Integration with VMware VAAI

Application Note

July 2022

ANNOUNCEMENT

Copyright

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

Trademarks

- QSAN, the QSAN logo, QSAN.com, XCubeFAS, XCubeSAN, XCubeNXT, XCubeNAS, XCubeDAS, XEVO, SANOS, and QSM are trademarks or registered trademarks of QSAN Technology, Inc.
- Microsoft, Windows, Windows Server, and Hyper-V are trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries.
- Linux is a trademark of Linus Torvalds in the United States and/or other countries.
- UNIX is a registered trademark of The Open Group in the United States and other countries.
- Mac and OS X are trademarks of Apple Inc., registered in the U.S. and other countries.
- Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
- VMware, ESXi, and vSphere are registered trademarks or trademarks of VMware, Inc. in the United States and/or other countries.
- Citrix and Xen are registered trademarks or trademarks of Citrix Systems, Inc. in the United States and/or other countries.
- Other trademarks and trade names used in this document to refer to either the entities claiming the marks and names or their products are the property of their respective owners.

TABLE OF CONTENTS

FIGURES

TABLES

NOTICES

Information contained in this document has been reviewed for accuracy. But it could include typographical errors or technical inaccuracies. Changes are made to the document periodically. These changes will be incorporated in new editions of the publication. QSAN may make improvements or changes in the products. All features, functionality, and product specifications are subject to change without prior notice or obligation. All statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

PREFACE

Executive Summary

In virtualization and cloud environments, the ever-increasing data production and demand continue to grow, resulting in an increasing demand for high-speed data transmission. Considering the consumption of server and network resources, budget and limited IT resources, it is necessary to find ways to optimize the existing IT resources within the organization.

The VAAI (VMware vSphere Storage APIs for Array Integration) supports direct data transfer in a compatible storage system without data transfer through the host. It can optimize system capacity and performance without increasing cost or complexity. With VAAI, servers can reduce the burden of daily data transmission tasks, thereby reducing the load on servers, SANs (Storage Area Networks), and NASs (Network Attached Storages).

VAAI reduces the burden on the server by using read/write operations to transfer data at the storage array level. Compared with the traditional data transmission method, it also greatly improves the transmission speed. Now, QSAN XCubeFAS, XCubeSAN, and XCubeNAS series are fully compatible with VAAI and has passed VMware Ready certification.

QSAN storage and VMware ESXi provide an efficient and cost-effective solution. It also optimizes IT resources and provides agility solutions for ever-increasing data.

Audience

This document is applicable for QSAN customers and partners who are familiar with QSAN products. Any settings which are configured with basic operations will not be detailed in this document. If there is any question, please refer to the user manuals of products, or contact QSAN support for further assistance.

Technical Support

Do you have any questions or need help trouble-shooting a problem? Please contact QSAN Support, we will reply to you as soon as possible.

- Via the Web: https://www.qsan.com/technical_support
- Via Telephone: +886-2-77206355
- (Service hours: 09:30 - 18:00, Monday - Friday, UTC+8)
- Via Skype Chat, Skype ID: qsan.support
- (Service hours: 09:30 - 02:00, Monday - Friday, UTC+8, Summer time: 09:30 - 01:00)
- Via Email: support@qsan.com

Information, Tip, and Caution

This document uses the following symbols to draw attention to important safety and operational information.

INFORMATION

INFORMATION provides useful knowledge, definition, or terminology for reference.

TIP

TIP provides helpful suggestions for performing tasks more effectively.

CAUTION

CAUTION indicates that failure to take a specified action could result in damage to the system.

---

1. Introduction to VMware VAAI

VAAI (VMware vSphere Storage APIs for Array Integration) is an API (Application Program Interface) framework that enables many storage tasks, such as Thin Provisioning, Full Copy, Block Zero, and Hardware Assisted Locking. In XCubeFAS, XCubeSAN, XCubeNXT, and XCubeNAS series products with VMware ESXi version 5.x or later, VAAI is supported and fully integrated. When performing storage-related tasks between the VMware ESXi hypervisor and QSAN storage products, this integration can save resources on the VMware ESXi server.

VAAI was introduced in VMware vSphere 4.1 with the following features implemented for achieving offload capabilities:

- Full Copy or Hardware Assisted Move
- Block Zero or Hardware-Assisted Zero
- Hardware Assisted Locking or Atomic Test and Set

Thin Provisioning was introduced in VMware vSphere 5.x. Detailed explanations of these features are presented as following.

1.1. Thin Provisioning

For scenarios where storage-based Thin Provisioning functions are used, VMware vSphere 5.x implements some VAAI enhancements, and QSAN storage products also support this function. The two main enhancements of VAAI Thin Provisioning are:

- Dead Space Reclamation (also known as UNMAP)
- Out of space conditions

1.1.1. Dead Space Reclamation

Traditionally, when a storage volume/LUN was mounted as a datastore, and there were virtual machines stored in the datastore, if any of virtual machines were deleted or migrated, the storage spaces which were occupied by the deleted/migrated virtual machines would still be treated as “in use” from the point of view of storage array. This may lead to a situation where the use of storage space is considered insufficient and the cost of purchasing disks has been wasted.

In the QSAN storage products, this problem can be avoided by providing the function of reclaiming the unused storage space (migrating or deleting virtual machines) when using Thin Provisioning volumes and reflecting it on the management interface of QSAN storage products.

Advantage

Using this function, the unused storage space can be accurately reported to the QSAN storage system, so that the space can be correctly reclaimed through the space reclamation on the Thin Provisioning volume/LUN in the QSAN storage system.

Theory

When using Thin Provisioning volumes/LUNs, after deleting the virtual disk or migrating the virtual disk to another datastore, or even after deleting the snapshot, VMware vSphere 5.x will use the SCSI UNMAP command to immediately release the physical space on the volume/LUN. .

1.1.2. Out of Space Condition

In a Thin Provisioning environment, if the space is insufficient, the datastore space will be over-provisioned by multiple virtual machines, which may lead to a catastrophic situation, which is caused by the lack of space.

In these situations, VMware vSphere 5.x or later will enhance the solution. If Thin Provisioning datastore reaches 100%, virtual machines that only require additional storage space blocks will be suspended, while other virtual machines will remain running.

Advantage

After the Thin Provisioning volume/LUN space is used up, the VMware ESXi server will temporarily suspend the virtual machines that require additional storage space. The administrator can then allocate more storage space by adding other RAID sets to the existing pool.

1.2. Full Copy

This feature helps storage array to make full copies of data within the array without letting VMware ESXi server physically read or write data to the storage array.

Effective Operations

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

Advantage

Reduces the CPU loads on VMware ESXi server, and prevent crowded I/Os between VMware ESXi server and storage array.

Theory

Without VAAI, when one of the above three operations is performed, the VMware ESXi server will read each block from the storage array and write it to a new location. During this period, a lot of server resources were consumed.

With support for VAAI and based on this feature, the VMware ESXi server sends a single SCSI (Extended Copy) command for a group of consecutive blocks to tell the storage array to copy these blocks from one location to another (new location). The commands on the network (if using iSCSI) are small, and the actual work will be performed within the storage array. This minimizes data transfer and speeds up the copy process. Please refer to the figure below to understand how to perform a Full Copy operation when trying to perform a VM cloning task from one datastore to another. Of course, these two datastores are based on two volumes/LUNs from the same QSAN storage.

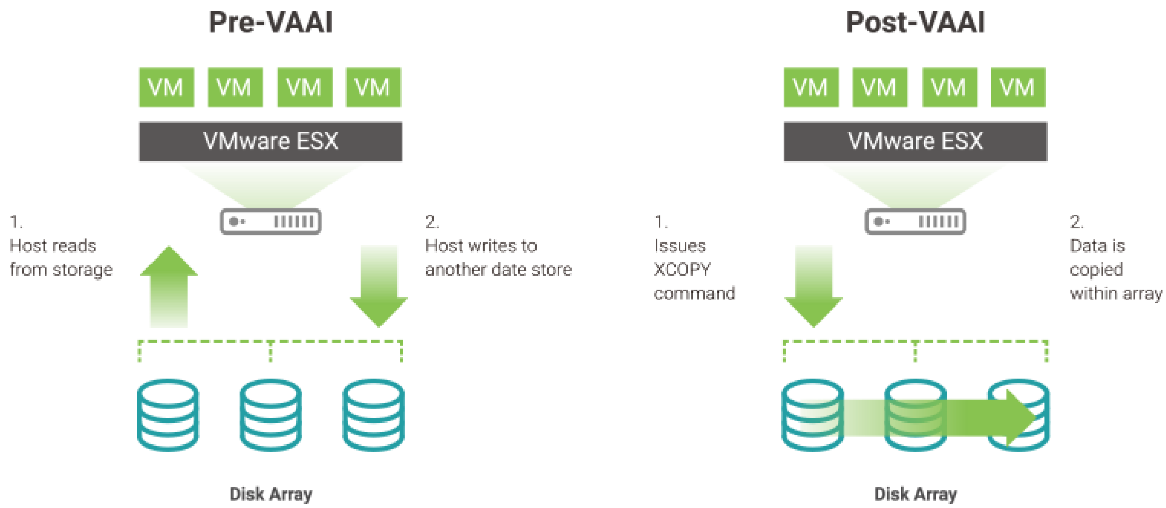

Figure 1-1 VAAI Full Copy

1.3. Block Zero

This feature helps storage array to zero out a large number of blocks for speeding up virtual machine configuration.

Effective Operations

- Create Thin Provisioning Eager Zero virtual disks (Thick Provisioning Eager Zero virtual disks are zeroed out when creating, and are not usable until the process is completed.)
- Write data to an unused area of a Thick Provisioning Lazy Zero virtual disk (Thick Provisioning Lazy Zero virtual disks can be used instantly after they are created.)

Advantage

Use this function to offload the process of writing zeros to the storage array. Eliminates repeated repetitive write commands to reduce the load on the VMware ESXi server, thereby greatly improving capacity allocation.

Theory

Without VAAI, zeroing disk blocks will send duplicated and repetitive write commands from the VMware ESXi server to each block on the storage array. The VMware ESXi server needs to wait for the completion of the previous write command before sending another command, which will result in huge resource costs and time consumed.

With VAAI enabled, VMware ESXi server uses SCSI Write Same command to tell storage array to write the same data to an amount of blocks. VMware ESXi server then doesn't need to send duplicated write command continuously; instead storage array will return the requesting service as though the process of writing zeros has been completed. QSAN storage finishes the zeroing out internally. Please refer to Figure-2 below, which shows the operation and process how Block Zero is performed between VMware ESXi server and QSAN storage.

After enabling VAAI, the VMware ESXi server uses the SCSI Write Same command to tell the storage array to write the same data to a certain number of blocks. In this way, the VMware ESXi server does not need to continuously send repeated write commands; instead, the storage array will return the requested service as if the process of writing zeros has been completed. QSAN storage finishes the zeroing out internally.

1.4. Hardware Assisted Locking

Hardware Assisted Locking, also called ATS (Atomic Test and Set), provides another way to protect the metadata of the VMFS cluster file system and improve the scalability of large ESXi servers that share VMFS datastore. ATS helps to lock the blocks in the volume/LUN instead of the entire volume/LUN added as datastore in the VMware ESXi server.

Effective Operations

- 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 (e.g., a snapshot file or a thin-provisioned virtual disk)

Advantage

When multiple VMware ESXi servers share the same datastore, Hardware Assisted Locking (or ATS) provides a more effective way to avoid retries for getting a lock. The locking mechanism is offloaded to the storage array, and the storage array performs locking at a granular level. This is helpful for scalability when sharing a datastore in a VMware cluster environment without compromising the integrity of the metadata in the VMFS shared storage pool.

Theory

Previously, VMware had a similar mechanism for locking virtual machines to prevent virtual machines from running on them. This mechanism can be modified by multiple VMware ESXi servers at the same time. It is based on the use of SCSI RESERVE and RELEASE commands. This protocol calls the unique access to an entire volume/LUN for the reserving ESXi server until this ESXi server sends a release. Under the protection of the SCSI RESERVE command, the ESXi server can update the metadata records on the storage array to reflect the usage without being disturbed by any other ESXi servers that also call the same part of the same storage array. Please refer to the Figure below, which shows the overall structure of this solution and affects the overall performance of the entire clustered VMware ESXi environment. The performance degradation caused by a large number of RESERVE and RELEASE commands is unacceptable in a VMware cluster environment. The VMware cluster environment accesses shared data storage from different virtual machines exponentially every day.

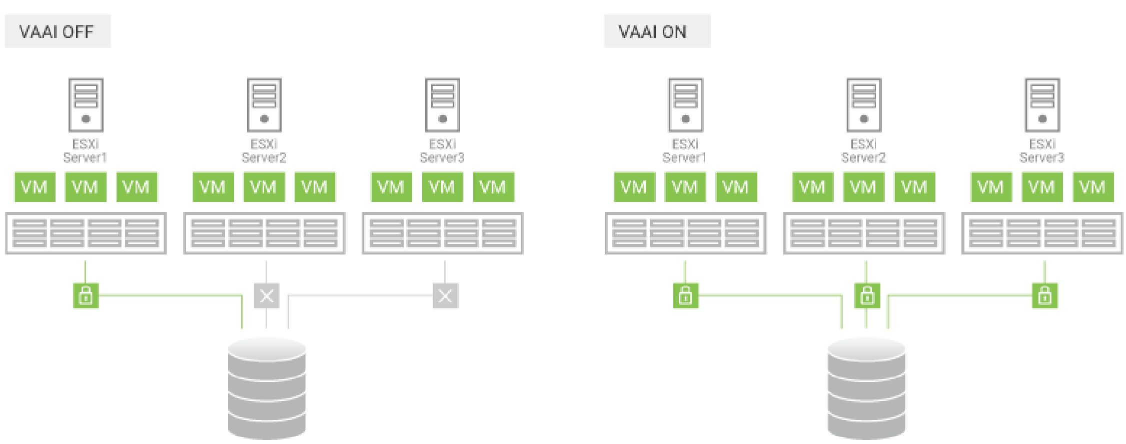

Figure 1-2 VAAI Hardware Assisted Locking

With VAAI, Hardware Assisted Locking provides a more granular method to protect VMFS metadata than the SCSI RESERVE and RELEASE commands. Hardware Assisted Locking uses the storage array ATS function to enable a fine-grained block-level locking mechanism. 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. Then, this new request only requires exclusion of other accesses to the target locked block, not the entire VMFS (which is volume/LUN) which contains the requested lock. When the virtual machine state changes, VMware uses this lock metadata update operation. This may be due to turning ON or OFF the power of the virtual machine, or modifying the configuration of the virtual machine, or even migrating the virtual machine from one ESXi server to another through vMotion.

1.5. Hardware Acceleration Support Status

After adding any storage volume/LUN through VMware vSphere Client, you can observe the status of hardware acceleration. Please navigate to the Configuration -> Hardware -> Storage, then click Datastores View, and check the Hardware Acceleration column displayed after each added datastore.

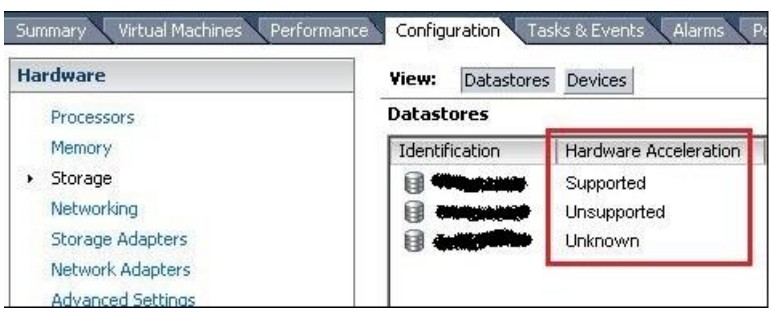

Figure 1-3 Hardware Acceleration Support Status

Table 1-1 Hardware Acceleration Status values

STATUS VALUE	DESCRIPTION
Supported	Storage devices support VAAI
Unsupported	Storage devices do not support VAAI
Unknown	Local datastores

2. TEST RESULTS

The integration of VAAI provides many benefits for improved performance. We have prepared tests and provided some experimental data to prove that VAAI is effective.

2.1. Test Environment

In this test, we use an example to build an environment that connects a VMware ESXi server with a QSAN XS5216D storage array to test the VAAI function.

2.1.1. Test Diagram

For the connection between XS5216D storage array and VMware ESXi server, please refer to the figure below. In this example, a brief environment will be provided.

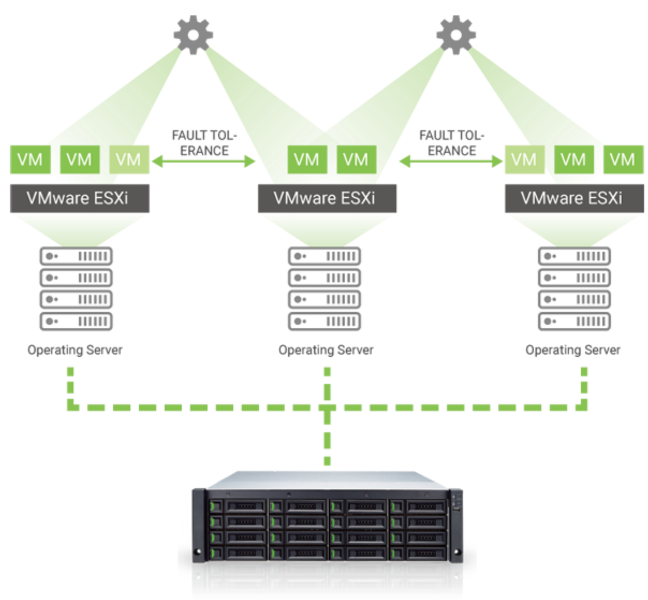

Figure 2-1 VAAI Test Diagram

2.1.2. Storage Configuration

By simulating two FC (Fibre Channel) volumes/LUNs from the XS5216D storage array as two VMFS data stores, this test was performed on a VMware ESXi server to simulate VM cloning and storage vMotion functions. The following Figure provides an idea of how to create pools and volumes.

When verifying the time taken for optimal data protection, the cache mode of these volumes is set to WT (Write-Through, the cache of the storage array on the volume is set to OFF).

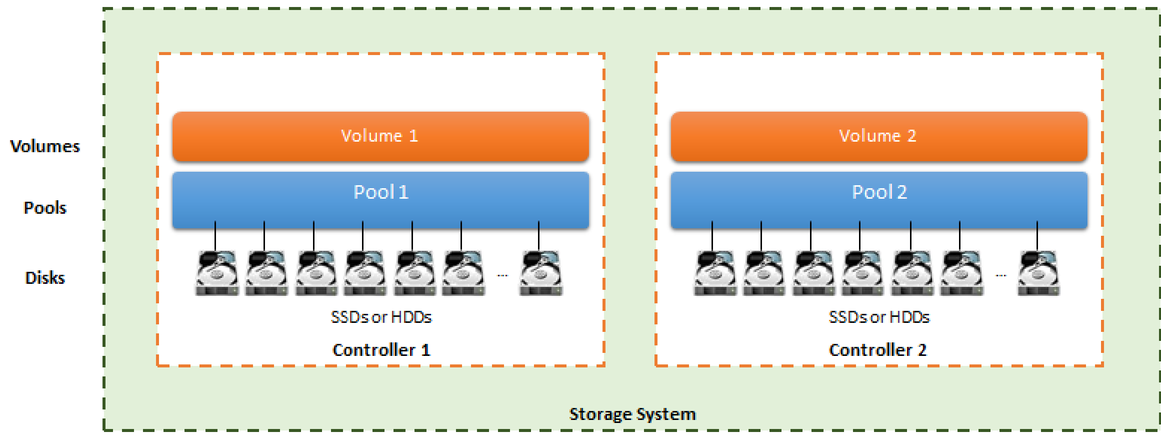

Figure 2-2 VAAI Storage Configuration

2.2. Test Results

Here are the test procedures and results of VAAI functions.

2.2.1. Thin Provisioning

Test Scenario

1. Create a Thin Provisioning volume/LUN on the XS5216D storage array, in which 200GB has been allocated.
2. Create a VMFS datastore on the connected VMware ESXi server.
3. Create a virtual machine based on this VMFS datastore, set the type of Disk Provision to Thin Provision, and set the size to 100GB.

4. Generate about 50GB data on the virtual machine. Observe the capacity consumed on VMFS datastore, which uses 50GB of 200GB.
5. Performed storage vMotion with VAAI ON to migrate the virtual machine to another VMFS datastore.
6. Observed the consumed capacity on the source VMFS datastore again, you will find that the used capacity is about 0GB of 200GB.
7. However, after the virtual machine is migrated, check the Available Capacity (GB) in the WebUI of XS5216D, it may still show 50GB. This is because the granularity of the QSAN XCubeSAN storage array is 1GB, and there is only a continuous 1GB as zero blocks can be reclaimed.
8. Please create a new virtual machine with Thick Provision Eager Zero on this VMFS datastore, and delete it after creation, then execute Space Reclamation in XS5216D storage array, the space shall be able to be reclaimed.

Summary

The supported granularity in the Thin Provisioning pool of QSAN XCubeSAN series products is 1GB. Although space reclamation can be enabled when creating a volume, it is sometimes necessary to manually fill zero blocks from the server so that unused blocks can be filled with zeros and reclaimed.

2.2.2. Full Copy

Test Scenario

1. Create a virtual machine with a 200GB Thick Provision Lazy Zero virtual disk on a VMFS datastore. The virtual disk is made of a FC (Fibre Channel) volume/LUN from the XS5216D storage array. The actual storage consumption on the datastore is about 77GB.
2. Migrate or clone the virtual machine from this datastore to another one which is made by another FC volume/LUN from the same XS5216D storage array.
3. Observe the time it takes to migrate or clone a virtual machine.
4. Repeat the above steps 1 to 3 with disabling VAAI, and compared the time taken.
5. The Table below shows the results of VAAI ON and VAAI OFF.

Test Result

Table 2-1 Time Taken for Full Copy

FULL COPY USE CASE	VAAI OFF	VAAI ON
Storage vMotion	26 min. 56 sec. = 1,616 sec.	6 min. 5 sec = 365 sec.
Virtual Machine Clone	25 min. 50 sec. = 1,550 sec.	5 min. 59 sec. = 359 sec.

Summary

Compared with VAAI ON and OFF, the performance is improved by 77.4% when testing storage vMotion. And the performance increased by about 76.8% when testing the virtual machine clone. So it reduces the CPU loads on VMware ESXi server, and prevent crowded I/Os between VMware ESXi server and storage array.

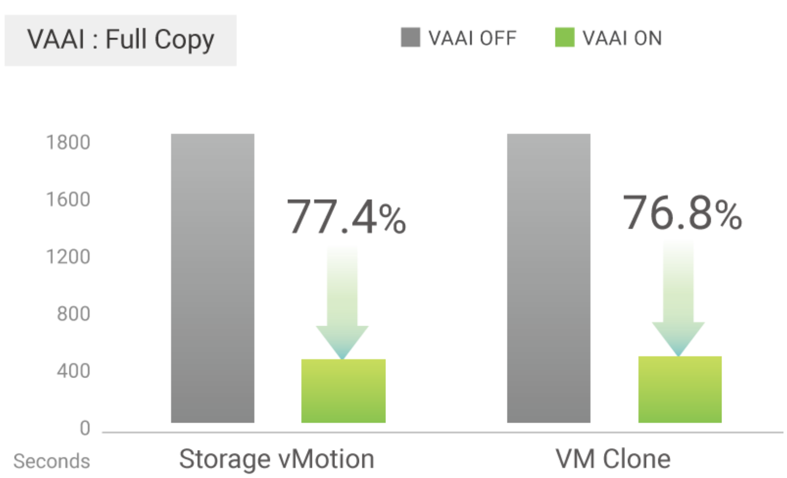

Figure 2-3 Time Saving of VAAI Full Copy

2.2.3. Block Zero

Test Scenario

1. Measured the time taken to create a 200GB Thick Provision Eager Zero virtual disk on a virtual machine.
2. Repeated the same step above in comparison with VAAI OFF.
3. The Table below shows the results of VAAI ON and VAAI OFF.

Table 2-2 Time Taken for Block Zero

BLOCK ZERO USE CASE	VAAI OFF	VAAI ON
Thick pool volume	9 min. 6 sec. = 546 sec.	4 min. 0 sec. = 240 sec.

Summary

When VAAI is enabled and trying to create a 200GB Thick Provision Eager Zero virtual disk, the performance improves by 56%. The virtual disk is stored in a thick pool in the XS5216D storage array.

3. CONCLUSION

The integration of VAAI in all QSAN storage series of products provides many benefits for improved performance and storage array management. The main features are:

- The Dead Space Reclamation capability of the Thin Provisioning feature can reclaim blocks from the thin provisioning volume/LUN on the QSAN storage products. In this way, you can avoid the lack of space by temporarily suspending the virtual machine. When the VMFS datastore space is used up, the virtual machine needs additional space. The administrator can then allocate more capacity by adding more RAID sets to the existing pool.
- The Full Copy feature accelerates the storage vMotion or virtual machine clone operations by transferring operations from the VMware ESXi server to the storage array itself, and greatly reduces resource usage when performing these operations.
- The Block Zero feature speeds up the deployment of Thick Provision Eager Zero virtual disks by offloading a large number of duplicate and repetitive zero blocks to the QSAN storage platform, helps to free the resources of VMware ESXi server for other tasks.
- The Hardware Assisted Locking feature delivers a more effective method to prevent retrying to obtain a lock when multiple ESXi servers share the same datastore. It can offload the locking mechanism to the QSAN storage array, which can be locked at a granular level. This improves the scalability of large ESXi servers that share the same datastore.

4. APPENDIX

4.1. Apply To

- XEVO firmware 2.0.0 and later
- SANOS firmware 2.0.0 and later
- QSM firmware 3.3.0 and later

4.2. Reference

VMware Documentations

- VMware vSphere Storage APIs – Array Integration (VAAI)