NetApp

Storage virtualization with VMware and Microsoft copy offload

ONTAP 9

NetApp
March 28, 2022

Table of Contents

Storage virtualization with VMware and Microsoft copy offload

Storage virtualization with VMware and Microsoft copy offload overview

VMware and Microsoft support copy offload operations to increase performance and network throughput. You must configure your system to meet the requirements of the VMware and Windows operating system environments to use their respective copy offload functions.

When using VMware and Microsoft copy offload in virtualized environments, your LUNs must be aligned. Unaligned LUNs can degrade performance.

Advantages of using a virtualized SAN environment

Creating a virtualized environment by using storage virtual machines (SVMs) and LIFs enables you to expand your SAN environment to all of the nodes in your cluster.

- Distributed management

  You can log in to any node in the SVM to administer all of the nodes in a cluster.

- Increased data access

  With MPIO and ALUA, you have access to your data through any active iSCSI or FC LIFs for the SVM.

- Controlled LUN access

  If you use SLM and portsets, you can limit which LIFs an initiator can use to access LUNs.

How LUN access works in a virtualized environment

In a virtualized environment, LIFs enable hosts (clients) to access LUNs through optimized and unoptimized paths.

A LIF is a logical interface that connects the SVM to a physical port. Although multiple SVMs can have multiple LIFs on the same port, a LIF belongs to one SVM. You can access LUNs through the SVMs LIFs.

Example of LUN access with a single SVM in a cluster

In the following example, Host 1 connects to LIF1.1 and LIF1.2 in SVM-1 to access LUN1. LIF1.1 uses the physical port node1:0c and LIF1.2 uses the node2:0c. LIF1.1 and LIF1.2 belongs only to SVM-1. If a new LUN is created on node 1 or node 2, for SVM-1, then it can use these same LIFs. If a new SVM is created, then new LIFs can be created using physical ports 0c or 0d on both the nodes.

Example of LUN access with multiple SVMs in a cluster

A physical port can support multiple LIFs serving different SVMs. Because LIFs are associated with a particular SVM, the cluster nodes can send the incoming data traffic to the correct SVM. In the following example, each node from 1 through 4 has a LIF for SVM-2 using the physical port 0c on each node. Host 1 connects to LIF1.1 and LIF1.2 in SVM-1 to access LUN1. Host 2 connects to LIF2-1 and LIF2-2 in SVM-2 to access LUN2. Both SVMs are sharing the physical port 0c on the nodes 1 and 2. SVM-2 has additional LIFs that Host 2 is using to access LUNs 3 and 4. These LIFs are using physical port 0c on nodes 3 and 4. Multiple SVMs can share the physical ports on the nodes.

Example of an active or optimized path to a LUN from a host system

In an active or optimized path, the data traffic does not travel over the cluster network; it travels the most direct route to the LUN. The active or optimized path to LUN1 is through LIF1.1 in node1, using physical port 0c. Host 2 has two active or optimized paths, one path to node1, LIF2.1, which is sharing physical port 0c and the other path to node4, LIF2.4, which is using physical port 0c.

Example of an active or unoptimized path (indirect) path to a LUN from a host system

In an active or unoptimized path (indirect) path, the data traffic travels over the cluster network. This issue occurs only if all the active or optimized paths from a host are unavailable to handle traffic. If the path from Host 2 to SVM-2 LIF2.4 is lost, then access to LUN3 and LUN4 traverses the cluster network. Access from Host 2 uses LIF2.3 on node3. Then the traffic enters the cluster network switch and backs up to node4 for access to the LUN3 and LUN4. It will then traverse back over the cluster network switch and then back out through LIF2.3 to Host 2. This active or unoptimized path is used until the path to LIF2.4 is restored or a new LIF is established for SVM-2 on another physical port on node 4.

Considerations for LIFs in cluster SAN environments

You must be aware of certain LIF considerations in a SAN environment.

- Initiators must use Multipath I/O (MPIO) and asymmetric logical unit access(ALUA) for failover capability for clusters in a SAN iSCSI or FC environment because SAN does not support automatic failover for LIFs.

- At least one SAN LIF of the appropriate protocol must be configured on each node that hosts a mapped LUN and the node's HA partner.

  You can configure two LIFs per node, one for each fabric being used with FC and to separate Ethernet networks for iSCSI.

- Some options are not applicable for iSCSI or FC.

  For example, you cannot use IP addresses with FC.

Improve VMware VAAI performance for ESX hosts

ONTAP supports certain VMware vStorage APIs for Array Integration (VAAI) features when the ESX host is running ESX 4.1 or later. These features help offload operations from the ESX host to the storage system and increase the network throughput. The ESX host enables the features automatically in the correct environment.

The VAAI feature supports the following SCSI commands:

* `EXTENDED_COPY`

  This feature enables the host to initiate the transfer of data between the LUNs or within a LUN without involving the host in the data transfer. This results in saving ESX CPU cycles and increasing the network throughput. The extended copy feature, also known as "copy offload," is used in scenarios such as cloning a virtual machine. When invoked by the ESX host, the copy offload feature copies the data within the storage system rather than going through the host network. Copy offload transfers data in the following ways:

  ◦ Within a LUN

  ◦ Between LUNs within a volume

  ◦ Between LUNs on different volumes within a storage virtual machine (SVM)

  ◦ Between LUNs on different SVMs within a cluster If this feature cannot be invoked, the ESX host automatically uses the standard READ and WRITE commands for the copy operation.

* `WRITE_SAME`

  This feature offloads the work of writing a repeated pattern, such as all zeros, to a storage array. The ESX host uses this feature in operations such as zero-filling a file.

* `COMPARE_AND_WRITE`

  This feature bypasses certain file access concurrency limits, which speeds up operations such as booting up virtual machines.

Requirements for using the VAAI environment

The VAAI features are part of the ESX operating system and are automatically invoked by the ESX host when you have set up the correct environment.

The environment requirements are as follows:

* The ESX host must be running ESX 4.1 or later.
* The NetApp storage system that is hosting the VMware datastore must be running ONTAP.
* (Copy offload only) The source and the destination of the VMware copy operation must be hosted on the same storage system within the same cluster.

  ⓘ The copy offload feature currently does not support copying data between VMware datastores that are hosted on different storage systems.

Determine if VAAI features are supported by ESX

To confirm whether the ESX operating system supports the VAAI features, you can check the vSphere Client or use any other means of accessing the host. ONTAP supports the SCSI commands by default.

You can check your ESX host advanced settings to determine whether VAAI features are enabled. The table indicates which SCSI commands correspond to ESX control names.

SCSI command	ESX control name (VAAI feature)
EXTENDED_COPY	`HardwareAcceleratedMove`
WRITE_SAME	`HardwareAcceleratedInit`
COMPARE_AND_WRITE	`HardwareAcceleratedLocking`

Microsoft Offloaded Data Transfer (ODX)

Microsoft Offloaded Data Transfer (ODX) overview

Microsoft Offloaded Data Transfer (ODX), also known as *copy offload*, enables direct data transfers within a storage device or between compatible storage devices without transferring the data through the host computer.

ONTAP supports ODX for both the SMB and SAN protocols.

In non-ODX file transfers, the data is read from the source and is transferred across the network to the host. The host transfers the data back over the network to the destination. In ODX file transfer, the data is copied directly from the source to the destination without passing through the host.

Because ODX offloaded copies are performed directly between the source and destination, significant performance benefits are realized, including faster copy time, reduced utilization of CPU and memory on the client, and reduced network I/O bandwidth utilization.

For SAN environments, ODX is only available when it is supported by both the host and the storage system. Client computers that support ODX and have ODX enabled automatically and transparently use offloaded file transfer when moving or copying files. ODX is used regardless of whether you drag-and-drop files through Windows Explorer or use command-line file copy commands, or whether a client application initiates file copy requests.

Requirements for using ODX

If you plan to use ODX for copy offloads, you need to be familiar with volume support considerations, system requirements, and software capability requirements.

To use ODX, your system must have the following:

- ONTAP

  ODX is automatically enabled in supported versions of ONTAP.

- Minimum source volume of 2 GB

  For optimal performance, the source volume should be greater than 260 GB.

- Deduplication

  ODX uses deduplication as part of the copy process. If you do not want deduplication on your SVM, you should disable ODX on that SVM.

- ODX support on the Windows client

  ODX is supported in Windows Server 2012 or later and in Windows 8 or later. The Interoperability Matrix contains the latest information about supported Windows clients.

  NetApp Interoperability Matrix Tool

- Copy application support for ODX

  The application that performs the data transfer must support ODX. Application operations that support ODX include the following:

  - Hyper-V management operations, such as creating and converting virtual hard disks (VHDs), managing Snapshot copies, and copying files between virtual machines
  - Windows Explorer operations
  - Windows PowerShell copy commands
  - Windows command prompt copy commands The Microsoft TechNet Library contains more information about supported ODX applications on Windows servers and clients.

- If you use compressed volumes, the compression group size must be 8K.

  32K compression group size is not supported.

ODX does not work with the following volume types:

- Source volumes with capacities of less than 2 GB
- Read-only volumes
- FlexCache volumes
- Semi-thick provisioned volumes

Use cases for ODX

You should be aware of the use cases for using ODX on SVMs so that you can determine under what circumstances ODX provides you with performance benefits.

Windows servers and clients that support ODX use copy offload as the default way of copying data across remote servers. If the Windows server or client does not support ODX or the ODX copy offload fails at any point, the copy or move operation falls back to traditional reads and writes for the copy or move operation.

The following use cases support using ODX copies and moves:

- Intra-volume

The source and destination files or LUNs are within the same volume.

- Inter-volume, same node, same SVM

  The source and destination files or LUNs are on different volumes that are located on the same node. The data is owned by the same SVM.

- Inter-volume, different nodes, same SVM

  The source and destination files or LUNs are on different volumes that are located on different nodes. The data is owned by the same SVM.

- Inter-SVM, same node

  The source and destination file or LUNs are on different volumes that are located on the same node. The data is owned by different SVMs.

- Inter-SVM, different nodes

  The source and destination file or LUNs are on different volumes that are located on different nodes. The data is owned by different SVMs.

- Inter-cluster

  The source and destination LUNs are on different volumes that are located on different nodes across clusters. This is only supported for SAN and does not work for SMB.

There are some additional special use cases:

- With the ONTAP ODX implementation, you can use ODX to copy files between SMB shares and FC or iSCSI attached virtual drives.

  You can use Windows Explorer, the Windows CLI or PowerShell, Hyper-V, or other applications that support ODX to copy or move files seamlessly using ODX copy offload between SMB shares and connected LUNs, provided that the SMB shares and LUNs are on the same cluster.

- Hyper-V provides some additional use cases for ODX copy offload:

  - You can use ODX copy offload pass-through with Hyper-V to copy data within or across virtual hard disk (VHD) files or to copy data between mapped SMB shares and connected iSCSI LUNs within the same cluster.

    This allows copies from guest operating systems to pass through to the underlying storage.

  - When creating fixed-sized VHDs, ODX is used for initializing the disk with zeros, using a well-known zeroed token.

  - ODX copy offload is used for virtual machine storage migration if the source and destination storage is on the same cluster.

    (i) To take advantage of the use cases for ODX copy offload pass-through with Hyper-V, the guest operating system must support ODX and the guest operating system's disks must be SCSI disks backed by storage (either SMB or SAN) that supports ODX. IDE disks on the guest operating system do not support ODX pass-through.

Special system file requirements

You can delete ODX files found in qtrees. You must not remove or modify any other ODX system files unless you are told by technical support to do so.

When using the ODX feature, there are ODX system files that exist in every volume of the system. These files enable point-in-time representation of data used during the ODX transfer. The following system files are in the root level of each volume that contains LUNs or files to which data was offloaded:

- `.copy-offload` (a hidden directory)

- `.tokens` (file under the hidden `.copy-offload` directory)

You can use the `copy-offload delete-tokens -path dir_path -node` *node_name* command to delete a qtree containing an ODX file.