

HP StorageWorks MSA2000 G1 or G2 and P2000 G3 FC MSA

Best practices

Technical white paper

Table of contents

About this document	3
Intended audience	3
The MSA2000 G1	3
Topics covered	3
Hardware overview	4
iSCSI, Fibre Channel, or SAS	5
Unified LUN Presentation (ULP)	5
Fault tolerance versus performance on the MSA2000fc G1	7
Choosing single or dual controllers	8
Choosing DAS or SAN attach	9
Dealing with controller failovers	10
Virtual disks	17
RAID levels	19
World Wide Name (WWN) naming conventions	21
Cache configuration	22
Fastest throughput optimization	26
Highest fault tolerance optimization	26
The MSA2000 G2	27
Topics covered	27
What's New in the MSA2000 G2	27
Hardware overview	28
Unified LUN Presentation (ULP)	29
Choosing single or dual controllers	30
Choosing DAS or SAN attach	31
Dealing with controller failovers	32
Virtual disks	35
Volume mapping	38
Configuring background scrub	38
RAID levels	39
Cache configuration	40
Fastest throughput optimization	42
Highest fault tolerance optimization	43
Boot from storage considerations	43
MSA70 considerations	43
Administering with HP SMU	44
MSA2000i G2 Considerations	44



The P2000 G3 MSA	45
Topics covered	45
What's New in the P2000 G3 MSA	45
Hardware overview	45
P2000 G3 MSA Settings	46
Disk Background Scrub, Drive Spin Down, and SMART	48
Cascading Array Enclosures	52
8 Gb Switches and SFP transceivers	54
Software	54
Versions	54
Description	54
HP StorageWorks P2000 Modular Smart Array Software Support/Documentation CD	55
Host Server Software	55
Best Practices for Firmware Updates	56
General P2000/MSA2000 Device Firmware Update Best Practices	56
P2000/MSA2000 Array Controller or I/O Module Firmware Update Best Practices	57
P2000/MSA2000 Disk Drive Firmware Update Best Practices	57
Summary	58
For more information	58

About this document

This white paper highlights the best practices for optimizing the HP StorageWorks MSA2000 G1, MSA2000 G2, and the P2000 G3 MSA, and should be used in conjunction with other HP StorageWorks Modular Smart Array manuals. Modular Smart Array (MSA) technical user documentations can be found at <http://www.hp.com/go/MSA2000>

Intended audience

This paper is intended for entry-level and mid-range HP StorageWorks MSA2000 G1, MSA2000 G2, and P2000 G3 MSA administrators and requires previous SAN knowledge. This document offers Modular Storage Array facts that can contribute to an MSA best customer experience.

This paper outlines a best practice approach to performance and configuration.

The paper is broken into three sections:

1. MSA2000 G1 best practices
2. MSA2000 G2 best practices
3. P2000 G3 MSA best practices

The MSA2000 G1

Topics covered

This section examines the following:

- Hardware overview
- Choosing between iSCSI, Fibre Channel, and SAS
- Fault Tolerance versus Performance
- Unified LUN Presentation (ULP)
- Choosing single or dual controllers
- Choosing DAS or SAN attach
- Dealing with controller failures
- Virtual disks
- RAID levels
- World Wide Name (WWN) naming conventions
- Cache configuration
- Fastest throughput optimization
- Highest fault-tolerance optimization

Hardware overview

HP StorageWorks 2000fc G1 Modular Smart Array

The MSA2000fc G1 is a 4 Gb Fibre Channel connected 2U storage area network (SAN) and direct-attach storage (DAS) solution designed for small to medium-sized deployments or remote locations. The G1 model comes standard with 12 Large Form Factor (LFF) drive bays, able to simultaneously accommodate enterprise-class SAS drives and archival-class SATA drives. Additional capacity can easily be added when needed by attaching up to three MSA2000 12 LFF bay drive enclosures. Maximum raw capacity ranges from 5.4 TB SAS or 12 TB SATA in the base cabinet, to over 21.6 TB SAS or 48 TB SATA with the addition of the maximum number of drive enclosures and necessary drives. The MSA2000fc G1 supports up to 64 single path hosts for Fibre Channel attach.

HP StorageWorks 2000i G1 Modular Smart Array

The MSA2000i is a 1 Gb Ethernet (1 GbE) iSCSI connected to 2U SAN array solution. The MSA2000i also, like the MSA2000sa, allows you to grow your storage as demands increase up to 21.6 TB SAS or 48 TB SATA, supporting up to 16 hosts for iSCSI attach.

The MSA2000i offers flexibility and is available in two models: A single controller version for lowest price with future expansion and a dual controller model for the more demanding entry-level situations that require higher availability. Each model comes standard with 12 drive bays that can simultaneously accommodate 3.5-inch enterprise-class SAS drives and archival-class SATA drives. Additional capacity can be easily added when needed, by attaching up to three MSA2000 12 bay drive enclosures.

HP StorageWorks 2000sa G1 Modular Smart Array

The MSA2000sa is a direct-attach 3-Gb SAS-connected 2U solution, designed for small to medium-sized deployments or remote locations. The MSA2000sa is also used as an integral part of the Direct Attach Storage for HP BladeSystem, bringing SAS direct attach storage to the HP BladeSystem c-Class enclosures. The MSA2000sa comes in two models—a basic single controller model for low initial cost with the ability to upgrade later and a model with dual controllers standard for the more demanding entry-level situations that require higher availability.

Each model comes standard with 12 drive bays that can simultaneously accommodate 3.5-inch enterprise-class SAS drives and archival-class SATA drives. Additional capacity can easily be added by attaching up to three MSA2000 12-bay LFF drive enclosures. Maximum raw capacity ranges from 5.4 TB SAS or 12 TB SATA in the base cabinet, to 21.6 TB SAS or 48 TB SATA with the addition of the maximum number of drives and drive enclosures. The MSA2000sa supports up to four hosts for SAS direct attach or 32 hosts for switch attach in the BladeSystem configuration.

iSCSI, Fibre Channel, or SAS

When choosing the right HP StorageWorks MSA2000 model, you should determine your budget and performance needs. Each model has unique features that should be weighed when making your decisions.

Each model has some distinct characteristics.

Characteristics of the MSA2000i G1 model:

- iSCSI uses the Transport Control Protocol (TCP) for moving data over Ethernet media
- Offers SAN benefits in familiar Ethernet infrastructure
- Lowers infrastructure cost
- Lowers cost of ownership

Characteristics of the MSA2000fc G1 model:

- Offers a faster controller for greater performance
- The Fibre Channel controllers support 4 Gb for better throughput
- Integrates easily into existing Fibre Channel infrastructure
- More scalability, greater number of LUNs, and optional snapshots

Characteristics of the MSA2000sa G1 model:

- Supports ULP (Unified LUN Presentation)
- No need to set interconnect settings
- Lower cost infrastructure

Unified LUN Presentation (ULP)

The MSA2000sa G1 uses the concept of ULP. ULP can expose all LUNs through all host ports on both controllers. ULP appears to the host as an active-active storage system where the host can choose any available path to access a LUN regardless of vdisk ownership.

ULP uses the T10 Technical Committee of the InterNational Committee for Information Technology Standards (INCITS) Asymmetric Logical Unit Access (ALUA) extensions, in SPC-3, to negotiate paths with aware host systems. Unaware host systems see all paths as being equal.

Overview

ULP presents all LUNS to all host ports

- Removes the need for controller interconnect path
- Presents the same World Wide Node Name (WWNN) for both controllers

Shared LUN number between controllers with a maximum of 512 LUNs

- No duplicate LUNs allowed between controllers
- Either controller can use any unused logical unit number

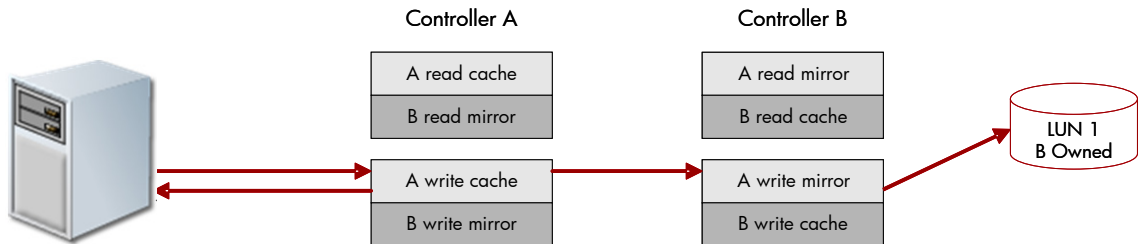
ULP recognizes which paths are “preferred”

- The preferred path indicates which is the owning controller per ALUA specifications
- “Report Target Port Groups” identifies preferred path
- Performance is slightly better on preferred path

Write I/O Processing with ULP

- Write command to controller A for LUN 1 owned by Controller B
- The data is written to Controller A cache and broadcast to Controller A mirror
- Controller A acknowledges I/O completion back to host
- Data written back to LUN 1 by Controller B from Controller A mirror

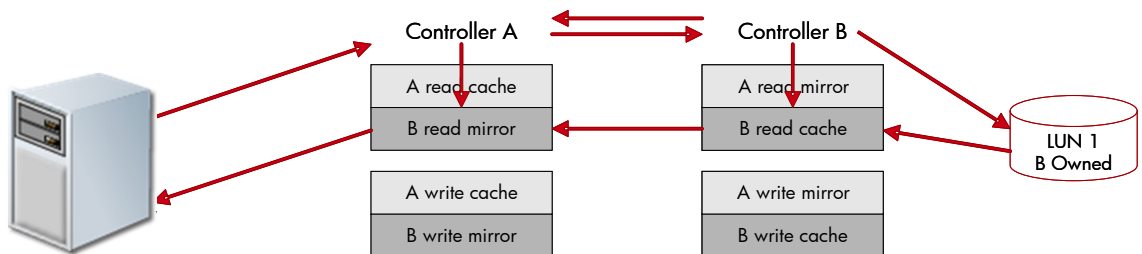
Figure 1: Write I/O Processing with ULP



Read I/O Processing with ULP

- Read command to controller A for LUN 1 owned by Controller B:
 - Controller A asks Controller B if data is in Controller B cache
 - If found, Controller B tells Controller A where in Controller B read mirror cache it resides
 - Controller A sends data to host from Controller B read mirror, I/O complete
 - If not found, request is sent from Controller B to disk to retrieve data
 - Disk data is placed in Controller B cache and broadcast to Controller B mirror
 - Read data sent to host by Controller A from Controller B mirror, I/O complete

Figure 2: Read I/O Processing with ULP



Fault tolerance versus performance on the MSA2000fc G1

Depending on whether performance or fault tolerance (where redundant components are designed for continuous processing) is more important to your solution, the host port interconnects need to be enabled or disabled through the HP Storage Management Utility (SMU). In an FC storage system, the host port interconnects act as an internal switch to provide data-path redundancy.

When availability is more important than performance, the host port interconnects should be enabled to connect the host ports in controller A to those in controller B. When the interconnects are enabled, the host has access to both controllers' mapped volumes. This dual access makes it possible to create a redundant configuration without using an external switch.

If one controller fails in this configuration, the interconnects remain active so hosts can continue to access all mapped volumes without the intervention of host-based failover software. The controllers accomplish this by means of FC target multi-ID, while a controller is failed over, each surviving controller host port presents its own port WWN and the port WWN of the interconnected, failed controller host port that was originally connected to the loop. The mapped volumes owned by the failed controller remain accessible until it is removed from the enclosure.

When the host port interconnects are disabled, volumes owned by a controller are accessible from its host ports only. This is the default setting.

When controller enclosures are attached directly to hosts and high availability is required, host port interconnects should be enabled. Host port interconnects are also enabled for applications where fault tolerance is required and performance is not, and when switch ports are at a premium.

When controller enclosures are attached through one or more switches, or when they are attached directly but performance is more important than fault tolerance, host port interconnects should be disabled.

Note:

The interconnect setting is available only for the MSA2000fc G1.
The MSA2000i G1 uses Ethernet switches for fault tolerance.
The MSA2000sa G1 employs ULP architecture that was discussed previously.

Tip:

It is a best practice to enable host port interconnects when controller enclosures are attached directly to hosts and high availability is required, or when switch ports are at a premium and fault tolerance is required.

Note:

Fault tolerance and performance are affected by cache settings as well. See "Cache Configuration" later in this paper for more information.

Choosing single or dual controllers

Although you can purchase a single-controller configuration, it is a best practice to use the dual-controller configuration to enable high availability and better performance. However, under certain circumstances, a single-controller configuration can be used as an overall redundant solution.

Dual controller

A dual-controller configuration improves application availability because in the unlikely event of a controller failure, the affected controller fails over to the surviving controller with little interruption to the flow of data. The failed controller can be replaced without shutting down the storage system, thereby providing further increased data availability. An additional benefit of dual controllers is increased performance as storage resources can be divided between the two controllers, enabling them to share the task of processing I/O operations. For the MSA2000 G1, a single-controller array is limited to 128 LUNs. With the addition of a second controller, the support increases to 256 LUNs. Controller failure results in the surviving controller:

- Taking ownership of all RAID sets
- Managing the failed controller's cache data
- Restarting data protection services
- Assuming the host port characteristics of both controllers

The dual-controller configuration takes advantage of mirrored cache. By automatically "broadcasting" one controller's write data to the other controller's cache, the primary latency overhead is removed and bandwidth requirements are reduced on the primary cache. Any power loss situation results in the immediate writing of cache data into both controllers' compact flash devices, removing any data loss concern. The broadcast write implementation provides the advantage of enhanced data protection options without sacrificing application performance or end-user responsiveness.

Single controller

A single-controller configuration provides no redundancy in the event that the controller fails; therefore, the single controller is a potential Single Point of Failure (SPOF). Multiple hosts can be supported in this configuration (up to two for direct attach). In this configuration, each host can have 1-Gb/sec (MSA2000i G1), 3-Gb/sec (MSA2000sa G1), or 2/4-Gb/sec (MSA2000fc G1) access to the storage resources. If the controller fails, or if the data path to a directly connected host fails, the host loses access to the storage until the problem is corrected and access is restored.

The single-controller configuration is less expensive than the dual-controller configuration. It is a suitable solution in cases where high availability is not required and loss of access to the data can be tolerated until failure recovery actions are complete. A single-controller configuration is also an appropriate choice in storage systems where redundancy is achieved at a higher level, such as a two-node cluster. For example, a two-node cluster where each node is attached to a controller enclosure with a single controller and the nodes do not depend upon shared storage. In this case, the failure of a controller is equivalent to the failure of the node to which it is attached.

Another suitable example of a high-availability storage system using a single controller configuration is where a host uses a volume manager to mirror the data on two independent single-controller storage systems. If one storage system fails, the other storage system can continue to serve the I/O operations. Once the failed controller is replaced, the data from the survivor can be used to rebuild the failed system.

Note:

When using a single-controller system, the controller must be installed in the slot A of the array.

Choosing DAS or SAN attach

There are two basic methods for connecting storage to data hosts: Direct Attached Storage (DAS) and Storage Area Networks (SAN). The option you select depends on the number of hosts you plan to connect and how rapidly you need your storage solution to expand.

Direct attach

DAS uses a direct connection between a data host and its storage system. The DAS solution of connecting each data host to a dedicated storage system is straightforward and the absence of storage switches can reduce cost. Like a SAN, a DAS solution can also share a storage system, but it is limited by the number of ports on the storage system. The MSA2000sa G1 only supports DAS. The MSA2000i G1 does not support direct attach, but does support an iSCSI SAN. The MSA2000fc G1 supports either direct attach or fabric switch attach configurations.

A powerful feature of the MSA2000fc G1 and MSA2000sa G1 storage systems are their ability to support four direct attach single-port data hosts, or two direct attach dual-port data hosts without requiring storage switches. The MSA2000fc G1 and MSA2000sa G1 can also support two single-connected hosts and one dual connected host for a total of three hosts.

The MSA2000sa G1 is also used as an integral part of the Direct Attach Storage for HP BladeSystem solution. In this configuration, the MSA2000sa G1 can support up to 32 blade server hosts attached to the storage array by means of a SAS switch that is integrated into the HP BladeSystem c-Class enclosure.

If the number of connected hosts is not going to change or increase beyond four then the DAS solution is appropriate. However, if the number of connected hosts is going to expand beyond the limit imposed by the use of DAS, it is best to implement a SAN. The SAN implementation is only supported on the MSA2000fc G1 and MSA2000i G1.

Tip:

It is a best practice to use a dual-port connection to data hosts when implementing a DAS solution to achieve a degree of redundancy.

Switch attach

A switch attach solution, or SAN, places a switch between the servers and storage systems. This strategy tends to use storage resources more effectively and is commonly referred to as storage consolidation. A SAN solution shares a storage system among multiple servers using switches, and reduces the total number of storage systems required for a particular environment, at the cost of additional element management (switches) and path complexity.

Host port interconnects are typically disabled. There is an exception to this rule; host port interconnects are enabled for applications where fault tolerance is required and highest performance is not required, and when switch ports are at a premium.

Using switches increases the number of servers that can be connected. Essentially, the maximum number of data hosts that can be connected to the SAN becomes equal to the number of available switch ports.

Note:

In a switched environment, the HP StorageWorks MSA2000fc G1 supports 64 hosts; the MSA2000sa G1 supports 32 (BladeSystem) hosts, while the HP StorageWorks MSA2000i G1 can support 16 hosts.

Tip:

It is a best practice to use a switched SAN environment anytime more than four hosts are used or when required storage or number of hosts is expected to grow.

Dealing with controller failovers

In the MSA2000fc G1 storage system, the host port interconnects act as an internal switch to provide data-path redundancy.

When the host port interconnects are enabled, port 0 on each controller is cross connected to port 1 on the other controller. This provides redundancy in the event of failover by making volumes owned by either controller accessible from either controller.

When the host port interconnects are disabled, volumes owned by a controller are accessible from its host ports only. This is the default configuration.

For a single-controller FC system, host port interconnects are almost always disabled.

For a dual-controller FC system in a direct-attach configuration, host port interconnects are typically enabled—except in configurations where fault tolerance is not required but better performance is required.

For a dual-controller FC system in a switch-attach configuration, host port interconnects are always disabled.

You cannot enable host port interconnects if any host port is set to point-to-point topology.

FC switch-attach configuration

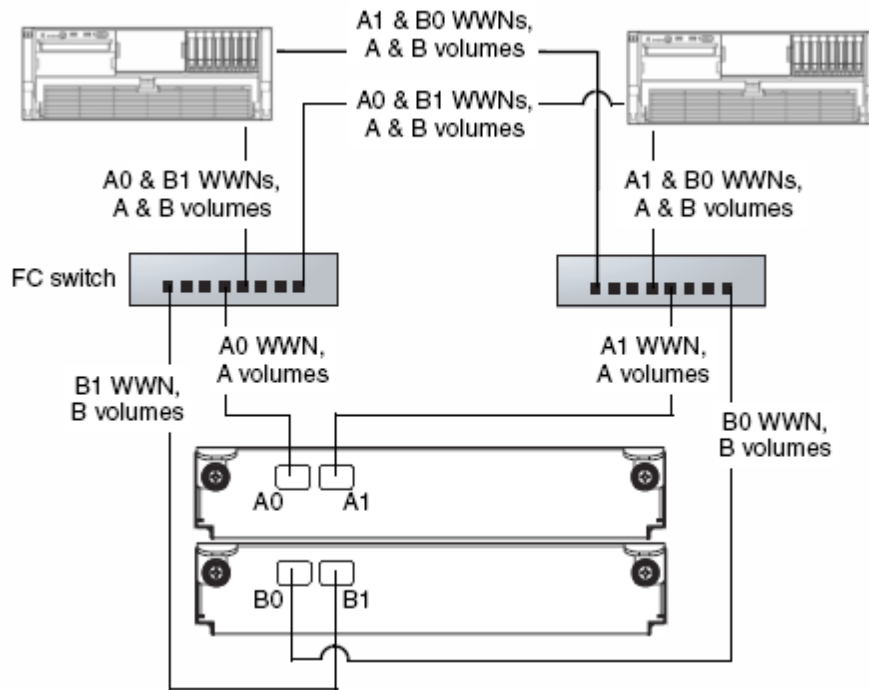
The topology only affects how mapped volumes and port WWNs are presented if one controller fails. Whichever topology is used, each data host has dual-ported access to volumes through both controllers.

- Failover in a switch-attach, loop configuration:
If one controller fails in a switch-attach configuration using loop topology, the host ports on the surviving controller present the port WWNs for both controllers. Each controller's mapped volumes remain accessible.
- Failover in a switch-attach, point-to-point configuration:
If one controller fails in a switch-attach configuration using point-to-point topology, the surviving controller presents its mapped volumes on its primary host port and the mapped volumes owned by the failed controller on the secondary port.

In a high-availability configuration, two data hosts connect through two switches to a dual-controller storage system and the host port interconnects are disabled.

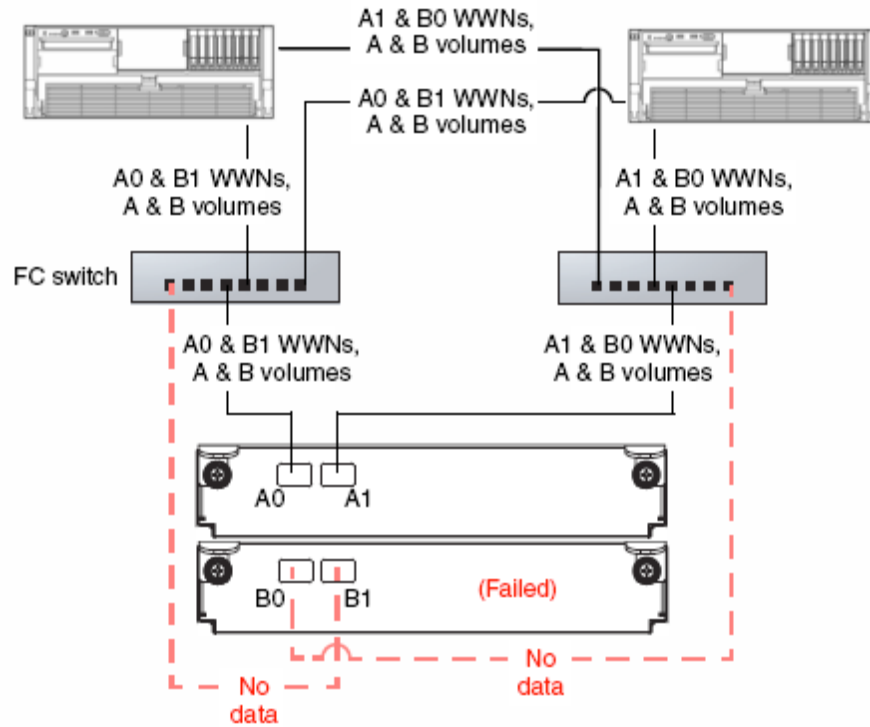
Figure 3 shows how port WWNs and mapped volumes are presented when both controllers are active.

Figure 3: FC storage presentation during normal operation (switch attach with two switches and two hosts)



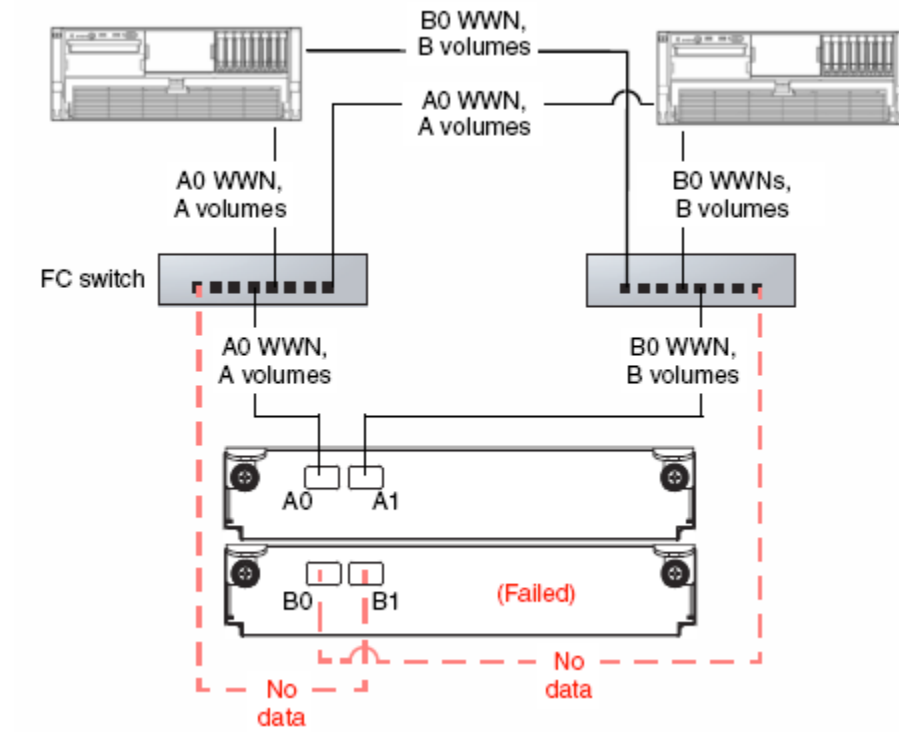
For a system using loop topology, Figure 4 shows how port WWNs and mapped volumes are presented if controller B fails.

Figure 4: FC storage presentation during failover (Switch attach, loop configuration)



For a system using point-to-point topology, Figure 5 shows how port WWNs and mapped volumes are presented if controller B fails.

Figure 5: FC storage presentation during failover (Switch attach, point-to-point configuration)



iSCSI switch-attach configuration

The high-availability configuration requires two gigabit Ethernet (GbE) switches. During active-active operation, both controllers' mapped volumes are visible to both data hosts.

A dual-controller MSA2012i G1 storage system uses port 0 of each controller as one failover pair and port 1 of each controller as a second failover pair. If one controller fails, all mapped volumes remain visible to all hosts. Dual IP-address technology is used in the failed-over state and is largely transparent to the host system. Figure 6 shows how port IP addresses and mapped volumes are presented when both controllers are active.

Figure 6: iSCSI storage presentation during normal operation

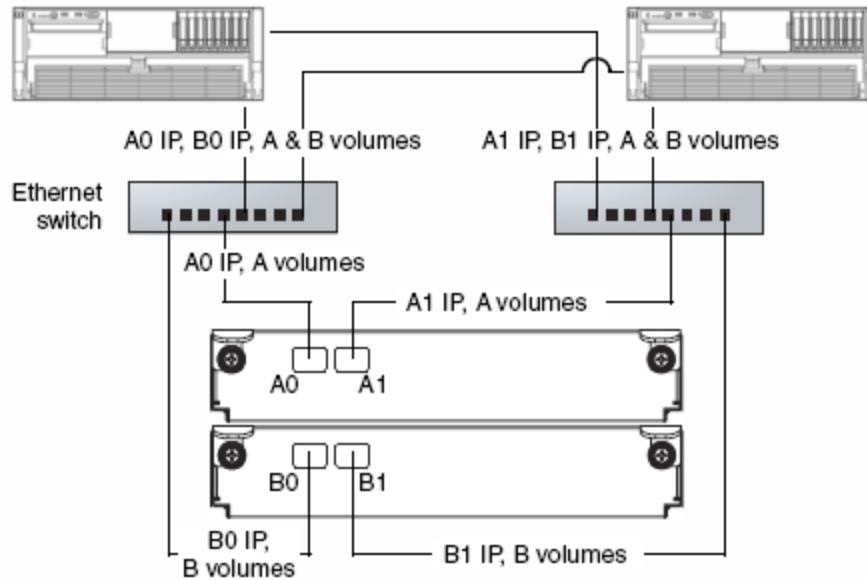
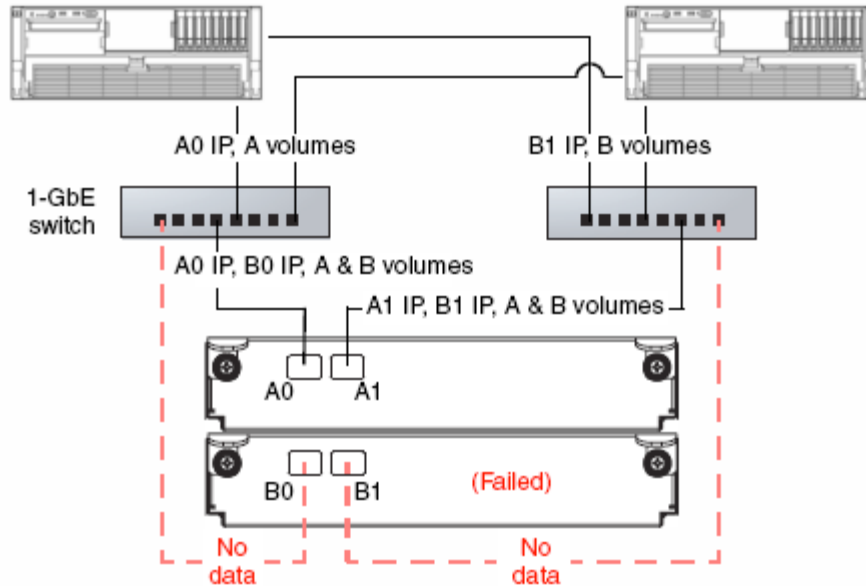


Figure 7 shows how port IP addresses and mapped volumes are presented if controller B fails.

Figure 7: iSCSI storage presentation during failover



SAS direct-attach configurations

The MSA2000sa G1 uses ULP. ULP is a controller software feature that enables hosts to access mapped volumes through both controllers' host ports (target ports) without the need for internal or external switches.

In a dual-controller SAS system, both controllers share a unique WWN so they appear as a single device to hosts. The controllers also share one set of LUNs for mapping volumes to hosts.

A host can use any available data path to access a volume owned by either controller. The preferred path, which offers slightly better performance, is through target ports on a volume's owning controller.

Note:

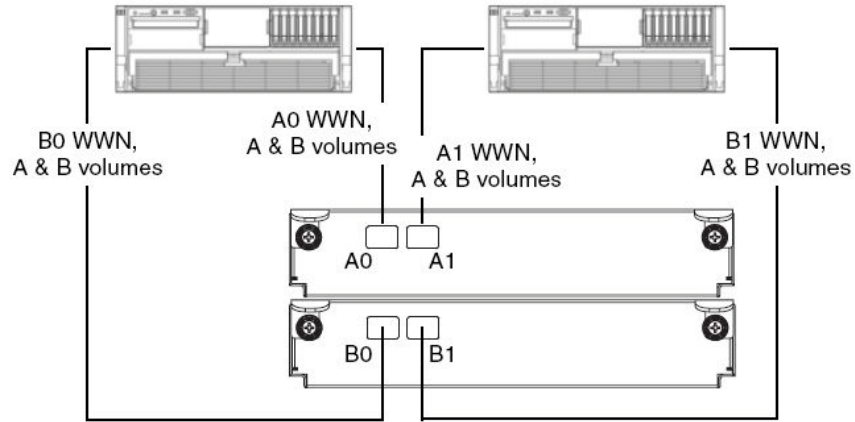
Ownership of volumes is not visible to hosts. However, in SMU you can view volume ownership and change the owner of a virtual disk and its volumes.

Note:

Changing the ownership of a virtual disk should never be done with I/O in progress. I/O should be quiesced prior to changing ownership.

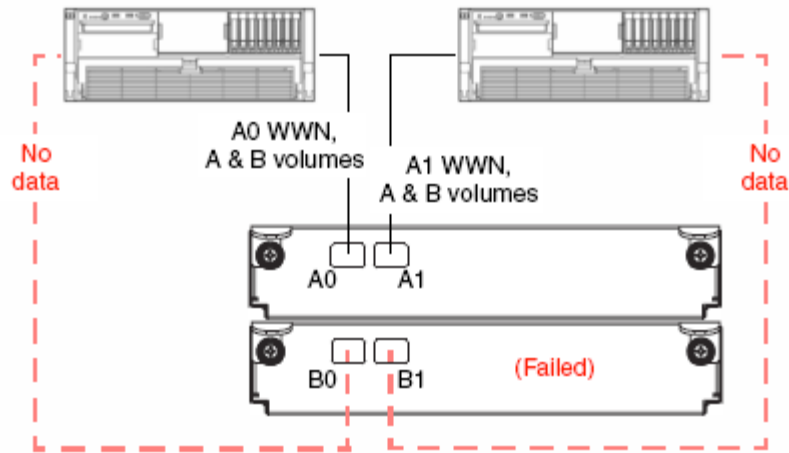
In the following configuration, both hosts have redundant connections to all mapped volumes.

Figure 8: SAS storage presentation during normal operation (high-availability, dual-controller, and direct attach with two hosts)



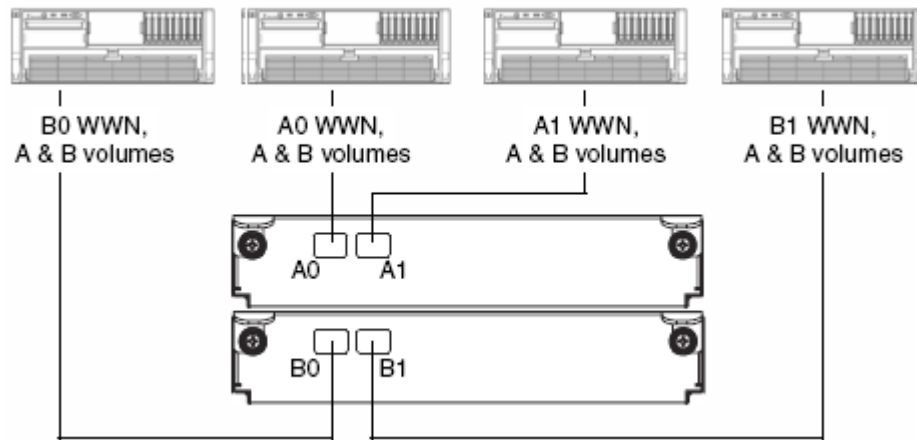
If a controller fails, the hosts maintain access to all of the volumes through the host ports on the surviving controller, as shown in the Figure 9.

Figure 9: SAS storage presentation during failover (high-availability, dual-controller, and direct attach with two hosts)



In the following configuration, each host has a non-redundant connection to all mapped volumes. If a controller fails, the hosts connected to the surviving controller maintain access to all volumes owned by that controller.

Figure 10: SAS storage presentation during normal operation (high-availability, dual-controller, and direct attach with four hosts)



Virtual disks

A virtual disk (vdisk) is a group of disk drives configured with a RAID level. Each virtual disk can be configured with a different RAID level. A virtual disk can contain either SATA drives or SAS drives, but not both. The controller safeguards against improperly combining SAS and SATA drives in a virtual disk. The system displays an error message if you choose drives that are not of the same type.

The HP StorageWorks MSA2000 system can have a maximum of 16 virtual disks per controller for a maximum of 32 virtual disks with a dual-controller configuration.

For storage configurations with many drives, it is recommended to consider creating a few virtual disks each containing many drives, as opposed to many virtual disks each containing a few drives. Having many virtual disks is not very efficient in terms of drive usage when using RAID 3 or RAID 5. For example, one 12-drive RAID 5 virtual disk has 1 parity drive and 11 data drives, whereas four 3-drive RAID 5 virtual disks each have 1 parity drive (4 total) and 2 data drives (only 8 total).

A virtual disk can be larger than 2 TB. This can increase the usable storage capacity of configurations by reducing the total number of parity disks required when using parity-protected RAID levels. However, this differs from using volumes larger than 2 TB, which requires specific operating system, Host Bus Adapter (HBA) driver, and application-program support.

Note:

The MSA2000 can support a maximum vdisk size of 16 TB.

Supporting large storage capacities requires advanced planning because it requires using large virtual disks with several volumes each or many virtual disks. To increase capacity and drive usage (but not performance), you can create virtual disks larger than 2 TB and divide them into multiple volumes with a capacity of 2 TB or less.

The largest supported vdisk is the number of drives allowed in a RAID set multiplied by the largest drive size.

- RAID 0, 3, 5, 6, 10 can support up to 16 drives (with 1 TB SATA drives that is 16 TB raw)
- RAID 50 can support up to 32 drives (with 1 TB SATA drives that is 32 TB raw)

Tip:

The best practice for creating virtual disks is to add them evenly across both controllers. With at least one virtual disk assigned to each controller, both controllers are active. This active-active controller configuration allows maximum use of a dual-controller configuration's resources.

Tip:

Another best practice is to stripe virtual disks across shelf enclosures to enable data integrity in the event of an enclosure failure. A virtual disk created with RAID 1, 10, 3, 5, 50, or 6 can sustain an enclosure failure without loss of data depending on the number of shelf enclosures attached. The design should take into account whether spares are being used and whether the use of a spare can break the original design. A plan for evaluation and possible reconfiguration after a failure and recovery should be addressed. Non-fault tolerant vdisks (RAID 0 or non-RAID) do not need to be dealt with in this context because a shelf enclosure failure with any part of a non-fault tolerant vdisk can cause the vdisk to fail.

Chunk size

When you create a virtual disk, you can use the default chunk size or one that better suits your application. The chunk (also referred to as stripe unit) size is the amount of contiguous data that is written to a virtual disk member before moving to the next member of the virtual disk. This size is fixed throughout the life of the virtual disk and cannot be changed. A stripe is a set of stripe units that are written to the same logical locations on each drive in the virtual disk. The size of the stripe is determined by the number of drives in the virtual disk. The stripe size can be increased by adding one or more drives to the virtual disk.

Available chunk sizes include:

- 16 KB
- 32 KB
- 64 KB (default)

If the host is writing data in 16 KB transfers, for example, then that size would be a good choice for random transfers because one host read would generate the read of exactly one drive in the volume. That means if the requests are random-like, then the requests would be spread evenly over all of the drives, which is good for performance.

If you have 16-KB accesses from the host and a 64 KB block size, then some of the host's accesses would hit the same drive; each stripe unit contains four possible 16-KB groups of data that the host might want to read.

Alternatively, if the host accesses were 128 KB in size, then each host read would have to access two drives in the virtual disk. For random patterns, that ties up twice as many drives.

Note:

On RAID 50 drives, the chunk size is displayed as:
<requested Chunk size> * (Num drives in sub vdisk—1)

For example: A requested chunk size of 32 KB with 4 drives in a sub array.
The chunk size is reported as 96 KB.

Using the formula: 32 K byte* (4-1) = 96 KB.

Tip:

The best practice for setting the chunk size is to match the transfer block size of the application.

RAID levels

Choosing the correct RAID level is important whether your configuration is for fault tolerance or performance. Table 1 gives an overview of supported RAID implementations highlighting performance and protection levels.

Note:

Non-RAID is supported for use when the data redundancy or performance benefits of RAID are not needed; no fault tolerance.

Table 1: An overview of supported RAID implementations

RAID Level	Cost	Performance	Protection Level
RAID 0 Striping	N/A	Highest	No data protection
RAID 1 Mirroring	High cost 2x drives	High	Protects against individual drive failure
RAID 3 Block striping with dedicated parity drive	1 drive	Good	Protects against individual drive failure
RAID 5 Block striping with striped parity drive	1 drive	Good	Protects against any individual drive failure; medium level of fault tolerance
RAID 6 Block striping with multiple striped parity	2 drives	Good	Protects against multiple (2) drive failures; high level of fault tolerance
RAID 10 Mirrored striped array	High cost 2x drives	High	Protects against certain multiple drive failures; high level of fault tolerance
RAID 50 Data striped across RAID 5	At least 2 drives	Good	Protects against certain multiple drive failures; high level of fault tolerance

Spares

When configuring virtual disks, you can add a maximum of four available drives to a redundant virtual disk (RAID 1, 3, 5, 6, and 50) for use as spares. If a drive in the virtual disk fails, the controller automatically uses the vdisk spare for reconstruction of the critical virtual disk to which it belongs. A spare drive must be the same type (SAS or SATA) as other drives in the virtual disk. You cannot add a spare that has insufficient capacity to replace the smallest drive in the virtual disk. If two drives fail in a RAID 6 virtual disk, two properly sized spare drives must be available before reconstruction can begin. For RAID 50 virtual disks, if more than one subdisk becomes critical, reconstruction, and use of vdisk spares occur in the order subvdisks are numbered.

You can designate a global spare to replace a failed drive in any virtual disk of the appropriate type for example, a SAS spare disk drive for any SAS vdisk or a vdisk spare to replace a failed drive in only a specific virtual disk. Alternatively, you can enable dynamic spares in HP SMU. Dynamic sparing enables the system to use any drive that is not part of a virtual disk to replace a failed drive in any virtual disk.

Tip:

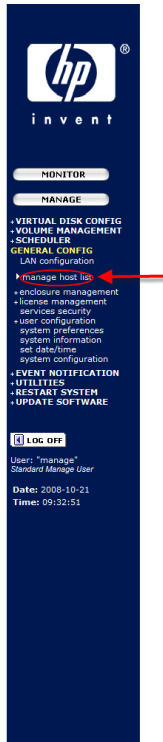
A best practice is to designate a spare disk drive for use if a drive fails. Although using a dedicated vdisk spare is the most secure way to provide spares for your virtual disks, it is also expensive to keep a spare assigned to each virtual disk. An alternative method is to enable dynamic spares or to assign one or more unused drives as global spares.

World Wide Name (WWN) naming conventions

A best practice for acquiring and renaming World Wide Names (WWN) for the MSA2000sa G1 is to plug-in one SAS cable connection at a time and then rename the WWN to an identifiable name.

Procedure:

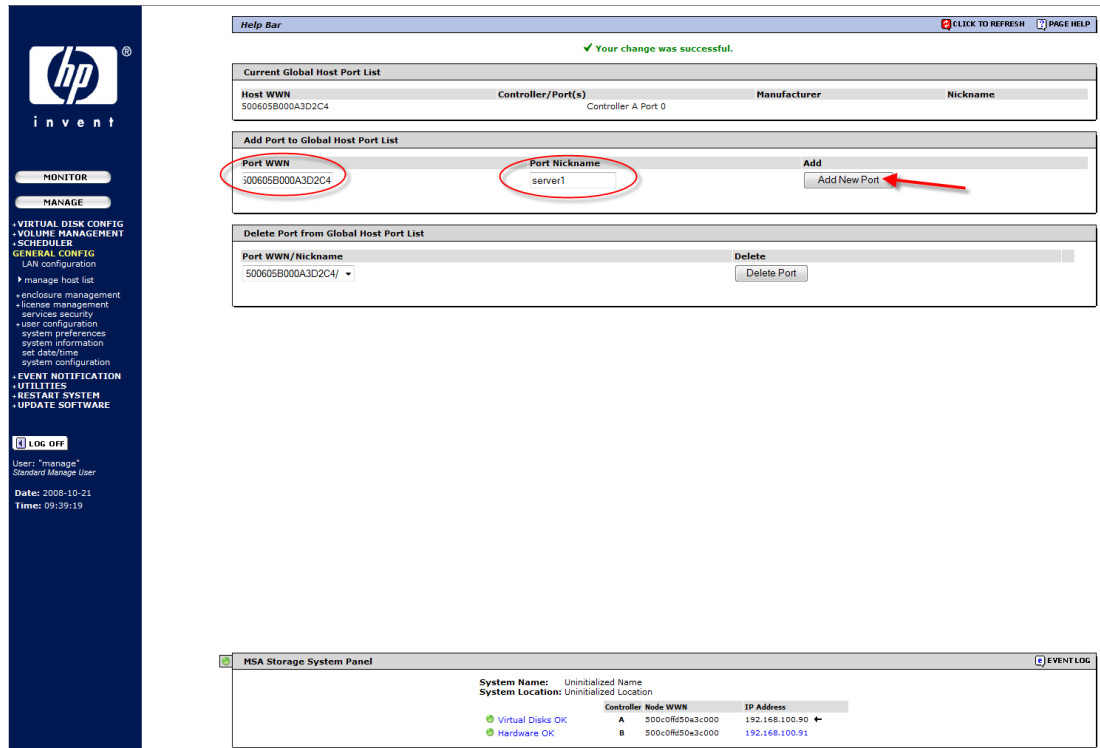
4. Open up the HP StorageWorks Storage Management Utility (SMU).
5. Click Manage → General Config. → manage host list.



The image shows the main content area of the SMU. At the top, there is a 'Help Bar' with a green checkmark and the text 'Your change was successful.' Below this is a table titled 'Current Global Host Port List' with columns for 'Host WWN', 'Controller/Port(s)', 'Manufacturer', and 'Nickname'. The table contains one row with the WWN '500605B000A3D2C4' and 'Controller A Port 0'. Below the table is a section titled 'Add Port to Global Host Port List' with input fields for 'Port WWN' and 'Port Nickname', and an 'Add' button. At the bottom is a section titled 'Delete Port from Global Host Port List' with a dropdown menu for 'Port WWN/Nickname' and a 'Delete' button.

The image shows the 'MSA Storage System Panel' at the bottom of the SMU interface. It displays system status information: 'System Name: Uninitialized Name', 'System Location: Uninitialized Location', and 'Virtual Disks OK'. Below this is a table for 'Controller Node WWN' with columns for 'Controller', 'Node WWN', and 'IP Address'. The table contains two rows: 'A' with WWN '500c0Rd50e3c000' and IP '192.168.100.90', and 'B' with WWN '500c0Rd50e3c000' and IP '192.168.100.91'. There is also an 'EVENT LOG' button in the top right corner of the panel.

6. Locate the WWN of the first SAS HBA under the “Current Global Host Port List” and type this WWN into the “Port WWN.” Type in a nickname for this port in the “Port Nickname” box.
7. Click “Add New Port” Click OK when the pop up window appears.



8. Plug in the SAS port of the HBA on the second server into the MSA2000sa controller port. Make sure the server is powered on.
9. Return to the Manage → General Config. → manage host list of the SMU. The new WWN should now appear.
10. Repeat steps 3–5 for the remaining servers.

Cache configuration

Controller cache options can be set for individual volumes to improve a volume’s fault tolerance and I/O performance.

Note:

To change the following cache settings, the user—who logs into the HP SMU—must have the “advanced” user credential. The manage user has the “standard” user credential by default. This credential can be changed using the HP SMU and going to the Manage → General Config. → User Configuration → Modify Users

Read-ahead cache settings

The read-ahead cache settings enable you to change the amount of data read in advance after two back-to-back reads are made. Read ahead is triggered by two back-to-back accesses to consecutive logical block address (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.

The default read-ahead size, which sets one chunk for the first access in a sequential read and one stripe for all subsequent accesses, works well for most users in most applications. The controllers treat volumes and mirrored virtual disks (RAID 1) internally as if they have a stripe size of 64 KB, even though they are not striped.

Caution: The read-ahead cache settings should only be changed if you fully understand how your operating system, application, and HBA (FC) or Ethernet adapter (iSCSI) move data so that you can adjust the settings accordingly. You should be prepared to monitor system performance using the virtual disk statistics and adjust read-ahead size until you find the right size for your application.

The Read Ahead Size can be set to one of the following options:

- **Default:** Sets one chunk for the first access in a sequential read and one stripe for all subsequent accesses. The size of the chunk is based on the block size used when you created the virtual disk (the default is 64 KB). Non-RAID and RAID 1 virtual disks are considered to have a stripe size of 64 KB.
- **Disabled:** 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. You can use the volume statistics read histogram to determine what size accesses the host is doing.
- **64, 128, 256, or 512 KB; 1, 2, 4, 8, 16, or 32 MB:** Sets the amount of data to read first and the same amount is read for all read-ahead accesses.
- **Maximum:** Let the controller dynamically calculate the maximum read-ahead cache size for the volume. For example, if a single volume exists, this setting enables the controller to use nearly half the memory for read-ahead cache.

Note:

Only use "Maximum" when host-side performance is critical and disk drive latencies must be absorbed by cache. For example, for read-intensive applications, you may want data that is most often read in cache so that the response to the read request is very fast; otherwise, the controller has to locate which disks the data is on, move it up to cache, and then send it to the host.

Note:

If there are more than two volumes, there is contention on the cache as to which volume's read data should be held and which has the priority; the volumes begin to constantly overwrite the other volume's data, which could result in taking a lot of the controller's processing power. Avoid using this setting if more than two volumes exist.

Cache optimization can be set to one of the following options:

- **Standard:** Works well for typical applications where accesses are a combination of sequential and random access. This method is the default.
- **Super-Sequential:** Slightly modifies the controller's standard read-ahead caching algorithm by enabling the controller to discard cache contents that have been accessed by the host, making more room for read-ahead data. This setting is not effective if random accesses occur; use it only if your application is strictly sequential and requires extremely low latency.

Write-back cache settings

Write back is a cache-writing strategy in which the controller receives the data to be written to disk, 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 drive. Write-back cache mirrors all of the data from one controller module cache to the other. Write-back cache improves the performance of write operations and the throughput of the controller.

When write-back cache is disabled, write-through becomes the cache-writing strategy. Using write-through cache, the controller writes the data to the disk before signaling the host operating system that the process is complete. Write-through cache has lower throughput and write operation performance than write back, but it is the safer strategy, with low risk of data loss on power failure. However, write-through cache does not mirror the write data because the data is written to the disk before posting command completion and mirroring is not required. You can set conditions that cause the controller to switch from write-back caching to write-through caching as described in “Auto-Write Through Trigger and Behavior Settings” later in this paper.

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

You can enable and disable the write-back cache for each volume, as volume write-back cache is enabled by default. Data is not lost if the system loses power because controller cache is backed by super capacitor technology. For most applications this is the correct setting, but because backend bandwidth is used to mirror cache, if you are writing large chunks of sequential data (as would be done in video editing, telemetry acquisition, or data logging) write-through cache has much better performance. Therefore, you might want to experiment with disabling the write-back cache. You might see large performance gains (as much as 70 percent) if you are writing data under the following circumstances:

- Sequential writes
- Large I/Os in relation to the chunk size
- Deep queue depth

If you are doing any type of random access to this volume, leave the write-back cache enabled.

Caution: Write-back cache should only be disabled if you fully understand how your operating system, application, and HBA (SAS) move data. You might hinder your storage system’s performance if used incorrectly.

Auto-write through trigger and behavior settings

You can set the trigger conditions that cause the controller to change the cache policy from write-back to write-through. While in write-through mode, system performance might be decreased.

A default setting makes the system revert to write-back mode when the trigger condition clears. To make sure that this occurs and that the system doesn’t operate in write-through mode longer than necessary, make sure you check the setting in HP SMU or the Command-line Interface (CLI).

You can specify actions for the system to take when write-through caching is triggered:

- **Revert when Trigger Condition Clears:** Switches back to write-back caching after the trigger condition is cleared. The default and best practice is Enabled.
- **Notify Other Controller:** In a dual-controller configuration, the partner controller is notified that the trigger condition is met. The default is Disabled.

Cache-mirroring mode

In the default active-active mode, data for volumes configured to use write-back cache is automatically mirrored between the two controllers. Cache mirroring has a slight impact on performance but provides fault tolerance. You can disable cache mirroring, which permits independent cache operation for each controller; this is called independent cache performance mode (ICPM).

The advantage of ICPM is that the two controllers can achieve very high write bandwidth and still use write-back caching. User data is still safely stored in non-volatile RAM, with backup power provided by super capacitors should a power failure occur. This feature is useful for high-performance applications that do not require a fault-tolerant environment for operation; that is, where speed is more important than the possibility of data loss due to a drive fault prior to a write completion.

The disadvantage of ICPM is that if a controller fails, the other controller may not be able to failover (that is, take over I/O processing for the failed controller). If a controller experiences a complete hardware failure, and needs to be replaced, then user data in its write-back cache is lost.

Data loss does not automatically occur if a controller experiences a software exception, or if a controller module is removed from the enclosure. If a controller should experience a software exception, the controller module goes offline; no data is lost, and it is written to disks when you restart the controller. However, if a controller is damaged in a non-recoverable way then you might lose data in ICPM.

Caution: Data might be compromised if a RAID controller failure occurs after it has accepted write data, but before that data has reached the disk drives. ICPM should not be used in an environment that requires fault tolerance.

Cache configuration summary

The following guidelines list the general best practices. When configuring cache:

- For a fault-tolerant configuration, use the write-back cache policy, instead of the write-through cache policy
- For applications that access both sequential and random data, use the standard optimization mode, which sets the cache block size to 32 KB. For example, use this mode for transaction-based and database update applications that write small files in random order
- For applications that access sequential data only and that require extremely low latency, use the super-sequential optimization mode, which sets the cache block size to 128 KB. For example, use this mode for video playback and multimedia post-production video- and audio-editing applications that read and write large files in sequential order

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 the following table. This section provides a basic starting point for fine-tuning your system, which should be done during performance baseline modeling.

Table 2: Optimizing performance for your application

Application	RAID level	Read ahead cache size	Cache optimization
Default	5 or 6	Default	Standard
HPC (High-Performance Computing)	5 or 6	Maximum	Standard
MailSpooling	1	Default	Standard
NFS_Mirror	1	Default	Standard
Oracle_DSS	5 or 6	Maximum	Standard
Oracle_OLTP	5 or 6	Maximum	Standard
Oracle_OLTP_HA	10	Maximum	Standard
Random1	1	Default	Standard
Random5	5 or 6	Default	Standard
Sequential	5 or 6	Maximum	Super-Sequential
Sybase_DSS	5 or 6	Maximum	Standard
Sybase_OLTP	5 or 6	Maximum	Standard
Sybase_OLTP_HA	10	Maximum	Standard
Video Streaming	1 or 5 or 6	Maximum	Super-Sequential
Exchange Database	10	Default	Standard
SAP	10	Default	Standard
SQL	10	Default	Standard

Fastest throughput optimization

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

- Host interconnects should be disabled when using the MSA2000fc G1.
- Host ports should be configured for 4 Gb/sec on the MSA2000fc G1.
- Host ports should be configured for 1 Gb/sec on the MSA2000i G1.
- Virtual disks 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.

Highest fault tolerance optimization

The following guidelines list the general best practices to follow when configuring your storage system for highest fault tolerance:

- Use dual controllers.
- Use two cable connections from each host.
- If using a direct attach connection on the MSA2000fc G1, host port interconnects must be enabled.
- If using a switch attach connection on the MSA2000fc G1, host port interconnects are disabled and controllers are cross-connected to two physical switches.
- Use Multipath Input/Output (MPIO) software.

The MSA2000 G2

Topics covered

This section examines the following:

- Hardware overview
- Unified LUN Presentation (ULP)
- Choosing single or dual controllers
- Choosing DAS or SAN attach
- Dealing with controller failures
- Virtual disks
- Volume mapping
- RAID levels
- Cache configuration
- Fastest throughput optimization
- Highest fault tolerance optimization
- Boot from storage considerations
- MSA70 considerations
- Administering with HP SMU
- MSA2000i G2 Considerations

What's New in the MSA2000 G2

- New Small Form Factor Chassis with 24 bays
- Support for Small Form Factor (SFF) SAS and SATA drives, common with ProLiant
- Support for attachment of three dual I/O MSA70 SFF JBODs (ninety-nine SFF drives)
- Increased support to four MSA2000 LFF disk enclosures (sixty LFF drives)
- Support for HP-UX along with Integrity servers
- Support for OpenVMS
- New high-performance controller with upgraded processing power
- Increased support of up to 512 LUNs in a dual controller system (511 on MSA2000sa G2)
- Increased optional snapshot capability to 255 snaps
- Improved Management Interface
- JBOD expansion ports changed from SAS to mini-SAS
- Optional DC-power chassis and a carrier-grade, NEBS certified solution
- Support for up to 8 direct attach hosts on the MSA2000sa G2
- Support for up to 4 direct attach hosts on the MSA2000i G2
- Support for up to 64 host port connections on the MSA2000fc G2
- Support for up to 32 host port connections on the MSA2000i G2
- Support for up to 32 hosts in a blade server environment on the MSA2000sa G2
- ULP (new for MSA2000fc G2 and MSA2000i G2 only)

Hardware overview

HP StorageWorks MSA2000fc G2 Modular Smart Array

The MSA2000fc G2 is a 4 Gb Fibre Channel connected 2U SAN or direct-connect solution designed for small to medium-sized departments or remote locations. The controller-less chassis is offered in two models—one comes standard with 12 3.5-inch drive bays, the other can accommodate 24 SFF 2.5-inch drives. Both are able to simultaneously support enterprise-class SAS drives and archival-class SATA drives.

Additional capacity can easily be added when needed by attaching either the MSA2000 12 bay drive enclosure or the MSA70 drive enclosure. Maximum raw capacity ranges from 5.4 TB SAS or 12 TB SATA in the base cabinet, to over 27 TB SAS or 60 TB SATA with the addition of the maximum number of drive enclosures and necessary drives. Configurations utilizing SFF drive chassis can grow to a total of 99 SFF drives. The LFF drive chassis can grow up to a total of 60 drives. The MSA2000fc G2 supports up to 64 single path hosts for Fibre Channel attach.

HP StorageWorks MSA2000i G2 Modular Smart Array

The MSA2000i G2 is an iSCSI GbE connected 2U SAN solution designed for small to medium-sized deployments or remote locations. The controller-less chassis is offered in two models—one comes standard with 12 LFF 3.5-inch drive bays, the other can accommodate 24 SFF 2.5-inch drives. Both are able to simultaneously support enterprise-class SAS drives and archival-class SATA drives. The chassis can have one or two MSA2000i G2 controllers.

The user can opt for the 24 drive bay SFF chassis for the highest spindle counts in the most dense form factor, or go for the 12 drive bay LFF model to max out total capacity. Choose a single controller unit for low initial cost with the ability to upgrade later; or decide on a model with dual controllers for the most demanding entry-level situations. Capacity can easily be added when needed by attaching additional drive enclosures. Maximum capacity ranges with LFF drives up to 27 TB SAS or 60 TB SATA with the addition of the maximum number of drive enclosures. Configurations utilizing the SFF drive chassis and the maximum number of drive enclosures can grow to 29.7 TB of SAS or 11.8 TB of SATA with a total of ninety-nine drives. The MSA2000i G2 has been fully tested up to 64 hosts.

HP StorageWorks 2000sa G2 Modular Smart Array

The MSA2000sa G2 is a 3 Gb SAS direct attach, external shared storage solution designed for small to medium-sized deployments or remote locations. The controller-less chassis is offered in two models—one comes standard with 12 LFF 3.5-inch drive bays, the other can accommodate 24 SFF 2.5-inch drives. Both are able to simultaneously support enterprise-class SAS drives and archival-class SATA drives. The chassis can have one or two MSA2300sa G2 controllers.

The user can opt for the 24-drive bay SFF chassis for the highest spindle counts in the most dense form factor, or go for the 12 drive bay LFF model to max out total capacity. Choose a single controller unit for low initial cost with the ability to upgrade later; or decide on a model with dual controllers for the most demanding entry-level situations. Capacity can easily be added when needed by attaching additional drive enclosures. Maximum capacity ranges with LFF drives up to 27 TB SAS or 60 TB SATA with the addition of the maximum number of drive enclosures. Configurations utilizing the SFF drive chassis and the maximum number of drive enclosures can grow to 29.7 TB of SAS or 11.8 TB of SATA with a total of ninety-nine drives. The MSA2000sa G2 has been fully tested up to 64 hosts.

Unified LUN Presentation (ULP)

The MSA2000 G2 uses the concept of ULP. ULP can expose all LUNs through all host ports on both controllers. The interconnect information is managed in the controller firmware and therefore the host port interconnect setting found in the MSA2000fc G1 is no longer needed. ULP appears to the host as an active-active storage system where the host can choose any available path to access a LUN regardless of vdisk ownership.

ULP uses the T10 Technical Committee of INCITS Asymmetric Logical Unit Access (ALUA) extensions, in SPC-3, to negotiate paths with aware host systems. Unaware host systems see all paths as being equal.

Overview:

ULP presents all LUNS to all host ports

- Removes the need for controller interconnect path
- Presents the same World Wide Node Name (WWNN) for both controllers

Shared LUN number between controllers with a maximum of 512 LUNs

- No duplicate LUNs allowed between controllers
- Either controller can use any unused logical unit number

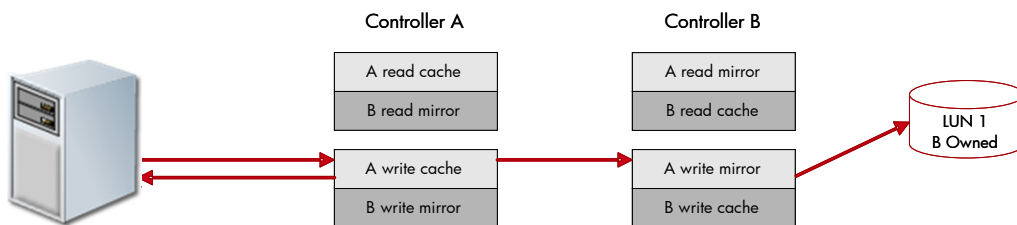
ULP recognizes which paths are “preferred”

- The preferred path indicates which is the owning controller per ALUA specifications
- “Report Target Port Groups” identifies preferred path
- Performance is slightly better on preferred path

Write I/O Processing with ULP

- Write command to controller A for LUN 1 owned by Controller B
- The data is written to Controller A cache and broadcast to Controller A mirror
- Controller A acknowledges I/O completion back to host
- Data written back to LUN 1 by Controller B from Controller A mirror

Figure 11: Write I/O Processing with ULP

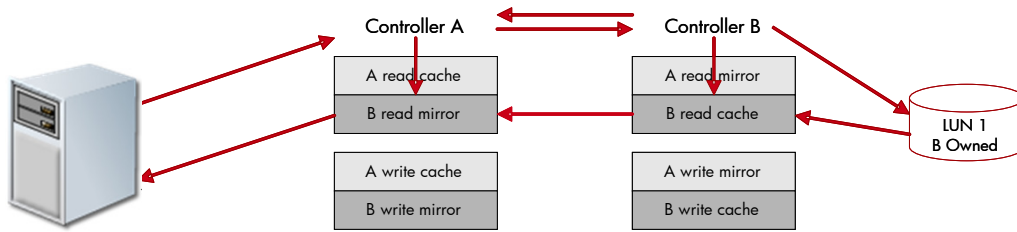


Read I/O Processing with ULP

- Read command to controller A for LUN 1 owned by Controller B:
 - Controller A asks Controller B if data is in Controller B cache
 - If found, Controller B tells Controller A where in Controller B read mirror cache it resides
 - Controller A sends data to host from Controller B read mirror, I/O complete
 - If not found, request is sent from Controller B to disk to retrieve data

- Disk data is placed in Controller B cache and broadcast to Controller B mirror
- Read data sent to host by Controller A from Controller B mirror, I/O complete

Figure 12: Read I/O Processing with ULP



Choosing single or dual controllers

Although you can purchase a single-controller configuration, it is best practice to use the dual-controller configuration to enable high availability and better performance. However, under certain circumstances, a single-controller configuration can be used as an overall redundant solution.

Dual controller

A dual-controller configuration improves application availability because in the unlikely event of a controller failure, the affected controller fails over to the surviving controller with little interruption to the flow of data. The failed controller can be replaced without shutting down the storage system, thereby providing further increased data availability. An additional benefit of dual controllers is increased performance as storage resources can be divided between the two controllers, enabling them to share the task of processing I/O operations. For the MSA2000fc G2, a single controller array is limited to 256 LUNs. With the addition of a second controller, the support increases to 512 LUNs. Controller failure results in the surviving controller by:

- Taking ownership of all RAID sets
- Managing the failed controller's cache data
- Restarting data protection services
- Assuming the host port characteristics of both controllers

The dual-controller configuration takes advantage of mirrored cache. By automatically "broadcasting" one controller's write data to the other controller's cache, the primary latency overhead is removed and bandwidth requirements are reduced on the primary cache. Any power loss situation will result in the immediate writing of cache data into both controllers' compact flash devices, reducing any data loss concern. The broadcast write implementation provides the advantage of enhanced data protection options without sacrificing application performance or end-user responsiveness.

Note:

When using dual controllers, it is highly recommended that dual-ported hard drives be used for redundancy. If you use single-ported drives in a dual controller system and the connecting path is lost, the data on the drives would remain unaffected, but connection to the drives would be lost until the path to them is restored.

Single controller

A single-controller configuration provides no redundancy in the event that the controller fails; therefore, the single controller is a potential Single Point of Failure (SPOF). Multiple hosts can be supported in this configuration (up to two for direct attach). In this configuration, each host can have access to the storage resources. If the controller fails, the host loses access to the storage.

The single-controller configuration is less expensive than the dual-controller configuration. It is a suitable solution in cases where high availability is not required and loss of access to the data can be tolerated until failure recovery actions are complete. A single-controller configuration is also an appropriate choice in storage systems where redundancy is achieved at a higher level, such as a two-node cluster. For example, a two-node cluster where each node is attached to an MSA2000fc G2 enclosure with a single controller and the nodes do not depend upon shared storage. In this case, the failure of a controller is equivalent to the failure of the node to which it is attached.

Another suitable example of a high-availability storage system using a single controller configuration is where a host uses a volume manager to mirror the data on two independent single-controller MSA2000fc G2 storage systems. If one MSA2000fc G2 storage system fails, the other MSA2000fc G2 storage system can continue to serve the I/O operations. Once the failed controller is replaced, the data from the survivor can be used to rebuild the failed system.

Note:

When using a single-controller system, the controller must be installed in the slot A of the array.

Choosing DAS or SAN attach

There are two basic methods for connecting storage to data hosts: Direct Attached Storage (DAS) and Storage Area Network (SAN). The option you select depends on the number of hosts you plan to connect and how rapidly you need your storage solution to expand.

Direct attach

DAS uses a direct connection between a data host and its storage system. The DAS solution of connecting each data host to a dedicated storage system is straightforward and the absence of storage switches can reduce cost. Like a SAN, a DAS solution can also share a storage system, but it is limited by the number of ports on the storage system.

A powerful feature of the storage system is its ability to support four direct attach single-port data hosts, or two direct attach dual-port data hosts without requiring storage switches. The MSA2000fc G2 can also support 2 single-connected hosts and 1 dual connected host for a total of 3 hosts.

If the number of connected hosts is not going to change or increase beyond four then the DAS solution is appropriate. However, if the number of connected hosts is going to expand beyond the limit imposed by the use of DAS, it is best to implement a SAN.

Tip:

It is a best practice to use a dual-port connection to data hosts when implementing a DAS solution. This includes using dual-ported hard drives for redundancy.

Switch attach

A switch attach solution, or SAN, places a switch between the servers and storage systems. This strategy tends to use storage resources more effectively and is commonly referred to as storage consolidation. A SAN solution shares a storage system among multiple servers using switches and reduces the total number of storage systems required for a particular environment, at the cost of additional element management (switches), and path complexity.

Using switches increases the number of servers that can be connected. Essentially, the maximum number of data hosts that can be connected to the SAN becomes equal to the number of available switch ports.

Note:

The HP StorageWorks MSA2000fc G2 supports 64 hosts.

Tip:

It is a best practice to use a switched SAN environment anytime more than four hosts or when growth in required or storage or number of hosts is expected.

Dealing with controller failovers

Since the MSA2000fc G2 uses Unified LUN Presentation, all host ports see all LUNs; thus failovers are dealt with differently than with the MSA2000fc.

FC direct-attach configurations

In a dual-controller system, both controllers share a unique node WWN so they appear as a single device to hosts. The controllers also share one set of LUNs to use for mapping volumes to hosts.

A host can use any available data path to access a volume owned by either controller. The preferred path, which offers slightly better performance, is through target ports on a volume's owning controller.

Note:

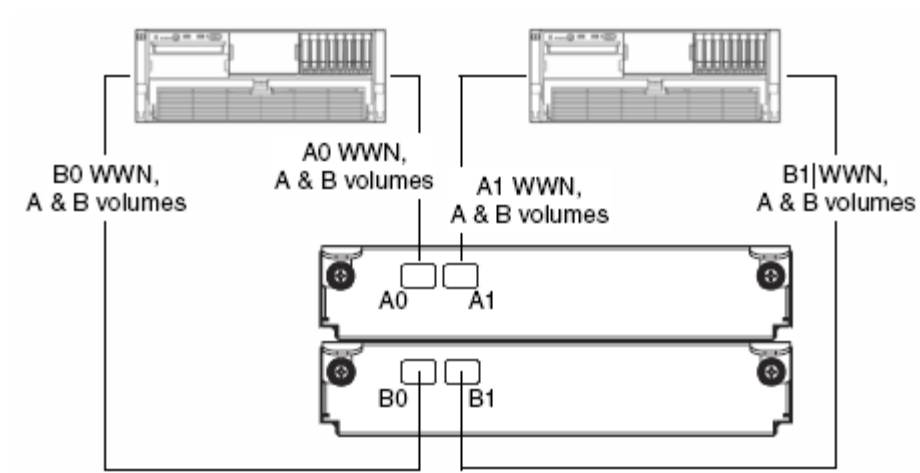
Ownership of volumes is not visible to hosts. However, in SMU you can view volume ownership and change the owner of a virtual disk and its volumes.

Note:

Changing the ownership of a virtual disk should never be done with I/O in progress. I/O should be quiesced prior to changing ownership.

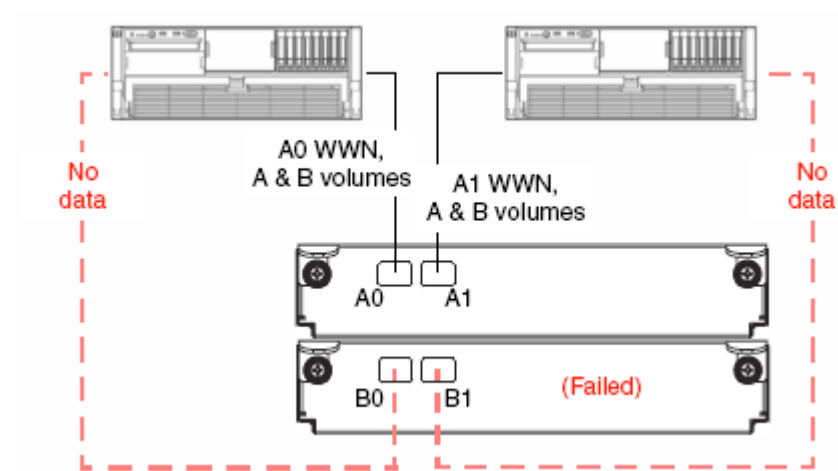
In the following configuration, both hosts have redundant connections to all mapped volumes.

Figure 13: FC storage presentation during normal operation (high-availability, dual-controller, and direct attach with two hosts)



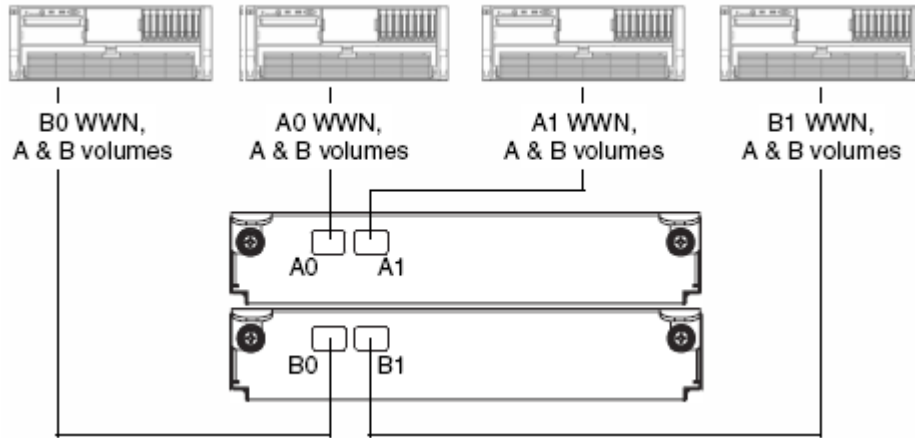
If a controller fails, the hosts maintain access to all of the volumes through the host ports on the surviving controller, as shown in the Figure 14.

Figure 14: FC storage presentation during failover (high-availability, dual-controller, and direct attach with two hosts)



In the following configuration, each host has a non-redundant connection to all mapped volumes. If a controller fails, the hosts connected to the surviving controller maintain access to all volumes owned by that controller. The hosts connected to the failed controller will lose access to volumes owned by the failed controller.

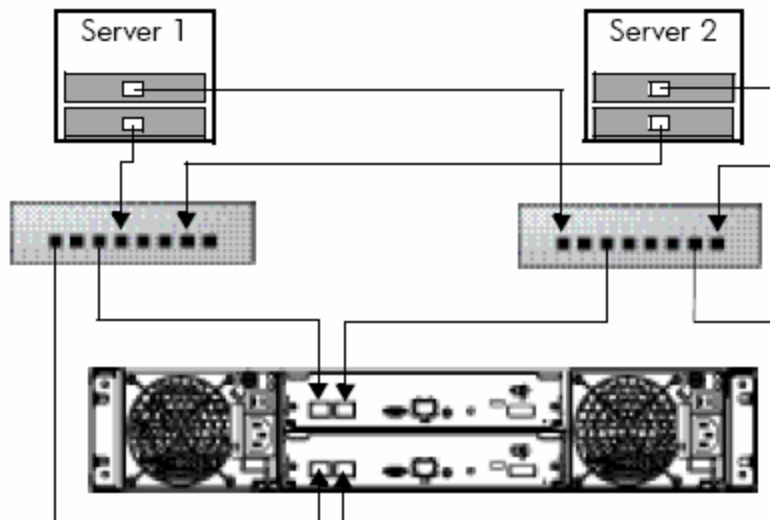
Figure 15: FC storage presentation during normal operation (High-availability, dual-controller, direct attach with four hosts)



FC switch-attach configuration

When using a switch configuration, it is important to have at least one port connected from each switch to each controller for redundancy. See Figure 16.

Figure 16: FC storage presentation during normal operation (high-availability, dual-controller, and switch attach with four hosts)



If controller B fails in this setup, the preferred path will shift to controller A and all volumes will be still accessible to both servers as in Figure 14. Each switch has a redundant connection to all mapped volumes; therefore, the hosts connected to the surviving controller maintain access to all volumes.

Virtual disks

A vdisk is a group of disk drives configured with a RAID level. Each virtual disk can be configured with a different RAID level. A virtual disk can contain SATA drives or SAS drives, but not both. The controller safeguards against improperly combining SAS and SATA drives in a virtual disk. The system displays an error message if you choose drives that are not of the same type.

The HP StorageWorks MSA2000 G2 system can have a maximum of 16 virtual disks per controller for a maximum of 32 virtual disks with a dual controller configuration.

For storage configurations with many drives, it is recommended to consider creating a few virtual disks each containing many drives, as opposed to many virtual disks each containing a few drives. Having many virtual disks is not very efficient in terms of drive usage when using RAID 3. For example, one 12-drive RAID-5 virtual disk has one parity drive and 11 data drives, whereas four 3-drive RAID-5 virtual disks each have one parity drive (four total) and two data drives (only eight total).

A virtual disk can be larger than 2 TB. This can increase the usable storage capacity of configurations by reducing the total number of parity disks required when using parity-protected RAID levels. However, this differs from using volumes larger than 2 TB, which requires specific operating system, HBA driver, and application-program support.

Note:

The MSA2000 G2 can support a maximum vdisk size of 16 TB.

Supporting large storage capacities requires advanced planning because it requires using large virtual disks with several volumes each or many virtual disks. To increase capacity and drive usage (but not performance), you can create virtual disks larger than 2 TB and divide them into multiple volumes with a capacity of 2 TB or less.

The largest supported vdisk is the number of drives allowed in a RAID set multiplied by the largest drive size.

- RAID 0, 3, 5, 6, 10 can support up to 16 drives (with 1 TB SATA drives that is 16 TB raw)
 - RAID 50 can support up to 32 drives (with 1 TB SATA drives that is 32 TB raw)
-

Tip:

The best practice for creating virtual disks is to add them evenly across both controllers. With at least one virtual disk assigned to each controller, both controllers are active. This active-active controller configuration allows maximum use of a dual-controller configuration's resources.

Tip:

Another best practice is to stripe virtual disks across shelf enclosures to enable data integrity in the event of an enclosure failure. A virtual disk created with RAID 1, 10, 3, 5, 50, or 6 can sustain an enclosure failure without loss of data depending on the number of shelf enclosures attached. The design should take into account whether spares are being used and whether the use of a spare can break the original design. A plan for evaluation and possible reconfiguration after a failure and recovery should be addressed. Non-fault tolerant vdisks do not need to be dealt with in this context because a shelf enclosure failure with any part of a non-fault tolerant vdisk can cause the vdisk to fail.

Chunk size

When you create a virtual disk, you can use the default chunk size or one that better suits your application. The chunk (also referred to as stripe unit) size is the amount of contiguous data that is written to a virtual disk member before moving to the next member of the virtual disk. This size is fixed throughout the life of the virtual disk and cannot be changed. A stripe is a set of stripe units that are written to the same logical locations on each drive in the virtual disk. The size of the stripe is determined by the number of drives in the virtual disk. The stripe size can be increased by adding one or more drives to the virtual disk.

Available chunk sizes include:

- 16 KB
- 32 KB
- 64 KB (default)

If the host is writing data in 16 KB transfers, for example, then that size would be a good choice for random transfers because one host read would generate the read of exactly one drive in the volume. That means if the requests are random-like, then the requests would be spread evenly over all of the drives, which is good for performance.

If you have 16-KB accesses from the host and a 64 KB block size, then some of the host's accesses would hit the same drive; each stripe unit contains four possible 16-KB groups of data that the host might want to read.

Alternatively, if the host accesses were 128 KB in size, then each host read would have to access two drives in the virtual disk. For random patterns, that ties up twice as many drives.

Tip:

The best practice for setting the chunk size is to match the transfer block size of the application

Vdisk initialization

During the creation of a vdisk, the manage user has the option to create a vdisk in online mode (default) or offline mode, only after the manage user has the advanced user type. By default, the manage user has the standard user type.

If the "online initialization" option is enabled, you can use the vdisk while it is initializing, but because the verify method is used to initialize the vdisk, initialization takes more time. Online initialization is fault tolerant.

If the "online initialization" option is unchecked ("offline initialization"), you must wait for initialization to complete before using the vdisk, but the initialization takes less time.

To assign the advanced user type to the manage user, log into the HP Storage Management Utility (SMU) and make sure the MSA23xx on the left frame is highlighted and then click the Configuration drop-down box. Then click Users → Modify User.

hp Storage Management Utility

System Status
System Time 2009-05-26 22:03:42
System Events 2 7 391

Configuration View
MSA2312fc
Logical
Vdisks
test disk (RAID10)
Hosts
Physical
Enclosure 1

MSA2312fc
View Provisioning Configuration Tools Wizards
MSA2312fc > View > Overview
System Overview
Select an entry from the table

System Overview

Health	Component	Capacity	Used Space
<input checked="" type="radio"/> OK	System	5.4TB	5.4TB
<input type="radio"/> OK	Enclosures	1	1.2TB / 4.2TB
<input type="radio"/> OK	Disks	12	1.2TB / 4.2TB
<input type="radio"/> OK	Vdisks	1	600.0GB
<input type="radio"/>	Volumes	2	75.0GB
<input type="radio"/>	Snap Pools	1	10.0GB
<input type="radio"/>	Snapshots	0	
<input type="radio"/>	Schedules	0	
<input type="radio"/>	Configuration Limits		
<input type="radio"/>	Licensed Features		
<input type="radio"/>	Versions		

System Information
Property Value

Click the radio button next to the manage user and type in the manage user password. From User Type, select "Advanced" and then to save the change, click the Modify User button → then OK.

MSA2312fc
View Provisioning Configuration Tools Wizards
MSA2312fc > Configuration > Users > Modify User
Modify User
Modify a system users settings

	Username	User Locale	WBI	CLI	FTP	Temperature Scale	Units	Base
<input type="radio"/>	monitor	English	x	x		Celsius	Auto	10
<input checked="" type="radio"/>	manage	English	x	x	x	Celsius	Auto	10
<input type="radio"/>	ftp	English			x	Celsius	Auto	10

User Name:*

Password:

Access Level:

User Type:
Standard
Advanced
Diagnostic

WBI Access:

CLI Access:

FTP Access:

Base Preference:

Precision Preference:

Unit Preference:

Temperature Preference:

Auto Sign Out:

Locale:

Volume mapping

It is a best practice to map volumes to the preferred path. The preferred path is both ports on the controller that owns the vdisk.

If a controller fails, the surviving controller will report it is now the preferred path for all vdisks. When the failed controller is back online, the vdisks and preferred paths switch back.

Best Practice

For fault tolerance, HP recommends mapping the volumes to all available ports on the controller.

For performance, HP recommends mapping the volumes to the ports on the controller that owns the vdisk. Mapping to the non-preferred path results in a slight performance degradation.

Note:

By default, a new volume will have the "all other hosts read-write access" mapping, so the manage user must go in and explicitly assign the correct volume mapping access.

Configuring background scrub

By default, the system background scrub or the MSA2000 G2 is enabled. However, you can disable the background scrub if desired. The background scrub continuously analyzes disks in vdisks to detect, report, and store information about disk defects.

Vdisk-level errors reported include:

Hard errors, medium errors, and bad block replacements (BBRs).

Disk-level errors reported include:

Metadata read errors, SMART events during scrub, bad blocks during scrub, and new disk defects during scrub.

For RAID 3, 5, 6, and 50, the utility checks all parity blocks to find data-parity mismatches. For RAID 1 and 10, the utility compares the primary and secondary disks to find mirror-verify errors. For NRAID and RAID 0, the utility checks for media errors.

You can use a vdisk while it is being scrubbed. Background scrub always runs at background utility priority, which reduces to no activity if CPU usage is above a certain percentage or if I/O is occurring on the vdisk being scrubbed. A background scrub may be in process on multiple vdisks at once. A new vdisk will first be scrubbed 20 minutes after creation. After a vdisk has been scrubbed, it will not be scrubbed again for 24 hours. When a scrub is complete, the number of errors found is reported with event code 207 in the event log.

Note:

If you choose to disable background scrub, you can still scrub selected vdisks by using Media Scrub Vdisk.

To change the background scrub setting:

In the Configuration View panel, right-click the system and select Configuration → Advanced Settings → System Utilities.

Either select (enable) or clear (disable) the Background Scrub option. The default is enabled. Click Apply.

Best Practice: Leave the default setting of Background Scrub ON in the background priority.

RAID levels

Choosing the correct RAID level is important whether your configuration is for fault tolerance or performance. Table 3 gives an overview of supported RAID implementations highlighting performance and protection levels.

Note:

Non-RAID is supported for use when the data redundancy or performance benefits of RAID are not needed; no fault tolerance.

Table 3: An overview of supported RAID implementations

RAID level	Cost	Performance	Protection level
RAID 0 Striping	N/A	Highest	No data protection
RAID 1 Mirroring	High cost— 2x drives	High	Protects against individual drive failure
RAID 3 Block striping with dedicated parity drive	1 drive	Good	Protects against individual drive failure
RAID 5 Block striping with striped parity drive	1 drive	Good	Protects against any individual drive failure; medium level of fault tolerance
RAID 6 Block striping with multiple striped parity	2 drives	Good	Protects against multiple (2) drive failures; high level of fault tolerance
RAID 10 Mirrored striped array	High cost 2x drives	High	Protects against certain multiple drive failures; high level of fault tolerance
RAID 50 Data striped across RAID 5	At least 2 drives	Good	Protects against certain multiple drive failures; high level of fault tolerance

Spares

You can designate a maximum of eight global spares for the system. If a disk in any redundant vdisk (RAID 1, 3, 5, 6, 10, and 50) fails, a global spare is automatically used to reconstruct the vdisk.

At least one vdisk must exist before you can add a global spare. A spare must have sufficient capacity to replace the smallest disk in an existing vdisk. If a drive in the virtual disk fails, the controller automatically uses the vdisk spare for reconstruction of the critical virtual disk to which it belongs. A spare drive must be the same type (SAS or SATA) as other drives in the virtual disk. You cannot add a spare that has insufficient capacity to replace the largest drive in the virtual disk. If two drives fail in a RAID 6 virtual disk, two properly sized spare drives must be available before reconstruction can begin. For RAID 50 virtual disks, if more than one sub-disk becomes critical, reconstruction and use of vdisk spares occur in the order sub-vdisks are numbered.

You can designate a global spare to replace a failed drive in any virtual disk, or a vdisk spare to replace a failed drive in only a specific virtual disk. Alternatively, you can enable dynamic spares in HP SMU. Dynamic sparing enables the system to use any drive that is not part of a virtual disk to replace a failed drive in any virtual disk.

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 virtual disk. At this point, the original drive slot position that failed is no longer part of the virtual disk. The original drive now becomes a “Leftover” drive.

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

1. Replace the failed drive with a new drive.
2. If the drive slot is still marked as “Leftover”, use the “Clear Disk Metadata” option found in the “Tools” submenu.
3. When the new drive is online and marked as “Available”, configure the drive as a global spare drive.
4. 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.
5. Replace the drive you manually removed from the enclosure.
6. If the drive is marked as “Leftover”, clear the metadata as in step 2 above.
7. Re-configure the drive as the new global spare.

Tip:

A best practice is to designate a spare disk drive for use if a drive fails. Although using a dedicated vdisk spare is the best way to provide spares for your virtual disks, it is also expensive to keep a spare assigned to each virtual disk. An alternative method is to enable dynamic spares or to assign one or more unused drives as global spares.

Cache configuration

Controller cache options can be set for individual volumes to improve a volume’s fault tolerance and I/O performance.

Note:

To change the following cache settings, the user—who logs into the HP SMU—must have the “advanced” user credential. The manage user has the “standard” user credential by default. This credential can be changed using the HP SMU and click on “Configuration,” then “Users,” then “Modify Users.”

Write-back cache settings

Write back is a cache-writing strategy in which the controller receives the data to be written to disk, 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 drive. Write-back cache mirrors all of the data from one controller module cache to the other. Write-back cache improves the performance of write operations and the throughput of the controller.

When write-back cache is disabled, write-through becomes the cache-writing strategy. Using write-through cache, the controller writes the data to the disk before signaling the host operating system that the process is complete. Write-through cache has lower throughput and write operation performance than write back, but it is the safer strategy, with low risk of data loss on power failure.

However, write-through cache does not mirror the write data because the data is written to the disk before posting command completion and mirroring is not required. You can set conditions that cause the controller to switch from write-back caching to write-through caching as described in “Auto-Write Through Trigger and Behavior Settings” later in this paper.

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

You can enable and disable the write-back cache for each volume. By default, volume write-back cache is enabled. Data is not lost if the system loses power because controller cache is backed by super capacitor technology. For most applications this is the correct setting, but because backend bandwidth is used to mirror cache, if you are writing large chunks of sequential data (as would be done in video editing, telemetry acquisition, or data logging) write-through cache has much better performance. Therefore, you might want to experiment with disabling the write-back cache. You might see large performance gains (as much as 70 percent) if you are writing data under the following circumstances:

- Sequential writes
- Large I/Os in relation to the chunk size
- Deep queue depth

If you are doing any type of random access to this volume, leave the write-back cache enabled.

Caution: Write-back cache should only be disabled if you fully understand how your operating system, application, and HBA (SAS) move data. You might hinder your storage system’s performance if used incorrectly.

Auto-write through trigger and behavior settings

You can set the trigger conditions that cause the controller to change the cache policy from write-back to write-through. While in write-through mode, system performance might be decreased.

A default setting makes the system revert to write-back mode when the trigger condition clears. To make sure that this occurs and that the system doesn’t operate in write-through mode longer than necessary, make sure you check the setting in HP SMU or the CLI.

You can specify actions for the system to take when write-through caching is triggered:

- **Revert when Trigger Condition Clears:** Switches back to write-back caching after the trigger condition is cleared. The default and best practice is Enabled.
- **Notify Other Controller:** In a dual-controller configuration, the partner controller is notified that the trigger condition is met. The default is Disabled.

Cache configuration summary

The following guidelines list the general best practices. When configuring cache:

- For a fault-tolerant configuration, use the write-back cache policy, instead of the write-through cache policy
- For applications that access both sequential and random data, use the standard optimization mode, which sets the cache block size to 32 KB. For example, use this mode for transaction-based and database update applications that write small files in random order
- For applications that access sequential data only and that require extremely low latency, use the super-sequential optimization mode, which sets the cache block size to 128 KB. For example, use this mode for video playback and multimedia post-production video- and audio-editing applications that read and write large files in sequential order

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 the following table. This section provides a basic starting point for fine-tuning your system, which should be done during performance baseline modeling.

Table 4: Optimizing performance for your application

Application	RAID level	Read ahead cache size	Cache optimization
Default	5 or 6	Default	Standard
HPC (High-Performance Computing)	5 or 6	Maximum	Standard
MailSpooling	1	Default	Standard
NFS_Mirror	1	Default	Standard
Oracle_DSS	5 or 6	Maximum	Standard
Oracle_OLTP	5 or 6	Maximum	Standard
Oracle_OLTP_HA	10	Maximum	Standard
Random1	1	Default	Standard
Random5	5 or 6	Default	Standard
Sequential	5 or 6	Maximum	Super-Sequential
Sybase_DSS	5 or 6	Maximum	Standard
Sybase_OLTP	5 or 6	Maximum	Standard
Sybase_OLTP_HA	10	Maximum	Standard
Video Streaming	1 or 5 or 6	Maximum	Super-Sequential
Exchange Database	5 for data; 10 for logs	Default	Standard
SAP	10	Default	Standard
SQL	5 for data; 10 for logs	Default	Standard

Note:

For Microsoft SQL 2008 and the MSA 2000, the recommended configuration is to assign two or more data virtual disks to one controller and two or more log virtual disks to the other controller.

Review the document entitled “SQL Server 2008 best practices for consolidation of multiple databases in an online transaction processing (OLTP) environment with HP MSA2000 storage” found at <http://h71019.www7.hp.com/ActiveAnswers/us/en/aa-categories.html>

For Microsoft Exchange Server and the MSA2000, the recommended configuration is to isolate exchange databases and their associated log workloads onto separate array virtual disks.

For further Best Practices on the Microsoft Exchange Server and MSA2000, search Active Answers at <http://h71019.www7.hp.com/ActiveAnswers/us/en/aa-categories.html>

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 for 4 Gb/sec on the MSA2000fc G2.

- Host ports should be configured for 1 Gb/sec on the MSA2000i G2.
- Host ports should be configured for 3 Gb/sec on the MSA2000sa G2.
- Virtual disks should be balanced between the two controllers.
- Disk drives should be balanced between the two controllers.
- Cache settings should be set to match [Table 4](#) (Optimizing performance for your application) for the application.

Highest fault tolerance optimization

The following guidelines list the general best practices to follow when configuring your storage system for highest fault tolerance:

- Use dual controllers
- Use two cable connections from each host
- Use Multipath Input/Output (MPIO) software

Boot from storage considerations

When booting from SAN, construct a separate virtual disk and volume that will be used only for the boot from SAN. Do not keep data and boot from SAN volumes on the same vdisk. This can help with performance. If there is a lot of I/O going to the data volume on a vdisk that shares a boot from SAN volume, there can be a performance drop in the I/O to the Operating System drives.

MSA70 considerations

Dual-domains

When using the MSA70 with dual-domains, dual I/O modules, make sure the following procedure is followed.

MSA70 systems with firmware earlier than 1.50:

If your MSA70 has installed firmware earlier than version 1.50, you must replace the chassis backplane before installing a second I/O module in the chassis. To determine your installed firmware version, use a server-based tool such as HP Systems Insight Manager or your Management Agents.

If installed firmware is earlier than 1.50, do the following:

1. Contact HP Support and order a replacement backplane:
MSA70: 430149-001

Caution:

Be sure to order the part number indicated in this notice, not the spare part number printed on your existing backplanes.

Be sure to order a quantity of two replacement kits.

2. Install the replacement backplane using instructions shipped with the backplane.
3. Install the additional I/O module using instructions shipped with the I/O module.

Firmware versions

If there are MSA70 enclosures connected to the MSA2000fc G2 (24 bay model only), make sure that the firmware on the enclosure is 2.18 or greater. If the MSA70 has a firmware version prior to 2.18, the MSA70 will be in a degraded state and virtual disks cannot be created or accessed from the MSA70.

Administering with HP SMU

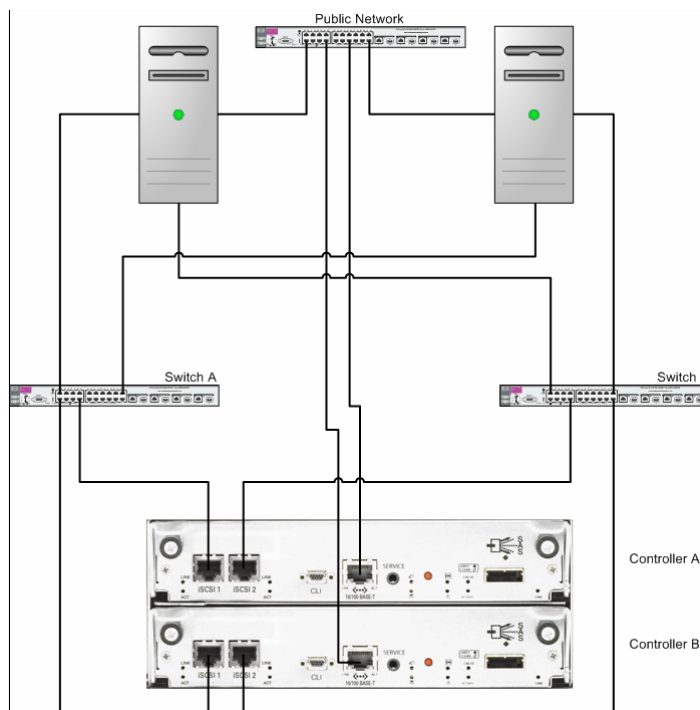
If you choose to use the HP StorageWorks Management Utility (SMU) for administration, it is best to use either the Firefox 3.0 or later or Internet Explorer 7 or later Web browsers.

MSA2000i G2 Considerations

When using the MSA2000i G2, 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 makes sure that the storage network is isolated from the other networks.

The private LAN is the network that goes from the server to the MSA2000i G2. This is the storage network. The storage network should be isolated from the Public network to improve performance. See Figure 17.

Figure 17: MSA2000i G2 Network



IP Address scheme for the controller pair

The MSA2000i G2 uses port 0 of each controller as one failover pair, and port 1 of each controller as a second failover pair. Therefore, port 0 of each controller must be in the same subnet, and port 1 of each controller should be in a second subnet.

For example (with a netmask of 255.0.0.0):

- Controller A port 0: 10.10.10.100
- Controller A port 1: 10.99.10.120
- Controller B port 0: 10.10.10.110
- Controller B port 1: 10.99.10.130

The P2000 G3 MSA

Topics covered

This section examines the following:

- What's new
- Hardware Overview
- P2000 G3 MSA Settings
- Disk Scrubbing, SMART, and Drive Spin Down
- Cascading Array Enclosures
- Software
- 8 Gb Switches and SFP transceivers

What's New in the P2000 G3 MSA

- Four port 6 Gb SAS controller or
- Two new 8 Gb Fibre Channel controllers with 2 GB cache memory each:
 - Standard model with two 8 Gb FC host ports each
 - Combo model with two 8 Gb FC host ports and two 1 GbE iSCSI ports each
- Increased support to seven P2000 LFF disk enclosures (96 LFF drives)
- Increased support to five D2700 SFF disk enclosures (149 SFF drives)
- 6 Gb SAS back end and HDD support
- 64 Snaps and clone capability come standard on G3 models
- 512 snapshots max (double the MSA2000 G2)
- Optional controller-based replication (Remote Snap)
- 512 Max LUN support
- Higher performance with upgraded processing power with increased I/O performance
- Improved System Management Utility (SMU) user interface
- Full support for G1/G2 to G3 upgrade, including cross-protocol upgrades

Hardware overview

HP StorageWorks P2000 G3 MSA Modular Smart Array

HP StorageWorks 2000 Family of storage arrays features P2000 G3 MSA arrays with the latest 8 Gb Fibre Channel and 6 Gb SAS connected models. The arrays are designed for entry-level customers and feature the latest in functionality and host-connect technology while offering excellent price/performance. The P2000 G3 MSA is ideal for companies with small budgets or limited IT expertise, and also for larger companies with departmental or remote requirements. Each solution is designed to be easy to deploy, to be secure, along with low management costs, while driving rapid return on investment.

The P2000 G3 FC is an 8 Gb Fibre Channel, while the P2000 G3 SAS is a 6 Gb SAS connected 2U storage area network (SAN) or direct connect solution (OS dependent) designed for small to medium-sized departments or remote locations.

The P2000 G3 MSA offers a choice of three controllers:

- A high-performance, Fibre Channel 8 Gb dual port model.
- A unique dual-purpose Combo controller with two 8 Gb Fibre Channel ports plus two 1 GbE iSCSI ports.
- A 6 Gb SAS controller with four ports per controller.

The controller-less chassis is offered in two models—one comes standard with twelve Large Form Factor (LFF) 3.5-inch drive bays, the other can accommodate twenty-four Small Form Factor (SFF) 2.5-inch drives (common with ProLiant). Both are able to simultaneously support enterprise-class SAS drives and archival-class SATA Midline drives. The SFF chassis also supports SAS Midline drives. Either chassis can have one or two P2000 G3 controllers.

The P2000 G3 gives great flexibility to choose the proper configuration to fit individual needs. The user can opt for the 24 drive bay SFF chassis for the highest spindle counts in the most dense form factor, or choose the 12 drive bay LFF model for maximum total capacity. In addition, LFF and SFF Disk Enclosures may be mixed. Choose a single controller unit for low initial cost with the ability to upgrade later; or decide on a model with dual controllers for the most demanding entry-level situations. Capacity can easily be added when needed by attaching additional drive enclosures. Maximum capacity ranges with LFF drives up to 57.6 TB SAS or 192 TB SATA with the addition of the maximum number of drive enclosures. Configurations utilizing the SFF drive chassis and the maximum number of drive enclosures can grow to 44.7 TB of SAS or 74.5 TB of SAS Midline or SATA Midline with a total of 149 SFF drives. The P2000 G3 FC has been fully tested up to 64 hosts.

The P2000 G3 SAS is the follow-on product to the MSA2000sa G2, adding the latest 6 Gb SAS technology to the four host ports per controller. It also features the same new scalability as the P2000 G3 FC model, and offers 6 Gb back-end transmission speed to drives and JBODs. The new G3 SAS array is designed for directly attaching up to four dual-path or eight single path servers. SAS array support for Blade Systems will continue to come from the MSA2000sa G2.

Note:

Larger hard drives are always in test; refer to <http://www.hp.com/go/p2000> to get the latest hard drive capacity limits.

The new generation G3 models have grown in other aspects too. 64 snapshot and clone capability come standard and optional snapshot capability has been increased to 512 snaps on the G3 models, optional Remote Snap capability has been added, while LUN support remains at up to 512 total volumes in a Dual Controller system. In addition to support for Windows and Linux on x86 and x64 server platforms, the P2000 G3 continues support for HP-UX, OpenVMS, Windows, and Linux on powerful Integrity servers.

P2000 G3 MSA Settings

The P2000 G3 MSA has the same capabilities as the MSA2000 G2 model.

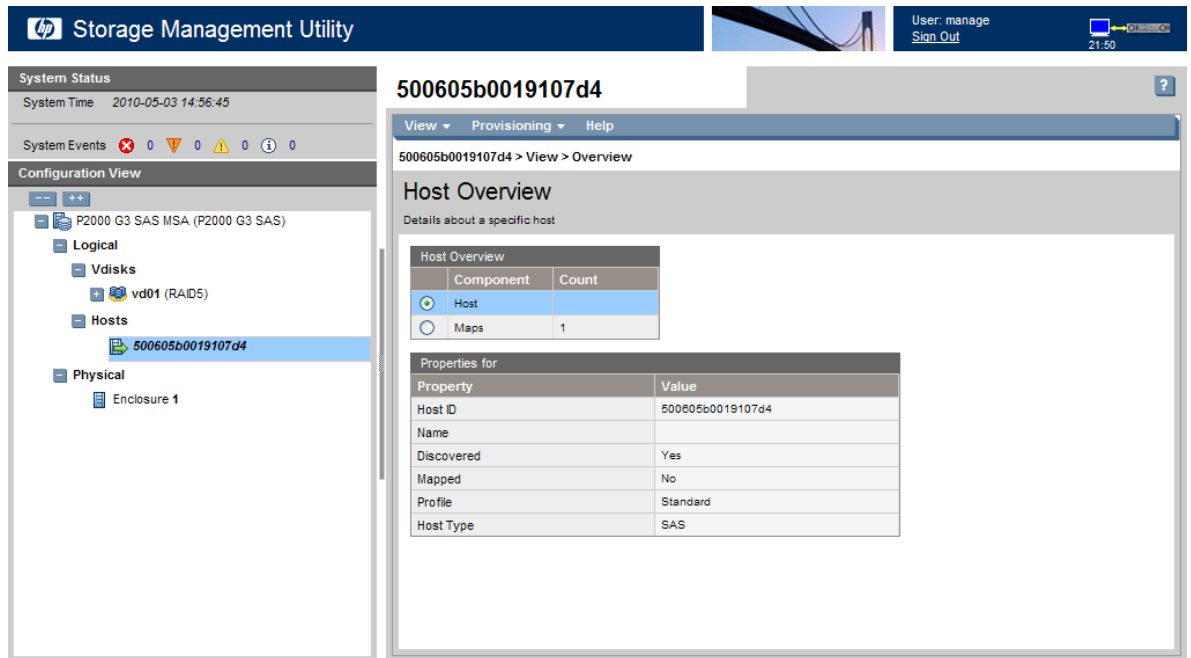
Please see: [MSA2000 G2 ULP section](#) through “[Cache configuration summary](#)” for list of these capabilities.

World Wide Name (WWN) naming conventions

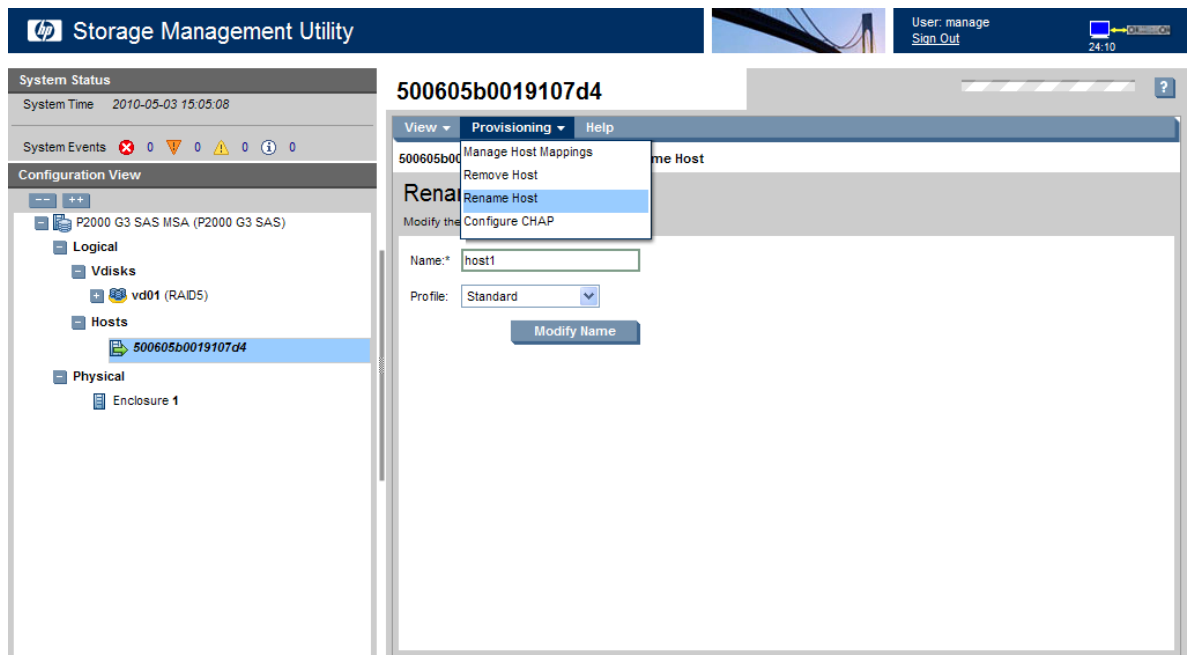
A best practice for acquiring and renaming World Wide Names (WWN) for the P2000 G3 SAS MSA is to plug-in one SAS cable connection at a time and then rename the WWN to an identifiable name.

Procedure:

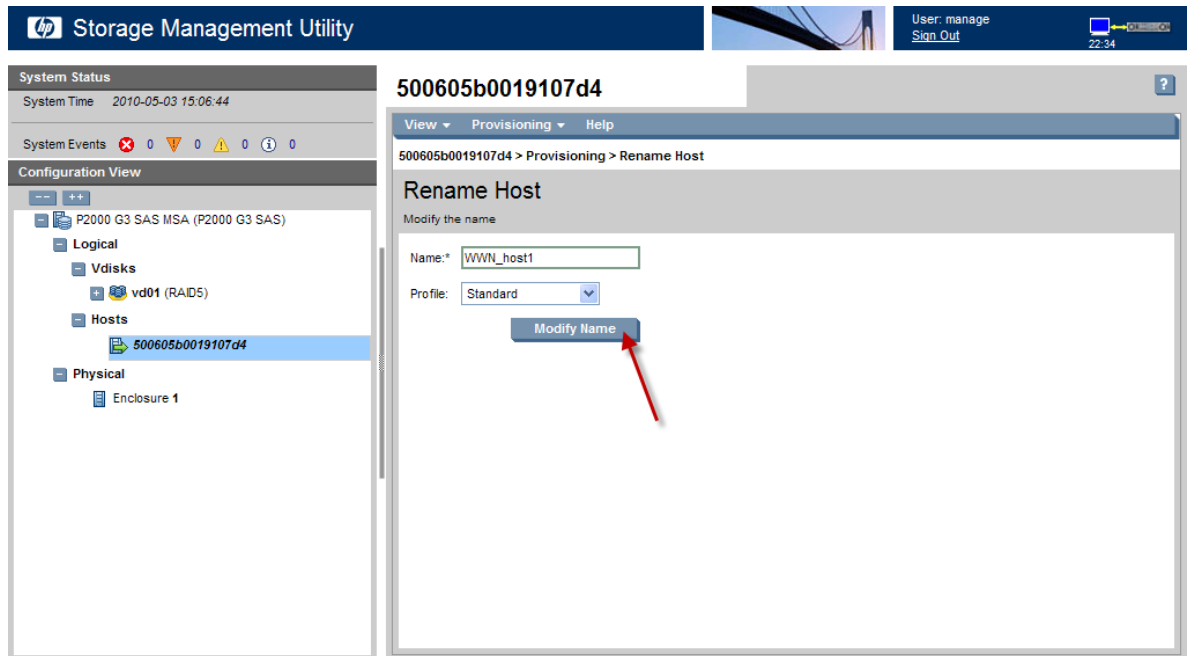
1. Login to the Storage Management Utility (SMU). The Status Summary page will be displayed.
2. Click “+” next to “Hosts” from the left Windows frame. This will expand the list to show all connected hosts.
3. Highlight the host in the list that you want to rename by clicking the WWN name.



4. On the right window frame, click Provisioning -> Rename Host.



5. Type in the host nickname and choose the Profile and then click Modify Name. Click OK on the pop-up window.



6. Plug in the SAS port of the HBA on the second server into the P2000 G3 SAS MSA controller port. Make sure the server is powered on.
7. Repeat steps 3–5 for the remaining servers.

Fastest throughput optimization

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

- Fibre Channel host ports should be configured for 8 Gb/sec on the P2000 G3 FC MSA.
- Virtual disks should be balanced between the two controllers.
- Disk drives should be balanced between the two controllers.
- Cache settings should be set to match [Table 4](#) for the MSA2000 G2 (Optimizing performance for your application) for the application.

Highest fault tolerance optimization

The following guidelines list the general best practices to follow when configuring your storage system for highest fault tolerance:

- Use dual controllers
- Use two cable connections from each host
- Use Multipath Input/Output (MPIO) software

Disk Background Scrub, Drive Spin Down, and SMART

The P2000 G3 MSA also uses the disk background scrubbing feature. You can scrub disk drives that are in a vdisk or have not yet been assigned to a vdisk whether or not they are in a vdisk.

The P2000 G3 MSA has now added the power saving feature called Drive Spin Down (DSD). The drive spin down is a cost and power saving tool. The drive spin down feature will stop virtual disks, available disk drives and global spares disk from spinning.

Self-Monitoring Analysis and Reporting Technology (SMART) technology can alert the controller of impending disk failure. When SMART is enabled, the system checks for SMART events one minute after a restart and every five minutes thereafter. SMART events are recorded in the event log.

Configuring background scrub for vdisks

You can enable or disable whether the system continuously analyzes disks in vdisks to detect, report, and store information about disk defects.

Vdisk-level errors reported include:

Hard errors, medium errors, and bad block replacements (BBRs).

Disk-level errors reported include:

Metadata read errors, SMART events during scrub, bad blocks during scrub, and new disk defects during scrub.

For RAID 3, 5, 6, and 50, the utility checks all parity blocks to find data-parity mismatches.

For RAID 1 and 10, the utility compares the primary and secondary disks to find data inconsistencies.

For NRAID (Non-RAID, non-striped) and RAID 0, the utility checks for media errors.

You can use a vdisk while it is being scrubbed. Background vdisk scrub runs at background utility priority, which reduces to no activity if CPU usage is above a certain percentage or if I/O is occurring on the vdisk being scrubbed. A vdisk scrub may be in process on multiple vdisks at once. A new vdisk will first be scrubbed 20 minutes after creation. After a vdisk is scrubbed, scrub will start again after the interval specified by the Vdisk Scrub Interval (hours) option. When a scrub is complete, the number of errors found is reported with event code 207 in the event log.

Note:

If you choose to disable background vdisk scrub, you can still scrub a selected vdisk by using Media Scrub Vdisk.

To configure background scrub for vdisks

1. In the Configuration View panel, right-click the system and select Configuration > Advanced Settings > System Utilities.
2. Set the options:
 - Either select (enable) or clear (disable) the Vdisk Scrub option.
 - Set the Vdisk Scrub Interval (hours), which is the interval between background vdisk scrub finishing and starting again, from 1–360 hours; the default is 24 hours.
3. Click Apply.

Configuring background scrub for disks not in vdisks

You can enable or disable whether the system continuously analyzes disks that are not in vdisks to detect, report, and store information about disk defects.

Errors reported include:

Metadata read errors, SMART events during scrub, bad blocks during scrub, and new disk defects during scrub.

The interval between background disk scrub finishing and starting again is 24 hours.

To configure background scrub for disks not in vdisks

1. In the Configuration View panel, right-click the system and select Configuration > Advanced Settings > System Utilities.
2. Either select (enable) or clear (disable) the Disk Scrub option.
3. Click Apply.

Configuring utility priority

You can change the priority at which the Verify, Reconstruct, Expand, and Initialize utilities run when there are active I/O operations competing for the system's controllers.

To change the utility priority

1. In the Configuration View panel, right-click the system and select Configuration > Advanced Settings > System Utilities.
2. Set Utility Priority to either:
 - **High:** Use when your highest priority is to get the system back to a fully fault-tolerant state. This causes heavy I/O with the host to be slower than normal. This value is the default.
 - **Medium:** Use when you want to balance data streaming with data redundancy.
 - **Low:** Use when streaming data without interruption, such as for a web server, is more important than data redundancy. This enables a utility such as Reconstruct to run at a slower rate with minimal effect on host I/O.
 - **Background:** Utilities run only when the processor has idle cycles.
3. Click Apply.

Best Practice: Leave the default setting of Background Scrub ON in the background priority.

Scheduling drive spin down for all vdisks

For all vdisks that are configured to use drive spin down (DSD), you can configure times to suspend and resume DSD so that vdisks remain spun-up during hours of frequent activity.

To configure DSD for a virtual disk

1. In the Configuration View panel, right-click a vdisk and select Configuration → Configure Vdisk Drive Spin Down.
2. Set the options:
 - Either select (enable) or clear (disable) the Enable Drive Spin Down option.
 - Set the Drive Spin Down Delay (minutes), which is the period of inactivity after which the vdisk's disks and dedicated spares automatically spin down, from 1–360 minutes. If DSD is enabled and no delay value is set, the default is 15 minutes. A value of 0 disables DSD.
3. Click Apply. When processing is complete a success dialog appears.
4. Click OK.

To configure DSD for available disks and global spares

1. In the Configuration View panel, right-click the local system and select Configuration → Advanced Settings → Disk.
2. Set the options:
 - Either select (enable) or clear (disable) the Available and Spare Drive Spin Down Capability option. If you are enabling DSD, a warning prompt appears; to use DSD, click Yes; to leave DSD disabled, click No.
 - Set the Drive Spin Down Delay (minutes), which is the period of inactivity after which available disks and global spares automatically spin down, from 1–360 minutes. If DSD is enabled and no delay value is set, the default is 15 minutes. The value 0 disables DSD.
3. Click Apply. When processing is complete a success dialog appears.
4. Click OK.

Note:

DSD affects disk operations as follows:

- Spun-down disks are not polled for SMART events.
 - Operations requiring access to disks may be delayed while the disks are spinning back up.
-

To change the SMART setting

1. In the Configuration View panel, right-click the system and select Configuration > Advanced Settings > Disk.
2. Set SMART Configuration to either:
 - **Don't Modify:** Allows current disks to retain their individual SMART settings and does not change the setting for new disks added to the system.
 - **Enabled:** Enables SMART for all current disks after the next rescan and automatically enables SMART for new disks added to the system. This option is the default.
 - **Disabled:** Disables SMART for all current disks after the next rescan and automatically disables SMART for new disks added to the system.
3. Click Apply.

Best Practice: HP recommends using the default value “Enabled.”

Cascading Array Enclosures

Since the P2000 G3 MSA can be upgraded from any MSA2000 G1/G2 model, the existing MSA70s and MSA2000 expansion array enclosures might hold data that you need. The MSA70 and MSA2000 expansion enclosures operate at 3 Gb, see the Figures 18 and 19.

Figure 18: Mixed 3 Gb and 6 Gb JBODs behind a P2000 G3 FC MSA array

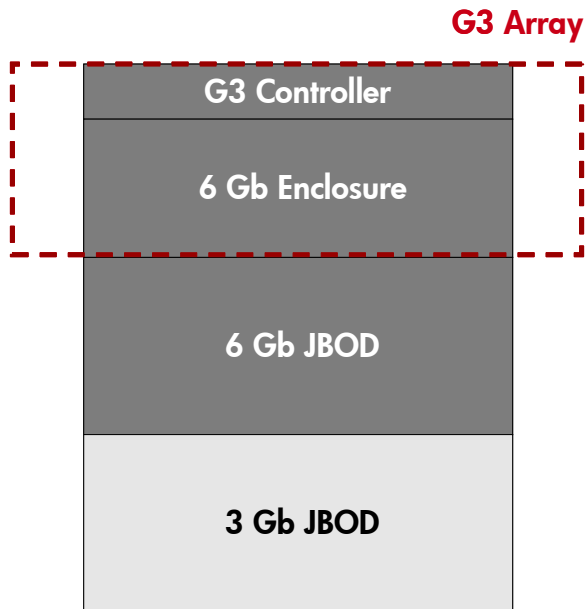
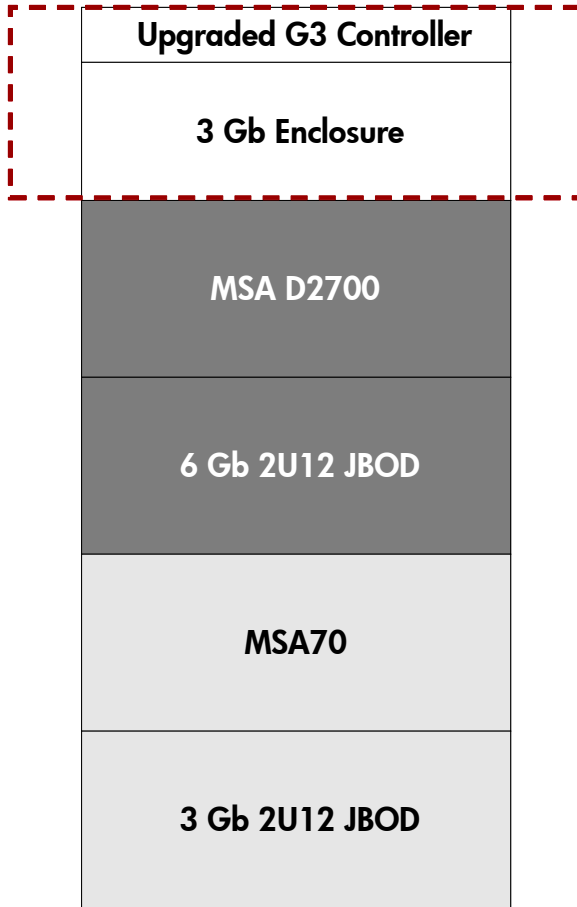


Figure 19: Mixed 3 Gb and 6 Gb JBODs behind an upgraded P2000 G3 FC MSA array



Note:

The 6 Gb MSA D2700 must come before the new 6 Gb MSA2000 2U12 Expansion Enclosure in the cascade chain.

If MSA2000 2U12 Expansion Enclosures are used in conjunction with D2700's, the MSA2000 2U12 Expansion Enclosures **MUST** come at the end of the cascade chain, and **ONLY** straight through cabling is allowed. Otherwise, either straight through cabling or reverse cabling is allowed. Note that access to 6 Gb enclosures following a 3 Gb enclosure is restricted to 3 Gb; therefore, if reverse cabling is used, to avoid the drop, place the 3 Gb enclosure in the middle of the cascade chain, and arrange Virtual Disks such that they do not span 6 Gb enclosures at the beginning and end of the cascade chain, and are owned by the controller closest to them in the cascade chain.

Table 5: SAS rates of supported enclosures for the P2000 G3 MSA

HP StorageWorks MSA System Model No.	Disk Form	Disk Quantity	SAS Rate
P2000 G3 MSA SAS SFF (controller enclosure)	2.5"	24-drive	6 Gb
P2000 G3 MSA SAS LFF (controller enclosure)	3.5"	12-drive	6 Gb
P2000 G3 MSA FC/iSCSI SFF (controller enclosure)	2.5"	24-drive	6 Gb
P2000 G3 MSA FC/iSCSI LFF (controller enclosure)	3.5"	12-drive	6 Gb
P2000 G3 MSA FC SFF (controller enclosure)	2.5"	24-drive	6 Gb
P2000 G3 MSA FC LFF (controller enclosure)	3.5"	12-drive	6 Gb
P2000 G3 6 Gb 3.5" 12-drive enclosure	3.5"	12-drive	6 Gb
D2700 6 Gb drive enclosure	2.5"	25-drive	6 Gb
MSA2000 3.5" 12-drive enclosure	3.5"	12-drive	3 Gb
MSA70 drive enclosure	2.5"	25-drive	3 Gb

For the P2000 G3 MSA cabling, consult the document titled "HP StorageWorks P2000 G3 MSA System Cable Configuration Guide" found at <http://www.hp.com/go/p2000>.

8 Gb Switches and SFP transceivers

The 8 Gb switches that HP offers use differing models of Small Form-Factor Pluggable (SFP) transceivers. The correct SFPs must be loaded into the correct supported 8 Gb switches when connecting to the P2000 G3 FC MSA. If the wrong SFPs are used in an unsupported 8 Gb switch, the storage on the P2000 G3 FC MSA will not be visible.

Here is a list of 2 SFPs and 1 supported switch for each:

- SFP part number AJ718A will work with the HP StorageWorks 8/20q Fibre Channel Switch (HP P\N: AM868A Revision: 0A) (20 ports)
- SFP part number AJ716A will work with the Brocade 8 Gb SAN Switch (HP P/N: AK242-63001 Revision: 0C) (24 ports)

Software

The section below introduces the HP StorageWorks P2000 Modular Smart Array Software Support/Documentation CD.

Versions

3.10—found in the shipping software kit

3.15—web launch version

Description

The HP StorageWorks P2000 Modular Smart Array Software Support/Documentation CD provides deployment and maintenance support for the HP StorageWorks P2000 G3 Modular Smart Array Family products which includes the P2000 G3 FC MSA, the P2000 G3 FC/iSCSI MSA, and the P2000 G3 SAS MSA.

HP StorageWorks P2000 Modular Smart Array Software Support/Documentation CD

HP StorageWorks P2000 Modular Smart Array Software Support/Documentation CD contains a common Windows/Linux navigation HTML framework to provide customers with a common installation experience.

This CD contains end user documents, host server software deliverables, and deployment and installation tools to simplify the setup and maintenance of your HP StorageWorks P2000 G3 Modular Smart Array Family product. The CD contains tabulated groups for Documents, Software, Firmware, Setup, Tools, and Service. This CD also contains the latest software drivers and user documents along with search links to secure the latest version from HP.com.

Here are some of the significant features of the HP StorageWorks P2000 Modular Smart Array Software Support/Documentation CD:

1. Provides step-by-step install instructions for each supported product with links to user documents included on the CD.
2. Contains Host software and drivers in various forms. Available only for the web launch CD (3.15)
 - a. OS specific Host Software bundles
 - b. Additional packages (Microsoft hot fixes and other drivers) that cannot be installed through the bundles
3. Contains the listing of all current P2000 G3 Modular Smart Array Family Firmware.
 - a. Provides search links to get the latest firmware from HP.com.
4. Contains all the P2000 G3 Modular Smart Array user documents in one place.
 - a. User documents included on the CD with Internationalized versions where available.
 - b. Links to additional documentation resources like White Papers are also available on the CD.
5. Provides additional tools that assist customers with various management tasks.
 - a. MSA Device Discovery Tool
 - i. Assists in discovering HP StorageWorks 2000 Modular Smart Array Family (MSA2000) and HP StorageWorks P2000 Modular Smart Array Systems that are direct attach or reachable over the network.
 - ii. Allows users to launch any of the management interfaces like SMU/Telnet/CLI/FTP for the selected device.
 - Provides an option to schedule log collection from the selected MSA device to pull the storage debug logs onto the host server at specified intervals.
 - Additionally, the MSA Device Discovery Tool can generate XML/Text output reports with inventory details of the local host and the discovered HP StorageWorks 2000 Modular Smart Array Family (MSA2000) and HP StorageWorks P2000 Modular Smart Array devices.
 - b. SNMP MIBs—The MSA2000 SNMP MIBs provide MIB browsers and other SNMP-aware software with the necessary information to query, update, and properly display SNMP variables on supported hardware.
 - c. Links to HP SMI-S documentation on the web
6. Provides various HP Support links for services like Product registration, Warranty, Service Care Packs, Learning Center, and so on.

Host Server Software

HP StorageWorks P2000 Modular Smart Array Software Support/Documentation CD provides various Host Software products corresponding to HP StorageWorks P2000 G3 Modular Smart Array Family products. It also contains links to hp.com where newer versions of these software products may be available.

The following are the key Host Software products contained in HP StorageWorks P2000 Modular Smart Array Software Support/Documentation CD:

Host Software Bundles

Separate Host Software bundles are provided for each of the supported OS platforms.

Windows bundles

Windows bundles are based on Windows ProLiant Support Packs. Separate bundles are available for Windows 2003 and 2008, and for each of the supported hardware architectures (x86, x64, and IA64). Bundles include installable Smart Components and the HP SUM engine. The HP SUM engine will have “pull from web updates,” so users can get the latest from web automatically.

Linux bundles

Linux Bundles are based on Linux ProLiant Support Packs. Separate bundles are available for RHEL4, RHEL5, SLES 10, SLES 11. Bundles include installable RPM packages and install script (install.sh)

Individual Host Software Smart Components

Individual Smart Components are available for each of the drivers contained in the bundles so customers can choose to install or update a specific driver without going through the bundle installation. Here again, the individual drivers are available for each of the supported OS platforms and hardware architectures. In addition to the drivers locally hosted on the CD, links are provided to hp.com where newer versions of the drivers may be available.

Microsoft Hot Fixes

Also available are the Microsoft Hot Fixes for various Microsoft dependent products like the Storport storage driver, VDS, and VSS. The Setup page provides detailed instructions on the sequence of steps required to install these hot fixes.

Best Practices for Firmware Updates

The sections below detail common firmware updates best practices for all generations of the MSA2000/P2000. This includes the MSA2000 G1, MSA2000 G2, and the P2000 G3 MSA.

General P2000/MSA2000 Device Firmware Update Best Practices

1. As with any other firmware upgrade, it is a recommended best practice to ensure that you have a full backup prior to the upgrade.
2. Before upgrading the firmware, ensure 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.
3. Do not cycle power 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 HP customer support.
4. After the device firmware update process is completed, confirm the new firmware version is displayed correctly via one of the MSA management interfaces—SMU GUI, MSA CLI, etc.

P2000/MSA2000 Array Controller or I/O Module Firmware Update Best Practices

1. The array controller (or I/O module) firmware can be updated in an online mode only in redundant controller systems.
2. When planning for a firmware upgrade, schedule an appropriate time to perform an online upgrade.
 - For single domain systems, I/O must be halted.
 - For dual domain 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.
3. 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.
4. When reverting to a previous version of the firmware, ensure 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 Partner Firmware Update (PFU) option and then downgrade the firmware on each of the controller separately (one after the other).
5. When performing FW updates to MSA70 drive enclosures, each enclosure will need to have a power cycle performed.

P2000/MSA2000 Disk Drive Firmware Update Best Practices

1. Disk drive upgrades on the HP StorageWorks P2000/MSA2000 storage systems is an offline process. All host and array I/O must be stopped prior to the upgrade.
2. If the drive is in a virtual disk, 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.
3. 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.

Summary

The HP StorageWorks MSA administrators should determine the appropriate levels of fault tolerance and performance that best suits their needs. Following the configuration options listed in this paper can help you make sure that the HP StorageWorks MSA family enclosure is optimized accordingly.

For more information

To learn more about the HP StorageWorks MSA2000, please visit <http://www.hp.com/go/msa2000>

To learn more about the HP StorageWorks P2000 G3 MSA, please visit <http://www.hp.com/go/p2000>

Share with colleagues     



Get connected

www.hp.com/go/getconnected

Current HP driver, support, and security alerts delivered directly to your desktop

