.

The Dell Compellent Storage Center in this lab, is a single pair of Series 40 Controllers with 48 146 GB 2.5 inch SAS drives making up Tier1, and 12 2 TB 3.5 inch drives for Tier 3. The Dell Compellent controllers communicate to the disk shelves via 6GB SAS.

The Dell Compellent array will create wide stripes of RAID10, and RAID5 over both tiers of disks in Tier1, but because of the large size of the disks being utilized in Tier3 exceeding 1 TB, the Dell Compellent array defaults to dual reduncancy using RAID6 instead of RAID5.

Two Dell R610 Servers were utilized as the vSphere hosts for this test with dual port Qlogic fiber channel HBA's providing the front end connectivity back to the Dell Compellent Storage array. The vSphere hosts HBA's were configured following the Dell Compellent Best Practices Guide for vSphere 5.0.





Test Lab Environment

All the tests were executed from the following hardware and software:

Storage

- Storage Center
  - Controllers: Series 40
  - Tier 1 Disks: 24 - 146GB SAS 15k RPM (23 Active, 1 hot spare)
  - Tier 2 Disks: 12 – 2Tb SAS 7.2k RPM (11 Active, 1 hot spare)
  - Front End Connectivity: 8 Gbps Fiber Channel

Server

- (2) Dell PowerEdge R610
  - Processors – 1 x 4 Intel Xeon E5240 @ 2.50 Ghz
  - Memory – 32 GB
  - LAN – 2 - Broadcom NetXtreme II BCM5708
  - I/O Card: Qlogic QLE2532 - Dual port 8 Gbps Fiber Channel Card

Hypervisor

- VMware ESXi 5.0 Build 515841
  - Path Selection Policy: Round Robin
  - Datastore format: VMFS-5

Virtual Machine Operating System

- Microsoft Windows Server 2008 R2 (64 bit)
  - Virtual Disk Size: 40 GB Thick Provision Eager Zeroed
  - Virtual Disk Size: 40 GB Thick Provision Lazy Zeroed

Test Overview

The following tests were conducted on the test lab system to validate the VMware VAAI primitives in real world scenarios.

Block Zero

- Create a Eager Zeroed VMDK with VMware - VAAI Enabled / Disabled

Copy Offload

- Clone a 40GB Eager Zeroed Windows 2008R2 Server – VAAI Enabled / Disabled
- Clone a 40GB Lazy Zeroed Windows 2008R2 Server – VAAI Enabled / Disabled
- Storage vMotion a 40GB Eager Zeroed Windows 2008R2 Server – VAAI Enabled / Disabled
- Storage vMotion a 40GB Lazy Zeroed Windows 2008R2 Server – VAAI Enabled / Disabled

Hardware Assisted Locking

- Power On 10 Virtual Machines – VAAI Enabled / Disabled

Data collected from VMware vCenter and Dell Compellent Enterprise manager will be used to validate the tests highlighted above.

Test Results

Block Zero - Create a 40 GB Eager Zeroed VMDK with VMware - VAAI Enabled / Disabled

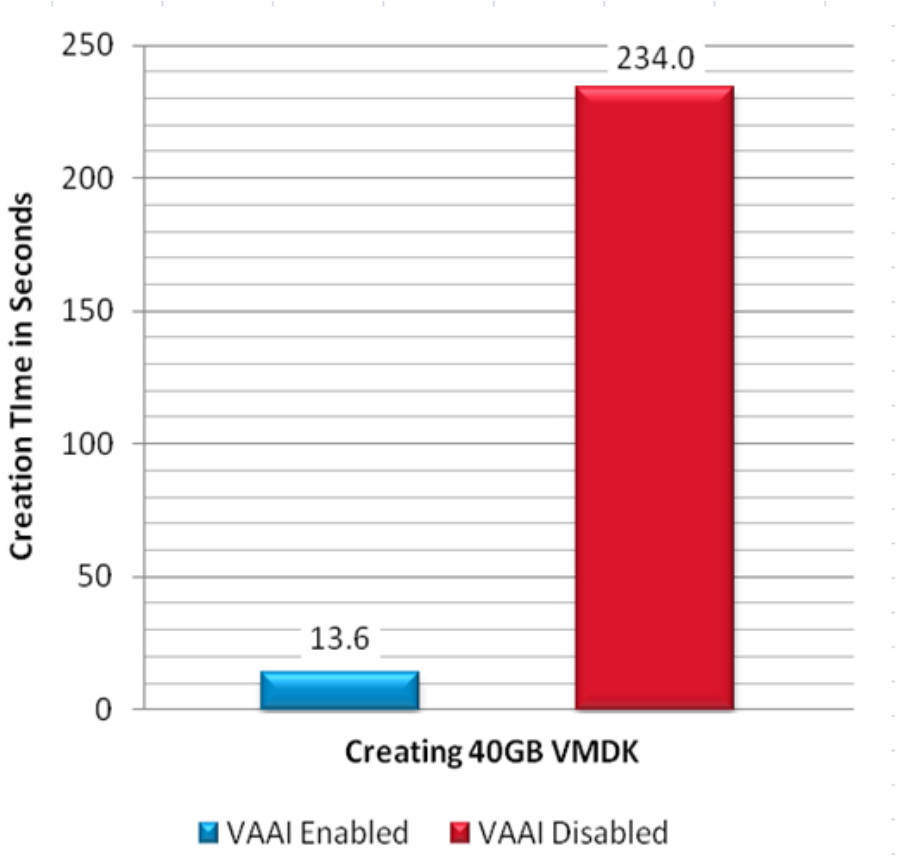
- Disable VAAI Primitive->ESXi->Advanced Setting->DataMover->DataMover.HardwareAcceleratedInit = 0
- time vmkfstools -c 40G /vmfs/volumes/VAAI-1/Test.vmdk -d eagerzeroedthick

VMware VAAI issues the Dell Compellent Storage Center a T10 SCSI WRITE SAME command 40960 times creating 1MB of Zeros in total producing a 40 GB Eager Zeroed Thick VMDK. VMware VAAI enables SCSI T10 (WRITE SAME) commands to be sent to the array to write 40GB of worth of zeros, thus offloading the payload of that task to the array instead of sending hundreds of megabytes of zeros over the storage network.

Utilizing the VAAISTATS ESXTOP counters, notice that the ZERO column increments up by Megabytes zeroed. In this example a 40 GB VMDK, which caused the VAAI ZERO counter to increase by 40960 MB. (40960 MB / 1024 MB = 40 GB)

The results of this test show that the VMware VAAI Block Zero primitive can significantly decrease the amount of time needed to create a VMware Thick Eager Zeroed VMDK. Using the VMware VAAI Block Zero showed a reduction of 1621% in the time it takes to create a Thick Eager Zeroed VMDK. The offload capabilities with VMware VAAI also offloaded the majority of the storage network throughput to the storage array freeing up cycles on the network storage fabric.

Size of VMDK (GB)	VAAI Enabled /sec	VAAI Disabled /sec	%Time Reduction
40	13.6	234.0	1621%



Copy Offload – Cloning and Storage vMotion of 40 GB Eager Zeroed and 40 GB Lazy Zeroed VMDK

- Disable VAAI Primitive->ESXi->Advanced Setting->DataMover->DataMover.HardwareAcceleratedMove = 0
- vCenter Right click on template source VM -> Clone -> New Destination Datastore

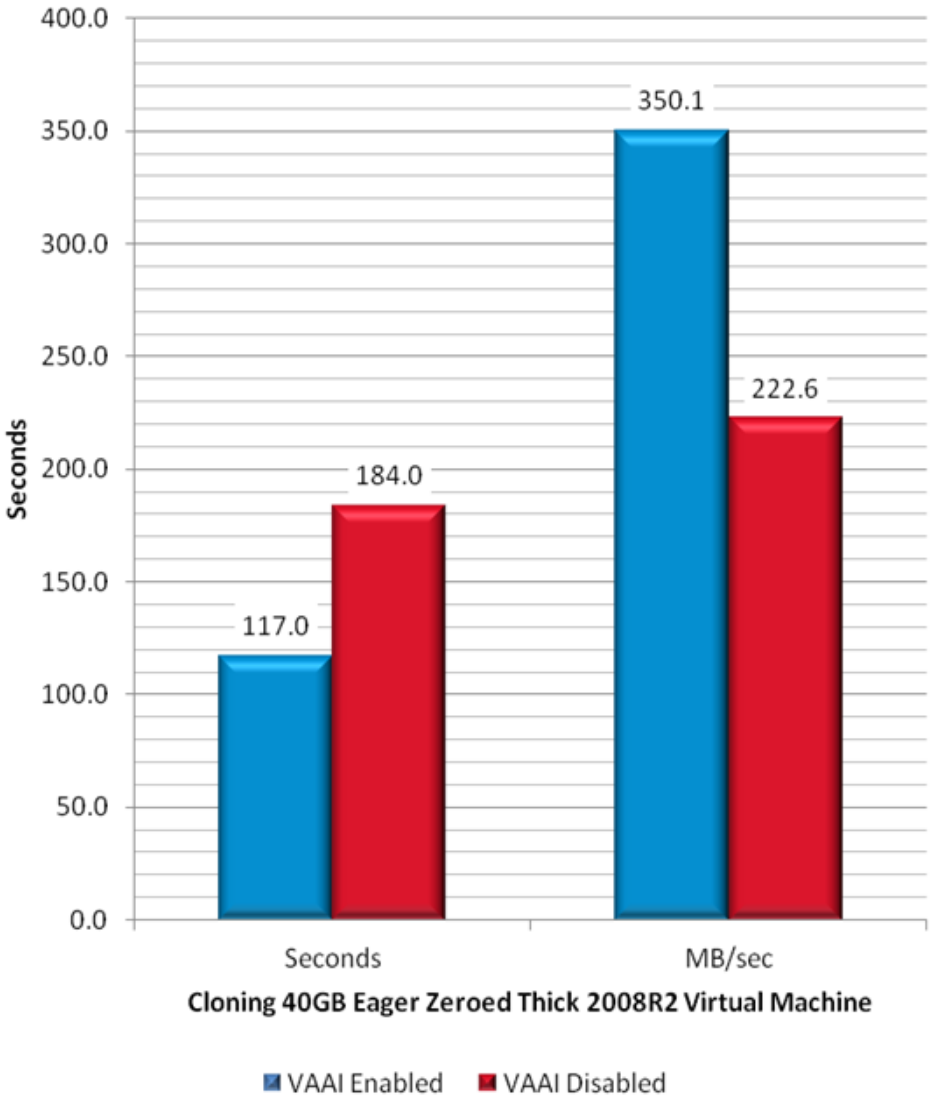
Validation of Copy Offload used the VAAISTATS ESXTOP counters CLONE_RD (Reads) and CLONE_WR (Writes) to determine if VMware Copy Offload VAAI primitive was being utilized to offload the copy of data. VMware vCenter task completion time was used to determine the total amount of time to complete the task and validate the test.

VMware VAAI issues the Dell Compellent Storage array T10 SCSI XCOPY commands when a telling the Dell Compellent array the source and destination volumes and source VMDK LBA's to be copied. Offloading the payload of the copy to the array removes the burden of the copy on the host CPU, memory, and storage network fabric.

The results of this test show that the VMware VAAI Copy Offload primitive can significantly decrease the amount of time needed to perform cloning and storage vMotion tasks within VMware. Using the VMware VAAI Copy Offload offers a significant decrease in the time it takes to clone or storage vMotion a Thick Eager Zeroed or a Thick Lazy Zeroed VMDK. The offload capabilities with VMware VAAI also offloaded the majority of the storage network throughput to the storage array freeing up cycles on the host CPU, memory, and storage network fabric.

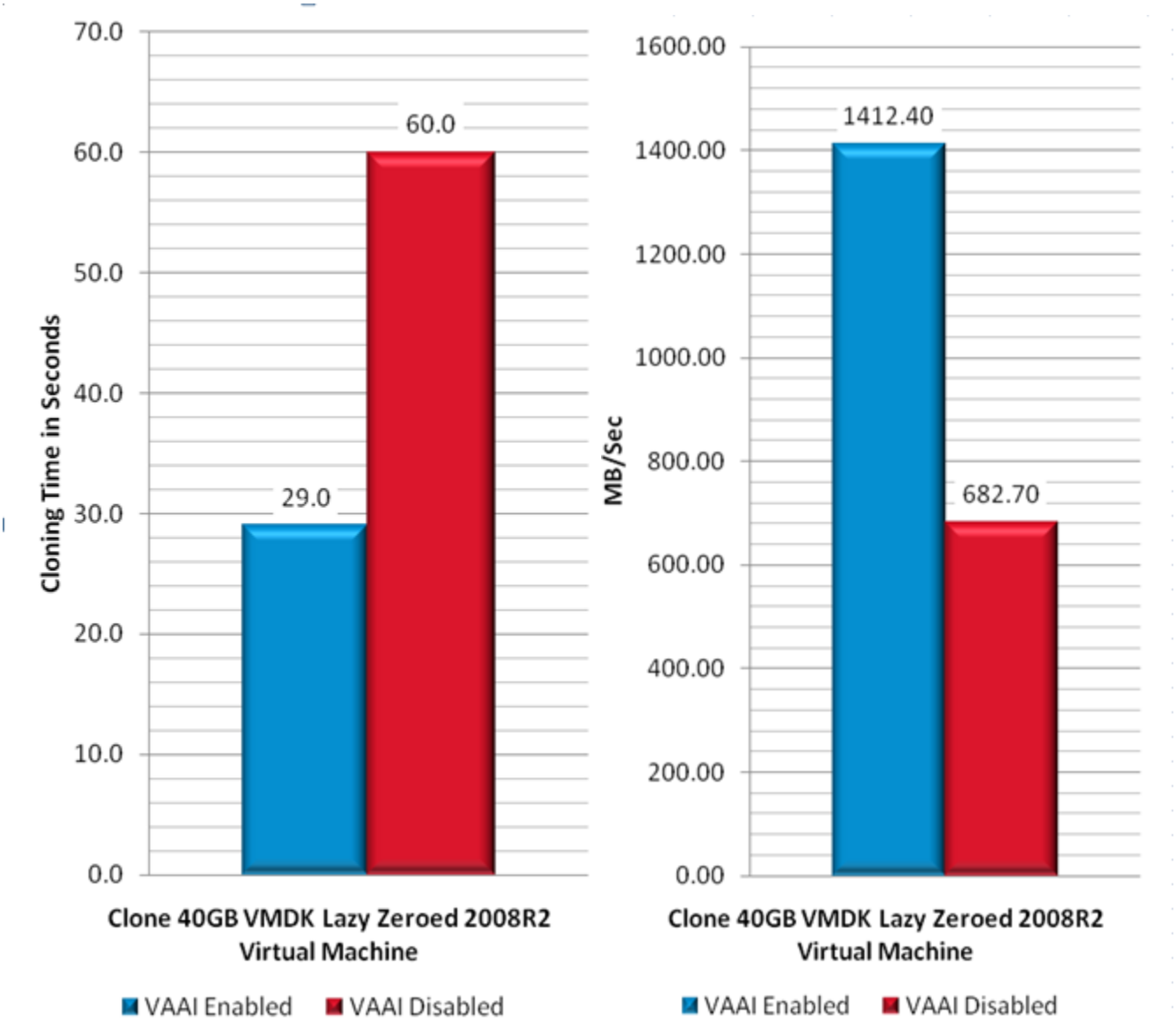
Cloning 40GB VMDK Eager Zeroed W2K8R2

Clone 40GB VMDK Eager Zeroed W2K8R2	Seconds	MB/sec
VAAI Enabled	117.0	350.1
VAAI Disabled	184.0	222.6



Cloning 40GB VMDK Lazy Zeroed W2K8R2

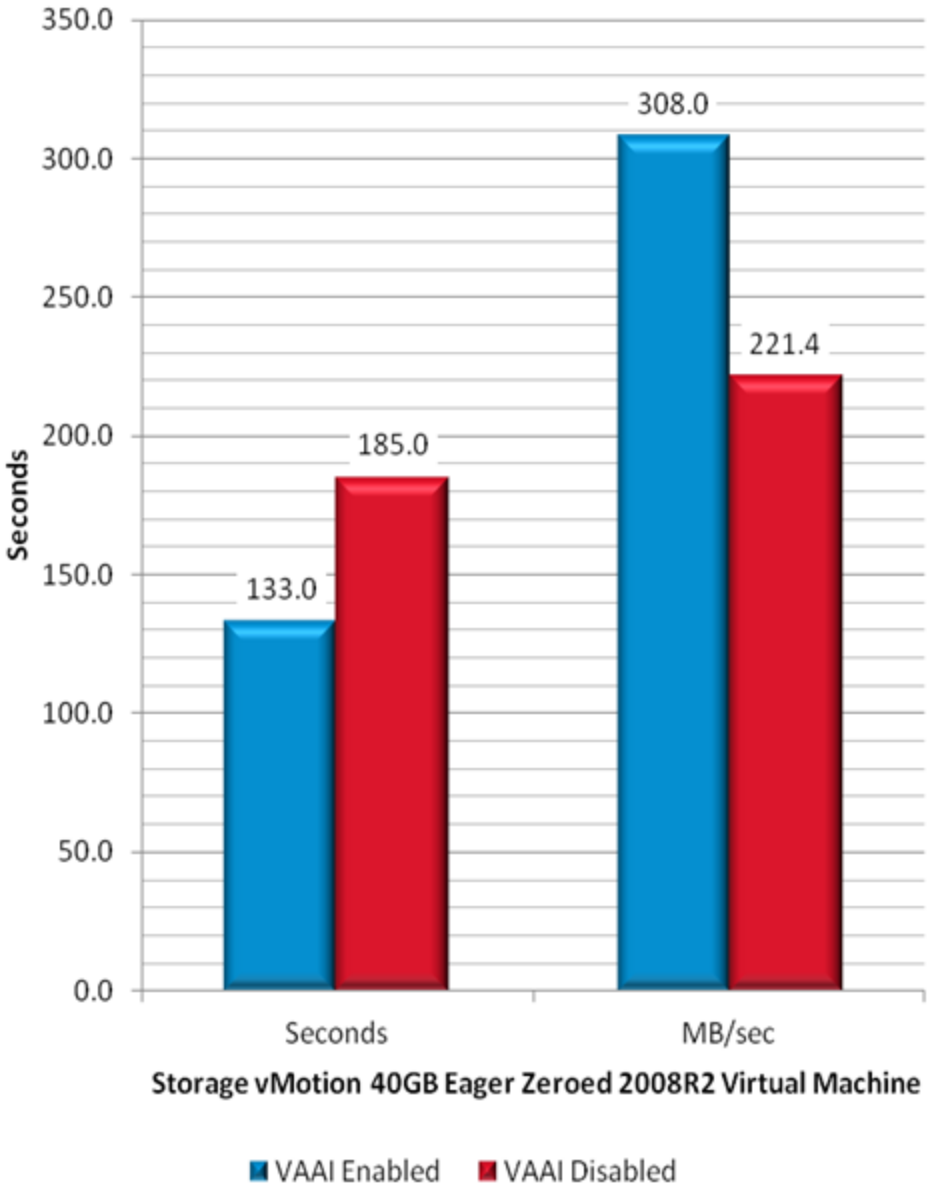
Clone 40GB VMDK Lazy Zeroed W2K8R2	Seconds	MB/sec
VAAI Enabled	29.0	1412.4
VAAI Disabled	60.0	682.7



Storage vMotion 40GB VMDK Eager Zeroed W2K8R2

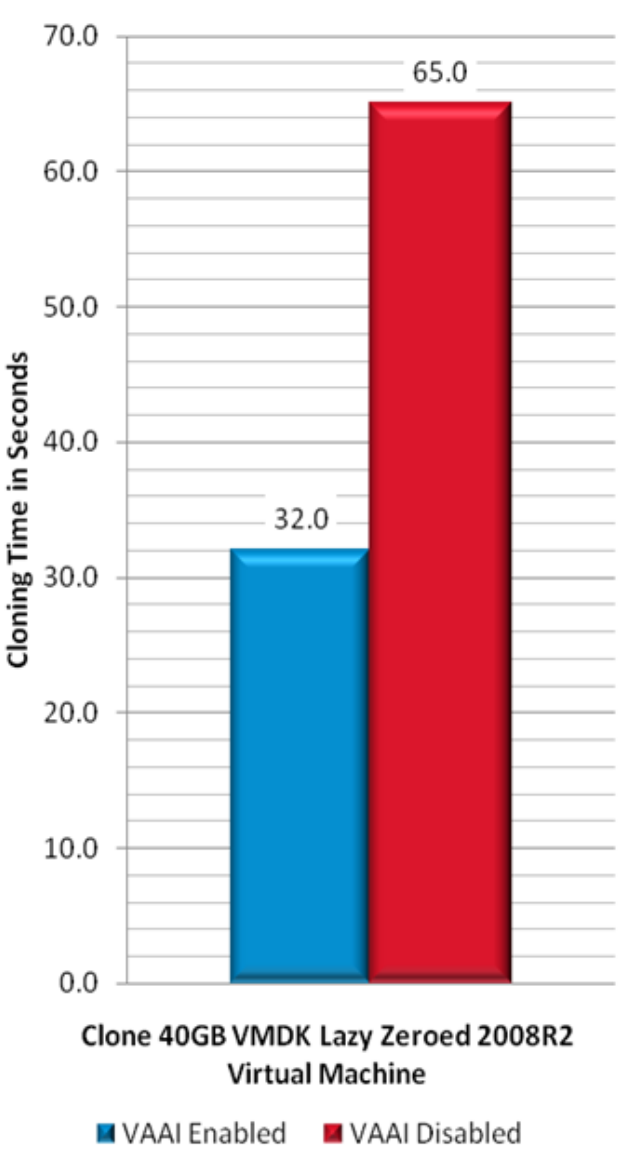
- Disable VAAI Primitive->ESXi->Advanced Setting->DataMover->DataMover.HardwareAcceleratedMove = 0
- vCenter Right click on template source VM -> Migrate -> Change Datastore -> New Destination Datastore

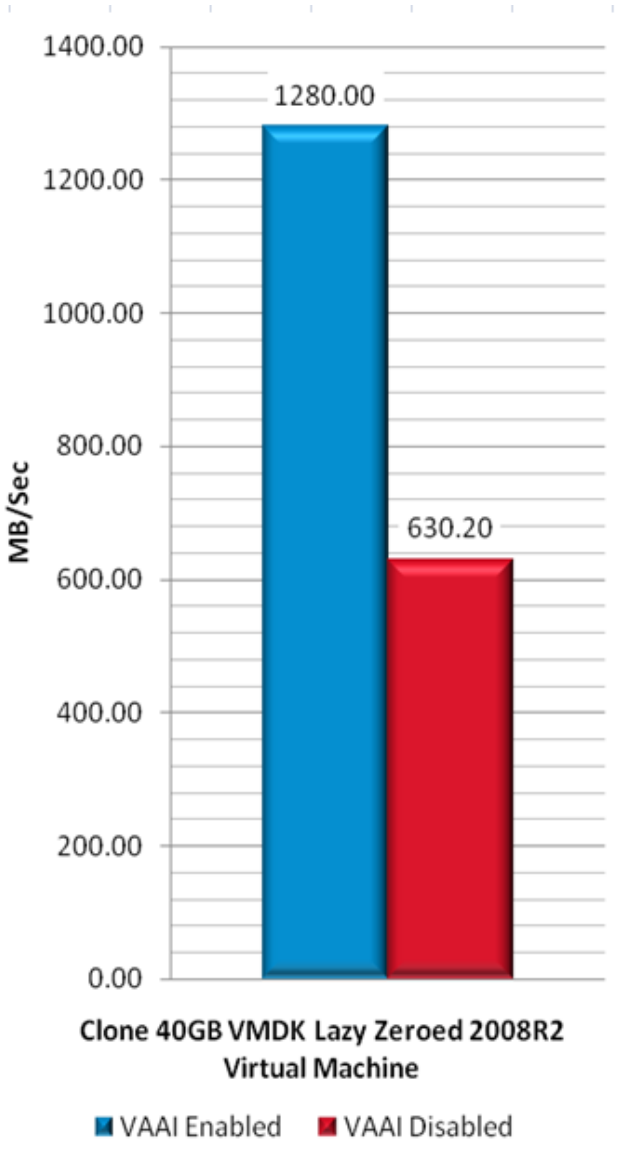
Storage vMotion 40 GB Eager Zeroed Thick VM	Seconds	MB/sec
VAAI Enabled	133.0	308.0
VAAI Disabled	185.0	221.4



Storage vMotion 40GB VMDK Lazy Zeroed W2K8R2

Storage vMotion 40GB VMDK Lazy Zeroed W2K8R2	Seconds	MB/sec
VAAI Enabled	32.0	1280.0
VAAI Disabled	65.0	630.2





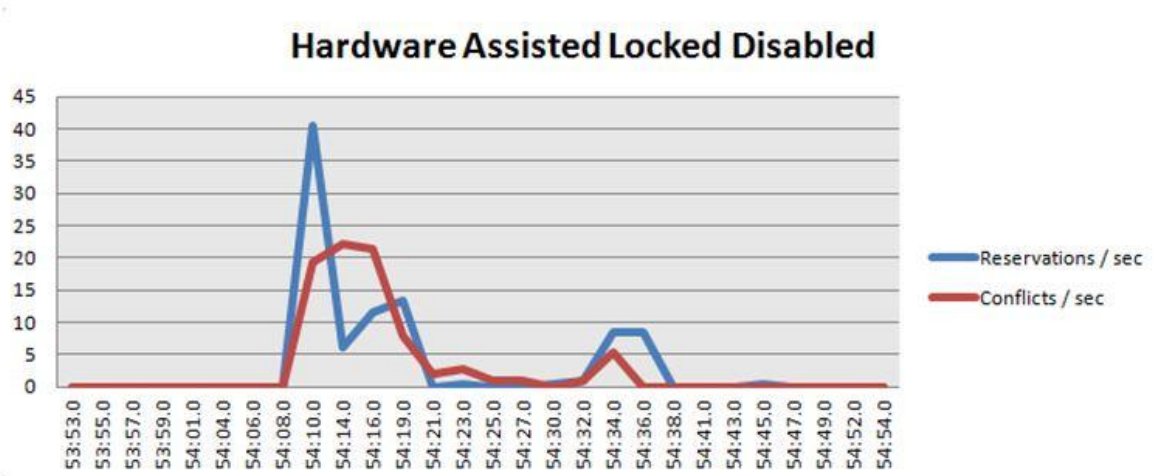
Hardware Assisted Locking – Power on 10 VM's simultaneously running on a single Datastore across 2 hosts.

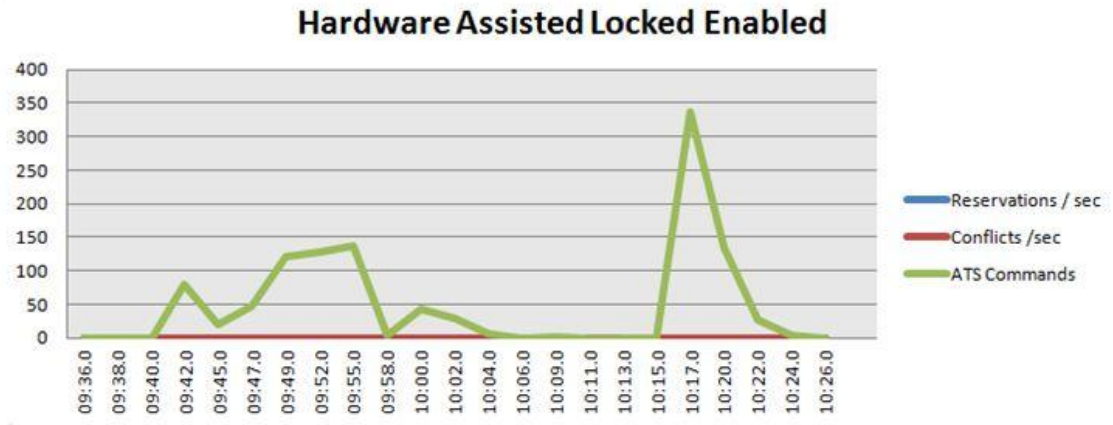
- Disable VAAI Primitive->ESXi->Advanced Setting->VMFS3->VMFS3.HardwareAcceleratedLocking = 0
- vCenter -> Right click on 10 guests -> Power -> Power On

VMware VAAI issues the Dell Compellent Storage Array T10 SCSI ATOMIC TEST AND SET commands locking only the small blocks of metadata that needs to be updated instead of locking the entire LUN. Allowing the small portion of metadata to be locked instead of the entire LUN offloads the task of locking the LUN to the Dell Compellent Storage Center eliminating the requirement from the vSphere host level.

Utilizing the VAAISTATS and RESV STATS ESXTOP counters, notice that the RESV/s (SCSI-2 reservations per second) column increments during the power-on of the ten VM's. Also notice that the CONS/s (conflicts per second) increases for a few moments while the ten VM's are powered on over the two vSphere hosts.

The results of this test show that the VMware VAAI Hardware Assisted Locking primitive can eliminate the SCSI-2 reservations required to lock a LUN when changing metadata for a VM. Using the VMware VAAI Hardware Assisted Locking notice that SCSI-2 reservations were eliminated during the power on phase of the ten VM's. Hardware Assisted Locking is a valuable asset in large vSphere host farms and alleviates the SCSI-2 reservation locking on shared datastores.





Conclusion

The Dell Compellent Storage Center along with VMware ESXi 5.0 and VAAI allows robust, efficient, and scalable virtual environments. VMware VAAI can be used for its tremendous offloading capabilities, but as shown in our lab validation, the offload also introduces performance efficiencies in addition to the offloading of tasks from the VMware hosts.

Block Zeroing has a tremendous upside for Dell Compellent VMware environments. Because of the speed and efficiency associated with creating eager zeroed thick disks on a Dell Compellent Storage Center, eager zeroed disks have little to no downside and offer significant upside for Tier 1 applications.

Copy Offload showed a significant increase in performance in cloning and Storage vMotion when working with VMware lazy zeroed VMDK's. Eager zeroed VMDK's also showed a significant increase but because of the need to copy all of the zeros in the VMDK the time overall is greater than the copy or Storage vMotion of a lazy zeroed VMDK. Without the speed improvement VAAI is a welcome relief to the VMware host, offloading the heavy lifting of the copies to the storage array.

Hardware Assisted Locking is another valuable tool to eliminate SCSI-2 reservation conflicts and allows the array to lock the small blocks of metadata to be updated during certain VMware tasks. Hardware Accelerated Locking does not address all of the scalability issues with VMFS disk format but does eliminated one of the most common contention issues with VMFS.

Appendix

VMware re-enables VAAI primitive Thin Provisioning

http://blogs.vmware.com/vsphere/2012/03/vaai-thin-provisioning-block-reclaimunmap-is-back-in-50u1.html

Dell Compellent VMware 5.0 Best Practices

http://kcint.compellent.com/Published%20Documents/Dell%20Compellent%20Best%20Practices%20with%20VMware%20vSphere%205.x.pdf

VMware vStorage APIs FAQ

http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1021976

Lazy Zeroed vs Eager Zeroed Performance Comparison

http://kcint.compellent.com/Published%20Documents/Lazy_vs_Eager_Zeroed_Performance_Comparison.pdf