HP StorageWorks MSA2000

Best practices

Table of contents

About this document

This white paper highlights the best practices for optimizing the HP Storage Works MSA2000 G1 and MSA2000 G2 pre-designed for use 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 MSA 2000 G1 and MSA2000 G2 administrators and requires previous SAN knowledge. This document offers common known Modular Storage Array facts that can contribute to an MSA best customer experience.

The key best practice considerations described in this paper are for performance and other more effective setup configurations.

This paper is broken into 2 sections:

1. MSA2000 G1 best practices
2. MSA2000 G2 best practices

Topics covered MSA2000 G1

This paper 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
- WWN naming
- 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 solution designed for small to medium size deployments or remote locations. It is the most scalable and highest performing array of the MSA2000 family. 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 hosts for Fibre Channel attach.

HP StorageWorks 2000i G1 Modular Smart Array

The MSA2000i is a 1 Gb Ethernet (1GbE) iSCSI connected to 2U SAN array solution. The MSA2000i also, like the MSA2000sa, allows customers to grow their 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, dual-ported SAS drives, and archival-class SATA drives. Additional capacity can easily be 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 and designed for small to medium size deployments or remote locations. The MSA2000sa product 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 twelve drive bays that can simultaneously accommodate 3.5-inch enterprise-class SAS drives and archival-class SATA drives. Additional capacity can easily be added when needed 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 over 21.6 TB SAS or 48 TB SATA with the addition of the maximum number of 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.

Following are some distinct characteristics of each model.

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
- Lower infrastructure cost
- Lower 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 G1use 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 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 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

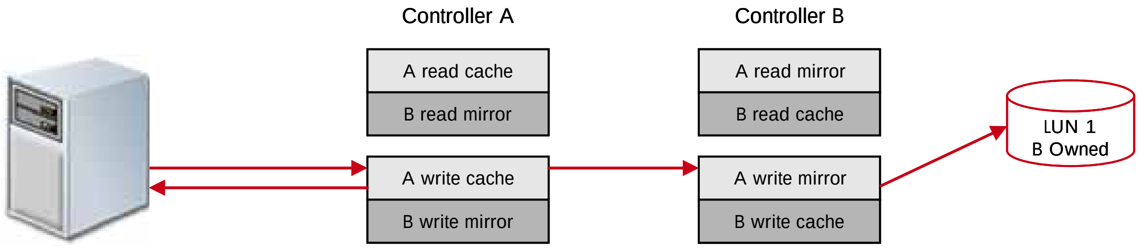
ULP recognizes which paths are “preferred”

- 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

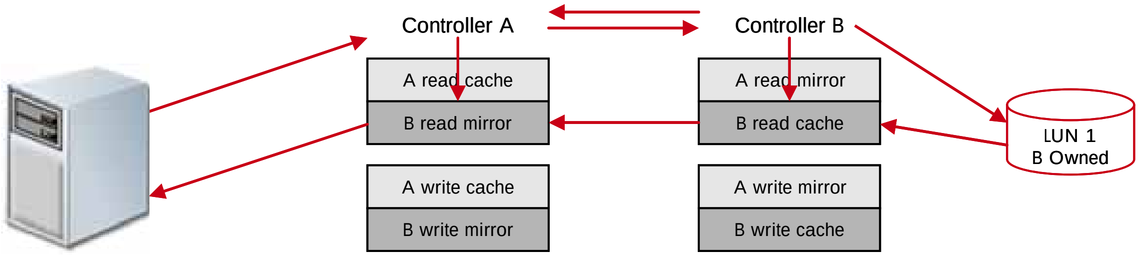
Figure 1: Write I/O Processing with ULP



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

Depending on whether fault tolerance (where redundant components are designed for continuous processing) or performance is more important to your solution, the host port interconnects need to be enabled or disabled, which is done 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 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 will result 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-Gbit/sec (MSA2000i G1), 3-Gbit/sec (MSA2000sa G1), or 2/4-Gbit/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.

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. DAS is currently the only supported method of connect for the MSA2000sa G1. The MSA2000i G1 does not support direct attach. The MSA2000fc G1 supports either direct attach or fabric switch attach configurations.

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 G1and 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 in integral part of the Direct Attach Storage for HP BladeSystem solution. In this configuration, the MSA2000sa G1can 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.

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 will be used or when growth in required storage or number of hosts is expected.

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

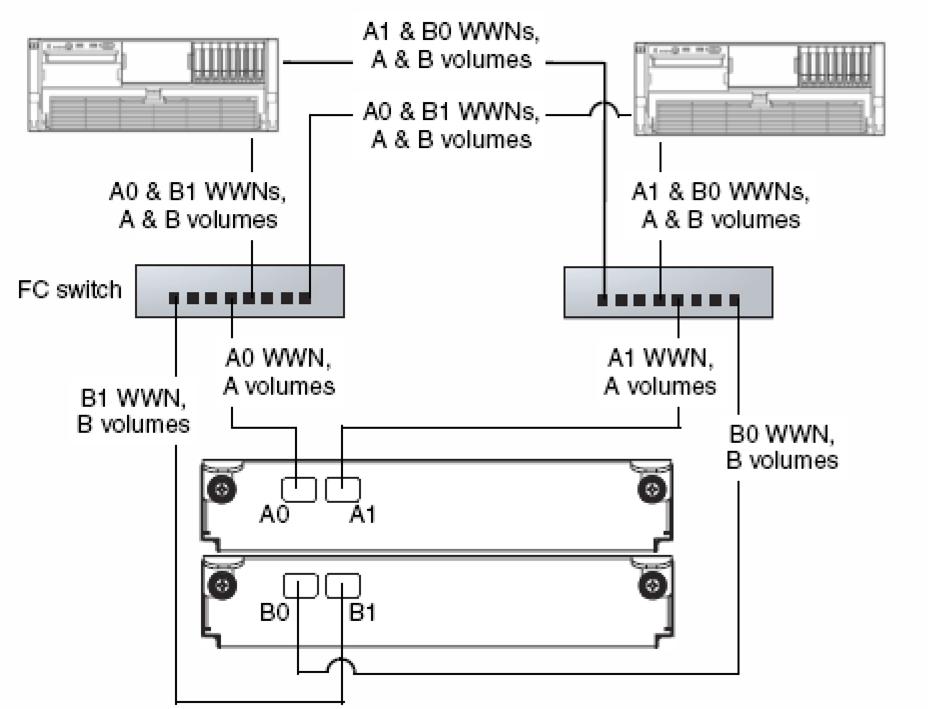
The topology only affects how mapped volumes and port WWNs are presented if one controller fails. Whichever topology is used, each data host has dual-ported access to volumes through both controllers.

- 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.

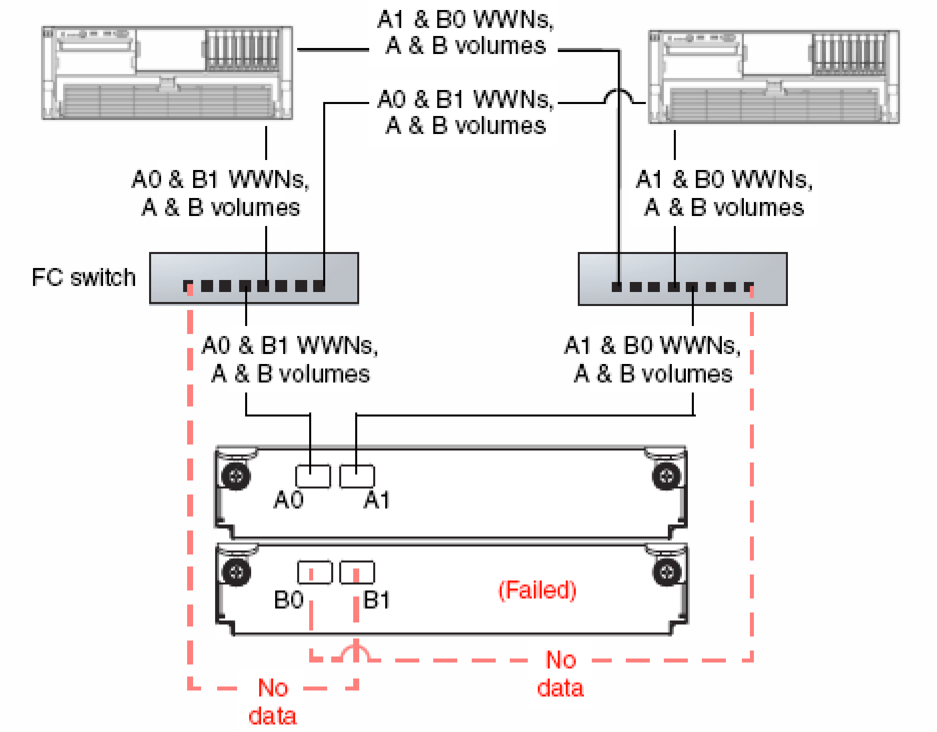
The following figure (Figure 3) shows how port WWNs and mapped volumes are presented when both controllers are active.

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



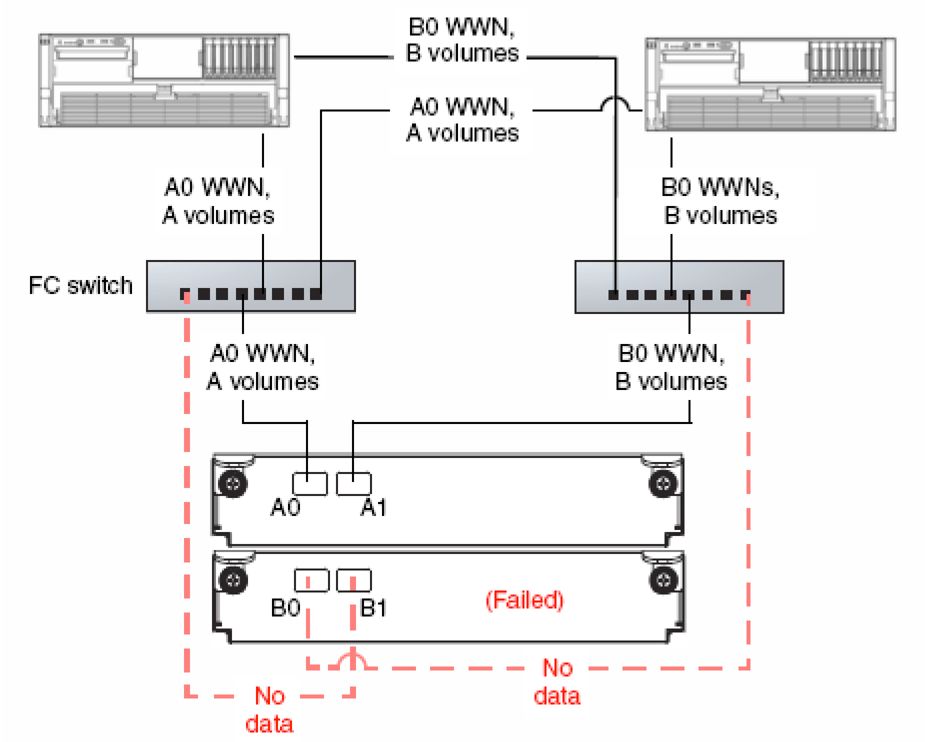
For a system using loop topology, the following figure (Figure 4) shows how port WWNs and mapped volumes are presented if controller B fails.

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



For a system using point-to-point topology, the following figure (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)

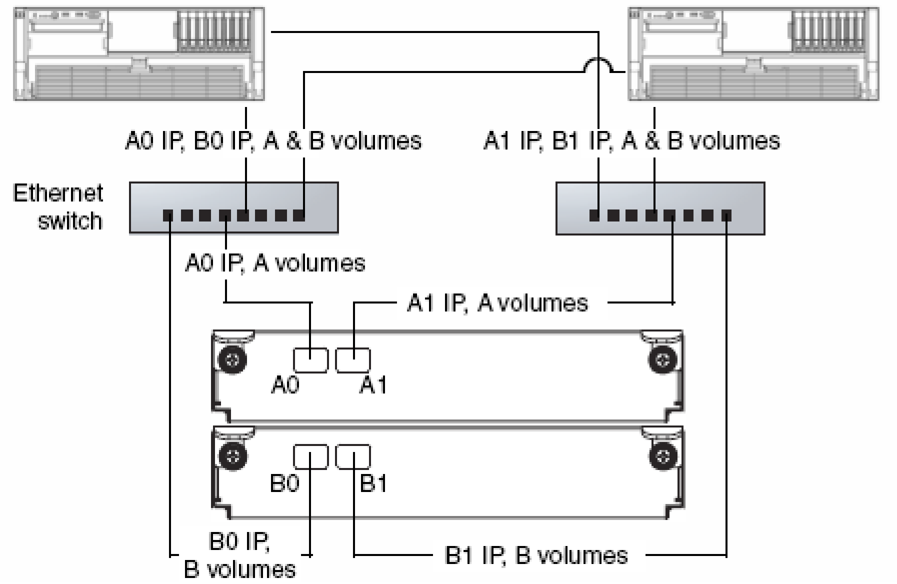


iSCSI switch-attach configuration

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

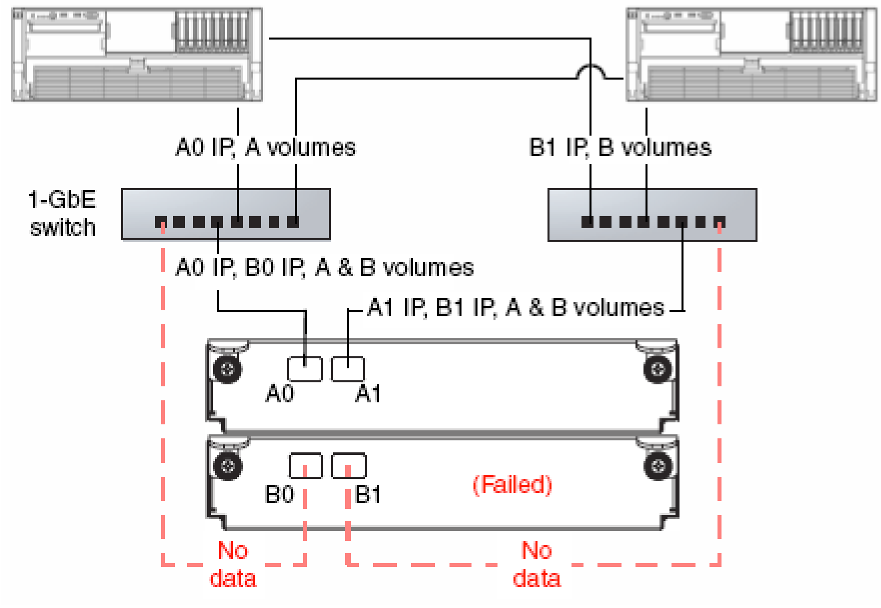
A dual-controller MSA2012i G1 storage system uses port 0 of each controller as one failover pair and port 1 of each controller as a second failover pair. If one controller fails, all mapped volumes remain visible to all hosts. Dual IP-address technology is used in the failed-over state and is largely transparent to the host system. The following figure (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



The following figure (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 node WWN so they appear as a single device to hosts. The controllers also share one set of LUNs for mapping volumes to hosts.

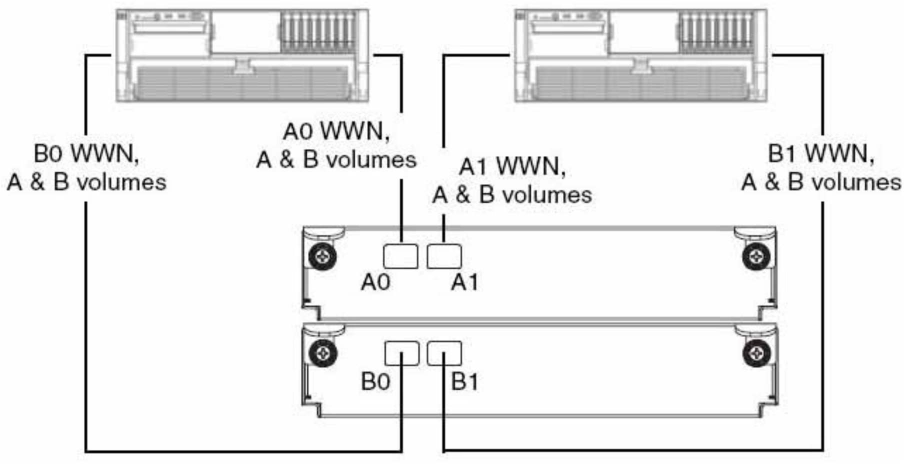
A host can use any available data path to access a volume owned by either controller. The preferred path, which offers slightly better performance, is through target ports on a volume's owning controller.

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.

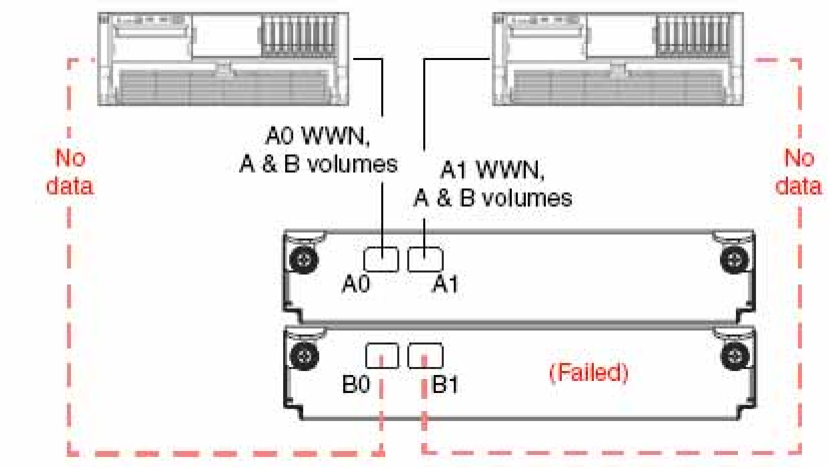
In the following configuration, both hosts have redundant connections to all mapped volumes.

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



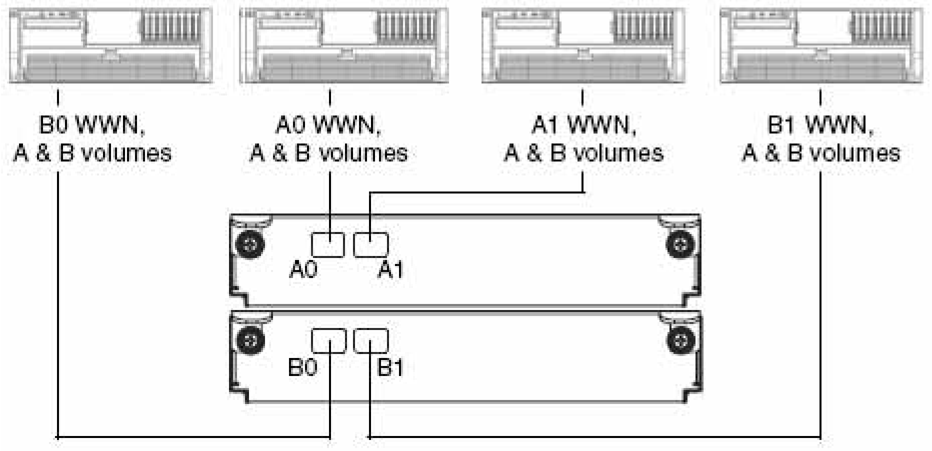
If a controller fails, the hosts maintain access to all of the volumes through the host ports on the surviving controller, as shown in the following figure.

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 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. 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 Tbyte. 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 Tbyte, which requires specific operating system, HBA driver, and application-program support.

Note:
The MSA2000 can support a maximum vdisk size of 16 Tbyte.

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 Tbyte and divide them into multiple volumes with a capacity of 2 Tbyte 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 Tbyte SATA drives that is 16 Tbyte raw)
- RAID 50 can support up to 32 drives (with 1 Tbyte SATA drives that is 32 Tbyte 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 will 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 will 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 Kbyte
- 32 Kbyte
- 64 Kbyte (default)

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

Alternatively, if the host accesses were 128 Kbyte 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 will be displayed as: <requested Chunk size> * (Num drives in sub vdisk - 1)"

For example: A requested chunk size of 32 Kbyte with 4 drives in a sub-array. The chunk size will be reported as 96k.

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

Tip:
The best practice for setting the chuck 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	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

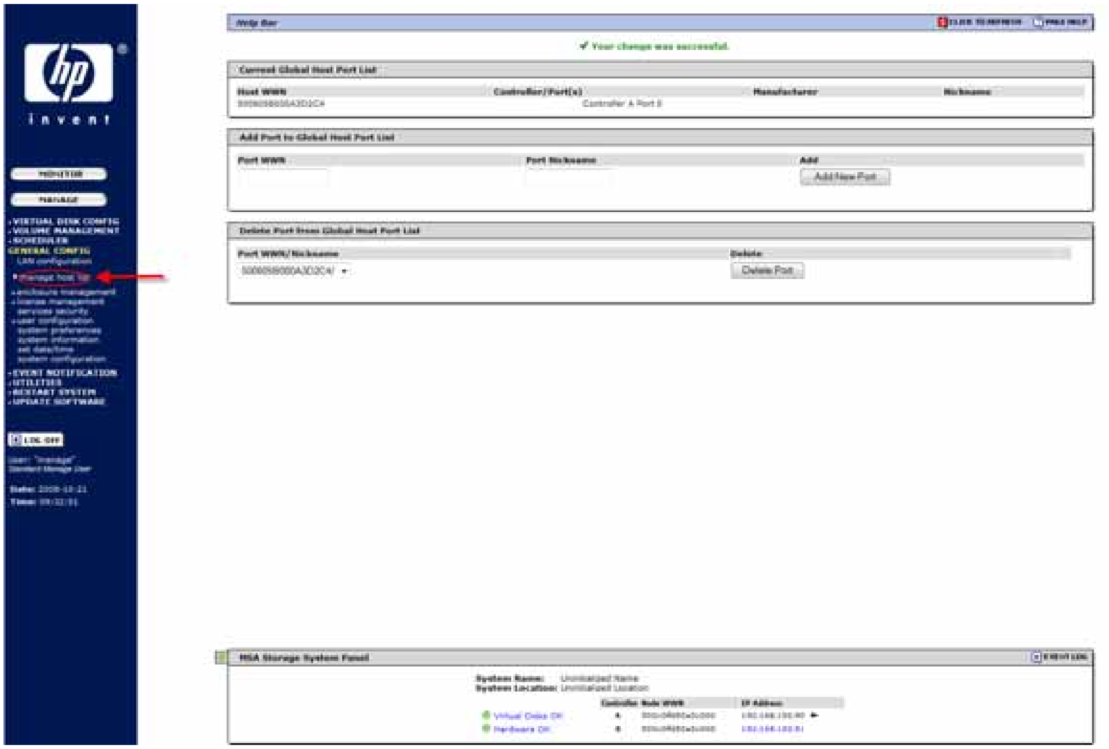
A best practice is to designate a spare disk drive for use if a drive fails. Although using a dedicated vdisk spare is the most secure way to provide spares for your virtual disks, it is also expensive to keep a spare assigned to each virtual disk. An alternative method is to enable dynamic spares or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

WWN naming

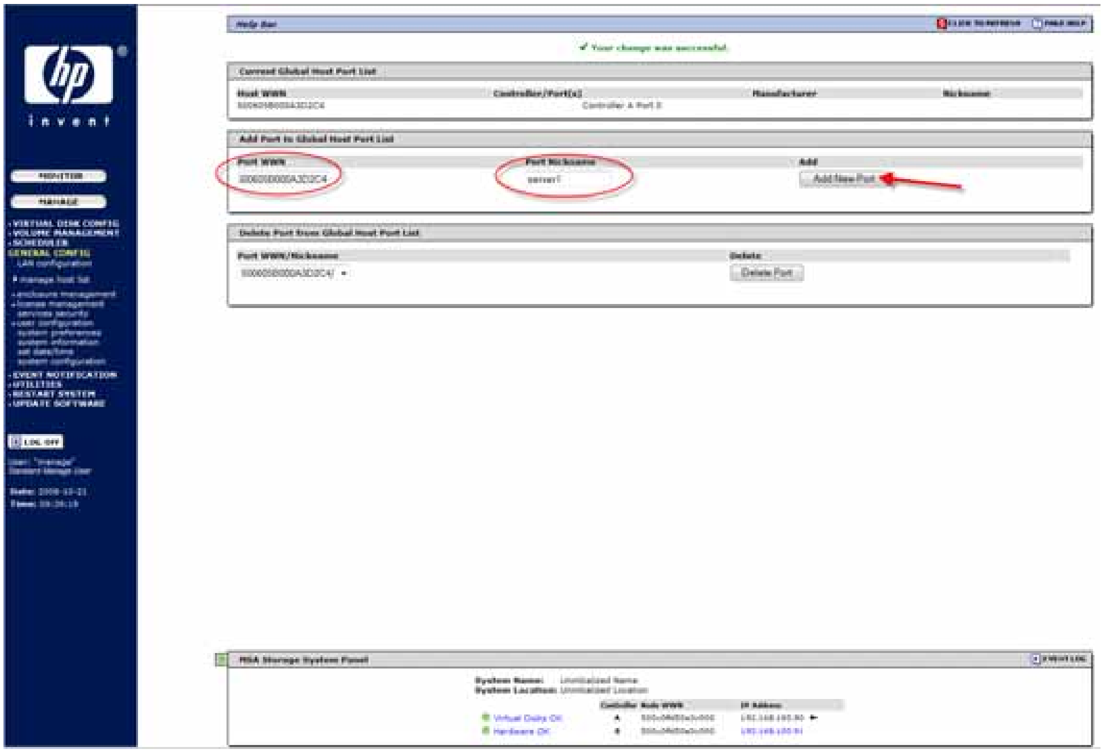
A best practice for acquiring and renaming World Wide Names (WWN) for the MSA2000sa G1 is to plug-in one SAS cable connection at a time and then rename the WWN to an identifiable name.

Procedure:

1. Open up the HP StorageWorks Storage Management Utility (SMU).
2. Click Manage -> General Config. -> manage host list.


Your change was successful.

3. 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.
4. Click “Add New Port” Click OK when the pop up window appears.



5. Plug in the SAS port of the HBA on the second server into the MSA2000sa controller port. Make sure the server is powered on.
6. Return to the Manage -> General Config -> manage host list of the SMU. The new WWN should now appear.
7. 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 Kbyte, 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 optimal 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 will 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 minimum 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. Backend bandwidth is used to mirror cache and because this mirroring uses backend bandwidth, 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-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 will 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 Kbtye. 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 Kbtye. 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 Gbit/sec on the MSA2000fc G1.
- Host ports should be configured for 1 Gbit/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.

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)

Topics covered on the MSA2000 G2

This paper examines the following:

- Hardware overview
- 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

Hardware overview

HP StorageWorks MSA2000fc G2 Modular Smart Array

The HP StorageWorks MSA2000fc G2 Modular Smart Array represents the newest fibre channel model of the MSA2000 family (which includes the MSA2000i iSCSI and the MSA2000sa SAS connected models), a new generation of HP storage arrays specifically designed for entry-level customers and features the very latest in functionality and technology.

The MSA2000fc G2 is a 4 Gb Fibre Channel connected 2U SAN or direct-connect solution designed for small to medium size 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 size 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 MSA2300i 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 3Gb SAS direct attach, external shared storage solution designed for small to medium size 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.

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 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

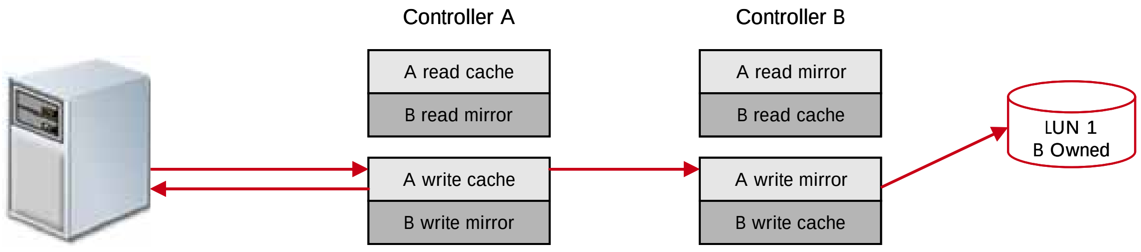
ULP recognizes which paths are “preferred”

- 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

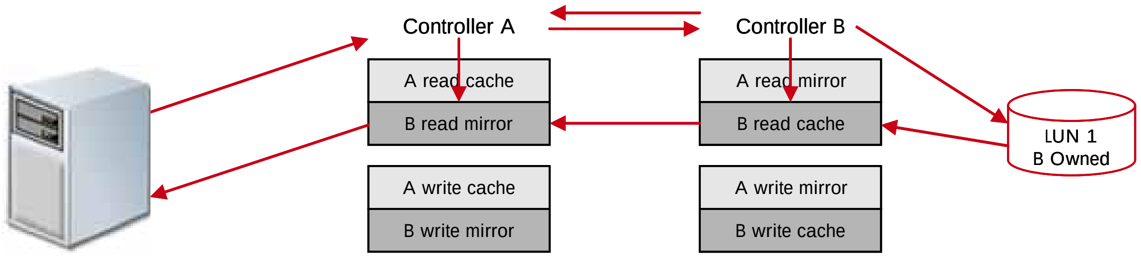
Figure 11: Write I/O Processing with ULP



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.

Choosing DAS or SAN attach

There are two basic methods for connecting storage to data hosts: DAS and 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.

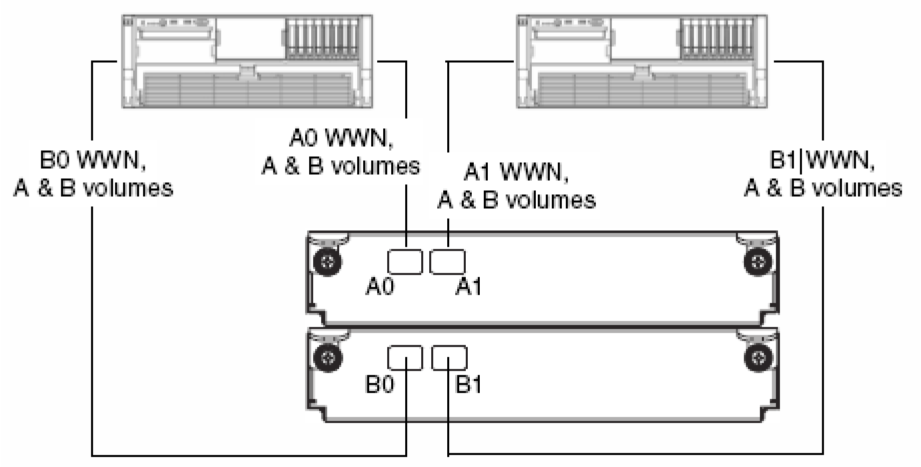
A host can use any available data path to access a volume owned by either controller. The preferred path, which offers slightly better performance, is through target ports on a volume’s owning controller.

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.

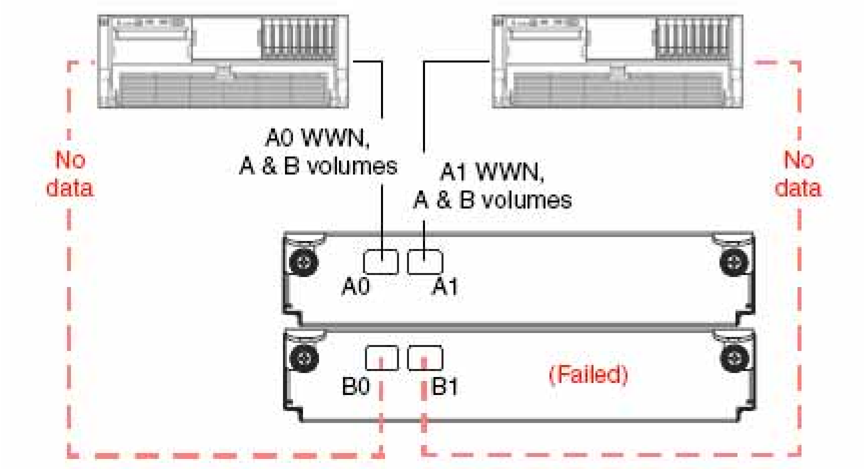
In the following configuration, both hosts have redundant connections to all mapped volumes.

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



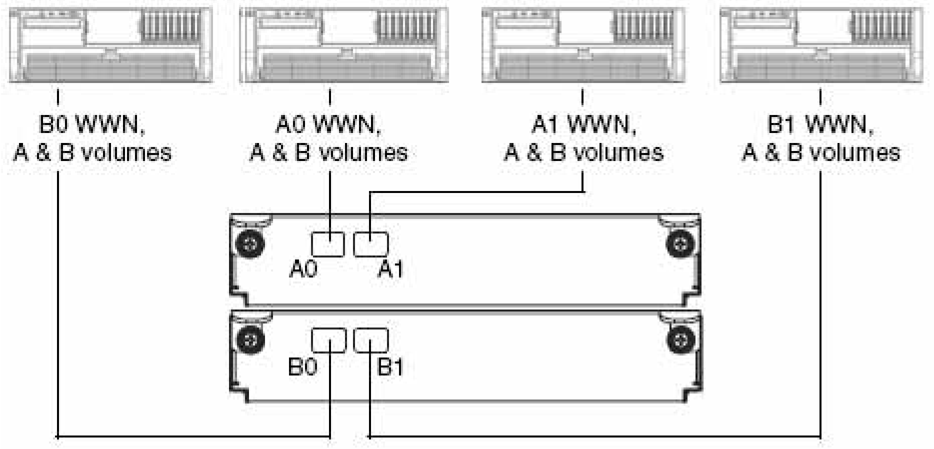
If a controller fails, the hosts maintain access to all of the volumes through the host ports on the surviving controller, as shown in the following figure.

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.

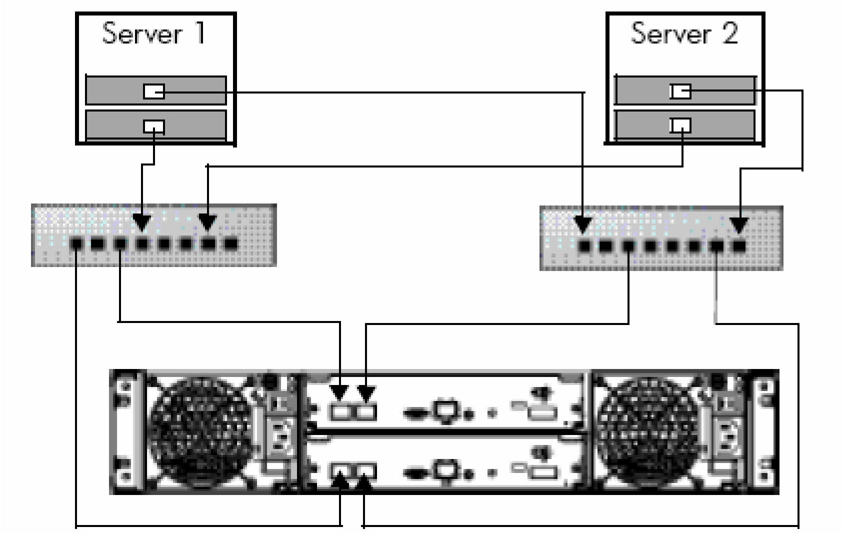
Figure 15: FC storage presentation during normal operation (High-availability, dual-controller, direct attach with four hosts)



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 Tbyte. 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 Tbyte, which requires specific operating system, HBA driver, and application-program support.

Note:
The MSA2000 G2 can support a maximum vdisk size of 16 Tbyte.

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 Tbyte and divide them into multiple volumes with a capacity of 2 Tbyte 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 Tbyte SATA drives that is 16 Tbyte raw)
- RAID 50 can support up to 32 drives (with 1 Tbyte SATA drives that is 32 Tbyte 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 will 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 will 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 Kbyte
- 32 Kbyte
- 64 Kbyte (default)

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

Alternatively, if the host accesses were 128 Kbyte 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 chuck size is to match the transfer block size of the application

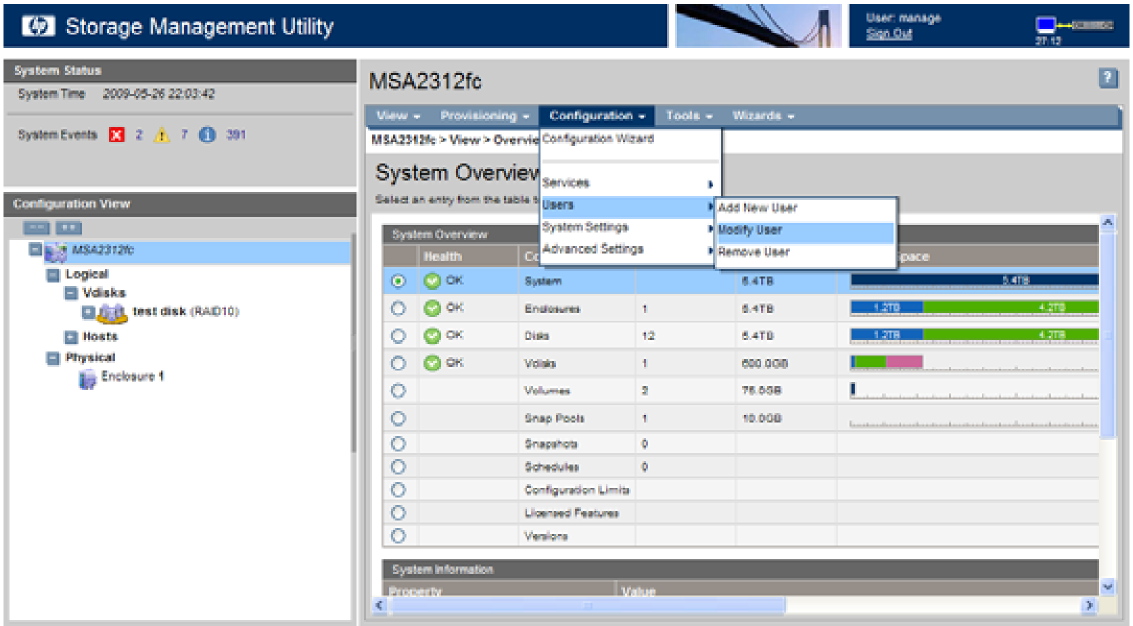
Vdisk initialization

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.



System Status

System Time 2009-05-26 22:03:42

System Events ☒ 2 ⚠ 7 ⓘ 391

MSA2312fc

MSA2312fc > View > Overvie

System Overview

Select an entry from the table t

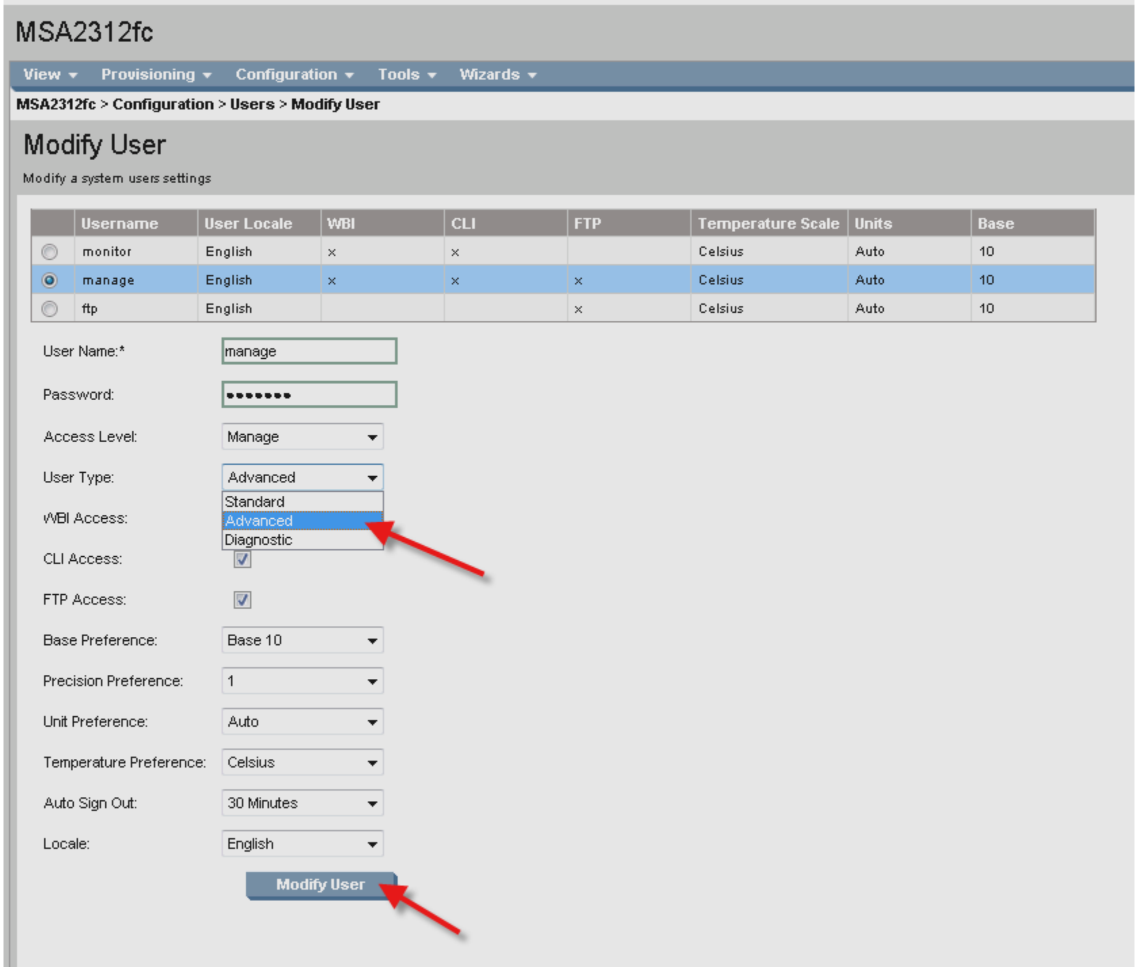
System Overview

	Health	Co			pace
●	OK	System		5.4TB	5.4TB
○	OK	Enclosures	1	5.4TB	1.2TB 4.2TB
○	OK	Disks	12	5.4TB	1.2TB 4.2TB
○	OK	Vdisks	1	600.0GB	
○		Volumes	2	75.5GB	
○		Snap Pools	1	10.0GB	
○		Snapshots	0		
○		Schedules	0		
○		Configuration Limits			
○		Licensed Features			
○		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

MSA2312fc > Configuration > Users > Modify User

Modify User

Modify a system users settings

	Username	User Locale	WBI	CLI	FTP	Temperature Scale	Units	Base
○	monitor	English	x	x		Celsius	Auto	10
●	manage	English	x	x	x	Celsius	Auto	10
○	ftp	English			x	Celsius	Auto	10

User Name:* manage

Password: •••••••

Access Level: Manage

User Type: Advanced

WBI Access:

CLI Access: ☑

FTP Access: ☑

Base Preference: Base 10

Precision Preference: 1

Unit Preference: Auto

Temperature Preference: Celsius

Auto Sign Out: 30 Minutes

Locale: English

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	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, 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.

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.

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 minimum 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. Backend bandwidth is used to mirror cache and because this mirroring uses backend bandwidth, 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 Kbtye. 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 Kbtye. 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	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 ports should be configured for 4 Gbit/sec on the MSA2000fc G2.
- Host ports should be configured for 1 Gbit/sec on the MSA2000i G2.
- Host ports should be configured for 3 Gbit/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 will 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 Web browsers.

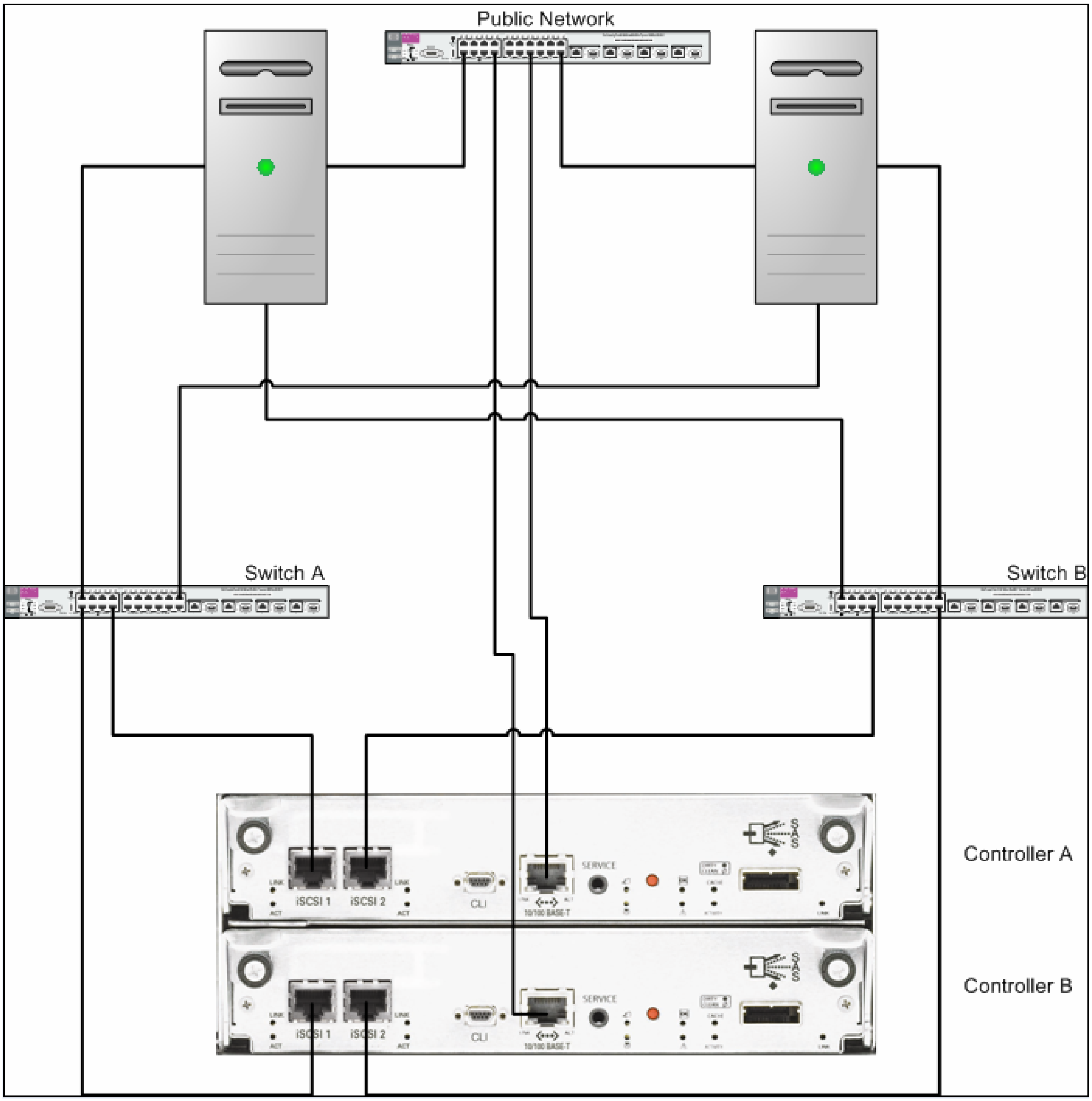
MSA2000i G2 Considerations

When using the MSA2000i G2, it is a best practice to use three network ports per server, two for the storage (Private) LAN and one for the Public LAN. This will ensure that the storage network is isolated from the other networks.

The private LAN is the network that goes from the server to the MSA2000i G2. This is the storage network. The storage network should be isolated from the Public network to ensure the best 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:

- Controller A port 0: 10.10.10.100
- Controller A port 1: 10.11.10.120
- Controller B port 0: 10.10.10.110
- Controller B port 1: 10.11.10.130

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 will make sure that the HP StorageWorks MSA2000 G1/G2 family enclosure will be optimized accordingly.

For more information

For more information, please visit http://www.hp.com/go/MSA2000

Technology for better business outcomes

4AA2-5019ENW Rev. 1, May 2009