Taking VMware vStorage API for Array Integration (VAAI) to the Next Dimension

Technical White Paper

FalconStor®

Taking VMware vStorage API for Array Integration (VAAI) to the Next Dimension

Technical White Paper

FalconStor Software, Inc.
2 Huntington Quadrangle, Suite 2S01
Melville, NY 11747
Phone: 631-777-5188
Fax: 631-501-7633
Website: www.falconstor.com

1.4.2012

Contents

Overview

Server virtualization has brought many benefits to data centers in recent times. One of the biggest resulting benefits is server consolidation, or the ability to host more guest operating systems using less hardware. This enables organizations to save on rack space, power, and cooling, for better overall resource utilization.

Another benefit of server virtualization is hardware abstraction, where each guest operating system runs in an emulated environment with a standard set of virtual hardware. This includes video cards, network cards, SCSI controllers, chipsets, and so on. Because this emulated virtual hardware is the same across any physical server running the same virtualization platform, a guest operating system can be easily relocated from one physical server to another, without an IT organization having to go through a re-installation or re-deployment. For the same reason, a single operating system (OS) image template for a Microsoft Windows or a Linux installation can be re-used and re-deployed as many times as necessary, across any x86 physical server make and model. The hardware abstraction aspect of server virtualization is extremely valuable for rapid deployments of new servers or for hardware upgrades.

VMware, in particular, has been a leader in introducing server virtualization features to enable transparent data and application movement both within a data center and beyond, extending into the cloud. For example, the VMware vSphere platform offers features such as VMware vMotion, VMware Dynamic Resource Scheduling (DRS), VMware Distributed Power Management (DPM), VMware Fault-Tolerance (FT), and VMware High Availability (HA), which allow virtual machines running in the same cluster group of vSphere hosts to operate optimally, with greater levels of protection, efficiency, and manageability.

With the availability of server motherboards supporting more CPU sockets, more RAM, and more I/O buses, along with AMD and Intel processors now sporting configurations of over eight (8) cores, the server consolidation ratio (the number of virtual machines that can be hosted per vSphere host) has increased significantly over the past several years.

However, with more virtual machines per host, a VMware vSphere hypervisor has to spend more compute resources handling network packet routing, as well as performing storage I/O routing. Not only do certain disk operations consume CPU and memory resources due to their constant stream of I/Os; but SAN data paths can become saturated due to their inherent high throughput. In addition, file-locking operations on the clustered VMware Virtual Machine File System (VMFS) become more frequent as the number of virtual machines increase in a VMware vSphere cluster, leading to severe performance impacts as well as occasional task failures. This is particularly true when expanding snapshot file space, expanding a thinly provisioned disk, or powering a virtual machine on or off. As a result, VMware technology can potentially present scalability challenges.

VAAI Primitives

In an attempt to alleviate scalability issues, VMware provides the VMware vStorage API for Array Integration (VAAI) for VMware vSphere v5.0. VMware VAAI is a storage-based hardware acceleration feature designed to offload burdens and bottlenecks from the hypervisor and onto supported storage subsystems.

VMware VAAI has three fundamental operations, known as “primitives”, along with a fourth optional one:

- **Hardware Assisted Locking or Atomic Test and Set (ATS)** – This primitive helps offload file-locking operations for a VMware VMFS volume, such as during a file creation or file expansion. With this feature, SCSI reservation locks at the LUN level are no longer used, and SCSI reservation conflicts are no longer an issue. As a result, many more virtual machines can be stored and managed on the same VMware VMFS volume without causing file-locking disruptions.

- **Full Copy** – This primitive allows a large file, such as a .vmdk disk image, to be copied from one location to another on the same array (same or different VMware VMFS data store), without having the hypervisor actually read and write every block between the source and the destination volume(s). This significantly reduces the time it takes for the file copy to complete, and alleviates the volume of I/O traffic between the VMware vSphere host, the SAN fabric, and the array controllers. The result is a reduction in I/O utilization on the SAN and CPU/RAM utilization on the VMware vSphere host during a virtual machine cloning task, a ‘Deploy from Template’ task, or VMware a Storage vMotion task. This frees up additional resources for other tasks or functions.

- **Block Zeroing** – Providing similar benefits as Full Copy, this primitive allows the hypervisor to send down the block list of a newly created “eager zeroed thick” virtual disk to the supported array. Creating an eager zeroed thick disk is similar to performing a low-level format on a physical disk, where every single block of the disk is written with a “0” to bring it to a “factory new” state. Without this primitive, the hypervisor would need to continue sending down countless amounts of write commands to the array just to instruct it to write zeros across the entire .vmdk image file. This process is time-consuming, and as is the case with a file copy, it actually can saturate the I/O path to the array with all zeros as the data value for each block to be overwritten. With this primitive enabled, only the block list itself is sent to the array; with the block list, the array itself can handle the entire “zeroing” of these blocks, and once complete, it can send back to the hypervisor the completion status for creating the “eager zeroed thick disk.”

- **Thin-Provisioning Stun –** This optional primitive allows proper handling of array-based thin-provisioned VMware VMFS volumes under situations such as .vmdk image file deletion (reclamation of storage space occupied by the deleted block list for the .vmdk file) or .vmdk image file expansion. For the latter condition, if a thin-provisioned LUN is about to run out of space, the array would report the low-space status to the vSphere host in advance, before the situation becomes critical.

VMware VAAI makes storage-based hardware acceleration extremely appealing in VMware vSphere environments. After all, who would say “no” to faster template deployments, shorter storage vMotion tasks, faster disk creation, higher VM consolidation ratios, and lower SAN data path utilization, along with lower CPU and memory consumption? However, it is important to realize that VMware VAAI has fundamental limitations that can limit its support scope.

For starters, the hardware acceleration features of the three main primitives only work with VMware VAAI-compliant storage arrays. Unless you are dealing with a tier-one array from one of the major enterprise-level storage vendors, chances are that your lower-end array will not support VMware VAAI. On the opposite side of the spectrum, in a large SAN environment, even if all of your SAN arrays were VMware VAAI-compliant, the three key primitives would only function within a single array enclosure. To clarify:

- You cannot use Hardware Assisted Locking if your array is not VMware VAAI-compliant and/or your VMware VMFS volume is spanning multiple VMFS extents spread across different arrays.
- You cannot use Full Copy if your array is not VMware VAAI-compliant and/or your source and destination VMware VMFS volumes are LUNs owned by different arrays.
- You cannot use Block-Zeroing if your array is not VMware VAAI-compliant

For further details about the primitives of VMware VAAI, and some of their limitations or constraints, please refer to the following Knowledge Base article (# 1021976) on the VMware KB website: http://kb.vmware.com/kb/1021976

FalconStor Value Proposition

The FalconStor® Network Storage Server (NSS) Gateway Appliance solution (v7.0) is a fully certified VMware Storage Virtualization Device (SVD) offering full compliance with all three key VMware VAAI primitives. This storage virtualization solution helps offload resource-intensive software-based operations onto any VMware HCL SAN array, even those who do not provide native VMware VAAI support. Moreover, the clustered FalconStor NSS Gateway Appliance solution helps extend the use of VMware VAAI storage hardware acceleration features across any heterogeneous SAN array environment, including a group of storage arrays, regardless of whether or not each array has native VMware VAAI support, and regardless of whether or not the VMware VMFS volumes involved are stored in the same array enclosure. VMware VAAI hardware acceleration is no longer constrained to a single array enclosure.

By supporting and complementing VMware VAAI, the FalconStor NSS Gateway Appliance enables the following capabilities, which would otherwise not be possible:

- Hardware-accelerated VMware Storage vMotion of a virtual machine between a VMware VMFS volume stored on array #1 and a VMware VMFS volume stored on array #2, even if neither array provides native VMware VAAI support.

- Cloning a virtual machine or deploying a virtual machine from a template stored on a VMFS volume on array #1 to a VMFS volume stored on array #2, even if neither array provides native VMware VAAI support.

- Hardware-accelerated creation of "Eager Zeroed Thick" virtual disk images on any SAN array, even if the array attached to the SVD offers no VMware VAAI support.

- Hardware-assisted locking of VMware VMFS file creation or expansion, resulting in much higher virtual machine consolidation ratios, even if the array has no native VMware VAAI support. This higher consolidation ratio is achieved by reducing reservation conflicts on VMware VMFS volumes suffering frequent file-locking activities. Reservation conflicts are known to cause severe task failures or task delays, such as during VMware Snapshot redo log file expansion (which is often witnessed when a VMware vStorage API for Data Protection-enabled backup [VADP backup] is in progress), during a VMware thin-provisioned disk file expansion, or when multiple virtual machines need to be powered on simultaneously.

- Support for VMware vCenter Site Recovery Manager, including certification for VMware vCenter Site Recovery Manager 5, across heterogeneous storage environments where the source and destination disk arrays are from different SAN storage vendors.

- Application-aware snapshot support for all Microsoft Windows and Linux virtual machines via the FalconStor® Snapshot Director for VMware Environments, for any array on the VMware SAN HCL (http://www.vmware.com/resources/compatibility/search.php?deviceCategory=san), and for any third-party application listed on the FalconStor Application and Database certification matrix: (http://www.falconstor.com/certification-matrix/applications-and-databases).

- Synchronous or asynchronous mirroring of VMware VMFS volumes and RDM disks across LUNs from two different array enclosures, including arrays from different vendors. The FalconStor NSS mirroring capability even doubles READ IOPS by alternating READ SCSI commands across the VMware VMFS primary and mirror LUNs.

- Advanced Read Cache and Write Cache I/O Acceleration algorithms via SafeCache™ and HotZone® technologies, which leverage standard solid-state (SSD) disks. Like other features of the FalconStor NSS solution, SafeCache and HotZone can be enabled on any VMware HCL-certified SAN array attached to the device, even if the SAN array does not provide built-in I/O acceleration support. These I/O acceleration features can be used not only for VMware View, but in VMware vCloud Director environments or standard VMware vSphere environments for virtual machines running high IOPS server applications or databases.

Conclusion

VMware VAAI addresses several performance issues introduced when the consolidation ratio of virtual machines to physical host servers increases. The VMware VAAI primitives have proven to be very efficient at boosting the performance of certain tasks, such as deployment from virtual machine templates, virtual machine cloning, and VMware Storage vMotion, as well as helping VMware vSphere clusters scale to higher numbers of virtual machines thanks to storage acceleration improvements made in file-locking. However, VMware VAAI only works with certain compliant storage arrays, and only when virtual machines have their data files stored on RDM or VMFS volumes coming from a single array.

The FalconStor NSS Gateway Appliance resolves all of these inherent limitations by offloading all VMware VAAI-supported operations using any combination of arrays listed on the VMware SAN HCL, even if these arrays are not VMware VAAI-compliant, and by allowing VMware VAAI-supported operations to span multiple storage arrays simultaneously. In addition, the clustered FalconStor NSS Gateway Appliance solution provides the unique ability to mirror an RDM or VMware VMFS volume across disks coming from different arrays for cross-cabinet redundancy, and supports SAN-based replication across data centers between heterogeneous array configurations.

Because virtual environments require the same level of protection as physical environments do, FalconStor is dedicated to providing the highest levels of data availability, consistency, scalability, and integration. FalconStor NSS provides storage virtualization in the same way VMware offers server virtualization: with vendor-agnostic, heterogeneous hardware support; and with features that improve availability, transparency, mobility, and data protection.