HITACHI
Inspire the Next

Hitachi Adaptable Modular Storage 2000 Family Best Practices for VMware Virtual Infrastructure

Best Practices Guide

By Patrick Flueckiger

February 2009

Hitachi Storage Solutions

Hitachi Data Systems

Summary

The many benefits of virtualization — reducing cost of operation, increasing responsiveness, improving availability, enhancing performance, creating more environmentally friendly data centers and improving return on investment — mean that more and more organizations are adopting it. Availability, balanced performance and ease of management are even more important than ever before with an increasing number of mission-critical applications being deployed in virtualized environments like VMware Virtual Infrastructure.

The Hitachi Adaptable Modular Storage 2000 family brings enterprise-class availability, performance and ease of management to organizations of all sizes at a midrange price. It offers a robust storage solution that reduces setup costs and eliminates performance bottlenecks thanks to advanced point-to point SAS-based architecture for concurrent I/O capacity and Hitachi Dynamic Load Balancing Controllers that use symmetric active-active architecture to dynamically spread I/O workloads across resources and allow I/O through any path.

Hitachi Data Systems has the expertise to help you maximize availability, performance, and ease-of-management capabilities of the 2000 family in a Virtual Infrastructure environment. The best practice recommendations and configuration details in this document help to ensure successful deployment and maintenance of Virtual Infrastructure environments that use 2000 family storage systems

Table of Contents

Hitachi Adaptable Modular Storage 2000 Family Best Practices for VMware Virtual Infrastructure

Best Practices Guide

By Patrick Flueckiger

The many benefits of virtualization — reducing cost of operation, increasing responsiveness, improving availability, enhancing performance, creating more environmentally friendly data centers and improving return on investment — mean that more and more organizations are adopting it. Availability, balanced performance and ease of management are even more important than ever before with an increasing number of mission-critical applications being deployed in virtualized environments like VMware Virtual Infrastructure.

The Hitachi Adaptable Modular Storage 2000 family brings enterprise-class availability, performance and ease of management to organizations of all sizes at a midrange price. It offers a robust storage solution that reduces setup costs and eliminates performance bottlenecks thanks to advanced point-to point SAS-based architecture for concurrent I/O capacity and Hitachi Dynamic Load Balancing Controllers that use symmetric active-active architecture to dynamically spread I/O workloads across resources and allow I/O through any path.

Hitachi Data Systems has the expertise to help you maximize availability, performance, and ease-of-management capabilities of the 2000 family in a Virtual Infrastructure environment. The best practice recommendations and configuration details in this document help to ensure successful deployment and maintenance of Virtual Infrastructure environments that use 2000 family storage systems

Hitachi Adaptable Modular Storage 2000 Family Overview

Hitachi Adaptable Modular Storage 2100, 2300 and 2500 systems are ideal for the most demanding application requirements in a VMware Virtual Infrastructure environment and deliver enterprise-class performance, capacity and functionality at a midrange price. With support native Fibre Channel SAN, iSCSI and NAS, storage requirements can be consolidated onto a single platform.

The Adaptable Modular Storage 2000 systems are the only midrange storage systems with the Hitachi Dynamic Load Balancing Controller, symmetric active-active controllers that provide integrated, automated hardware-based, front to back-end I/O load balancing, thus eliminating many complex and time-consuming tasks that storage administrators typically face. This ensures that I/O traffic to back-end disk devices is dynamically managed, balanced and shared equally across both controllers. No other midrange storage products that scale beyond 100TB have a serial attached SCSI (SAS) drive interface. The new point-to-point back-end design virtually eliminates I/O transfer delays and contention associated with Fibre Channel arbitration and provides significantly higher bandwidth and I/O concurrency.

These systems are configured and managed from the same easy-to-use software that manages Hitachi Simple Modular Storage products. Combining the Adaptable Modular Storage 2000 family systems with the Hitachi Universal Storage Platform™ VM brings midrange customers even more enterprise features, such as storage virtualization, dynamic tiered storage, dynamic provisioning and advanced data replication. Table 1 provides an overview of the 2000 family hardware.

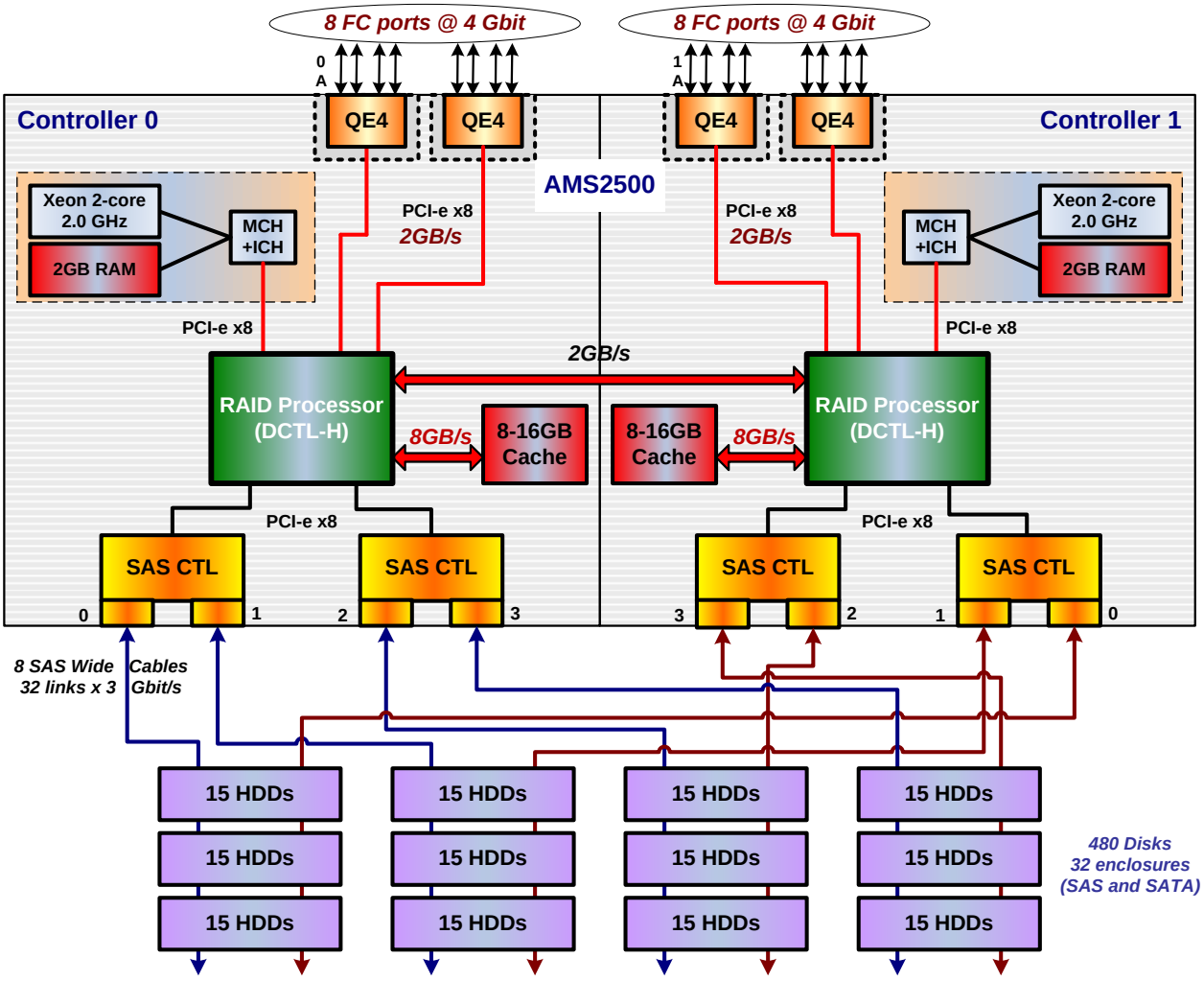
Table 1. Hitachi Adaptable Modular Storage 2000 Family Overview

Feature	*Adaptable Modular Storage 2100*	*Adaptable Modular Storage 2300*	*Adaptable Modular Storage 2500*
Maximum number of disk drives supported	120 SATA and SAS	240 SATA and SAS	480 SATA and SAS
Maximum cache	8GB	16GB	32GB
Maximum number of attached hosts	512	1,024	2,048
Front-end interfaces	4 x 4 Gb/s Fibre Channel or 4 x 1 GB/s iSCSI	8 x 4 Gb/s Fibre Channel or 4 x 1 GB/s iSCSI	16 x 4 Gb/s Fibre Channel or 8 x 1 GB/s iSCSI or 8 Fibre Channel + 4 iSCSI
Back-end disk drive connections	16 x 3 Gb/s SAS Links	16 x 3 Gb/s SAS Links	32 x 3 Gb/s SAS Links
Bandwidth to disks	4800MB/s	4800MB/s	9600MB/s
Connections per disk tray	2 SAS wide cables	2 SAS wide Cables	2 SAS wide cables
Active loops per number of disk drives	8 SAS links/15	8 SAS links/15	8 SAS links/15

Symmetric Active-Active Controller Architecture

As data storage requirements escalate and IT departments begin to reach the limits of virtualization's ability to control the necessary hardware, they must examine new approaches to storing and managing data. Controller technology offers a revolutionary, automated solution to data storage, but not all controller technology is created equal. Figure 1 illustrates the symmetric active-active controller architecture of the 2500.

Figure 1. Hitachi Adaptable Modular Storage 2500 Symmetric Active-Active Controller Architecture



A storage controller manages the flow of information between the server and the data, assigning two paths, in case one of the paths fails or is overloaded. *Active* and *passive* refer to the relative ability of a controller to assume an application workload — whether it sits idle until it is called upon to balance a load, or it is actively available at all times. Asymmetric controller architecture requires the administrator to define a preferred path and offers faster performance for an I/O following the preferred path, while a symmetric controller does not require preferred path definitions and offers the same level of performance regardless of the path.

Despite many advances in storage technologies, no host or network balancing algorithm is perfect. Round robin is a scheduling algorithm that assigns time slices to each process in equal portions and in order — handling all processes without priority. However, the round robin approach does not balance server performance, and if scheduled jobs vary in size, the round robin algorithm favors large jobs over others.

For optimum performance and availability, every layer of technology must be balanced: at the host, network and storage levels. Host-level balancing generally requires installing and updating host software as well as determining configuration and compatibility dependencies based on applications, OS type, cluster configuration and virtualization requirements. Similarly, a balanced network eliminates bottlenecks and reroutes I/Os to keep data moving efficiently.

Imbalances at the host and network levels must be absorbed and mitigated at the storage system level. As a result of these dependencies and complexities, it is difficult to develop a well-balanced, highly available, high-performance system. For example, if a network has one large block of data that is retrieved as an I/O request,

it is limited to the bandwidth of a single link, causing a load imbalance. Even if link aggregation is used on both the server and the storage system, any traffic flow between a single server and the storage system cannot exceed the bandwidth of a single link in the aggregate. Because none of these balancing technologies are perfect, administrators must rely on the solid foundation of the SAN to compensate. These limitations increase administration costs or at least make it impossible to get efficient use of a SAN without a SAN administrator.

Nonstop and Nondisruptive Microcode Upgrades

Microcode upgrades on the Hitachi Adaptable Modular Storage 2000 family are nonstop and nondisruptive. This means that a host never loses a path to the storage system and its LUs during a microcode upgrade.

Hitachi Dynamic Load Balancing

In a Virtual Infrastructure environment, the symmetric active-active controller's automatic load balancing reduces setup costs and eliminates performance bottlenecks. Virtual Infrastructure offers multipathing policies of most recently used (MRU), fixed and round robin (with experimental support). The fixed multipathing policy uses the configured, preferred path to a LU in preference to any other. In the event of a path error, Virtual Infrastructure fails to the next available path. When the preferred path recovers, Virtual Infrastructure routes all activity back through the preferred path for that LU. In this way, using fixed multipathing with symmetric active-active load balancing allows a configuration to dynamically repair itself.

A symmetric active-active controller does not rely on third-party host software for failover and load balancing; instead, those functions are native to the operating system. Removing third-party host software from the equation makes it easier to optimize SAN performance and perform maintenance. For example, storage controller microcode updates can be installed quickly and easily with little impact to operations because only the individual microprocessor is reported instead of the entire controller. This allows the application to continue running smoothly, even if the server has only one path to the storage controller.

Optimal performance is achieved with minimal input from storage administrators and back-end load balancing occurs automatically and evens out the utilization rates of both controllers.

Table 2. Virtual Infrastructure Multipathing Policies with Symmetric Active-Active Controller Architecture and Dynamic Load Balancing

Multipathing Policy	*HDS Recommendation*	*Advantage*	*Disadvantage*
Most recently used	Not recommended	None.	Manually set path policy for every LUN on every ESX host. No control over workload distribution across SAN fabrics and storage ports.
Fixed	Recommended	This is the default path policy for VMware ESX 3.5 for any Hitachi storage system. No manual changes needed.	Workload must still be balanced across host bus adapters and SAN fabrics.
Round robin	Experimental support only from VMware	Workload is automatically balanced across host bus adapters and SAN fabrics.	Manually set path policy for every LUN on every ESX host.

Symmetric active-active controller architecture combined with Hitachi Dynamic Load Balancing and Virtual Infrastructure's round robin multipathing policy dramatically simplifies the setup and management of a Virtual Infrastructure environment because the workload is automatically balanced across the following:

- Host bus adapters and SAN fabrics (due to multipathing policy round robin)
- Storage systems front-end FC ports (due to multipathing policy round robin and symmetric active-active controller architecture)
- RAID groups and LUs (due to symmetric active-active controller architecture and Hitachi Dynamic Load Balancing)

Hitachi Global Solution Services

Complex Virtual Infrastructure environments need a full analysis to determine the specific performance and capacity requirements of the storage systems. Hitachi Data Systems Global Solution Services helps you maximize investments and efficiencies with a suite of professional services. For the Adaptable Modular Storage 2000 family, Global Solution Services can assist with the following important tasks:

- Remote copy planning and design
- Implementation of Hitachi TrueCopy® Synchronous and Extended Distance software
- Implementation of Hitachi ShadowImage® Replication software
- Data migration planning, design and implementation

For more information about services to help meet regulatory compliance requirements, protect data, reduce total cost of ownership or develop a disaster recovery plan, contact your Hitachi Data Systems representative or visit www.hds.com.

Basic Storage System Setup

The Hitachi Adaptable Modular Storage System 2000 family has no system parameters that need to be set specifically for a Virtual Infrastructure environment. With the new symmetric active-active controller architecture and Dynamic Load Balancing, the LUN-ownership concept no longer exists and the associated parameters (for example, LUN ownership change disable mode) that are available on the predecessor storage systems are obsolete.

Host Groups per HBA Versus Host Groups per ESX Hosts or VMware Cluster

To present a set of common, shared LUs to multiple ESX hosts or to a VMware cluster, host groups can be created either per HBA port (that is, per WWPN) or per a group of ESX hosts or VMware cluster.

A host group created on an HBA port basis contains the HBA's WWPN and a set of common, shared LUs (that is, only one WWPN, multiple LUs). A host group created per group of ESX hosts or per VMware cluster contains at least one WWPN from every ESX host and multiple LUs (that is, multiple WWPNs, multiple LUs). Every LU must be presented with the same host LU ID to every host or VMware treats the LU as a snapshot LU and disables access to the VMFS by default.

Although both concepts are supported, Hitachi Data Systems recommends creating host groups per HBA port (that is, per WWPN).

Figure 2. Host Groups Created per HBA Port

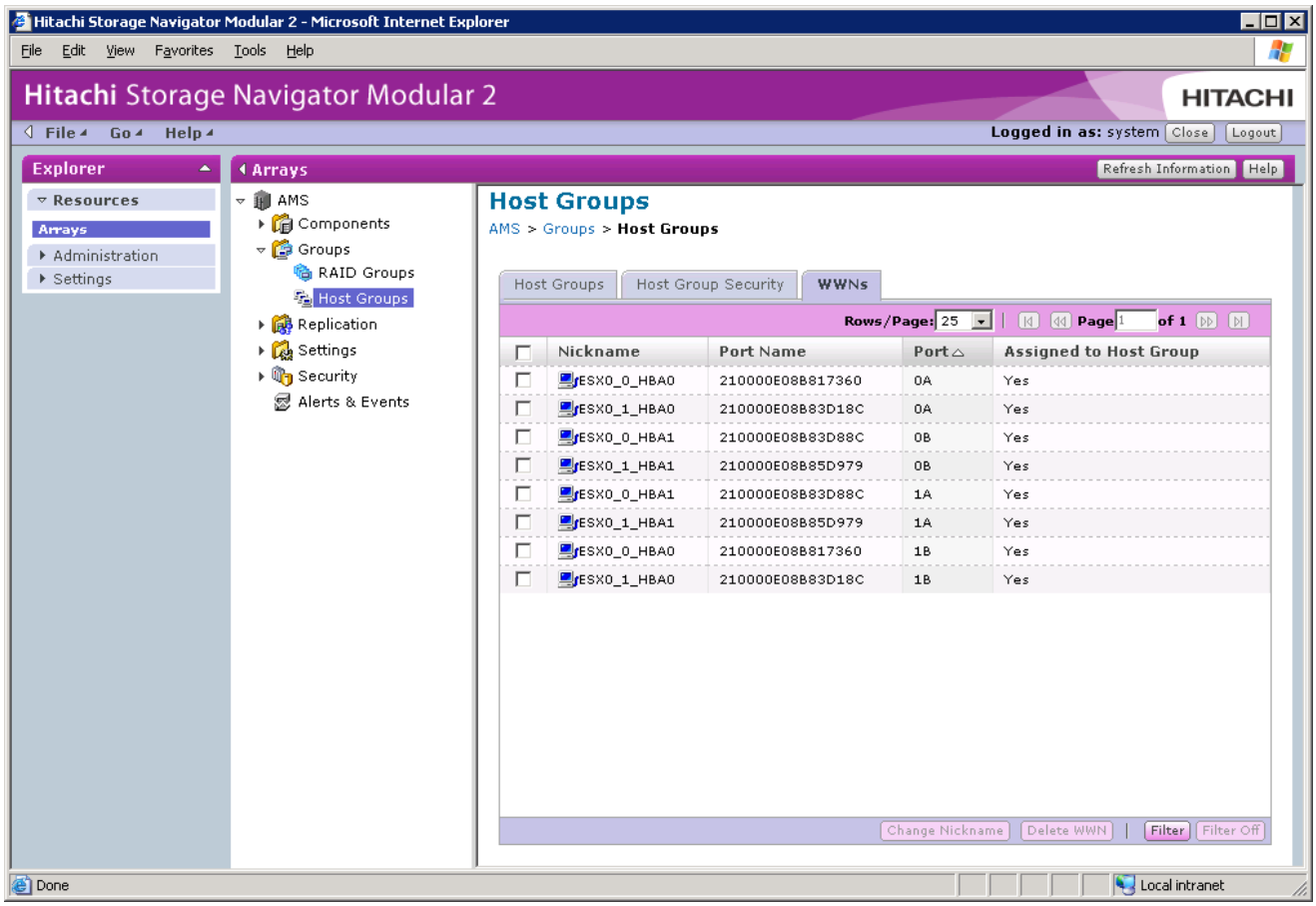


Table 3. Host Group Recommendations

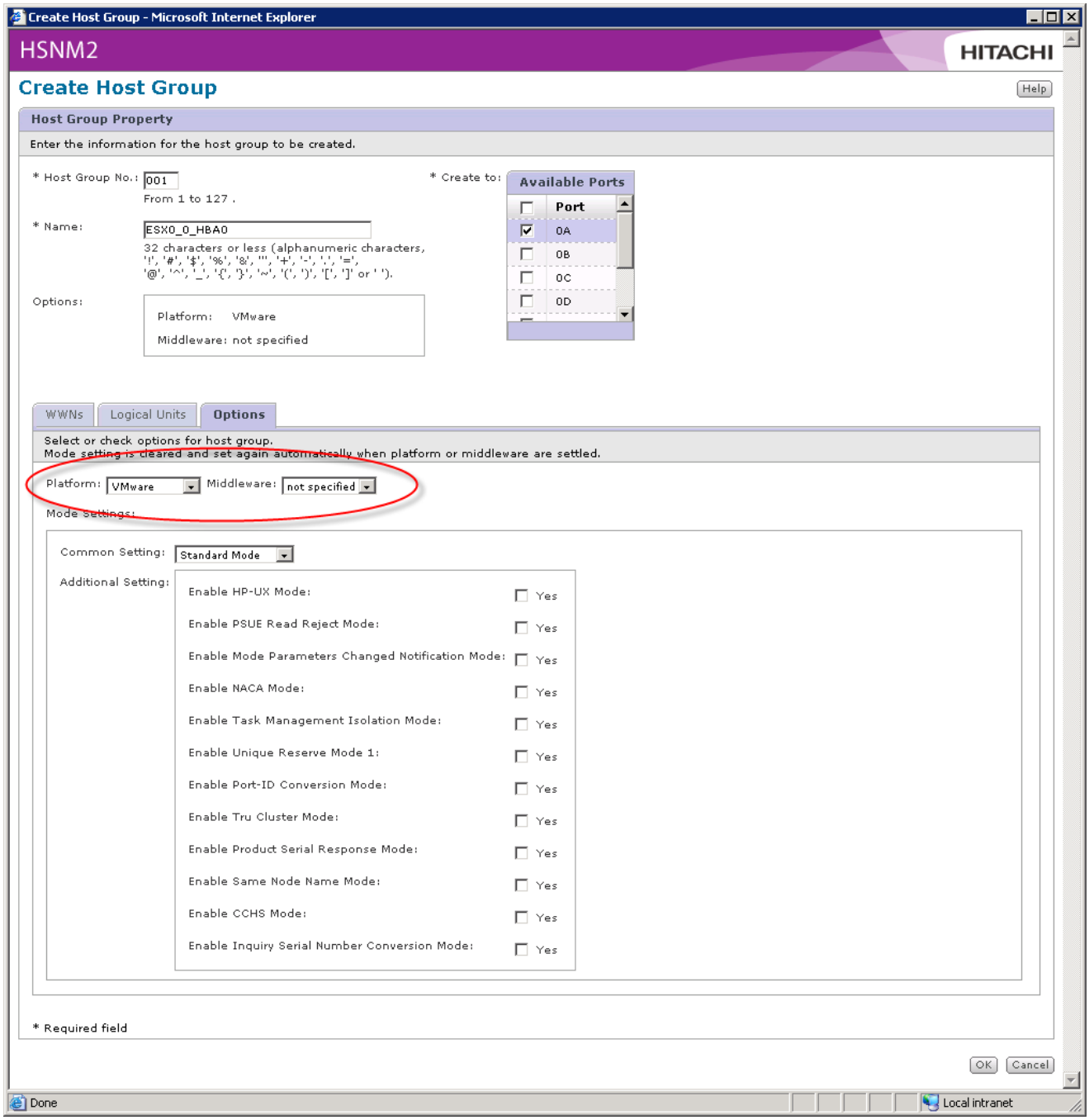
The following table lists some advantages and disadvantages of creating host groups per HBA port versus host groups per ESX hosts or VMware cluster.

Host Group Configuration	*Advantages*	*Disadvantages*
One host group per HBA (HBA port)	Allows boot from SAN. Better flexibility. Changes to the host group affect only one host. Better troubleshooting possibilities and log analysis.	More host groups to be created and maintained.
One host group per ESX hosts or VMware cluster	Fewer host groups to be created and maintained.	Does not allow ESX hosts to boot from SAN. Less flexibility (for example, an additional LU cannot be presented to only one ESX host). Changes to the host group might affect all hosts.

Host Group Options

To create host groups for ESX hosts, choose **VMware** on the **Platform** drop-down menu and **not specified** on the Middleware drop-down menu. These settings automatically set the correct modes as shown in Figure 3: For **Common Setting**, select Standard Mode. For **Additional Setting**, uncheck or deselect every option. This is a user-friendly option offered by the Storage Navigator Modular 2 GUI that is not available for Storage Navigator Modular 2 command-line interface (CLI). Using the Storage Navigator Modular 2 CLI, you must set the Platform and Middleware mode settings separately or leave them unchanged because these are the default settings.

Figure 3. Host Groups Options



Storage Provisioning

Capacity and performance cannot be considered independently. Performance always depends on and affects capacity and vice versa. That's why it's very difficult or impossible in real-life scenarios to provide best practices for the best LU size, the number of virtual machines that can run on a single VMFS and so on without knowing capacity and performance requirements. However, several factors must be considered when planning storage provisioning for a Virtual Infrastructure environment.

Size of LU

When determining the right LU size, consider the factors listed in Table 4. These factors are especially important from a storage system perspective. In addition, the individual virtual machine's capacity and performance requirements (basic virtual disk requirements, virtual machine swap space, spare capacity for virtual machine snapshots, and so on) must also be considered.

Table 4. LU Size Considerations

Factor	*Comment*
Minimum LU size	The smallest LU that can be created on the Hitachi Adaptable Modular Storage 2000 family is one block (512 bytes). The smallest VMFS that can be created on an ESX host is approx. 1GB.
Maximum LU size	The largest LU that can be created on the Hitachi Adaptable Modular Storage 2000 family is 60TB. The largest VMFS that can be created on an ESX host is 2TB (without using extents, 64TB when using extents).
Disk queues	More LUs offer more disk queues and therefore better performance.
SCSI reservations	Because SCSI reservations are set per LU, more but smaller LUs typically lead to less SCSI reservation conflicts and therefore better performance.
Modular volume migration	Smaller LUs can be migrated using Hitachi Modular Storage Migration to a broader range of possible target RAID groups.
Data replication	Using more but smaller LUs offers better flexibility and granularity when replication a VMFS within a storage system (ShadowImage Replication software, Hitachi Copy-on-Write Snapshot software) or across storage systems (TrueCopy Synchronous or Extended Distance software).
VMware VMFS upgrades	Using more but smaller LUs offers better flexibility and granularity when upgrading a VMFS.
VMware vCenter Site Recovery Manager	Using more but smaller LUs offers better flexibility and granularity when creating protection groups and recovery plans in VMware vCenter Site Recovery Manger.

Considering all of the factors in Table 4, a good LU size to start with is 256GB or 512GB. This number might vary from environment to environment.

Number of Virtual Machines per VMFS, per RAID Group

The number of virtual machines that can run simultaneously on a VMFS depends on the aggregated capacity and performance requirements of the virtual machines. Because all LUs on a particular RAID group share the performance and capacity offered by the RAID group, Hitachi Data Systems recommends dedicating RAID groups to an ESX host or a group of ESX hosts (for example, a ESX cluster) and not assigning LUs from the same RAID group to the Virtual Infrastructure environment and other non-ESX hosts. This prevents the ESX I/O from impacting or being affected by other applications and LUs on the same RAID group and makes management cleaner.

Follow these best practices:

- Create and dedicate RAID groups to your ESX hosts (that is, VMware cluster)
- Create LUs of 256 GB or 512 GB in size, or adjust the LU size considering the factors listed in Table 4. Always present LUs with the same H-LUN if they are shared with multiple hosts.
- Create VMFSs on the LUs as needed.
- Create and run virtual machines on the VMFS files.
- Monitor and measure the capacity and performance usage of the RAID group with VMware vCenter Server, Hitachi Tuning Manager software and Hitachi Performance Monitor software.

Monitoring and measuring the capacity and performance usage of the RAID group results in one the following cases:

- If all of the capacity offered by the RAID group is used but performance of the RAID group is still good, add RAID groups and therefore more capacity. In this case, consider migrating the LUs to a different RAID group with less performance using Hitachi Modular Volume Migration or Hitachi Tiered Storage Manager.
- If all of the performance offered by the RAID group is used but capacity is still available, do not use the remaining capacity by creating more LUs and VMFSs because this leads to even more competition on the RAID group and overall performance for the virtual machines residing on this RAID group is affected. In this case, leave the capacity unused and add more RAID groups and therefore more performance resources. Also consider migrating the LUs to a different RAID group with better performance.

In a real environment, it is not possible to use 100 percent of both capacity and performance of a RAID group, but the usage ratio can be optimized by actively monitoring the systems and moving data to the appropriate storage tier if needed using Hitachi Modular Volume Migration or Hitachi Tiered Storage Manager. An automated solution using these applications from the Hitachi Storage Command Suite helps to reduce the administrative overhead and optimize storage utilization.

Fibre Channel Front-end Ports

Provision storage on two Fibre Channel front-end ports (on one port per controller) is sufficient for redundancy on the Hitachi Adaptable Modular Storage Systems 2000 Family. This results in two paths to each LU from an ESX host's point of view.

Basic ESX Host Setup

Hitachi Data Systems recommends applying the VMware patches that change the default multipathing policy for Hitachi modular storage systems to fixed mode. For more information, see VMware's Knowledge Base article 1003518 for VMware ESX 3.0.x and article 1004161 for VMware ESX 3.5.x Search by article number to find the appropriate information.

SCSI Reservations

Many operations — such as virtual machine power on or off, VMotion, virtual machines running with virtual disk snapshots or growing a file — require getting a file lock or a metadata lock in VMFS, resulting in a temporary SCSI reservation. During this short period of time, the entire LU is exclusively reserved and locked by a host and is therefore not accessible by other hosts. Excessive SCSI reservations by one or multiple ESX hosts cause SCSI reservation conflicts and can cause performance degradation on other ESX hosts accessing the same VMFS or LU.

You can search VMkernel log files for SCSI reservation conflict warnings. For more information about how to reduce SCSI reservations and how to avoid SCSI reservation conflicts, see VMware's Fibre Channel SAN Configuration Guide.

From a storage provisioning point of view, SCSI reservation conflicts can be reduced by creating more but smaller LUs to achieve a given capacity.

Queue Depth

The Hitachi Adaptable Modular Storage System 2000 family supports a maximum queue depth per Fibre Channel front-end port of 512. There is no queue depth limit on the LU. This value determines the maximum queue depth setting on the VMware ESX hosts. A large queue depth does not provide better performance per se. Instead, monitor the queued commands on the hosts and adjust the HBA queue depth to ensure that the front-end port does not exceed the maximum queue depth of 512 per Fibre Channel port. Use the `esxtop` command to determine the number of commands queued per ESX host.

Table 5. Queue Depth Settings for Hitachi Adaptable Modular Storage 2000 Family

Path Policy	*Recommended Maximum Setting*
Fixed / MRU Round Robin	512 divided by the total number of concurrent active paths to a given Fibre Channel port Example: 8 ESX hosts, 2 HBAs each, 1 path from each host to the same port maximum queue depth = 512 / 8 = 64

VMware Snapshot LUs

At the moment that an LU is formatted with the VMFS by an ESX host, it writes a signature that contains the following information to the disk:

- Serial number of the storage system
- Internal LU ID
- External LU ID (host LUN)

Whenever an ESX host has access to a VMFS, it reads and compares the signature with its current view. If it does not match, the host disables access to the VMFS (not the LU) by default and generates the following warning:

```
ALERT: LVM: 4941: vmhba0:0:1:1 may be a snapshot: disabling access. See
resignaturing section in the SAN config guide.
```

In this case, access to a LU can be enabled with one of the following ESX advanced parameters:

- `LVM.EnableResignature`
- `LVM.DisableSnapshotLUN`

The default values of these parameters disable access to the VMFS.

Table 6 lists the situations where an ESX host might treat a VMFS as a snapshot LUN. In these situations host lists and recognizes the LU in the Storage Adapter screen of Virtual Infrastructure Client. A SAN rescan on the HBA might be necessary to discover the LU if it is newly mapped. The VMFS might appear in the Storage screen of Virtual Infrastructure client depending on the situation and the settings of the advanced ESX parameters.

Table 6. ESX Host Treats VMFS as Snapshot LU Situations

Situation	*ESX Behavior*
In-system data replication (ShadowImage, Copy-on-Write)	The LU is treated as a snapshot and access to the VMFS is disabled by default. Reason: At least the internal LU ID changed. It is possible that the external LU ID also changed.
Cross-system data replication (TrueCopy Synchronous, TrueCopy Extended Distance)	The LU is treated as a snapshot and access to the VMFS is disabled by default. Reason: At least the serial number of the storage system changed. Note: VMware vCenter Site Recovery Manager uses cross-system data replication and handles the situation.
Present LU to multiple ESX hosts with different host LUNs	The LU is treated as a snapshot by some or all ESX hosts and access to the VMFS is disabled by default. Reason: The external LU ID (host LUN) has changed for some or all ESX hosts.
Modular Volume Migration	The LU is not treated as a snapshot, the VMFS is accessible. Reason: Modular Volume Migration is transparent and non-disruptive to the host. Neither serial number of the storage system, internal LU ID nor external LU ID (host LUN) changed.
Online RAID group expansion	Does not affect the VMFS.
Online LUSE	Does not affect the VMFS.

Hitachi Data Systems Corporation

Corporate Headquarters 750 Central Expressway, Santa Clara, California 95050-2627 USA
Contact Information: + 1 408 970 1000 www.hds.com / info@hds.com

Asia Pacific and Americas 750 Central Expressway, Santa Clara, California 95050-2627 USA
Contact Information: + 1 408 970 1000 www.hds.com / info@hds.com

Europe Headquarters Sefton Park, Stoke Poges, Buckinghamshire SL2 4HD United Kingdom
Contact Information: + 44 (0) 1753 618000 www.hds.com / info.uk@hds.com

AS-004-00 February 2009