Hewlett Packard
Enterprise

HPE MSA 1040/2040

Best practices

Technical white paper

Contents

About this document

This white paper highlights the best practices for optimizing the HPE MSA 1040/2040, and should be used in conjunction with other HPE Modular Smart Array (MSA) manuals. MSA technical user documentations can be found at hp.com/go/msa1040 and hp.com/go/msa2040.

Intended audience

This white paper is intended for HPE MSA 1040/2040 administrators with previous storage area network (SAN) knowledge. It offers MSA practices that can contribute to an MSA best customer experience.

This paper is also designed to convey best practices in the deployment of the HPE MSA 1040/2040 array.

Prerequisites

Prerequisites for using this product include knowledge of:

- Networking
- Storage system configuration
- SAN management
- Connectivity methods such as direct attach storage (DAS), Fibre Channel, and serial attached SCSI (SAS)
- Internet SCSI (iSCSI) and Ethernet protocols

Related documentation

In addition to this guide, please refer to other documents or materials for this product:

- HPE MSA System Racking Instructions
- HPE MSA 1040 Installation Guide
- HPE MSA 1040 System Cable Configuration Guide
- HPE MSA 1040 User Guide
- HPE MSA 1040 SMU Reference Guide
- HPE MSA 1040 CLI Reference Guide
- HPE MSA 2040 Installation Guide
- HPE MSA 2040 System Cable Configuration Guide
- HPE MSA 2040 User Guide
- HPE MSA 2040 SMU Reference Guide
- HPE MSA 2040 CLI Reference Guide

You can find the HPE MSA 1040 documents at: hp.com/go/msa1040

You can find the HPE MSA 2040 documents at: hp.com/go/msa2040

Introduction

The HPE MSA 1040 is designed for entry-level market needs featuring 8 Gb Fibre Channel, 6/12 Gb SAS, 1GbE, and 10GbE iSCSI protocols. The MSA 1040 arrays leverages new 4th-generation controller architecture with a new processor, 2-host ports per controller and 4 GB cache per controller.

An outline of the MSA 1040 features:

- New controller architecture with a new processor
- 6 GB cache per controller (Data [Read/Write] cache = 4 GB and Metadata and System OS memory = 2 GB)
- Support for solid state drives (SSDs)
- 6 Gb/12 Gb SAS connectivity
- Support for MSA Fanout SAS cables
- 2 host ports per controller
- 4 Gb/8 Gb FC connectivity
- 1GbE/10GbE iSCSI connectivity
- Support for up to 4 disk enclosures including the array enclosure
- Support for up to 99 small form factor (SFF) drives
- Support for Thin Provisioning; requires a license[1]
- New Web Interface[1]
- Support for Sub-LUN Tiering; requires a license[1]
- Support for Read Cache[1]
- Support for Performance Tier[1]; requires a license
- Wide Striping[1]; requires a license; Wide Striping allows more hard drives behind a single volume to improve performance (e.g., >16 drives for a volume).

The HPE MSA 2040, a high-performance storage system designed for HPE customers desiring 8 and/or 16 Gb Fibre Channel, 6 Gb SAS and/or 12 Gb SAS, and 1GbE and/or 10GbE iSCSI connectivity with 4 host ports per controller. The MSA 2040 storage system provides an excellent value for customers needing performance balanced with price to support initiatives such as consolidation and virtualization.

The MSA 2040 delivers this performance by offering:

- New controller architecture with a new processor
- 6 GB cache per controller (Data [Read/Write] cache = 4 GB and Metadata and System OS memory = 2 GB)
- Support for solid state drives
- 4 host ports per controller
- 4 Gb/8 Gb/16 Gb FC connectivity
- 6 Gb/12 Gb SAS connectivity
- 1GbE/10GbE iSCSI connectivity
- Support for both FC and iSCSI in a single controller
- Support for up to 8 disk enclosures including the array enclosure
- Support for up to 199 small form factor (SFF) drives
- Support for Full Drive Encryption (FDE) using Self-Encrypting Drives (SED)[2]

[1] With GL200 and greater Firmware. With GL220 and greater Firmware for MSA 1040 SSD features. Creation of an SSD Virtual Disk Group for both read and write capabilities requires a Performance Auto Tiering License (D4T79A/D4T79AAE).
[2] SED drives are only supported in the MSA 2040.

- Support for Thin Provisioning[3]
- Support for Sub-LUN Tiering[3]
- Support for Read Cache[3]
- Support for Performance Tier[3]; requires a license
- New Web Interface[3]
- Wide Striping[3]; Wide Striping allows more hard drives behind a single volume to improve performance (e.g., >16 drives for a volume).

The HPE MSA 1040/2040 storage system brings the performance benefits of SSDs to MSA array family customers. This array has been designed to maximize performance by using high-performance drives across all applications sharing the array.

The HPE MSA 2040 storage systems are positioned to provide an excellent value for customers needing increased performance to support initiatives such as consolidation and virtualization.

The HPE MSA 1040/2040 storage systems ship standard with a license for 64 Snapshots and Volume Copy for increased data protection. There is also an optional license for 512 snapshots. The HPE MSA 1040/2040 can also replicate data between arrays (P2000 G3, MSA 1040, or MSA 2040 SAN model using FC [linear volumes only] or iSCSI) with the optional Remote Snap feature.

Note
With Remote Snap, the MSA array can replicate linear volumes between the MSA 1040/MSA 2040 and P2000 G3 or MSA 1040/MSA 2040, and virtual volumes between two MSA 1040/MSA 2040 arrays running GL220 FW and greater.

Terminology

Virtual Disk (Vdisk): The Vdisk nomenclature is being replaced by Disk Group. For Linear Storage and in the Storage Management Utility (SMU) Version 2 you will still see references to Vdisk; for Virtual Storage and the SMU Version 3 you will see Disk Group. Vdisk and Disk Group are essentially the same. Vdisks (Linear Disk Groups) have additional RAID types; NRAID, RAID 0, and RAID 3 are available only in the CLI, and RAID 50 is available in both the CLI and SMU.

Linear Storage: Linear Storage is the traditional storage that has been used for the four MSA generations. With Linear Storage, the user specifies which drives make up a RAID Group and all storage is fully allocated.

Virtual Storage: Virtual Storage is an extension of Linear Storage. Data is virtualized not only across a single disk group, as in the linear implementation, but also across multiple disk groups with different performance capabilities and use cases.

Page: An individual block of data residing on a physical disk. For Virtual Storage, the page size is 4 MB. A page is the smallest unit of data that can be allocated, de-allocated, or moved between virtual disk groups in a tier or between tiers.

Disk Group: A Disk Group is a collection of disks in a given redundancy mode (RAID 1, 5, 6, or 10 for Virtual Disk Groups and NRAID and RAID 0, 1, 3, 5, 6, 10, or 50 for Linear Disk Groups). A Disk Group is equivalent to a Vdisk in Linear Storage and utilizes the same proven fault tolerant technology used by Linear Storage. Disk Group RAID level and size can be created based on performance and/or capacity requirements. With GL200 and greater firmware multiple Virtual Disk Groups can be allocated into a Storage Pool for use with the Virtual Storage features; while Linear Disk Groups are also in Storage Pools, there is a one-to-one correlation between Linear Disk Groups and their associated Storage Pools.

Storage Pools: The GL200 firmware and greater introduces Storage Pools which are comprised of one or more Virtual Disk Groups or one Linear Disk Group. For Virtual Storage, LUNs are no longer restricted to a single disk group as with Linear Storage. A volume's data on a given LUN can now span all disk drives in a pool. When capacity is added to a system, users will benefit from the performance of all spindles in that pool.

[3] With GL200 and greater Firmware. With GL220 and greater Firmware for MSA 1040 for SSD features Creation of an SSD Virtual Disk Group for both read and write capabilities requires a Performance Auto Tiering License (D4T79A/D4T79AAE).

When leveraging Storage Pools, the MSA 1040/2040 supports large, flexible volumes with sizes up to 128 TB and facilitates seamless capacity expansion. As volumes are expanded data automatically reflows to balance capacity utilization on all drives.

Logical Unit Number (LUN): The MSA 1040/2040 arrays support 512 volumes and up to 512 snapshots in a system. All of these volumes can be mapped to LUNs. Maximum LUN sizes are up to 128 TB and the LUNs sizes are dependent on the storage architecture: Linear vs. Virtualized. Thin Provisioning allows the user to create the LUNs independent of the physical storage.

Thin Provisioning: Thin Provisioning allows storage allocation of physical storage resources only when they are consumed by an application. Thin Provisioning also allows over-provisioning of physical storage pool resources allowing ease of growth for volumes without predicting storage capacity upfront.

Thick Provisioning: All storage is fully allocated with Thick Provisioning. Linear Storage always uses Thick Provisioning.

Tiers: Disk tiers are comprised of aggregating 1 or more Disk Groups of similar physical disks. The MSA 1040/2040 support 3 distinct tiers:

1. A Performance tier with SSDs
2. A Standard SAS tier with Enterprise SAS HDDs
3. An Archive tier utilizing Midline SAS HDDs

Prior to GL200 firmware, the MSA 1040/2040 operated through manual tiering, where LUN level tiers are manually created and managed by using dedicated Vdisks and volumes. LUN level tiering requires careful planning such that applications requiring the highest performance be placed on Vdisks utilizing high performance SSDs. Applications with lower performance requirements can be placed on Vdisks comprised of Enterprise SAS or Midline SAS HDDs. Beginning with GL200 and greater firmware, the MSA 1040/2040 supports Sub-LUN Tiering and automated data movement between tiers.

The MSA 1040/2040 automated tiering engine moves data between available tiers based on the access characteristics of that data. Frequently accessed data contained in pages will migrate to the highest available tier delivering maximum I/O's to the application. Similarly, "cold" or infrequently accessed data is moved to lower performance tiers. Data is migrated between tiers automatically such that I/O's are optimized in real time.

The Archive and Standard Tiers are provided at no charge on the MSA 2040 platform beginning with GL200 and greater firmware. The MSA 1040 requires a license when using the Archive and Standard Tiers. A Performance Tier utilizing a fault tolerant SSD Disk Group is a paid feature that requires a license for both the MSA 1040 and MSA 2040. Without the Performance Tier license installed, SSDs can still be used as Read Cache with the Sub-LUN Tiering feature. Sub-LUN Tiering from SAS MDL (Archive Tier) to Enterprise SAS (Standard Tier) drives is provided at no charge for the MSA 2040.

Note
The MSA 1040 requires a license to enable Sub-LUN Tiering and other Virtual Storage features such as Thin Provisioning.

Read Cache: Read Cache is an extension of the controller cache. Read Cache allows a lower cost way to get performance improvements from SSD drives.

Sub-LUN Tiering: Sub-LUN Tiering is a technology that allows for the automatic movement of data between storage tiers based on access trends. In the MSA 1040/2040, Sub-LUN Tiering places data in a LUN that is accessed frequently in higher performing media while data that is infrequently accessed is placed in slower media.

General best practices

Implement Virtual Disk Groups

Beginning with the release of the GL200 firmware, storage administrators can implement features such as Thin Provisioning, Sub-LUN Tiering, Read Cache, and Wide Striping.

HPE recommends utilizing virtual storage to take advantage of the advanced virtualization features of the firmware.

Use version 3 of the Storage Management Utility

With the release of the GL200 firmware, there is an updated version of the Storage Management Utility (SMU). This new Web Graphical User Interface (GUI) allows the user to use the new features of the GL200 firmware. This is version 3 of the SMU (V3).

SMU V3 is the recommended Web GUI. SMU V3 can be accessed by adding "/v3" to the IP address of the MSA array:
https://<MSA array IP>/v3

SMU V3 must be used for virtual volume replication.

The required Web GUI is SMU V2 if you are using the replication features of the MSA 1040/2040 for linear volumes. SMU V2 can be accessed by adding "/v2" to the IP address of the MSA array: **https://<MSA array IP>/v2**

Become familiar with the array by reading the manuals

The first recommended best practice is to read the corresponding guides for either the HPE MSA 1040 or HPE MSA 2040. These documents include the User Guide, the Storage Management Utility (SMU) Reference Guide, or the Command Line Interface (CLI) Reference Guide. The appropriate guide will depend on the interface that you will use to configure the storage array. Always operate the array in accordance with the user manual. In particular, never exceed the environmental operation requirements.

Other HPE MSA 1040 and HPE MSA 2040 materials of importance to review are:

- The HPE MSA Remote Snap Technical white paper located at: h20195.www2.hpe.com/v2/GetPDF.aspx/4AA1-0977ENW.pdf

Stay current on firmware

Use the latest controller, disk, and expansion enclosure firmware to benefit from the continual improvements in the performance, reliability, and functionality of the HPE MSA 1040/2040. For additional information, see the release notes and release advisories for the respective MSA products.

This information can be located at: hp.com/go/msa1040 or hp.com/go/msa2040

Use tested and supported configurations

Deploy the MSA array only in supported configurations. Do not risk the availability of your critical applications to unsupported configurations. HPE does not recommend nor provide HPE support for unsupported MSA configurations.

HPE's primary portal used to obtain detailed information about supported HPE Storage product configurations is single point of connectivity knowledge (SPOCK). An HPE Passport account is required to enter the SPOCK website.

SPOCK can be located at: hpe.com/storage/spock

Understand what a host is from the array perspective

An initiator is analogous to an external port on a host bus adapter (HBA). An initiator port does not equate to a physical server, but rather a unique connection on that server. For example, a dual port FC HBA has two ports and therefore there are two unique initiators, and the array will show two separate initiators for that HBA.

With the new GL200 firmware, there is a new definition for host. A host is a collection of 1 or more initiators. GL200 firmware also supports more initiators than in previous versions of MSA 1040/2040 firmware. Previous versions of firmware were limited to supporting only 64 hosts with

1 initiator per host; the latest firmware can support 512 hosts with multiple initiators per host.

In the GL200 firmware, the array supports the grouping of initiators under a single host and grouping hosts into a host group. Grouping of initiators and hosts allows simplification of the mapping operations.

Rename hosts to a user friendly name

Applying friendly names to the hosts enables easy identification of which hosts are associated with servers and operating systems. A recommended method for acquiring and renaming Worldwide Name (WWN) is to connect one cable at a time and then rename the WWN to an identifiable name.

The procedure below outlines the steps needed to rename hosts using version 3 of the SMU.

1. Log into the SMU and click “Hosts” from the left frame.
2. Locate and highlight the WWN (ID) you want to name.
3. From the Action button, click Modify Initiator.
4. Type in the initiator nickname and click OK.
5. Repeat for additional initiator connections.

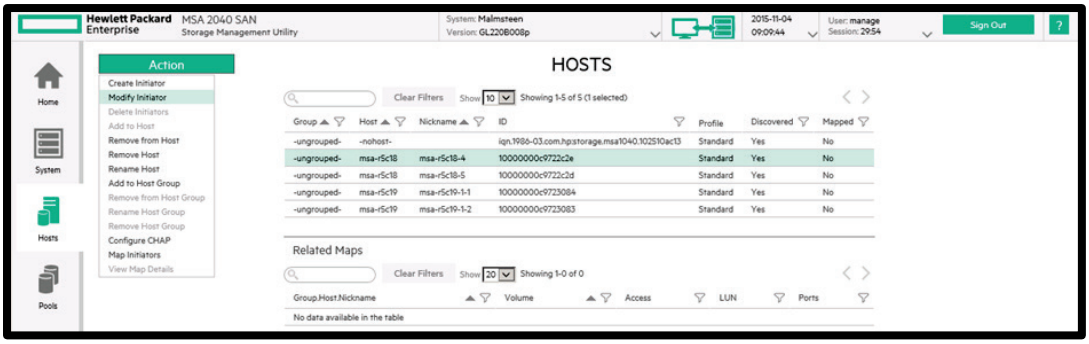

Figure 1. Renaming hosts

The recommended practice would be to use initiator nicknaming as outlined in figure 1, host aggregating of initiators and the grouping of hosts using V3 SMU.

Disk Group initialization for Linear Storage

During the creation of a Disk Group for Linear Storage, the user has the option to create a Disk Group in online mode (default) or offline mode. If the “online initialization” option is enabled, you can use the Disk Group while it is initializing. Online initialization takes more time because parity initialization is used during the process to initialize the Disk Group. Online initialization is supported for all HPE MSA 1040/2040 RAID levels except for RAID 0 and NRAID. Online initialization does not impact fault tolerance.

If the “online initialization” option is unchecked, which equates to “offline initialization”, you must wait for initialization to complete before using the Disk Group for Linear Storage, but the initialization takes less time to complete.

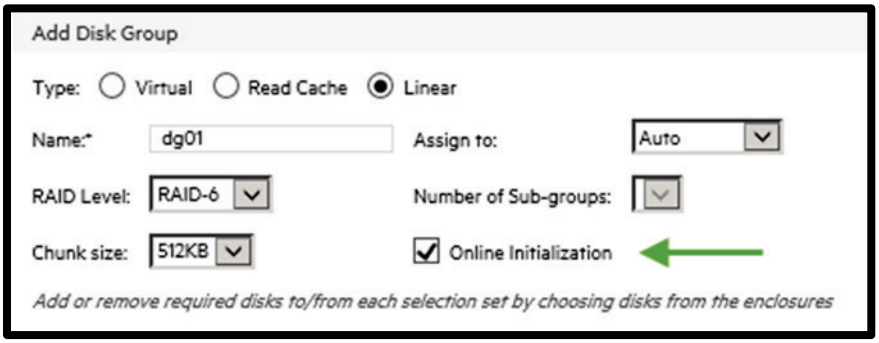

Figure 2. Choosing online or offline initialization

Best practice for monitoring array health

Setting up the array to send notifications is important for troubleshooting and log retention.

Configure email and SNMP notifications

The Storage Management Utility (SMU) version 3 is the recommended method for setting up email and SNMP notifications. Setting up these services is easily accomplished by using a Web browser; to connect; type in the IP address of the management port of the HPE MSA 1040/2040.

Email notifications can be sent to up to as many as three different email addresses. In addition to the normal email notification, enabling managed logs with the "Include logs as an email attachment" option enabled is recommended. When the "Include logs as an email attachment" feature is enabled, the system automatically attaches the system log files to the managed logs email notifications sent. The managed logs email notification is sent to an email address which will retain the logs for future diagnostic investigation.

The MSA 1040/2040 storage system has a limited amount of space to retain logs. When this log space is exhausted, the oldest entries in the log are overwritten. For most systems this space is adequate to allow for diagnosing issues seen on the system. The managed logs feature notifies the administrator that the logs are nearing a full state and that older information will soon start to get overwritten. The administrator can then choose to manually save off the logs. If "Include logs as an email attachment" is also checked, the segment of logs which is nearing a full state will be attached to the email notification. Managed logs attachments can be multiple MB in size.

Enabling the managed logs feature allows log files to be transferred from the storage system to a log-collection system to avoid losing diagnostic data. The "Include logs as an email attachment" option is disabled by default.

HPE recommends enabling SNMP traps. Version 1 SNMP traps can be sent to up to three host trap addresses (i.e., HPE SIM Server or other SNMP server). To send version 3 SNMP traps, create a SNMPv3 user with the Trap Target account type. Use SNMPv3 traps rather than SNMPv1 traps for greater security. SNMP traps can be useful in troubleshooting issues with the MSA 1040/2040 array.

To configure email and version 1 SNMP settings in the SMU, click **Home -> Action -> Set Up Notifications**. Enter the correct information for email, SNMP, and Managed Logs. See figure 4.

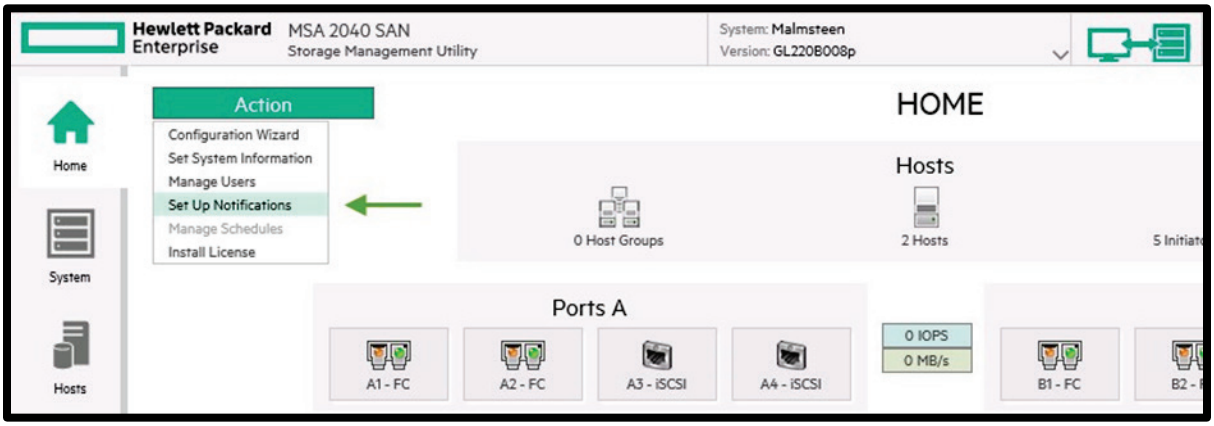

Figure 3. Setting Up Management services

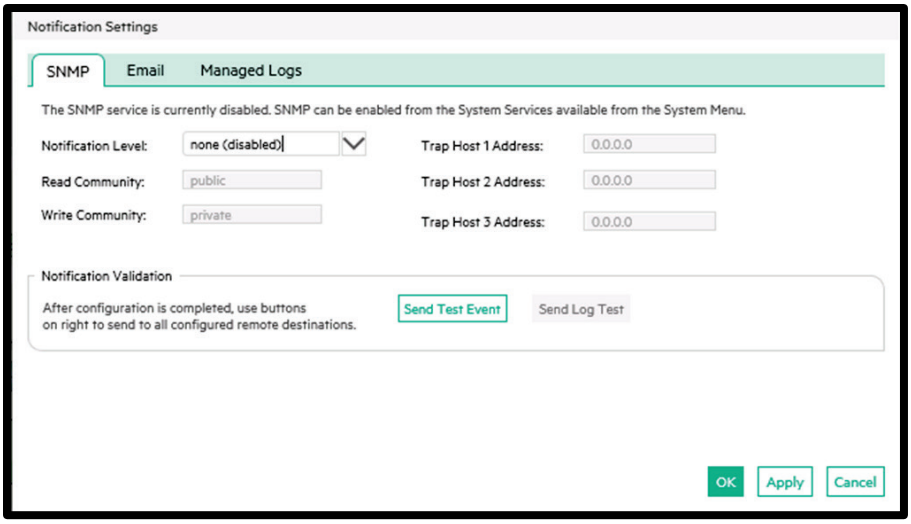

Figure 4. SNMP, Email, and Managed Logs Notification Settings

To configure SNMPv3 users and trap targets, click **Home | Action | Manage Users**. See figure 5.

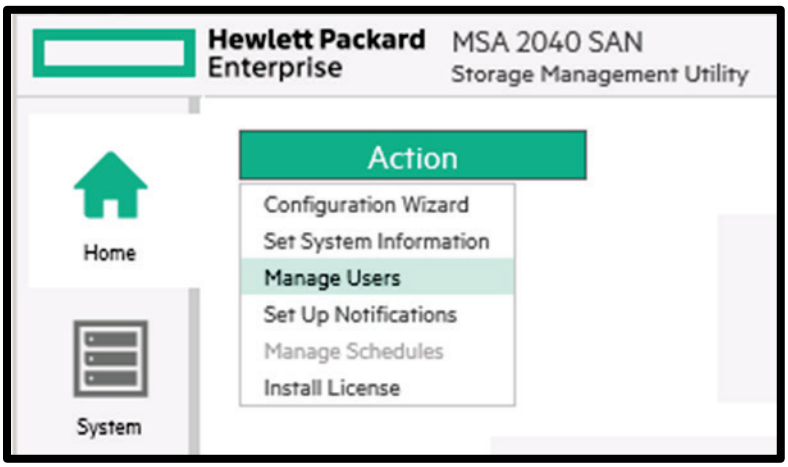

Figure 5. Manage Users

Enter the correct information for SNMPv3 trap targets. See figure 6.

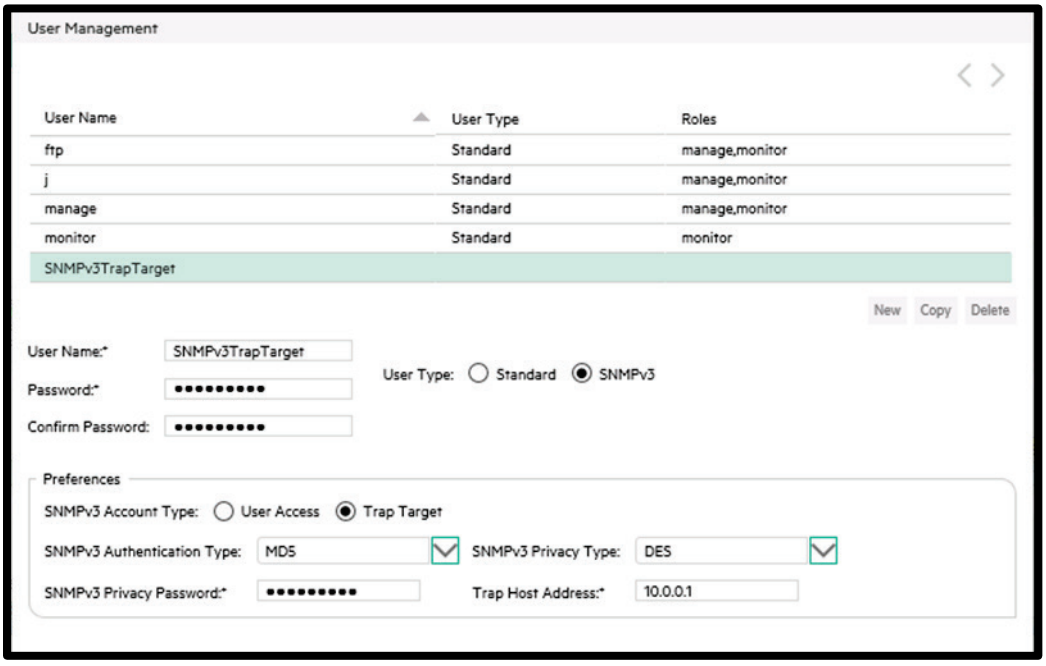

Figure 6. User Management

Setting the notification level for email and SNMP

Setting the notification level to Warning, Error, or Critical on the email and SNMP configurations will ensure that events of that level or above are sent to the destinations (i.e., SNMP server, SMTP server) set for that notification. HPE recommends setting the notification level to Warning.

HPE MSA 1040/2040 notification levels:

- Warning will send notifications for all Warning, Error, or Critical events.
- Error will only send Error and Critical events.
- Critical will only send Critical events.

Sign up for proactive notifications for the HPE MSA 1040/2040 array

Sign up for proactive notifications to receive MSA product advisories. Applying the suggested resolutions can enhance the availability of the product.

Sign up for the notifications at: hpe.com/go/myadvisory

Best practices for provisioning storage on the HPE MSA 1040/2040

The release of the GL200 firmware for the MSA 1040/2040 introduced virtual storage features such as Thin Provisioning and Sub-LUN Tiering. The section below will assist in the best methods for optimizing these features for the MSA 1040/2040.

Thin Provisioning

Thin Provisioning is a storage allocation scheme that automatically allocates storage as your applications need it.

Thin Provisioning dramatically increases storage utilization by removing the equation between allocated and purchased capacity. Traditionally, application administrators purchased storage based on the capacity required at the moment and for future growth. This resulted in over-purchasing capacity and unused space.

With Thin Provisioning, applications can be provided with all of the capacity to which they are expected to grow but can begin operating on a smaller amount of physical storage. As the applications fill their storage, new storage can be purchased as needed and added to the array's storage pools. This results in a more efficient utilization of storage and a reduction in power and cooling requirements.

Thin Provisioning is enabled by default for virtual storage. The overcommit setting only applies to virtual storage and simply lets the user oversubscribe the physical storage (i.e., provision volumes in excess of physical capacity). If a user disables overcommit, they can only provision virtual volumes up to the available physical capacity. The overcommit setting is not applicable on traditional linear storage.

Overcommit is performed on a per pool basis and using the "Change Pool Settings" option. To change the Pool Settings to overcommit disabled:

1. Open V3 of the SMU and select "Pools"
2. Click "Change Pool Settings"
3. Uncheck the "Enable overcommitment of pool?" by clicking the box. See figures 7 and 8.

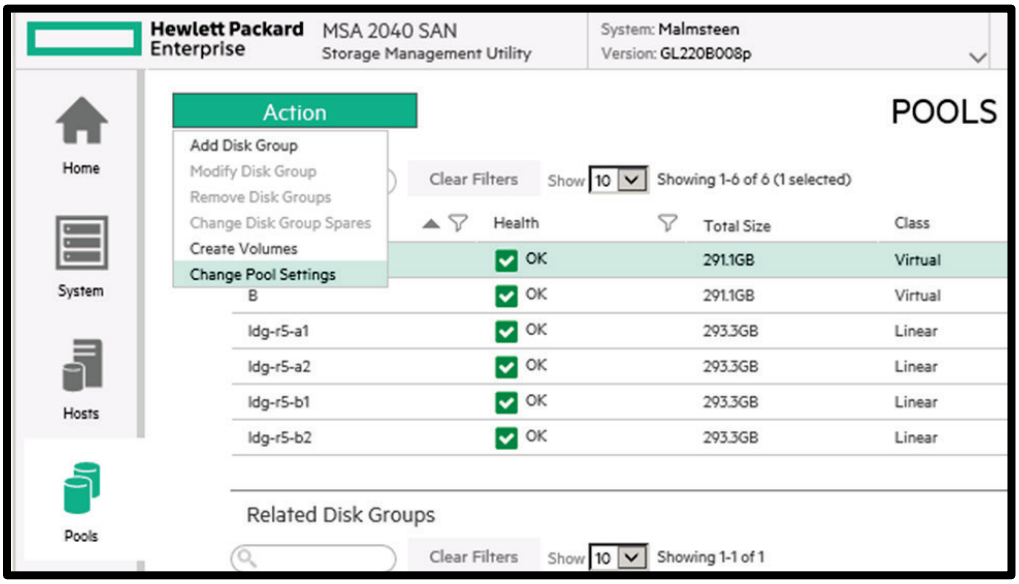

Figure 7. Changing Pool Settings

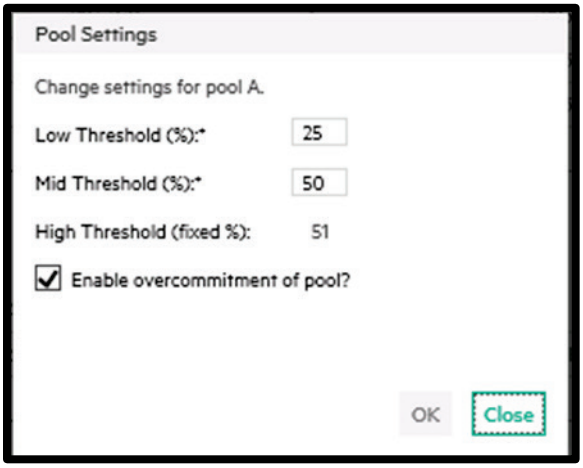

Figure 8. Disabling the overcommit of the pool

Thresholds and Notifications

If you use Thin Provisioning, monitor space consumption and set notification thresholds appropriately for the rate of storage consumption. The thresholds and notifications below can help determine when more storage needs to be added.

Users with a manage role can view and change settings that affect the thresholds and corresponding notifications for each storage pool.

- Low Threshold—When this percentage of pool capacity has been used, Informational event 462 is generated to notify the administrator. This value must be less than the Mid Threshold value. The default is 25 percent.
- Mid Threshold—When this percentage of pool capacity has been used, Warning event 462 is generated to notify the administrator to add capacity to the pool. This value must be between the Low Threshold and High Threshold values. The default is 50 percent. If the over-commitment setting is enabled, the event has informational severity; if the over-commitment setting is disabled, the event has Warning severity.
- High Threshold—When this percentage of pool capacity has been used, Warning event 462 is generated to alert the administrator that it is critical to add capacity to the pool. This value is automatically calculated based on the available capacity of the pool minus reserved space. This value cannot be changed by the user.

See figures 7 and 8 above on how to set the thresholds.

T10 unmap for Thin Reclaim

Unmap is the ability to reclaim thinly provisioned storage after the storage is no longer needed. There are procedures to reclaim unmap space when using Thin Provisioning and ESX.

The user should run the unmap command with ESX 5.0 Update 1 or higher to avoid performance issues. In ESX 5.0, unmap is automatically executed when deleting or moving a Virtual Machine.

In ESX 5.0 Update 1 and greater, the unmap command was decoupled from auto reclaim; therefore, use the VMware® vSphere CLI command to run unmap command.

See VMware documentation for further details on the unmap command and reclaiming space.

Pool Balancing

Creating and balancing storage pools properly can help with performance of the MSA array. Hewlett Packard Enterprise recommends keeping pools balanced from a capacity utilization and performance perspective. Pool balancing will leverage both controllers and balance the workload across the two pools.

Assuming symmetrical composition of storage pools, create and provision storage volumes by the workload that will be used. For example, an archive volume would be best placed in a pool with the most available Archive Tier space. For a high performance volume, create the Disk Group on the pool that is getting the least amount of I/O on the Standard and Performance Tiers.

Determining the pool space can easily be viewed in V3 of the SMU. Simply navigate to “Pools” and click the name of the pool.

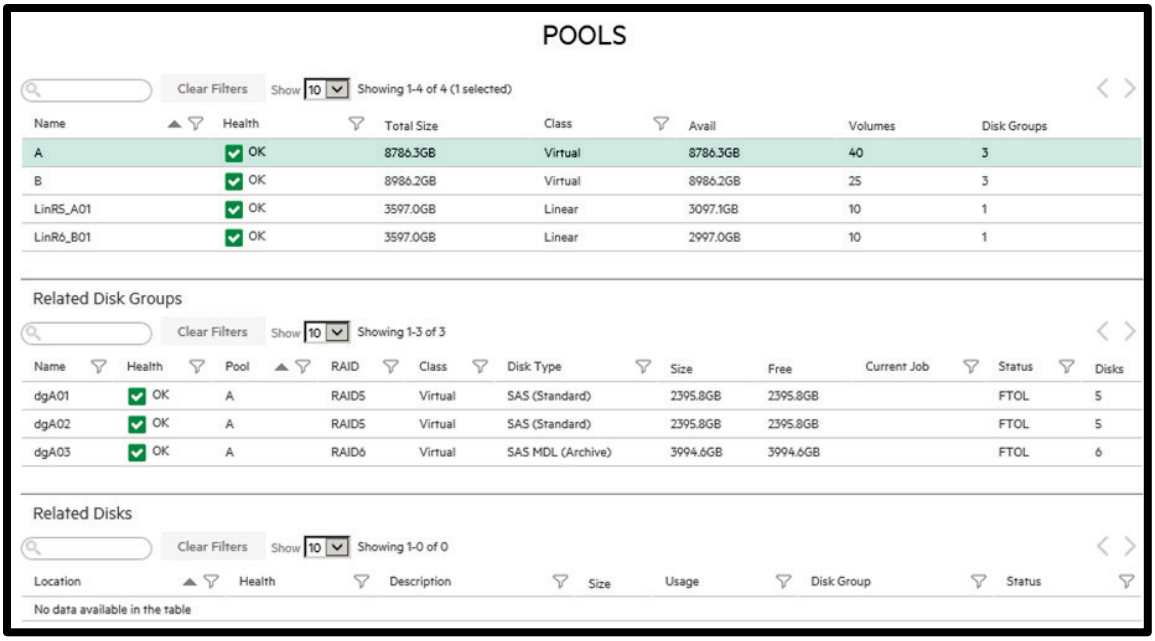

Figure 9. MSA Pool A screen

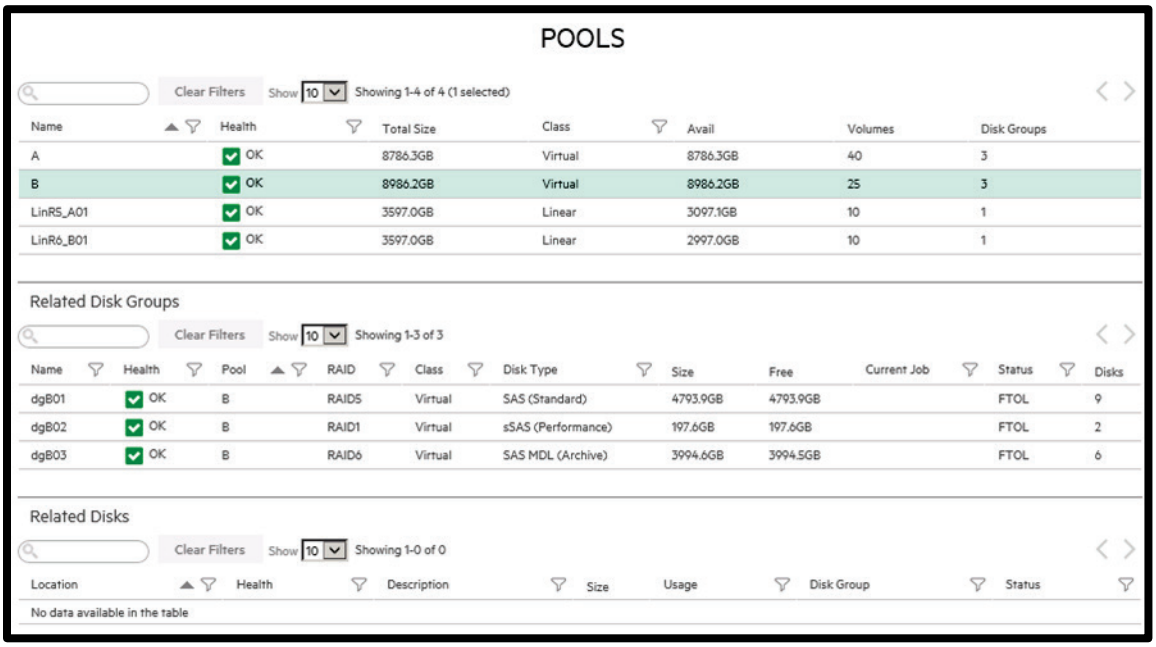

Figure 10. MSA Pool B screen

Viewing the performance of the pools or Virtual Disk Groups can also assist in determining where to place the Archive Tier space.

From V3 of the SMU, navigate to “Performance” then click “Virtual Pools” from the “Show:” drop-down box. Next, click the pool and for real time data, click “Show Data”. For Historical Data, click the “Historical Data” box and “Set time range”.

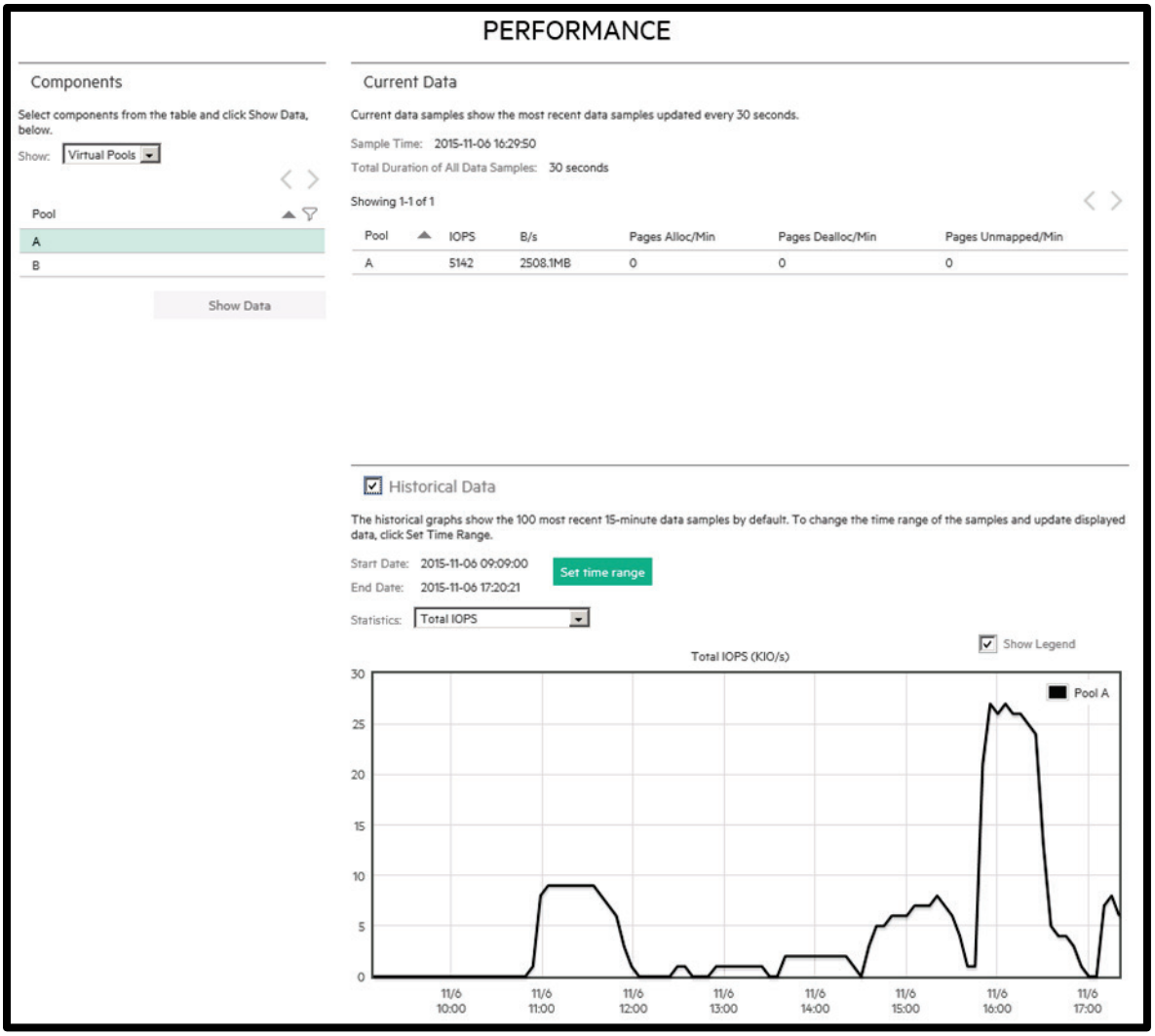

Figure 11. MSA Virtual Pools Performance Screen

Tiering

A Tier is defined by the disk type in the Virtual Disk Groups.

- Performance Tier contains SSDs
- Standard Tier contains 10k rpm/15k rpm Enterprise SAS drives
- Archive Tier contains MDL SAS 7.2k rpm drives

Disk Group Considerations

With the GL200 and greater firmware on the MSA, allocated pages are evenly distributed between disk groups in a tier; therefore, create all disk groups in a tier with the same RAID type and number of drives to ensure uniform performance in the tier.

Consider an example where the first Disk Group in the Standard Tier consists of five 15K Enterprise SAS drives in a RAID 5 configuration. To ensure consistent performance in the tier, any additional disk groups for the Standard Tier should also be a RAID 5 configuration. Adding a new disk group configured with four 10K Enterprise SAS drives in a RAID 6 configuration will produce inconsistent performance within the tier due to the different characteristics of the disk groups.

For optimal write performance, parity based disk groups (RAID 5 and RAID 6) should be created with "The Power of 2" method. This method means that the number of data (non-parity) drives contained in a disk group should be a power of 2. See table 1.

Table 1. Power of 2 Method

RAID TYPE	TOTAL DRIVES PER DISK GROUP	DATA DRIVES	PARITY DRIVES
RAID 5	3	2	1
RAID 5	5	4	1
RAID 5	9	8	1
RAID 6	4	2	2
RAID 6	6	4	2
RAID 6	10	8	2

Due to the limitation of Disk Groups in a pool, which is 16, RAID type should be considered when creating new Disk Groups. For example, instead of creating multiple RAID 1 Disk Groups, consider using a larger RAID 10 Disk Group.

Drive Type and Capacity Considerations when using Tiering

All hard disk drives in a tier should be the same type. For example, do not mix 10k rpm and 15k rpm drives in the same Standard Tier. If you have a Performance Tier, consider sizing the Performance Tier to be 5%–10% the capacity of the Standard Tier.

Disk Group RAID Type Considerations

RAID 6 is recommended when using large capacity Midline (MDL) SAS drives in the Archive Tier. The added redundancy of RAID 6 will protect against data loss in the event of a second disk failure with large MDL SAS drives.

RAID 5 is commonly used for the Standard Tier where the disks are smaller and faster resulting in shorter rebuild times. RAID 5 is used in workloads that typically are both random and sequential in nature.

See the Best practices for SSDs section for RAID types used in the Performance Tier and Read Cache.

Global Spares with Tiers

Using Global spares is recommended for all tiers based on spinning media. When using these global spares, make sure to use the same drive types as the Disk Group. The drive size must be equal or larger than the smallest drive in the tier.

Expanding Virtual Volumes

There might come a time when the Virtual Disk Group in a pool will start to fill up. To easily add more space, the MSA implements Wide Striping to increase the size of the virtual volumes. The recommended method to increase the volume size is to add a new Virtual Disk Group with the same amount of drives and RAID type as the existing Virtual Disk Group has.

For example, a Virtual Disk Group in pool A is filling up. This Disk Group is a five 300 GB drive, 15k rpm, RAID 5 Disk Group. The recommended procedure would be to create a new Virtual Disk Group on pool A that also has five, 300 GB 15K disk drives in a RAID 5 configuration.

Volume Tier Affinity

Volume tier affinity is a settable attribute that allows a storage administrator to define Quality of Service (QoS) preferences for virtual volumes in a tiered environment. There are three volume tier affinity options-Archive, Performance, and No Affinity. A setting of "Archive" will prefer the lowest tier of service, "Performance" will prefer the higher tiers of service, and "No Affinity" will use the standard tiering strategy. Tier Affinity is not the same as Tier Pinning and does not restrict data to a given tier and capacity. Data on a volume with "Archive" affinity can still to be promoted to a performance tier if that data becomes in demand to the host application.

Note

The "Performance" affinity does not require an SSD tier and will use the highest performance tier available.

Mechanics of Volume Tier Affinity

Volume tier affinity acts as a guide to the system on where to place data from a given volume in the available tiers.

The standard strategy is to prefer the highest spinning disk (non-SSD) tiers for new sequential writes and the highest tier available (including SSD) for new random writes. As data is later accessed by the host application, data will be moved to the most appropriate tier based on demand with "hot" data being promoted up towards the highest performance tier and "cold" data being demoted downwards to the lower spinning disk based tiers. This standard strategy will be followed for data on volumes set to "No Affinity".

For data on volumes set to the "Performance" affinity, the standard strategy will be followed for all new writes; however, subsequent access to that data will have a lower threshold for promotion upwards making it more likely for that data to be available on the higher performance tiers. Preferential treatment will be provided to "hot" data that has "Performance" affinity at the SSD tier making it more likely for "Archive" or "No Affinity" data to be demoted out of the SSD tier to make room. This is useful for volumes where you know the data will be in demand and want to ensure that it has priority treatment for promotion to and retention in the highest performance tier.

For volumes that are set to the "Archive" affinity, all new writes will be initially placed in the archive tier as long as space is available. If no space is available, new writes will be placed on the next higher tier available. Subsequent access to that data will allow for its promotion to the performance tiers as it becomes "hot"; however, the data will have a lower threshold for demotion and will be moved out of the highest performance SSD tier if there is a need to promote "hot" data up from a lower tier.

Volume Tier Affinity Impact in Existing Environments

After upgrading to GL220 and greater firmware, all existing virtual volumes will have a default setting of "No Affinity" and will continue to use the standard tiering strategy.

If the affinity of an existing volume is changed to the "Performance" affinity, there is no immediate change made. The current data on the volume will be promoted with the "Performance" strategy outlined in the previous section based on host application needs.

If the affinity of an existing volume is changed to "Archive", a background operation will be performed to begin moving the data for the affected volume down in to the Archive tier. This process is done at a low priority to have as minimal effect on host I/O performance as possible. As data in a volume with an "Archive" affinity becomes "hot", it will be promoted up to the higher tiers automatically. New data written to the volume will be targeted to the Archive tiers based on the strategy outlined in the previous section.

Configuring Volume Tier Affinity

For new virtual volumes, at volume creation time set the affinity from the "Preference" drop down. Note that the default is "No Affinity".

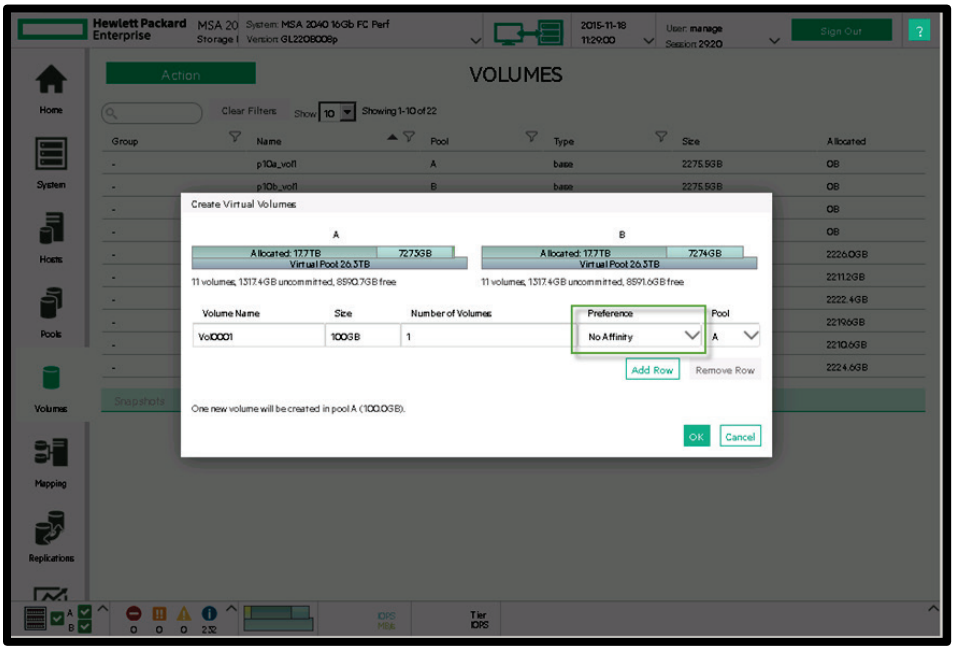

Figure 12. Setting Virtual Volume Tier Affinity

For existing virtual volumes, from the SMU select the "**Volumes**" tab, select the volume you want to set an affinity on and from the "**Action**" menu select "**Modify Volume**" and set the affinity from the "**Preference**" drop down.

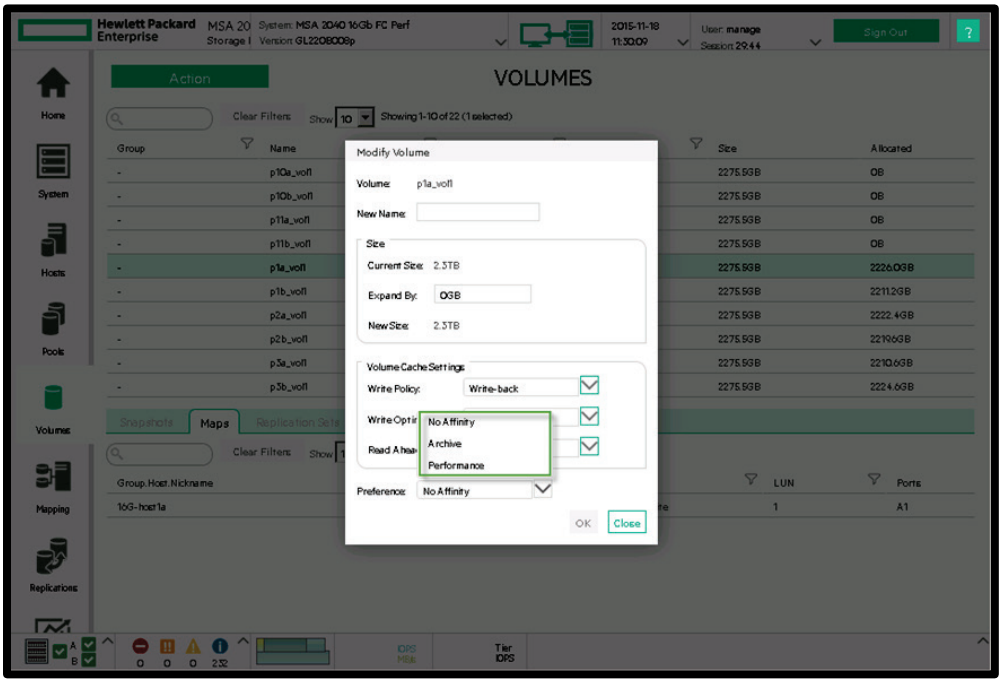

Figure 13. Modifying Virtual Volume Tier Affinity

Hewlett Packard Enterprise recommends the default "No Affinity" option for most configurations. This setting attempts to balance the frequency of data access, disk cost, and disk availability by moving this volume's data to the appropriate tier.

If the virtual volume uses mostly random or bursty low latency workloads such as Online Transaction Processing (OLTP), Virtual Desktop Infrastructure (VDI), or Virtualization environments, Hewlett Packard Enterprise recommends setting the preference to "Performance". This setting keeps as much of this volume's data in the Performance tier for as long a period as possible.

If the virtual volume contains infrequently accessed workloads such as backup data or email archiving, Hewlett Packard Enterprise recommends setting the preference to "Archive". This option will keep as much of this volume's data in the Archive tier for as long a period as possible.

Best practices when choosing drives for HPE MSA 1040/2040 storage

The characteristics of applications and workloads are important when selecting drive types for the HPE MSA 1040/2040 array.

Drive types

The HPE MSA 1040 array supports SSDs, SAS Enterprise drives, and SAS Midline (MDL) drives. The HPE MSA 2040 array supports SSDs, SAS Enterprise drives, SAS Midline (MDL) drives, and Self-Encrypting Drives (SED). See the Full Disk Encryption section for more information on SED drives. The HPE MSA 1040/2040 array does not support Serial ATA (SATA) drives. Choosing the correct drive type is important; drive types should be selected based on the workload and performance requirements of the volumes that will be serviced by the storage system. For sequential workloads, SAS Enterprise drives or SAS MDL drives provide a good price-for-performance tradeoff over SSDs. If more capacity is needed in your sequential environment, SAS MDL drives are recommended. SAS Enterprise drives offer higher performance than SAS MDL and should also be considered for random workloads when performance is a premium. For high performance random workloads, SSDs would be appropriate.

SAS MDL drives are not recommended for constant high workload applications. SAS MDL drives are intended for archival purposes.

Best practices to improve availability

There are many methods to improve availability when using the HPE MSA 1040/2040 array. High availability is always advisable to protect your assets in the event of a device failure. Outlined below are some options that will help you in the event of a failure.

Volume mapping

Using volume mapping correctly can provide high availability from the hosts to the array. For high availability during a controller failover, a volume must be mapped to at least one port accessible by the host on both controllers. Mapping a volume to ports on both controllers ensures that at least one of the paths is available in the event of a controller failover, thus providing a preferred/optimal path to the volume.

In the event of a controller failover, the surviving controller will report that it is now the preferred path for all Disk Groups. When the failed controller is back online, the Disk Groups and preferred paths switch back to the original owning controller.

Best practice is to map volumes to two ports on each controller to take advantage of load balancing and redundancy to each controller.

Mapping a port will make a mapping to each controller; thus, mapping port 1 will map host ports A1 and B1. Mapping to port 2 will map host ports A2 and B2.

With this in mind, make sure that physical connections are set up correctly on the MSA, so that a server has a connection to both controllers on the same port number. For example, on a direct attach MSA 2040 SAS with multiple servers, make sure that ports A1 and B1 are connected to server A, ports A2 and B2 are connected to server B, and so on.

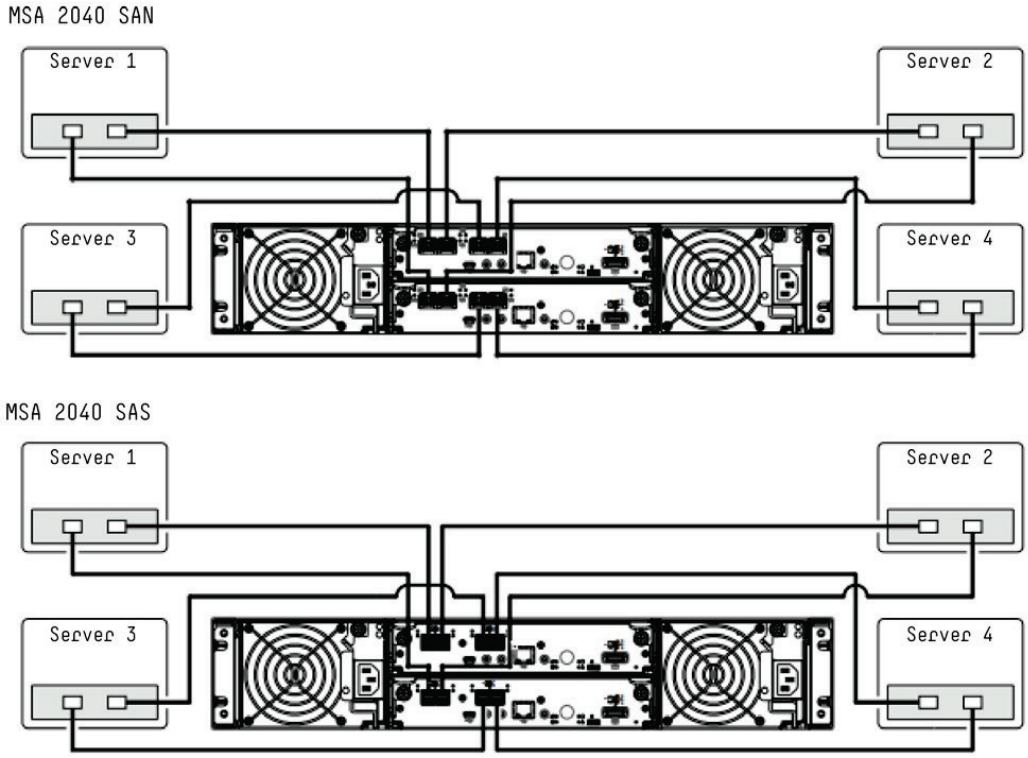

Figure 14. Direct Attach Cabling

It is not recommended to enable more than 8 paths to a single host, i.e., 2 HBA ports on a physical server connected to 2 ports on the A controller and 2 ports on the B controller. Enabling more paths from a host to a volume puts additional stress on the operating system's multipath software which can lead to delayed path recovery in very large configurations.

Note
Volumes should not be mapped to multiple servers at the same time unless the operating systems on the servers are cluster aware. However, since a server may contain multiple unique initiators, mapping a volume to multiple unique initiators (that are contained in the same server) is supported and recommended. Recommended practice is to put multiple initiators for the same host into a host and map the host to the LUNs, rather than individual maps to initiators.

Redundant paths

To increase the availability of the array to the hosts, multiple, redundant paths should be used along with multipath software. Redundant paths can also help in increasing performance from the array to the hosts (discussed later in this paper). Redundant paths can be accomplished in multiple ways. In the case of a SAN attach configuration, best practice would be to have multiple, redundant switches (SANs) with the hosts having at least one connection into each switch (SAN), and the array having one or more connections from each controller into each switch. In the case of a direct attach configuration, best practice is to have at least two connections to the array for each server. In the case of a direct attach configuration with dual controllers, best practice would be to have at least one connection to each controller.

Snapshots

HPE MSA snapshot services enable increased data protection by creating recovery points for your data by taking a "picture" of your data at a specific point in time. Snapshots are then maintained even as data continues to change. In the event of a failure, you can recover to any previous snapshot. Snapshots are a great complement to tape or disk backup strategy.

MSA Snapshot functionality is controller-based, so host resources are not used. The MSA 1040/2040 snapshot services utilize copy-on-write capabilities when operating on linear volumes and redirect-on-write capabilities when operating on virtualized volumes.

Hewlett Packard Enterprise recommends utilizing the snapshot functionality for data protection.

Review the snapshot space management guidelines below when using MSA snapshots.

Snapshots on Linear Volumes

How do you use snapshots on linear volumes?

Occasionally—If you already maintain your snap pool space, then you probably do not need to change anything. The system automatically sets the snap pool limit to either 20% of the volume size or the minimum snap-pool size (5GiB) and only notifies you if a threshold is crossed.

Regularly—If you want the system to remove old snapshots, you might want to consider changing the limit policy to delete. The oldest and lowest priority snapshots are deleted first. See the help for set priorities for default priorities for snapshot retention.

What is the rate of change of your data?

Based on the rate of change in your data and the desired snapshot retention, adjust the snap pool space limit and thresholds accordingly.

Snapshots on Virtual Volumes

How do you use snapshots on virtual volumes?

Occasionally—If you already maintain your snapshot space, then you probably do not need to change anything. The system automatically sets the limit at 10% of the pool and only notifies you if a threshold is crossed.

Regularly—If you want the system to remove old snapshots, you might want to consider changing the limit policy to delete. Only unmapped snapshots that are leaves of a snapshot tree are considered for deletion. The oldest and lowest priority snapshots are deleted first.

What is the rate of change of your data?

Based on the rate of change in your data and the desired snapshot retention, adjust the snapshot space limit and thresholds accordingly.

Note
Linear retention policies are based on snapshot type and apply to a snap pool. Virtual retention policies apply to individual volumes and are inherited. If you set the retention level of a base volume, this will not affect existing snapshots of that volume, only the snapshots created after setting the retention level.

Multipath software

To fully utilize redundant paths, multipath software should be installed on the hosts. Multipath software allows the host operating system to use all available paths to volumes presented to the host; redundant paths allow hosts to survive SAN component failures. Multipath software can increase performance from the hosts to the array. Table 2 lists supported multipath software by operating systems.

Note
More paths are not always better. Enabling more than 8 paths to a single volume is not recommended.

Table 2. Multipath and operating systems

OPERATING SYSTEM	MULTIPATH NAME	VENDOR ID	PRODUCT ID
Windows® 2008/2012	Microsoft® multipath I/O (MPIO)	HPE	MSA 2040 SAN MSA 2040 SAS MSA 1040 SAN MSA 1040 SAS
Linux®	Device mapper/multipath	HPE	MSA 2040 SAN MSA 2040 SAS MSA 1040 SAN MSA 1040 SAS
VMware	Native multipath (NMP)	HPE	MSA 2040 SAN MSA 2040 SAS MSA 1040 SAN MSA 1040 SAS

Installing MPIO on Windows Server® 2008 R2/2012

Microsoft has deprecated server manager cmd for Windows Server 2008 R2 so you will use the ocsetup command instead.

1. Open a command prompt window and run the following command:

```
C:\>ocsetup MultiPathIO /norestart
C:\>mpclaim -n -i -d "HP      MSA 1040 SAN"
```

Note
There are 6 spaces between HPE and MSA in the mpclaim command.

The mpclaim –n option avoids rebooting. Reboot is required before MPIO is operational.

The MPIO software is installed. When running the mpclaim command, type in the correct product ID for your MSA product. See table 2 above.

2. If you plan on using MPIO with a large number of LUNs, configure your Windows Server Registry to use a larger PDORemovePeriod setting.
   a. If you are using a Fibre Channel connection to a Windows server running MPIO, use a value of 90 seconds.
   b. If you are using an iSCSI connection to a Windows server running MPIO, use a value of 300 seconds. See "Long Failover Times When Using MPIO with Large Numbers of LUNs" below for details.

Once the MPIO DSM is installed, no further configuration is required; however, after initial installation, you should use Windows Server Device Manager to ensure that the MPIO DSM has installed correctly as described in "Managing MPIO LUNs."

Long Failover Times When Using MPIO with Large Numbers of LUNs

Microsoft Windows servers running MPIO use a default Windows Registry **PDORemovePeriod** setting of 20 seconds. When MPIO is used with a large number of LUNs, this setting can be too brief, causing long failover times that can adversely affect applications.

The Microsoft Technical Bulletin **Configuring MPIO Timers**, describes the **PDORemovePeriod** setting:

"This setting controls the amount of time (in seconds) that the multipath pseudo-LUN will continue to remain in system memory, even after losing all paths to the device. When this timer value is exceeded, pending I/O operations will be failed, and the failure is exposed to the application rather than attempting to continue to recover active paths. This timer is specified in seconds. The default is 20 seconds. The max allowed is MAXULONG."

Workaround: If you are using MPIO with a large number of LUNs, edit your registry settings so that **HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\mpio\Parameters\PDORemovePeriod** is set to a higher value.

- If you are using a Fibre Channel connection to a Windows server running MPIO, use a value of 90 seconds.
- If you are using an iSCSI connection to a Windows server running MPIO, use a value of 300 seconds.

For more information, refer to Configuring MPIO Timers at: technet.microsoft.com/en-us/library/ee619749%28WS.10%29.aspx

Managing MPIO LUNs

The Windows Server Device Manager enables you to display or change devices, paths, and load balance policies, and enables you to diagnose and troubleshoot the DSM. After initial installation of the MPIO DSM, use Device Manager to verify that it has installed correctly.

If the MPIO DSM was installed correctly, each MSA 1040/2040 storage volume visible to the host will be listed as a multi-path disk drive as shown in the following example.

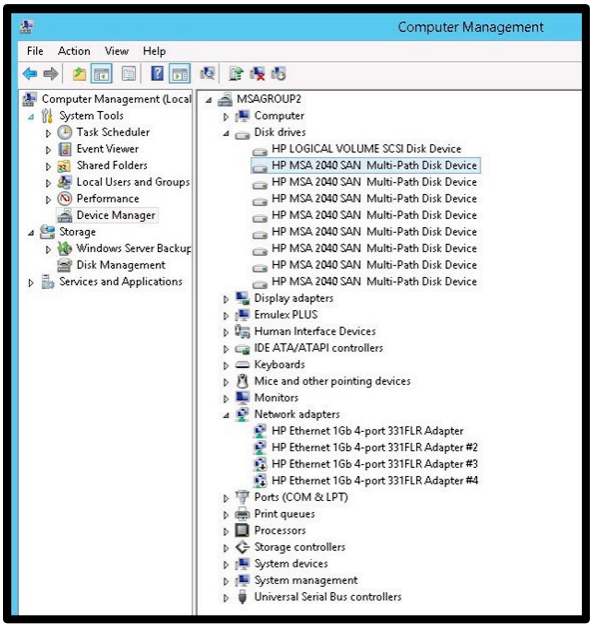

Figure 15. MSA Multi-Path Disk Device

To verify that there are multiple, redundant paths to a volume, right-click the Multi-Path Disk Device and select Properties.

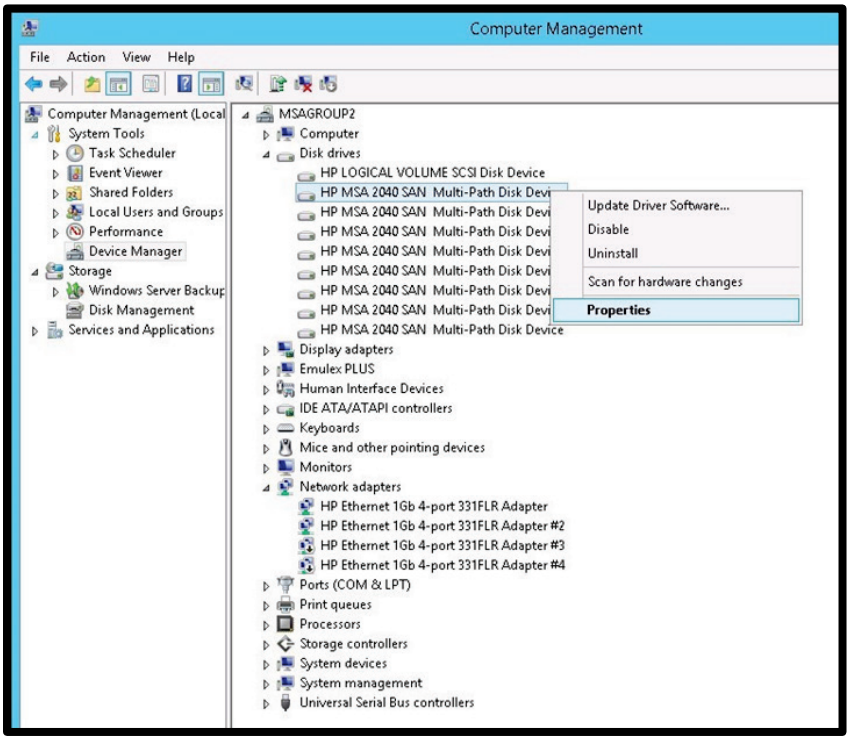

Figure 16. Selecting Properties of MSA Multi-Path Disk Device

Click the MPIO tab to view the MPIO property sheet, which enables you to view or change the load balance policy and view the number of paths and their status.

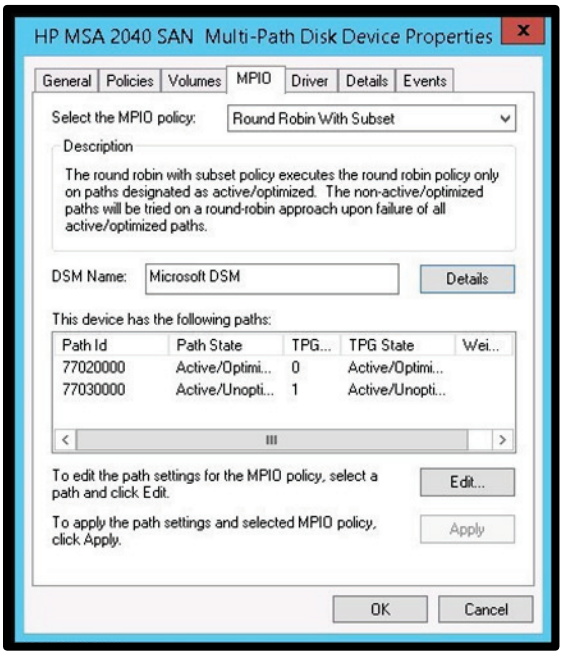

Figure 17. MSA Multi-Path Disk Device Properties MPIO Tab

The Details tab shows additional parameters.

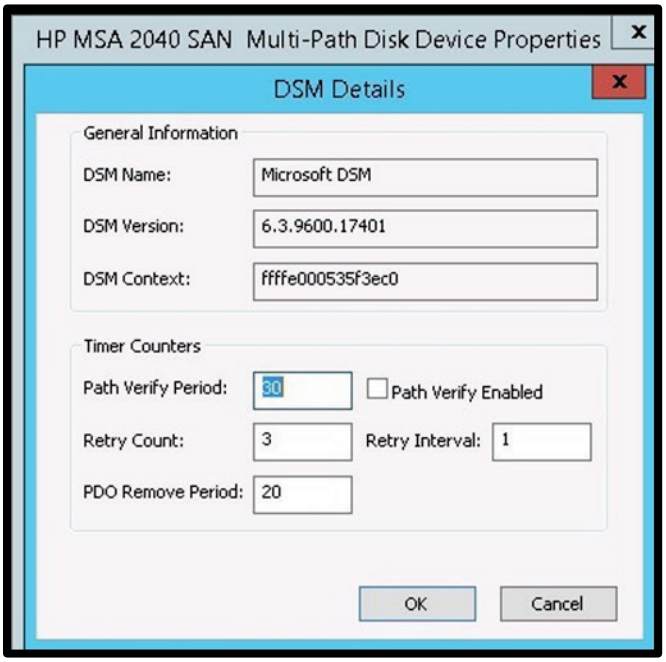

Figure 18. MSA Multi-Path Disk Device Properties DSM Details

Dual power supplies

The HPE MSA 1040/2040 chassis and supported expansion enclosures ship with dual power supplies. At a minimum, connect both power supplies in all enclosures. For the highest level of availability, connect the power supplies to separate power sources.

Dual controllers

The HPE MSA 2040 can be purchased as a single or dual controller system; the HPE MSA 1040 is sold only as a dual controller system. Utilizing a dual controller system is best practice for increased reliability for two reasons. First, dual controller systems will allow hosts to access volumes during a controller failure or during firmware upgrades (given correct volume mapping discussed above). Second, if the expansion enclosures are cabled correctly, a dual controller system can withstand an expansion I/O Module (IOM) failure, and in certain situations a total expansion enclosure failure.

Reverse cabling of expansion enclosures

The HPE MSA 1040/2040 firmware supports both fault tolerant (reverse cabling) and straight-through SAS cabling of expansion enclosures. Fault tolerant cabling allows any expansion enclosure to fail or be removed without losing access to other expansion enclosures in the chain. For the highest level of fault tolerance, use fault tolerant (reverse) cabling when connecting expansion enclosures.

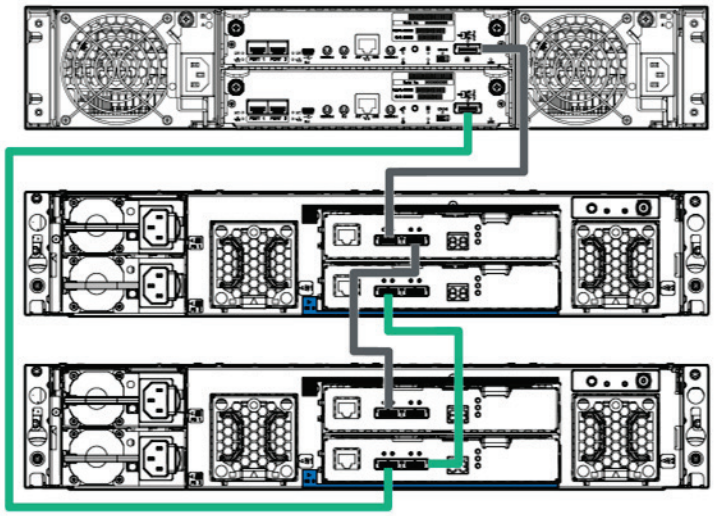

Figure 19. Reverse cabling example using the HPE MSA 1040 system

See the MSA Cable Configuration Guide for more details on cabling the HPE MSA 1040/2040.

The HPE MSA 1040/2040 Cable Configuration Guides can be found on the MSA support pages. For MSA 1040: hpe.com/support/msa1040 For MSA 2040: hpe.com/support/msa2040

Create Disk Groups across expansion enclosures

The HPE recommendation is to stripe Disk Groups across shelf enclosures to enable data integrity in the event of an enclosure failure. A Disk Group created with RAID 1, 10, 3, 5, 50, or 6 can sustain one or more expansion enclosure failures without loss of data depending on RAID type. Disk Group configuration should take into account MSA drive sparing methods such as dedicated, global, and dynamic sparing.

Drive sparing

Drive sparing, sometimes referred to as hot spares, is recommended to help protect data in the event of a disk failure in a fault tolerant Disk Group (RAID 1, 3, 5, 6, 10, or 50) configuration. In the event of a disk failure, the array automatically attempts to reconstruct the data from the failed drive to a compatible spare. A compatible spare is defined as a drive that has sufficient capacity to replace the failed disk and is the same media type (i.e., SAS SSD, Enterprise SAS, Midline SAS, or SED drives). The HPE MSA 1040/2040 supports dedicated, global, and dynamic sparing. The HPE MSA 1040/2040 will reconstruct a critical or degraded Disk Group.

Important
An offline or quarantined Disk Group is not protected by sparing.

Supported spare types:

- **Dedicated spare**—reserved for use by a specific Disk Group to replace a failed disk. This method is the most secure way to provide spares for Disk Groups. The array supports up to 4 dedicated spares per Disk Group. Dedicated spares are only applicable to Linear Storage.
- **Global spare**—reserved for use by any fault tolerant Disk Group to replace a failed disk. The array supports up to 16 global spares per system. At least one Disk Group must exist before you can add a global spare. Global Spares are applicable to both Virtual and Linear Storage.
- **Dynamic spare**—all available drives are available for sparing. If the MSA has available drives and a Disk Group becomes degraded any available drive can be used for Disk Group reconstruction. Dynamic spares are only applicable to Linear Storage.

Sparing process

When a disk fails in a redundant Disk Group, the system first looks for a dedicated spare for the Disk Group. If a dedicated spare is not available or the disk is incompatible, the system looks for any compatible global spare. If the system does not find a compatible global spare and the dynamic spares option is enabled, the system uses any available compatible disk for the spare. If no compatible disk is available, reconstruction cannot start.

During reconstruction of data, the effected Disk Group will be in either a degraded or critical status until the parity or mirror data is completely written to the spare, at which time the Disk Group returns to fault tolerant status. For RAID 50 Disk Groups, if more than one Sub-Disk Group becomes critical, reconstruction and use of spares occurs in the order Sub-Disk Groups are numbered. In the case of dedicated spares and global spares, after the failed drive is replaced, the replacement drive will need to added back as a dedicated or global spare.

Best practice for sparing is to configure at least one spare for every fault tolerant Disk Group in the system.

Drive replacement

In the event of a drive failure, replace the failed drive with a compatible drive as soon as possible. As noted above, if dedicated or global sparing is in use, mark the new drive as a spare (either dedicated or global), so it can be used in the future for any other drive failures.

Working with Failed Drives and Global Spares

When a failed drive rebuilds to a spare, the spare drive now becomes the new drive in the Disk Group. At this point, the original drive slot position that failed is no longer part of the Disk Group. The original drive should be replaced with a new drive.

In order to get the original drive slot position to become part of the Disk Group again, do the following:

1. Replace the failed drive with a new drive.
2. When the new drive is online and marked as “Available”, configure the drive as a global spare drive.
3. Fail the drive in the original global spare location by removing it from the enclosure. The RAID engine will rebuild to the new global spare which will then become an active drive in the RAID set again.
4. Replace the drive you manually removed from the enclosure.
5. If the drive is marked as “Leftover”, clear the disk metadata.
6. Re-configure the drive as the new global spare.

Virtual Storage only uses Global sparing. Warnings alerts are sent out when the last Global spare is used in a system.

Implement Remote Snap replication

The HPE MSA 1040/2040 storage system Remote Snap feature is a form of asynchronous replication that replicates block-level data from a volume on a local system to a volume on the same system or on a second independent system. The second system may be at the same location as the first, or it may be located at a remote site. Replicating from a volume on a local system to a volume on the same system is only supported when using linear replication.

Linear replication or virtual replication can exist on an array, but not both at the same time. Also, there is a one-to-one system replication limit for virtual replication.

Best practice is to implement Remote Snap replication for disaster recovery.

Use the secured web access (HTTPS) when utilizing Remote Snap replication on the MSA.

Note

Remote Snap requires a purchasable license in order to implement.

To obtain a Remote Snap license, go to: myenterpriselicense.hpe.com

See the HPE MSA Remote Snap Technical white paper: h20195.www2.hpe.com/v2/GetPDF.aspx/4AA1-0977ENW.pdf

Use VMware Site Recovery Manager with Remote Snap replication

VMware vCenter Site Recovery Manager (SRM) is an extension to VMware vCenter that delivers business-continuity and disaster-recovery solution that helps you plan, test, and execute the recovery of vCenter virtual machines. SRM can discover and manage replicated datastores, and automate migration of inventory from one vCenter to another. Site Recovery Manager integrates with the underlying replication product through a storage replication adapter (SRA).

SRM is currently supported on the MSA 1040/2040 in linear mode only.

For best practices with SRM and MSA Remote Snap replication, see the "HPE MSA 2040 Storage Configuration and Best Practices for VMware vSphere" technical white paper: h20195.www2.hp.com/V2/GetPDF.aspx%2F4AA4-7060ENW.pdf

Best practices to enhance performance

This section outlines configuration options for enhancing performance for your array.

Cache settings

One method to tune the storage system is by choosing the correct cache settings for your volumes. Controller cache options can be set for individual volumes to improve a volume's I/O performance.

Caution

Only disable write-back caching if you fully understand how the host operating system, application, and adapter move data. If used incorrectly, you might hinder system performance.

Using write-back or write-through caching

By default, volume write-back cache is enabled. Because controller cache is backed by super-capacitor technology, if the system loses power, data is not lost. For most applications, write-back caching enabled is the best practice. With the transportable cache feature, write-back caching can be used in either a single or dual controller system. See the MSA 1040/2040 User Guide for more information on the transportable cache feature.

You can change a volume's write-back cache setting. Write-back is a cache-writing strategy in which the controller receives the data to be written to disks, stores it in the memory buffer, and immediately sends the host operating system a signal that the write operation is complete, without waiting until the data is actually written to the disk. Write-back cache mirrors all of the data from one controller module cache to the other unless cache optimization is set to no-mirror. Write-back cache improves the performance of write operations and the throughput of the controller. This is especially true in the case of random I/O, where write-back caching allows the array to coalesce the I/O to the Disk Groups.

When write-back cache is disabled, write-through becomes the cache-writing strategy. Using write-through cache, the controller writes the data to the disks before signaling the host operating system that the process is complete. Write-through cache has lower write operation and throughput performance than write-back, but all data is written to non-volatile storage before confirmation to the host. However, write-through cache does not mirror the write data to the other controller cache because the data is written to the disk before posting command completion and cache mirroring is not required. You can set conditions that cause the controller to change from write-back caching to write-through caching. Please refer to the HPE MSA 1040/2040 User Guide for ways to set the auto write through conditions correctly. In most situations, the default settings are acceptable.

In both caching strategies, active-active failover of the controllers is enabled.

Optimizing read-ahead caching

You can optimize a volume for sequential reads or streaming data by changing its read ahead, cache settings. Read ahead is triggered by sequential accesses to consecutive LBA ranges. Read ahead can be forward (that is, increasing LBAs) or reverse (that is, decreasing LBAs). Increasing the read-ahead cache size can greatly improve performance for multiple sequential read streams. However, increasing read-ahead size will likely decrease random read performance.

- Adaptive—this option works well for most applications: it enables adaptive read-ahead, which allows the controller to dynamically calculate the optimum read-ahead size for the current workload. This is the default.
- Stripe—this option sets the read-ahead size to one stripe. The controllers treat non-RAID and RAID 1 Disk Groups internally as if they have a stripe size of 512 KB, even though they are not striped.

- Specific size options—these options let you select an amount of data for all accesses.
- Disabled—this option turns off read-ahead cache. This is useful if the host is triggering read ahead for what are random accesses. This can happen if the host breaks up the random I/O into two smaller reads, triggering read ahead.

Caution
Only change read-ahead cache settings if you fully understand how the host operating system, application, and adapter move data so that you can adjust the settings accordingly.

Optimizing cache modes

You can also change the optimization mode for each volume.

- Standard—this mode works well for typical applications where accesses are a combination of sequential and random; this method is the default. For example, use this mode for transaction-based and database update applications that write small files in random order.
- No-mirror—in this mode each controller stops mirroring its cache metadata to the partner controller. This improves write I/O response time but at the risk of losing data during a failover. Unified LUN Presentation (ULP) behavior is not affected, with the exception that during failover any write data in cache will be lost. In most conditions no-mirror is not recommended, and should only be used after careful consideration.

Parameter settings for performance optimization

You can configure your storage system to optimize performance for your specific application by setting the parameters as shown in table 3. This section provides a basic starting point for fine-tuning your system, which should be done during performance baseline modeling.

Table 3. Optimizing performance for your application

APPLICATION	RAID LEVEL	READ-AHEAD CACHE SIZE	CACHE WRITE OPTIMIZATION
Default	5 or 6	Adaptive	Standard
High-Performance Computing (HPC)	5 or 6	Adaptive	Standard
Mail spooling	1	Adaptive	Standard
NFS_Mirror	1	Adaptive	Standard
Oracle_DSS	5 or 6	Adaptive	Standard
Oracle_OLTP	5 or 6	Adaptive	Standard
Oracle_OLTP_HA	10	Adaptive	Standard
Random 1	1	Stripe	Standard
Random 5	5 or 6	Stripe	Standard
Sequential	5 or 6	Adaptive	Standard
Sybase_DSS	5 or 6	Adaptive	Standard
Sybase_OLTP	5 or 6	Adaptive	Standard
Sybase_OLTP_HA	10	Adaptive	Standard
Video streaming	1 or 5 or 6	Adaptive	Standard
Exchange database	5 for data; 10 for logs	Adaptive	Standard
SAP®	10	Adaptive	Standard
SQL	5 for data; 10 for logs	Adaptive	Standard

Other methods to enhance array performance

There are other methods to enhance performance of the HPE MSA 1040/2040. In addition to the cache settings, the performance of the HPE MSA 1040/2040 array can be maximized by using the following techniques.

Place higher performance SSD and SAS drives in the array enclosure

The HPE MSA 1040/2040 controller is designed to have a single SAS link per drive in the array enclosure and only four SAS links to expansion enclosures. Placing higher performance drives (i.e., SSD and Enterprise SAS drives) in the storage enclosure allows the controller to utilize the performance of those drives more effectively than if they were placed in expansion enclosures. This process will help generate better overall performance.

Fastest throughput optimization

The following guidelines list the general best practices to follow when configuring your storage system for fastest throughput:

- Host ports should be configured to match the highest speed your infrastructure supports.
- Disk Groups should be balanced between the two controllers.
- Disk drives should be balanced between the two controllers.
- Cache settings should be set to match table 2 ("Optimizing performance for your application") for the application.
- In order to get the maximum sequential performance from a Disk Group, you should only create one volume per Disk Group. Otherwise you will introduce randomness into the workload when multiple volumes on the Disk Group are being exercised concurrently.
- Distribute the load across as many drives as possible.
- Distribute the load across multiple array controller host ports.

Creating Disk Groups

When creating Disk Groups, best practice is to add them evenly across both controllers when using linear storage or across both pools when using virtual storage. With at least one Disk Group assigned to each controller, both controllers are active. This active-active controller configuration allows maximum use of a dual-controller configuration's resources.

Choosing the appropriate RAID levels

Choosing the correct RAID level when creating Disk Groups can be important for performance. However, there are some trade-offs with cost when using the higher fault tolerant RAID levels.

See table 4 below for the strengths and weaknesses of the supported HPE MSA 1040/2040 RAID types.

Table 4. HPE MSA 1040/2040 RAID levels

RAID LEVEL	MINIMUM DISKS	ALLOWABLE DISKS	DESCRIPTION	STRENGTHS	WEAKNESSES
NRAID	1	1	Non-RAID, non-striped mapping to a single disk	Ability to use a single disk to store additional data	Not protected, lower performance (not striped)
0	2	16	Data striping without redundancy	Highest performance	No data protection: if one disk fails all data is lost
1	2	2	Disk mirroring	Very high performance and data protection; minimal penalty on write performance; protects against single disk failure	High redundancy cost overhead: because all data is duplicated, twice the storage capacity is required
3	3	16	Block-level data striping with dedicated parity disk	Excellent performance for large, sequential data requests (fast read); protects against single disk failure	Not well-suited for transaction-oriented network applications; write performance is lower on short writes (less than 1 stripe)
5	3	16	Block-level data striping with distributed parity	Best cost/performance for transaction-oriented networks; very high performance and data protection; supports multiple simultaneous reads and writes; can also be optimized for large, sequential requests; protects against single	Write performance is slower than RAID 0 or RAID 1

6	4	16	Block-level data striping with double distributed parity	Best suited for large sequential workloads; non-sequential read and sequential read/write performance is comparable to RAID 5; protects against dual disk failure	Higher redundancy cost than RAID 5 because the parity overhead is twice that of RAID 5; not well-suited for transaction-oriented network applications; non-sequential write performance is slower than RAID 5
10 (1+0)	4	16	Stripes data across multiple RAID 1 Sub-Disk Groups	Highest performance and data protection (protects against multiple disk failures)	High redundancy cost overhead: because all data is duplicated, twice the storage capacity is required; requires minimum of four disks
50 (5+0)	6	32	Stripes data across multiple RAID 5 Sub-Disk Groups	Better random read and write performance and data protection than RAID 5; supports more disks than RAID 5; protects against multiple disk failures	Lower storage capacity than RAID 5

Note
RAID types NRAID, RAID 0, and RAID 3 can only be created using the Command Line Interface (CLI) and are not available in the SMU. When using Virtual Storage, only non-fault tolerant RAID types can be used in the Performance, Standard, and Archive and Tiers. NRAID and RAID 0 are used with Read Cache as the data in the Read Cache SSDs is duplicated on either the Standard or Archive Tier.

Volume mapping

For increased performance, access the volumes from the ports on the controller that owns the Disk Group, which would be the preferred path. Accessing the volume on the non-preferred path results in a slight performance degradation.

Optimum performance with MPIO can be achieved with volumes mapped to multiple paths on both controllers. When the appropriate MPIO drivers are installed on the host, only the preferred (optimized) paths will be used. The non-optimized paths will be reserved for failover.

Best practices for SSDs

SSDs are supported in both the MSA 1040 and MSA 2040 array systems. The performance capabilities of SSDs are a great alternative to traditional spinning hard disk drives (HDDs) in highly random workloads. SSDs cost more in terms of dollars per GB throughput than spinning hard drives; however, SSDs cost much less in terms of dollars per IOP. Keep this in mind when choosing the numbers of SSDs per MSA 1040/2040 array.

Use SSDs for randomly accessed data

The use of SSDs can greatly enhance the performance of the array. Since there are no moving parts in the drives, data that is random in nature can be accessed much faster.

Data such as database indexes and TempDB files would best be placed on a volume made from an SSD based Disk Group since this type of data is accessed randomly.

Another good example of a workload that would benefit from the use of SSDs is desktop virtualization, for example, Virtual Desktop Infrastructure (VDI) where boot storms require high performance with low latency.

SSD and performance

There are some performance characteristics which can be met with linear scaling of SSDs. There are also bandwidth limits in the MSA 1040/2040 controllers. There is a point where these two curves intersect. At the intersecting point, additional SSDs will not increase performance. See figure 20.

The MSA 1040/2040 reaches this bandwidth at a low number of SSDs. For the best performance using SSDs on the MSA 1040/2040, use a minimum of 4 SSDs with 1 mirrored pair of drives (RAID 1) per controller. RAID 5 and RAID 6 are also good choices for SSDs, but require more drives using the best practice of having one Disk Group owned by each controller. This would require 6 SSDs for RAID 5 and 8 SSDs for RAID 6. All SSD volumes should be contained in fault tolerant Disk Groups for data integrity.

Base the number of SSDs to use on the amount of space that is needed for your highly random, high performance data set. For example, if the amount of data that is needed to reside in the SSD volumes exceeds a RAID 1 configuration, use a RAID 5 configuration.

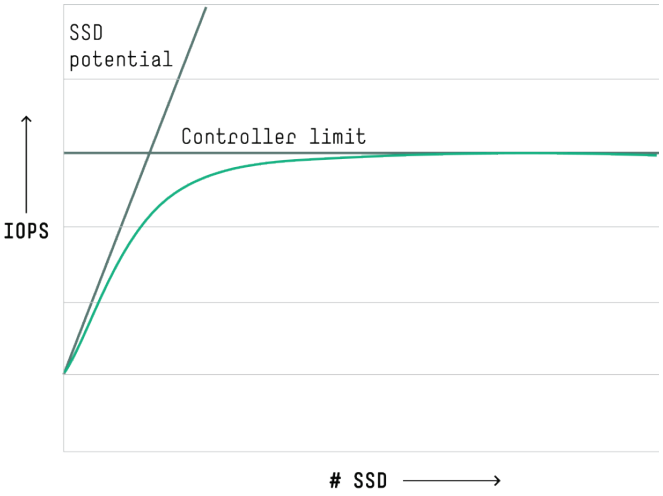

Figure 20. SSD performance potential vs. MSA 1040/2040 controller limit

Note
There is no limit to the number of SSDs that can be used in the MSA 1040/2040 array system.

SSD Read Cache

SSD Read Cache is a feature that extends the MSA 1040/2040 controller cache.

Read cache is most effective for workloads that are high in random reads. The user should size the read cache capacity based on the size of the hot data being randomly read. A maximum of 2 SSD drives per pool can be added for read cache.

HPE recommends beginning with 1 SSD assigned per storage pool for read cache. Monitor the performance of the read cache and add more SSDs as needed.

There is a 4 TB maximum limit per pool for read cache.

Note
You can have SSDs in a fault tolerant Disk Group as a Performance Tier or as a non-fault tolerant (up to 2 disks) Disk Group as Read Cache. But neither pool can have both a Performance Tier and a Read Cache. For example, pool A can have a Performance Tier and pool B can have a Read Cache.

SSD wear gauge

SSDs have a limited number of times they can be written and erased due to the memory cells on the drives. The SSDs in the HPE MSA 2040 come with a wear gauge as well as appropriate events that are generated to help detect the failure. Once the wear gauge reaches 0 percent, the integrity of the data is not guaranteed.

Best practice is to replace the SSD when the events and gauge indicate <5 percent life remaining to prevent data integrity issues.

Full Disk Encryption

Full Disk Encryption (FDE) is a data security feature used to protect data on disks that are removed from a storage array. The FDE feature uses special Self-Encrypting Drives (SEDs) to secure user data. FDE functionality is only available on the MSA 2040.

The SED is a drive with a circuit built into the drive's controller chipset which encrypts/decrypts all data to and from the media automatically. The encryption is part of a hash code which is stored internally on the drive's physical medium. In the event of a failure of the drive or the theft of a drive, a proper key sequence needs to be entered to gain access to the data stored within the drive.

Full Disk Encryption on the MSA 2040

The MSA 2040 storage system uses a passphrase to generate a lock key to enable securing the entire storage system. All drives in a Full Disk Encryption (FDE) secured system are required to be SED (FDE Capable). By default, a system and SED drive are not secured and all data on the disk may be read/written by any controller. The encryption on the SED drive conforms to FIPS 140-2.

To secure an MSA 2040, you must set a passphrase to generate a lock key and then FDE secure the system. Simply setting the passphrase does not secure the system. After an MSA 2040 system has been secured, all subsequently installed disks will automatically be secured using the system lock key. Non-FDE capable drives will be unusable in a secured MSA 2040 system.

Note
The system passphrase should be saved in a secure location. Loss of the passphrase could result in loss of all data on the MSA 2040 Storage System.

All MSA 2040 storage systems will generate the same lock key with the same passphrase. It is recommended that you use a different passphrase on each FDE secured system. If you are moving the entire storage system, it is recommended to clear the FDE keys prior to system shutdown. This will lock all data on the disks in case of loss during shipment. Only clear the keys after a backup is available and the passphrase is known. Once the system is in the new location, enter the passphrase and the SED drives will be unlocked with all data available.

SED drives which fail in an FDE secured system can be removed and replaced. Data on the drive is encrypted and cannot be read without the correct passphrase.

Best practices for Disk Group expansion

With the ever changing storage needs seen in the world today, there comes a time when storage space gets exhausted quickly. The HPE MSA 1040/2040 gives you the option to grow the size of a LUN to keep up with your dynamic storage needs.

A Disk Group expansion allows you to grow the size of a Disk Group in order to expand an existing volume or create volumes from the newly available space on the Disk Group. Depending on several factors, Disk Group expansion can take a significant amount of time to complete. For faster alternatives, see the "Disk Group expansion recommendations" section.

Note
Disk Group Expansion is not supported with Virtual Storage. If you have Virtual Storage and are running out of storage space, the procedure to get more storage space would be to add another Disk Group to a pool.

The factors that should be considered with respect to Disk Group expansion include but are not limited to:

- Physical disk size
- Number of disks to expand (1–4)
- I/O activity during Disk Group expansion

Note
Disk Group Expansion is only available when using Linear Storage.

During Disk Group expansion, other disk utilities are disabled. These utilities include Disk Group Scrub and Rebuild.

Disk Group expansion capability for supported RAID levels

The chart below gives information on the expansion capability for the HPE MSA 2040 supported RAID levels.

Table 5. Expansion capability for each RAID level

RAID LEVEL	EXPANSION CAPABILITY	MAXIMUM DISKS
NRAID	Cannot expand	1
0, 3, 5, 6	Can add 1–4 disks at a time	16
1	Cannot expand	2
10	Can add 2 or 4 disks at a time	16
50	Can expand the Disk Group one RAID 5 Sub-Disk Group at a time. The added RAID 5 Sub-Disk Group must contain the same number of disks as each original Sub-Disk Group	32

Important
If during the process of a Disk Group expansion one of the disk members of the Disk Group fails, the reconstruction of the Disk Group will not commence until the expansion is complete. During this time, data is at risk with the Disk Group in a DEGRADED or CRITICAL state.

If an expanding Disk Group becomes DEGRADED (e.g., RAID 6 with a single drive failure) the storage administrator should determine the level of risk of continuing to allow the expansion to complete versus the time required to backup, re-create the Disk Group (see "Disk Group expansion recommendations") and restore the data to the volumes on the Disk Group.

If an expanding Disk Group becomes CRITICAL (e.g., RAID 5 with a single drive failure) the storage administrator should immediately employ a backup and recovery process. Continuing to allow the expansion places data at risk of another drive failure and total loss of all data on the Disk Group.

Disk Group expansion can be very time consuming. There is no way to reliably determine when the expansion will be complete and when other disk utilities will be available.

Follow the procedure below.

1. Backup the current data from the existing Disk Group.
2. Using the SMU or CLI, start the Disk Group expansion.
3. Monitor the Disk Group expansion percentage complete.

Note
Once a Disk Group expansion initiates it will continue until completion or until the Disk Group is deleted.

Disk Group expansion recommendations

Before expanding a Disk Group, review the information below to understand the best alternative method for allocating additional storage to hosts.

Allocate "quiet" period(s) to help optimize Disk Group expansion

Disk Group expansion can take a few hours with no data access for smaller capacity hard drives and may take several days to complete with larger capacity hard drives. Priority is given to host I/O or data access over the expansion process during normal array operation. While the

system is responding to host I/O or data access requests, it may seem as if the expansion process has stopped. When expanding during "quiet" periods, expansion time is minimized and will allow quicker restoration of other disk utilities.

This method of expansion utilizes the expand capability of the system and requires manual intervention from the administrator. The procedure below outlines the steps to expand a Disk Group during a "quiet" period.

In this context, a "quiet" period indicates a length of time when there is no host I/O or data access to the system. Before starting the Disk Group expansion:

1. Stop I/O to existing volumes on the Disk Group that will be expanded.
2. Backup the current data from the existing volumes on the Disk Group.
3. Shutdown all hosts connected to the HPE MSA 1040/2040 system.
4. Label and disconnect host side cables from the HPE MSA 1040/2040 system.

Start and monitor Disk Group expansion:

1. Using the SMU or CLI, start the Disk Group expansion.
2. Monitor the Disk Group expansion percentage complete.

When expansion is complete or data access needs to be restored:

1. Re-connect host side cables to the HPE MSA 1040/2040 system.
2. Restart hosts connected to the HPE MSA 1040/2040 system.

If additional "quiet" periods are required to complete the Disk Group expansion:

1. Shutdown all hosts connected to the HPE MSA 1040/2040 system.
2. Label and disconnect host side cables from the HPE MSA 1040/2040 system.
3. Monitor the Disk Group expansion percentage complete.

Re-create the Disk Group with additional capacity and restore data

This method is the easiest and fastest method for adding additional capacity to a Disk Group. The online Disk Group initialization allows a user to access the Disk Group almost immediately and will complete quicker than the expansion process on a Disk Group that is also servicing data requests. The procedure below outlines the steps for recreating a Disk Group with additional capacity and restoring data to that Disk Group.

Procedure:

1. Stop I/O to existing volumes on the Disk Group that will be expanded.
2. Backup the current data from the existing volumes on the Disk Group.
3. Delete the current Disk Group.
4. Using the SMU or CLI, create a new Disk Group with the available hard drives using online initialization.
5. Create new larger volumes as required.
6. Restore data to the new volumes.

Best practices for firmware updates

The sections below detail common firmware update best practices for the MSA 1040/2040.

General MSA 1040/2040 device firmware update best practices

- As with any other firmware upgrade, it is a recommended best practice to ensure that you have a full backup prior to the upgrade.
- Before upgrading the firmware, make sure that the storage system configuration is stable and is not being reconfigured or changed in any way. If any configurations changes are in progress, monitor them using the SMU or CLI and wait until they are completed before proceeding with the upgrade.
- Do not power cycle or restart devices during a firmware update. If the update is interrupted or there is a power failure, the module could become inoperative. Should this happen, contact HPE customer support.
- After the device firmware update process is completed, confirm the new firmware version is displayed correctly via one of the MSA management interfaces—e.g., SMU or CLI.

MSA 1040/2040 array controller or I/O module firmware update best practices

- The array controller (or I/O module) firmware can be updated in an online mode only in redundant controller systems.
- When planning for a firmware upgrade, schedule an appropriate time to perform an online upgrade.
  - For single controller systems, I/O must be halted.
  - For dual controller systems, because the online firmware upgrade is performed while host I/Os are being processed, I/O load can impact the upgrade process. Select a period of low I/O activity to ensure the upgrade completes as quickly as possible and avoid disruptions to hosts and applications due to timeouts.
- When planning for a firmware upgrade, allow sufficient time for the update.
  - In single-controller systems, it takes approximately 10 minutes for the firmware to load and for the automatic controller restart to complete.
  - In dual-controller systems, the second controller usually takes an additional 20 minutes, but may take as long as one hour.
  - When reverting to a previous version of the firmware, ensure that the management controller (MC) Ethernet connection of each storage controller is available and accessible before starting the downgrade.
  - When using a Smart Component firmware package, the Smart Component process will automatically first disable partner firmware update (PFU) and then perform downgrade on each of the controllers separately (one after the other) through the Ethernet ports.
  - When using a binary firmware package, first disable the PFU option and then downgrade the firmware on each of the controller separately (one after the other).

MSA 1040/2040 disk drive firmware update best practices

- Disk drive upgrades on the HPE MSA 1040/2040 storage systems is an offline process. All host and array I/O must be stopped prior to the upgrade.
- If the drive is in a Disk Group, verify that it is not being initialized, expanded, reconstructed, verified, or scrubbed. If any of these tasks is in progress, before performing the update wait for the task to complete or terminate it. Also verify that background scrub is disabled so that it doesn't start. You can determine this using SMU or CLI interfaces. If using a firmware smart component, it would fail and report if any of the above pre-requisites are not being met.
- Disk drives of the same model in the storage system must have the same firmware revision. If using a firmware smart component, the installer would ensure all the drives are updated.

Miscellaneous best practices

Using Linear and Virtual Disk Groups

There are some considerations to be aware of when using Linear and Virtual Disk Groups together on the same MSA 1040/2040 storage array. See the User Guide for more information on Linear and Virtual Disk Groups.

Scenario 1:
You have Linear Disk Groups and want to keep using only Linear Disk Groups.

You can continue to use the MSA array and create the Linear Disk Groups in the same manner used before upgrading. You can use either Version 2 or Version 3 of the SMU.

Scenario 2:
You have Linear Disk Groups and want to start using Virtual Disk Groups and keep the remaining Linear Disk Groups.

You will be able to have both Linear and Virtual Disk Groups on the same MSA array. In order to utilize the Virtual Disk Groups, there must be available drives to construct the Disk Groups. This process might require purchasing new drives if no unused drives exist.

Use Version 3 of the SMU to create the new Virtual Disk Groups. Version 2 of the SMU does not have the capability to create Virtual Disk Groups.

Scenario 3:
You have Linear Disk Groups and want to start using Virtual Disk Groups and want to migrate the existing Linear Disk Groups to use the MSA virtualization features.

Host-based migration is necessary to move the linear volumes to virtual volumes.

To migrate the data, create new Virtual Disk Groups and then copy the data from the Linear Disk Groups onto the newly created Virtual Disk Group using a file copy.

Again, this process might require purchasing new drives for use with the Virtual Disk Groups if no unused drives exist.

Use Version 3 of the SMU to create the new Virtual Disk Groups. Version 2 of the SMU does not have the capability to create Virtual Disk Groups.

Boot from storage considerations

When booting from SAN, the best option is to create a linear Disk Group and allocate the entire Disk Group as a single LUN for the host boot device. This can improve performance for the boot device and avoid I/O latency in a highly loaded array. Booting from LUNs provisioned from pools where the volumes share all the same physical disks as the data volumes is also supported, but is not the best practice.

8 Gb/16 Gb switches and small form-factor pluggable transceivers

The HPE MSA 2040 storage system uses specific small form-factor pluggable (SFP) transceivers that will not operate in the HPE 8 Gb and 16 Gb switches. Likewise, the HPE Fibre Channel switches use SFPs which will not operate in the HPE MSA 2040.

The HPE MSA 2040 controllers do not include SFPs. Qualified SFPs for the HPE MSA 2040 are available for separate purchase in 4 packs. Both 8G and 16G SFPs are available to meet the customer need and budget constraints. All SFPs in an HPE MSA 2040 should conform to the installation guidelines given in the product QuickSpecs. SFP speeds and protocols can be mixed, but only in the specified configurations.

In the unlikely event of an HPE MSA 2040 controller or SFP failure, a field replacement unit (FRU) is available. SFPs will need to be moved from the failed controller to the replacement controller.

Please see the HPE Transceiver Replacement Instructions document for details found at hpe.com/support/msa2040/manuals.

The MSA 1040 8 Gb Dual Controller FC arrays include 8 Gb FC SFPs in all ports. These are the same 8 Gb FC SFPs available for the MSA 2040 and will only function in MSA arrays.

In the unlikely event of an HPE MSA 1040 controller or SFP failure, a field replacement unit (FRU) is available. SFPs will need to be moved from the failed controller to the replacement controller.

MSA 1040/2040 iSCSI considerations

When using the MSA 2040 SAN controller in an iSCSI configuration or using the MSA 1040 1GbE or 10GbE iSCSI storage systems, it is a best practice to use at least three network ports per server, two for the storage (private) LAN and one or more for the public LAN(s).This will ensure that the storage network is isolated from the other networks.

The Private LAN is the network that goes from the server to the MSA 1040 iSCSI or MSA 2040 SAN controller. This Private LAN is the storage network and the Public LAN is used for management of the MSA 1040/2040. The storage network should be isolated from the Public LAN to improve performance.

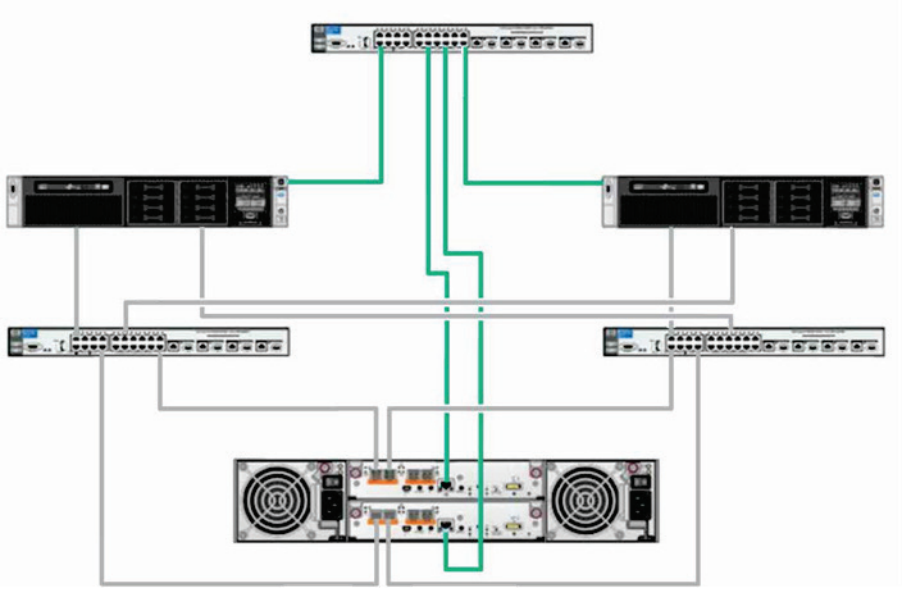

Figure 21. MSA 2040 SAN iSCSI Network

IP address scheme for the controller pair

The MSA 2040 SAN controller in iSCSI configurations or the MSA 1040 iSCSI should have ports on each controller in the same subnets to enable preferred path failover. The suggested means of doing this is to vertically combine ports into subnets. See examples below.

Example with a netmask of 255.255.255.0:

MSA 2040 SAN:

Controller A port 1: 10.10.10.100

Controller A port 2: 10.11.10.110

Controller A port 3: 10.10.10.120

Controller A port 4: 10.11.10.130

Controller B port 1: 10.10.10.140

Controller B port 2: 10.11.10.150

Controller B port 3: 10.10.10.160

Controller B port 4: 10.11.10.170

MSA 1040 iSCSI:

Controller A port 1: 10.10.10.100

Controller A port 2: 10.11.10.110

Controller B port 1: 10.10.10.120

Controller B port 2: 10.11.10.130

Jumbo frames

A normal Ethernet frame can contain 1500 bytes whereas a jumbo frame can contain a maximum of 9000 bytes for larger data transfers. The MSA reserves some of this frame size; the current maximum frame size is 1400 for a normal frame and 8900 for a jumbo frame. This frame maximum can change without notification. If you are using jumbo frames, make sure to enable jumbo frames on all network components in the data path.

Summary

HPE MSA 1040/2040 administrators should determine the appropriate levels of fault tolerance and performance that best suits their needs. Understanding the workloads and environment for the MSA SAN is also important. Following the configuration options listed in this paper can help optimize the HPE MSA 1040/2040 array accordingly.

Learn more at

hp.com/go/MSA

4AA4-6892ENW, February 2016, Rev. 6

Hewlett Packard Enterprise