# Adaptive Optimization for HP StoreVirtual

## Table of contents

- Executive summary ................................................................................................................................. 2
- Terminology ............................................................................................................................................... 2
- Features and benefits ............................................................................................................................... 3
  - Overview of Adaptive Optimization ....................................................................................................... 3
  - Performance benefits of Adaptive Optimization .................................................................................... 4
- Enabling Adaptive Optimization ............................................................................................................... 4
  - Enabling a storage cluster for AO ........................................................................................................ 4
  - Using Adaptive Optimization with volumes, snapshots, and smartClone volumes ............................ 6
- Adaptive Optimization status using the Command Line Interface .......................................................... 8
- New AO parameter for creating and editing volumes ............................................................................. 9
- Adaptive Optimization process ............................................................................................................. 10
  - Adaptive Optimization algorithm ...................................................................................................... 10
- Data path description and behavior ....................................................................................................... 11
  - Network RAID technology ................................................................................................................ 11
  - Preferred page ..................................................................................................................................... 11
  - VIP Load Balancing and the StoreVirtual DSM for MPIO ................................................................. 12
- Tiered storage devices for StoreVirtual VSA .......................................................................................... 14
  - VSA recommendations for tiered storage devices ............................................................................ 14
  - Assigning storage devices to StoreVirtual VSA ................................................................................ 15
- Monitoring Adaptive Optimization ......................................................................................................... 16
  - Using Adaptive Optimization volume capacity metrics to view utilization .................................... 17
  - Monitoring Adaptive Optimization performance metrics ................................................................. 17
  - Performance monitoring process ........................................................................................................ 18
- Conclusion ............................................................................................................................................... 19
Executive summary

With new auto-tiering technologies for HP StoreVirtual Storage come opportunities to optimize the cost and performance of StoreVirtual storage clusters. By utilizing Adaptive Optimization technology to migrate data between storage tiers within individual storage systems, StoreVirtual products provide a unique method for balancing performance versus capacity within a storage pool and lowering overall cost.

Automated tiering intelligently tracks in-flight use statistics at a granular level and migrates highly accessed data to higher performing storage devices in real-time. An important balance between internal data movement and application IO is maintained using advanced comparative usage metrics to ensure that data is not moved unnecessarily or too often.

All of this is achieved at a cost point lower than SSD storage systems but with much improved performance over traditional SAS based storage systems.

This paper provides a comprehensive look at the technology and practical application for storage architects and administrators involved in planning and implementing a StoreVirtual iSCSI storage infrastructure.

Terminology

Adaptive Optimization: A technology that allows data to be stored on storage system media that have different performance characteristics. Adaptive Optimization allows for the movement of data between storage tiers based on access trends.

AO Capable Cluster: The cluster includes at least one multi-tiered storage system and no volumes are set to AO Permitted.

AO Enabled Cluster: The cluster includes at least one multi-tiered storage system and has at least one volume set to AO Permitted.

AO Cluster Not Capable: The cluster does not include any storage systems with more than one tier of storage.

Cluster: A grouping of storage systems that resides in a Management Group. A cluster provides a single pool of storage from which volumes can be created and presented to client servers. The volumes seamlessly span the storage systems in the cluster. The cluster capacity can be expanded by adding additional storage systems to the cluster.

Dynamic Optimization Tier: StoreVirtual clusters made of storage systems with a specific disk configuration. For example, one cluster contains nodes with SAS disk while another cluster contains nodes with SSD and SAS disks that have a higher IOPS capability.

LeftHand Operating System: The storage server operating system running on all StoreVirtual storage systems. The OS running on StoreVirtual VSA is identical to that on the dedicated hardware platforms and provides the same features.

Management Group: A logical entity for organizing StoreVirtual clusters and storage nodes. It is the top level container which can contain one or more clusters or storage nodes.

Network RAID: StoreVirtual technology for maintaining replicated copies of data across a cluster. Network RAID (greater than Network RAID 0) ensures high availability of data.

Page: An individual block of data residing on a physical disk. For StoreVirtual devices, the page size is 256 KB.

Peer Motion: Technology that allows volumes to be migrated between storage clusters residing in the same management group. Migration occurs online without interruption to the application server.

Storage System: An individual instance of the StoreVirtual LeftHand Operating System running on a single storage server or as a virtual machine (in the case of the StoreVirtual VSA).

Tier: A grouping of like RAID devices internal to a StoreVirtual storage node (server or virtual machine) that is provided a hierarchical designation based on the storage device’s performance capability.

- Tier 0—The fastest tier of storage in a hybrid/auto-tiering solution.
- Tier 1—The slower tier of storage in a hybrid/auto-tiering solution.

The StoreVirtual Virtual Storage Appliance (VSA) requires the administrator to assign any storage type to Tier 0 or Tier 1 when using Adaptive Optimization. It is recommended to utilize either SSD and SAS together or SAS and Nearline drives together.

VSA: The HP StoreVirtual Virtual Storage Appliance. A StoreVirtual storage appliance that runs the LeftHand OS as a virtual machine and is available for both VMware vSphere and Hyper-V. Allows use of local storage to create a highly available distributed iSCSI storage area network.
Features and benefits

The primary benefit of Adaptive Optimization is the ability to realize a higher level of performance by adding a small amount of Tier-0 storage compared to what could be achieved in a system consisting only of Tier-1 storage. This is accomplished at a much lower price point than a system consisting of only Tier-0 storage.

Up to this point, tiering for StoreVirtual systems has been achieved via Dynamic Optimization, creating separate storage pools (clusters) with differing storage characteristics and moving the volumes between the performance tiers using Peer Motion online volume migration technology. This technology will continue to be important and Adaptive Optimization appliance based storage systems can be thought of as a new classification of StoreVirtual device that is positioned in the highest tier within the Dynamic Optimization space ahead of SAS and Near Line rotational disk based clusters.

Figure 1. Dynamic Optimization tiering moves data between storage clusters using Peer Motion online volume migration

Adaptive Optimization works within a Dynamic Optimization tier providing performance enhancement within each storage system comprising the storage cluster.

Overview of Adaptive Optimization

Adaptive Optimization is an innovative technology that greatly increases the efficient use of faster storage devices by intelligently moving data between storage devices with different performance characteristics within a single storage system.

These differing storage devices, known herein as tiers, have different speeds and costs. This feature detects the most frequently accessed data and in real time, migrates it to the highest performing tier, displacing less frequently accessed data to slower, less expensive disk storage. The net result of this selective provisioning is to offer performance approaching pure solid-state at a much lower cost. The feature adapts to changing workloads, re-provisioning storage to optimize the performance of a new workload. Each StoreVirtual storage system performs this re-provisioning independently based on its own workload.

Traditional tiering solutions have suffered from slow reaction times due to the scheduled transfers of data between tiers. While this has helped to mitigate performance degradation from excessive data movement, this type of tiering requires detailed planning and added management complexity.

The StoreVirtual Adaptive Optimization feature reduces planning complexity by utilizing innovative algorithms to pinpoint hot data at a highly granular level and re-provision data in real time while maintaining a balance between background IO operations and application IO requirements. The result is an effective solution that increases performance substantially without administrator intervention.
Performance benefits of Adaptive Optimization

By combining SSD and SAS drives the effective performance of the storage systems can approach that of an all SSD system in terms of both IOPS and latency. However, the system also benefits from the capacity provided by the SAS drives which drives down the cost considerably.

The following figure shows an example of IOPS and latency for each individual tier within a storage system and the combined performance under various workloads. The Tier 0 only performance data typically sets the upper bound while the Tier 1 only performance data typically sets the lower bound. Combining Tier 0 (SSD) and Tier 1 (HDD) typically has a minimal impact on IOPS but a significant positive impact on latency.

Table 1. Performance results of both standalone Tiers and combined Tiers on a StoreVirtual Storage System

<table>
<thead>
<tr>
<th>Workload</th>
<th>Tier 0 Only</th>
<th>Tier 1 Only</th>
<th>Tier 0 + Tier 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>8K Random Read</td>
<td>145,000 IOPS</td>
<td>2,200 IOPS</td>
<td>141,500 IOPS</td>
</tr>
<tr>
<td></td>
<td>2ms</td>
<td>20ms</td>
<td>2.4ms</td>
</tr>
<tr>
<td>8K Random OLTP</td>
<td>37,500 IOPS</td>
<td>880 IOPS</td>
<td>33,500 IOPS</td>
</tr>
<tr>
<td></td>
<td>2ms</td>
<td>18.5ms</td>
<td>3.3ms</td>
</tr>
<tr>
<td>8K Random Write</td>
<td>16,500 IOPS</td>
<td>1,200 IOPS</td>
<td>4,600 IOPS</td>
</tr>
<tr>
<td></td>
<td>2ms</td>
<td>10ms</td>
<td>2.3ms</td>
</tr>
</tbody>
</table>

The combined result is quite dramatic in that total IOPS and latency remain close to Tier 0 performance capability for both 8K Random Read and 8K Random OLTP tests. In the case of 8K Random Write tests the slower numbers reflect the use of Network RAID 10 which is a technology that replicates writes in the storage cluster for high availability. The duplication of writes involving spinning media does have an impact on overall write performance. These numbers demonstrate a best case example of how Adaptive Optimization can accelerate performance when Tier-0 is 100 percent utilized. The degree to which Tier-0 is utilized in specific cases will depend on the degree to which applications access particular areas of data locality within larger data sets.

Enabling Adaptive Optimization

The ability to provide Adaptive Optimization services is evaluated at a cluster, storage system and volume level.

An individual storage system becomes AO capable as soon as it has two (2) different types of storage devices configured within it and is running at version 11.0 or greater of the LeftHand OS. For hardware based offerings the two types of storage will be pre-defined. The SSD disk tier will use a pre-defined RAID configuration and storage administrators will have various RAID types to choose from for the SAS disk tier.

Storage Devices for AO tiering on VSA are defined during the installation process with multiple drive configurations possible as outlined in the section on System Requirements. Storage tiers can also be configured or edited on VSAs after the initial configuration using the HP StoreVirtual Centralized Management Console (CMC).

Enabling a storage cluster for AO

Cluster enablement for Adaptive Optimization is based on the existence of AO capable storage nodes in the cluster and the version of the LeftHand OS that is running on the storage nodes in the management group.

The requirements for AO to be enabled in a cluster are as follows:

1. All storage nodes in the management group must, at a minimum, run LeftHand OS version 11.0 or greater.
2. The cluster must contain at least one storage system that is AO capable.
3. The StoreVirtual Centralized Management Console must be version 11.0 or greater.
4. There is a valid license for Adaptive Optimization. The AO feature is included in the license for all StoreVirtual products with the exception of the 1 TB and 4 TB version of the StoreVirtual VSA.
It is a recommended best practice that AO be used in clusters where all nodes are AO capable. In storage clusters containing less than 100 percent AO capable storage systems there are several factors that mitigate performance benefits and make this a non-optimal configuration.

- Given that data pages are distributed across all storage systems in the clusters, only pages elevated to a faster tier on an AO capable node will experience faster access.
- Pages accessed on one storage system will be accessed from the same storage system on subsequent requests. If this system is not AO enabled no performance benefit will be seen.
- When using Network RAID 10, two (2) copies of the data are written to two (or more) storage systems in the cluster. Acknowledgement of the write back to the application server is based on the slower of the two. If one storage system receiving the written page is AO capable and the other is not there may not be a performance benefit seen. This is also dependent on the saturation of the controller write cache for that particular system. If the controller write cache is not saturated the fact that a system is not AO capable will not affect performance for the given write.

**Note:**
While it is possible to have a mixture of AO and non-AO capable storage systems in the same cluster this is not a recommended practice. Best performance results will always been seen in storage clusters containing identical storage systems.

**Figure 2.** Indicators displaying Adaptive Optimization capability are available at the Storage System and Cluster within the StoreVirtual Centralized Management Console
Using Adaptive Optimization with volumes, snapshots, and smartClone volumes

By default Adaptive Optimization is set to Permitted for all new primary volumes created in an AO enabled storage cluster. This is the recommended setting in most cases as only pages within the volume with the appropriate high level of usage will be promoted to a faster tier. It is easy for system administrators to misidentify where hotspots will occur on a specific volume. AO is designed to alleviate this with an accurate tiering algorithm that elevates only the most important data.

However, AO can be set to Not-Permitted for any volume on the Advanced tab of both the New Volume dialog box or the Edit Volume dialog box. Administrators may consider setting volumes to Not-Permitted if they do not want the volume to ever be able to access Tier-0 storage. In these cases administrators may decide that, even though a volume may be accessed frequently in some cases, its performance is not of sufficient performance to warrant promotion of data within the volume to Tier-0 storage. Also, administrators may consider migrating this type of volume to a separate Dynamic Optimization tier such as a cluster utilizing only rotational disk.

Figure 3. Setting Adaptive Optimization attributes for a volume on the New Volume dialog box

Note:
The fact that a volume is set to AO Permitted does not necessarily mean that the volume is currently using Adaptive Optimization. The setting is available regardless of the storage system status. The storage system must be checked to determine that it provides Adaptive Optimization.

Setting AO to Not Permitted on a volume will not immediately cause data to be moved down to the slower tier. Instead the volume will simply cease to accrue further counters for page accesses. As the existing counts decay over time the data pages will be displaced naturally as newly promoted data takes its place in the faster tier. It should be noted, however, that changing the AO permissions on a volume does cause an immediate update of all of the metadata pointers associated with the pages on the volume. These updates may impact overall system resources and as such, it is not recommended to change AO settings for volumes frequently or during times when the system is extremely busy.

Volumes with snapshots
Snapshots inherit the AO permissions of the parent volume in all cases. Therefore a volume with AO Not Permitted that is then changed to Permitted will cascade this change to all child snapshots. This makes sense given that a StoreVirtual snapshot is a layer within the overall volume itself and therefore an indispensable portion of the volume's structure.

When a snapshot is taken of a volume the data between the current time and the time the last snapshot was taken is encapsulated in a new snapshot layer and marked as read only. A new layer is then created at the top level of the volume to accept new data and changes to existing data.
AO use with SmartClone volumes
SmartClone volumes are fully writable volumes based on a snapshot within an existing volume. As such they will always inherit the AO permission of the snapshot that they are based on. A SmartClone volume can be created from any existing snapshot and can be individually managed just like other volumes. SmartClone volume data will be dependent on the point-in-time snapshot it was created from and any other snapshots from the same volume that were taken at earlier times. In addition a new writable layer is created to accommodate new data and changes to existing data.

Figure 5. Illustration of a SmartClone volume with the clone point based off of the second snapshot of a volume

Note:
Although SmartClone volumes inherit the AO setting of its sibling volume, it is possible to change AO Permitted/not Permitted behavior from the Edit Volume dialog box of the SmartClone volume. However, doing so will also change the AO permissions for all sibling volumes and all associated snapshots.
Figure 6. A warning dialog when changing AO permission from a SmartClone volume that tells of impending changes to the sibling volume and associated snapshots

It is possible to create multiple SmartClone volumes from a single snapshot or from different snapshots within the same volume. Keep in mind that changing the AO permissions for a volume will change the AO permissions for all associated snapshots and therefore all SmartClone volumes created from those snapshots.

**Adaptive Optimization status using the Command Line Interface**

The status of AO for storage systems, clusters, and volumes as outlined above can also be viewed using commands for the StoreVirtual Command Line Interface. The following commands are available from the CLI and now include AO specific information.

- **getVolumeInfo**—Displays whether or not the volume is AO Permitted as well as whether the parent cluster is AO capable.
- **getClusterInfo**—Displays whether or not the cluster is AO capable.
- **getGroupInfo**—Displays complete volume and cluster information.
- **getNsmInfo**—Displays the tier count of the storage systems and whether or not the storage system is AO capable.

Figure 7. Output from running the getVolumeInfo CLI command shows that AO is Permitted for the volume and that the cluster is AO capable
**New AO parameter for creating and editing volumes**

In addition to the status commands there are multiple CLI commands that can now make use of a new `adaptiveOptimization` parameter. The parameter modifies the volume-level flag indicating whether Adaptive Optimization is Permitted or Not Permitted. The parameter is `adaptiveOptimization = 0/1`

- **0** means the flag is set to off. The volume is **Not Permitted** to make use of Adaptive Optimization.
- **1** means the flag is set to on. The volume is **Permitted** and will make use of Adaptive Optimization should the cluster support it.

If the `adaptiveOptimization` parameter is not specified then the default of **0** or Not Permitted will be used. There it is always an optional parameter.

The following commands now make use of the `adaptiveOptimization` parameter:

- `createVolume`
- `makePrimary`
- `makeRemote`
- `modifyVolume`
- `provisionVolume`

For detailed information on using the command line interface please refer to the HP StoreVirtual LeftHand OS Command Line Interface User Guide.
Adaptive Optimization process

This section describes the internal algorithm used for re-provisioning of data pages within AO capable storage nodes.

StoreVirtual Storage Systems use 256 KB data page block size providing very granular usage tracking and migration of highly accessed pages. This small size reduces background workload when moving data between tiers by efficiently targeting only high priority data.

Adaptive Optimization algorithm

The data movement algorithm manages timely reactions to changes in data access patterns while avoiding overreaction to events which can cause the system to move data unnecessarily and therefore degrade performance.

In order to achieve this, data access must be carefully tracked and evaluated against newly emerging patterns. Doing so requires multiple components associated with both incoming and outgoing operations. Adaptive Optimization is enabled for both read and write operations and both types of page accesses are treated the same in terms of tracking the number of accesses in the heat map.

There are 3 primary components to the algorithm.

1. **Heat Map:** For each tier a heat map is stored in the form of a Least Recently Used (LRU) list that tracks the number of page accesses. Each page accesses triggers a calculation which is used to determine whether data should be moved to a faster tier. Every time a page is hit it will be evaluated for promotion.

2. **Move Request Handler:** This process collects move requests and executes moves in order of importance using the difference in page counts. The system must, by necessity, be very reactive to ongoing page accesses. As a result a large number of move requests will be granted but the ability of a page to move to a higher tier is determined by how many backlogged accepted move requests have yet to be moved. The more requests that are accepted (inside a node) the more difficult it becomes to grant a move of a specific page. This effectively allows only the most important data to be moved all of the time.

3. **Decay Algorithm:** As the heat map is built up it must be kept relevant in time. Periodically, due to time elapsed or the amount of traffic the system has received the heat maps are readjusted and decayed a half life. Pages that are accessed frequently must not build up so much history that newer hot pages cannot be moved into a higher (faster) tier.

A key benefit of the system is that all pages Permitted for AO tiering are treated equally in the system. Therefore as the number of page accesses for hot data increases, the overall bar is raised as to which blocks will win promotion for re-allocation to a faster tier. On the other hand the decay algorithm ensures that newly hot data can eventually meet the requirement by halving the counters when the average number of counters in Tier-0 meet a pre-defined threshold or when a pre-defined amount of time elapses. It is this balance that ensures the most important data always resides at the fastest tier while not excessively promoting data that is not accessed enough times to trigger a promotion.

These features also allow AO to be enabled for all volumes on a system with confidence that data which is only accessed occasionally will always reside at the slower tier.

In a cluster recently enabled for AO, free space will initially exist in Tier 0. New writes will always arrive at the fastest tier whenever space exists. Once Tier 0 is full, new writes will be written to Tier 1. This is because a newly written page only has a count of 1 assigned to it in the heat map and does not yet meet the requirements for elevation to the faster tier. Subsequent page accesses to this data will trigger a calculation to be made that compares page access counts and may qualify the page to displace an existing page in the next fastest tier.

At all times client IO requests are given priority in the storage systems. The number of pages that are currently being moved between tiers is dynamically adjusted based on system load.
Data path description and behavior

This section will cover technologies and considerations that affect data layout with the storage cluster and access to that data.

Storage virtualization technology used in StoreVirtual systems distributes data evenly across all storage systems and physical storage devices in a storage cluster. This distribution of data works to enhance the benefits of Adaptive Optimization by distributing portions of an individual volume to faster tiers across multiple storage systems.

The data layout within a distributed cluster also implies design considerations that should be taken into account for iSCSI initiator access to targets. Obtaining access to data with the fewest number of network hops is the goal. SSD provides very low latency page access and therefore attention should be given to designing data path access with the most direct path to data. This becomes particularly important in stretch multi-site clusters where there are some options for decreasing network utilization between the sites.

Network RAID technology

StoreVirtual Network RAID technology can provide replicated copies of data pages within the storage cluster to provide protection from individual storage system failures as well as complete site outages. 256 KB data pages are laid out across the storage systems in such a way that if any one system goes offline a complete copy of the data remains on other storage systems to continue servicing IO requests. In addition storage systems can be organized in logical "sites" that ensure a copy of all data resides at each site, protecting data access from a complete site outage or disruption of communication between sites.

Network RAID 10 is the most common Network RAID type and recommended for most applications due to its resiliency and high performance.

Figure 9. Example of page layout for Network RAID 10. If any one storage system goes offline any given page can still be accessed from another system

While a complete description of Network RAID technology will not be discussed here it is necessary to provide an overview as it pertains to the next section on Preferred Page and because the differences in performance between Network RAID types is a consideration when using low latency disk types such as SSD.

For Adaptive Optimization clusters it is recommended to only use Network RAID 10 and its variants (NR 10+1, NR 10+2, etc.). The encoding of parity bits for Network RAID 5/6 introduces additional latency which undermines the purpose for using Adaptive Optimization. Network RAID 10 provides the best balance of performance and high availability.

Preferred page

When using Network RAID there will be multiple copies of the data on one or more storage systems in the cluster. Preferred Page refers to a technology that ensures subsequent read accesses to the same page of data will occur from the same storage system. In an Adaptive Optimization enabled cluster this is beneficial as it ensures that the same page of data replicated across multiple storage systems will only be elevated to a faster tier on a single storage system. This ensures efficient use of higher cost resources.

Due to the equal distribution of data within the storage cluster a volume with a portion of data residing in a faster tier will have an equal amount of that portion in the faster tier on each storage system in the cluster.

Example:
In a cluster with 4 storage systems and a volume with 40 percent of its logical data in Tier 0, an equal portion of that 40 percent will reside in Tier 0 on each storage system. The end result will be that a unique 10 percent of the volume will reside in Tier 0 on each storage systems to make up the 40 percent of the volume that resides in Tier 0 across the cluster.

Keep in mind that when using Network RAID 10, 40 percent of the logical volume equates to 20 percent of it physical data when taking into account replicated copies of all pages. The following diagram helps to clarify this concept.
VIP Load Balancing and the StoreVirtual DSM for MPIO

The load balancing type and configuration for the application server will have an impact on performance for very fast disk types such as SSD in certain scenarios. In order to discuss this it is important to first have an understanding of the methods available for host based multi-pathing with StoreVirtual products.

VIP Load Balancing—Description and Considerations

VIP Load Balancing describes a method for distributing the client iSCSI sessions in a StoreVirtual cluster in such a way that the number of sessions are distributed equally across all of the storage systems in the cluster. A Virtual IP address is assigned to the cluster as a single point of contact and requests for connectivity from iSCSI initiator to target are redirected to a storage system in the cluster. This storage system then acts as the point of contact into the cluster for all subsequent IO data transfer for that particular session and volume. We will refer to this point of contact as the Gateway System.

Subsequent iSCSI initiator requests associated with other volumes in the cluster will be redirected to a different storage system in a round robin fashion so that iSCSI sessions are distributed equally across storage systems.

For best performance in AO enabled clusters utilizing SSD devices where very low latency is a primary concern, location of the Gateway System becomes important when we consider the topology of a multi-site cluster.

Multi-site clusters maintain a copy of data at two or more distinct physical locations. By assigning storage systems at a physical site to a logical site within the StoreVirtual Centralized Management Console we are also specifying in which physical locations copies of the data will reside. For example, creating a Volume using Network RAID 10 in a multi-site cluster with two logical sites specified, instructs the StoreVirtual systems to maintain a distinct copy of the data at each site.

Application servers can also be assigned to one of the logical sites. This is known as Site Preference. By specifying Site Preference for an application server we ensure that the Gateway System for an assigned volume will reside at the same site. This means that communication to the Storage System coordinating the retrieval of data is local and does not traverse the network link between sites.

Figure 11. Site membership showing assignment of storage systems and servers in a multi-site cluster.
Figure 12. Shows a multi-site cluster and an application server with site preference established, accessing a volume. The redirected iSCSI session resolves to a Gateway System at the same site.

With VIP Load Balancing some data will always be retrieved from the remote site but Site Preference minimizes the number of network hops and thus the latency for volumes connected to applications servers at the local site. In the case of shared volumes, such as those used with server clustering, both application servers must access the volume through the same Gateway System.

StoreVirtual DSM for Microsoft® MPIO—description and considerations

The StoreVirtual Device Specific Module (DSM) for Microsoft Multipath I/O is an HP developed multipath solution for the MPIO framework that contains hardware-specific information for optimizing connectivity with StorageVirtual clusters. The DSM is designed to work exclusively with the Microsoft iSCSI Initiator and MPIO.

Rather than the single Gateway approach discussed previously with VIP Load Balancing, the DSM allows, for each connected volume, a unique iSCSI session for data transfer to be initiated with each StoreVirtual storage system in a cluster. A passive algorithm embedded in the DSM code provides instruction on the correct location of data pages within the cluster and maintains the correct Network RAID data layout. Rather than the need to proxy I/O request through storage systems, data is read directly from each storage system. In the case of written pages the first page will land on a storage system it is intended for and a subsequent page will then be passed on from the storage system to a second system (or more) based on the Network RAID type in use.

Figure 13. Illustrates a multi-site StoreVirtual cluster and an application server with the HP DSM for MPIO installed. In this case Round Robin is enabled in MPIO.
When using The HP DSM in Adaptive Optimization enabled clusters, lower latency may result because no proxy of pages occur through a Storage Gateway system for reads. Direct access to and from low latency SSD drives preserves the low latency requirement of enterprise applications.

Another aspect of using the HP DSM for MPIO with AO is to consider the impact of traversing the inter-site link in multi-site clusters. A StoreVirtual multi-site cluster operates as a single cohesive unit so it is not possible to isolate all traffic to only one site. However, when using the HP DSM for MPIO it is possible to assign the application server to its local site in the StoreVirtual Centralized Management Console and isolate the read traffic to the storage systems on the local site. There are tradeoffs in doing so but it does eliminate the inter-site link latency for read page accesses that might otherwise impede the benefit of data stored on the faster SSD tier. Data is striped evenly across all disk within a tier so the trade off is that read accesses from spinning disk devices will be limited to the total IOP capability of just the disks that reside at the local site.

Keep in mind that replicated write pages for Network RAID must still be copied to storage systems at the remote site in all cases.

**Figure 14.** In a multi-site cluster with the application server assigned to the local site the HP DSM for MPIO restricts iSCSI sessions to only the storage systems at the local site for read access

**Tiered storage devices for StoreVirtual VSA**

Adaptive Optimization requires at least one storage system with two distinct storage devices. For StoreVirtual appliance hardware platforms this will be SSD and SAS rotational disks. In the StoreVirtual VSA the underlying storage is determined by the storage administrator.

While it is possible to use AO in clusters with only some of the storage systems AO capable, best performance results will be seen when all storage systems in the cluster are AO capable storage systems.

For detailed information on hardware requirements and installation for VSA please refer to the [HP StoreVirtual Storage VSA Installation and Configuration Guide](#).

**VSA recommendations for tiered storage devices**

The StoreVirtual VSA requires the storage administrator to specify the two (2) separate storage devices used for Tier 0 and Tier 1 during the installation process. There are three (3) combinations of drive types recommended. For each drive type combination a minimum percentage of Tier 0 to Tier 1 capacity is recommended for acceptable performance. However, in some cases this is completely dependent on the workload requirements for the storage systems.

**VSA with SSD as Tier 0 and SAS Drives as Tier 1**

StoreVirtual VSAs can be configured to use SSD as Tier 0 and SAS drives as Tier 1. This is an excellent choice and provides the best possible performance for Adaptive Optimization enabled storage systems. In this configuration it is recommended to use capacity of between 10 percent to 15 percent in the SSD Tier 0 and between 85 percent to 90 percent capacity in the SAS Tier 1.

This configuration is recommended when at least some meaningful performance capability is required from Tier 1. That is, data in Tier 1 is still actively requested and requires excellent performance but not as demanding as that which can be provided from SSD.
VSA with SAS drives as Tier 0 and Nearline Drives as Tier 1
StoreVirtual Virtual Storage Appliances can be configured to use SAS drives as Tier 0 and Nearline drives as Tier 1. In this configuration, it is recommended to size Tier 0 with a number of SAS drives capable of handling all IOP requirements for the storage system. Nearline drives are recommended for archive data and less demanding workloads such as file shares and should not be computed in performance sizing. An appropriate number of Nearline disks can then be added to address additional capacity requirements.

This is a good choice for moderate workload needs and to speed up systems handling file serving tasks.

VSA with SAS drives/RAID 10 as Tier 0 and SAS drives/RAID 5 as Tier 1
StoreVirtual Virtual Storage Appliances can be configured to use SAS drives configured with RAID 10 as Tier 0 and SAS drives configured with RAID 5 as Tier 1. In this configuration, performance differences between the tiers will not be as profound as other recommended drive configurations. Sizing will be dependent on the workload requirements of the system, and there is no specific recommendation for the percentage of each drive configuration that should be assigned to each tier.

This choice provides additional flexibility and performance enhancement to standard RAID 5 SAS implementations.

VSA with SSD as Tier 0 and Nearline Drives as Tier 1
This particular option is not generally recommended. Given that Nearline drives should not be used for performance intensive applications, this configuration would require enough SSD capacity to handle all IO operations for the storage system. This will drive up cost and make this option less attractive in many cases. It is still possible to use this option in rare corner case scenarios.

Assigning storage devices to StoreVirtual VSA
Multiple VSA virtual machines use local or SAN based storage devices presented from the Hypervisor to create a distributed storage cluster that provides high availability iSCSI devices that can survive server and site outages.

The StoreVirtual Virtual Storage Appliance (VSA) allows the storage administrators to select the storage devices that will be used for the two (2) tiers.

It is very important that the correct (higher performing) storage device is assigned to Tier 0 and lower performing storage device assigned to Tier 1. A mismatch in storage device assignment across storage nodes will hinder storage performance.

Figure 15. During VSA installation, storage devices for Tier 0 and Tier 1 are chosen from the available configured devices.
It is possible to reconfigure the storage devices on the VSA through the Centralized Management Console if they have been incorrectly assigned. Go to the **Storage** sub-section under the storage node and then to the **RAID Setup** tab. Under RAID Setup Tasks select Reconfigure Tier…. Here you can edit the Tier for each attached storage device. Selecting OK will restart an internal process on the VSA and then the VSA must complete data synchronization before the next VSA can be reconfigured.

**Figure 16.** For VSA assignment of storage devices to each tier is available under the Storage Node in the CMC

---

### Monitoring Adaptive Optimization

The StoreVirtual Performance Monitor provides real-time performance monitoring of storage clusters down to 5 second intervals. A large number of metrics are available that characterize performance latencies, IOPs, queue depth, page size, and many other counters critical to understanding and evaluating performance. Data can be viewed in the Centralized Management Console as counters, graphically via charts and can also be exported via log files (csv files) and SNMP traps.

Adaptive Optimization adds a new set of performance and usage counters to the HP StoreVirtual Centralized Management Console. AO statistics have their own monitor located under the cluster Performance Monitor area of the management tree. Many of the same performance metrics available for volumes are present but now broken down by each tier for individual volumes and storage systems.

The Adaptive Optimization monitor is modeled after the cluster Performance Monitor and located directly under the Performance Monitor. The “Adaptive Optimization” monitor is accessed under the “Performance Monitor” node in the left hand pane of the CMC for clusters that are AO enabled. Each console can be opened in its own separate windows for easy side-by-side analysis and comparison.

**Figure 17.** Location of the Adaptive Optimization monitor in the CMC

AO metrics include both capacity utilization and performance counters. Capacity of each tier as well as the percentage and capacity of volumes by tier are shown. Performance counters show total volume metrics as well as metrics for volumes by each tier.

For a complete glossary of performance counters available in StoreVirtual environments please refer to the HP StoreVirtual Storage User Guide

In general volume metrics are of the most interest to understanding tier utilization and characterizing application performance. Storage System metrics show overall system utilization and can be important in evaluating overall cluster utilization, individual storage system health and for troubleshooting specific performance related problems.
Using Adaptive Optimization volume capacity metrics to view utilization

Volume level metrics available in the Adaptive Optimization console include both capacity utilization of each tier as well each volume by tier. The following list of capacity metrics show the information available.

- Percent Volume Space All Tiers Used: % of all tiers that volume is currently consuming.
- Percent Volume Space Tier 0 Used: % of volume that currently resides on Tier 0.
- Percent Volume Space Tier 1 Used: % of volume that currently resides on Tier 1.
- Volume Space All Tiers Used: Total physical capacity consumed by volume in GB.
- Volume Space Tier 0 Used: Total physical capacity of Tier 0 consumed by volume in GB.
- Volume Space Tier 1 Used: Total physical capacity of Tier 1 consumed by volume in GB.

These capacity amounts show the amount of physical space consumed after taking into account the Network RAID type used. For example, if a volume shows 50 GB of consumption at the application server and Network RAID 10 is being used for the volume on the cluster, the total physical capacity of the cluster consumed would be 100 GB.

If the Percent Volume Space Tier 0 Used is 5 percent this equates to 5 GB of physical capacity consumed within Tier 0 for the volume. Due to Preferred Page technology described earlier in this paper 5 GB consumed in Tier 0 means another 5 GB would be consumed in Tier 1 to hold the replicated Network RAID 10 copy.

Figure 18. Volume information including capacity consumed by tier from the Volume Use tab

In the Centralized Management console Consumed Space refers to physical capacity after taking into account Network RAID. It should be noted that the total Consumed Space will be different than the sum of Tier 0 and Tier 1 Consumed Space. This is because the Consumed Space number is the provisioned space, whereas the tier breakdown is based on the written pages.

Monitoring Adaptive Optimization performance metrics

This section provides some key metrics and usage examples for evaluating performance of individual volumes that span multiple tiers. For a detailed explanation of the Performance Monitor, Adaptive Optimization monitor, and all performance counters available, please refer to the HP StoreVirtual Storage User Guide.

The following performance metrics are available for monitoring individual volumes as well as aggregate metrics for all volumes assigned to a given application server.

- IO Latency Reads: Average time, in milliseconds to service read request.
- IO Latency Writes: Average time, in milliseconds to service read and write requests.
- IO Latency Total: Average time, in milliseconds to service write request.
- IO’s Reads Percent: Percent of IO’s for the sample interval that are reads.
- IO’s Writes Percent: Percent of IO’s for the sample interval that are writes.
- IOPS Reads: Average read requests per second for the sample interval.
- IOPS Writes: Average write requests per second for the sample interval.
- IOPS Total: Average read+write requests per second for the sample interval.
- Tier 0 IO Reads Percent: Percent of Tier 0 IO’s for the sample interval that are reads.
- Tier 0 IO Writes Percent: Percent of Tier 0 IO’s for the sample interval that are writes.
- Tier 0 IOPS Reads: Average Tier 0 read requests per second for the sample interval.
- Tier 0 IOPS Writes: Average Tier 0 write requests per second for the sample interval.
- Tier 0 IOPS Total: Average Tier 0 read+write requests per second for the sample interval.
- Tier 1 IO Reads Percent: Percent of Tier 1 IO’s for the sample interval that are reads.
- Tier 1 IO Writes Percent: Percent of Tier 1 IO’s for the sample interval that are writes.
- Tier 1 IOPS Reads: Average Tier 1 read requests per second for the sample interval.
- Tier 1 IOPS Writes: Average Tier 1 write requests per second for the sample interval.
- Tier 1 IOPS Total: Average Tier 1 read+write requests per second for the sample interval.
After an AO Permitted volume has been operating for a period of time it is possible to analyze how much highly utilized or hot data there is compared to moderate or under-utilized data on a volume. Doing so will provide insight on the effectiveness of tiering for a given application and provide specific information that shows the utilization or IOP rates of each of the tiers that the volume spans. In addition, confirming that latency remains low for IO access to the faster tier provides assurance that the most critical application data is being given the highest priority.

To understand the workload of a volume and the division of workload between the two (2) tiers, there are some key metrics that can be used.

**Performance monitoring process**

First, to understand where the data resides and what percentage of the data resides on each tier view the following metrics. Understanding the data layout provides some context for then analyzing the performance characteristics of the volume. It is common for only a small percentage of a volumes data to be highly accessed and therefore reside in Tier 0.

- Volume Space Tier 0 Used, Volume Space Tier 1 Used
- Percent Volume Space Tier 0 Used, Percent Volume Space Tier 1 Used

Next is to understand the utilization rates and where the IO is being serviced. We can view the IOs per second for the entire volume and for each tier. Given that large numbers of page accesses provides the criteria for elevation to a faster tier it is common to see higher utilization in Tier 0 data even though much less capacity may reside there.

- IOPS Total, IOPS Reads, IOPS Writes for entire volume
- Tier 0 IOPS Total, Tier 0 IOPS Read, Tier 0 IOPS Write
- Tier 1 IOPS Total, Tier 1 IOPS Read, Tier 1 IOPS Write

Finally we want to understand how fast the IO is being serviced as this is often the most critical metric for many applications. This can be viewed through a variety of latency counters. At this time latency if available at the volume level and not individually for each tier.

- IO Latency Total, IO Latency Reads, IO Latency Writes for entire volume

These metrics provides a clear view of how efficiently the workload is being serviced by the storage cluster and allows pinpointing the location of data and where it is being serviced. By viewing the average latency of both the cluster and individual volumes the system can be verified to be working correctly and efficiently.
Conclusion

Storage tiering solutions provide value to organizations seeking to increase the performance to cost ratio of continually growing storage demands. However, many such solutions introduce additional management complexity and are limited by lack of responsiveness or excessive data movement. HP StoreVirtual storage products now feature an advanced sub-LUN auto-tiering technology that addresses the traditional challenges of many traditional tiered storage solutions. Adaptive Optimization for StoreVirtual is a fully automated solution that tracks in-flight use statistics and re-provisions only the most critical data to low latency SSD devices. No user intervention is ever required and the solution integrates seamlessly with existing StoreVirtual storage solutions.

Adaptive Optimization is designed to be an “autonomic” or self-regulating system which adapt to changing workloads without user intervention. An important balance between internal data movement and application IO is maintained using comparative usage metrics and ensures data is not moved unnecessarily or too often. Data is tracked and moved at a granular 256 KB page size isolating only high priority data and in turn increasing system efficiency.

Adaptive Optimization technology introduces a new way to address high performance use cases with HP StoreVirtual Storage. The flexibility to implement this technology with StoreVirtual VSA provides a unique ability to design customized, software based tiered storage solutions to meet a wide range of application specific performance needs without needing to over-provision expensive high performance storage to satisfy the most demanding workloads.

Learn more at
hp.com/go/storevirtual4000