

Study Guide for

HH0-120

Hitachi Data Systems Certified Professional – Modular

Author: Eric Vanderburg

Date: 3-6-2009

Table of Contents

Overview	3
Essential Terminology	4
History	5
Table 1.1: HDS Product Line Specifications	5
AMS 2000 Series	6
Table 1.2: AMS 1000 and AMS 2000 product line comparison	6
Features	6
Service Oriented Storage Solutions	8
Host Groups	9
Software	10
Hitachi Dynamic Link Manager	10
Storage Navigator	11
Storage Navigator Commands	12
Software Features	12
Device Manager	13
Tuning Manager	14
Tiered Storage Manager	16
NAS	17
Content Archive Platform	18
ShadowImage Replication	19
Copy on Write Snapshot	21
TrueCopy Remote Replication	22
RAID Manager and CCI	23

Overview

In order to obtain the Hitachi Data Systems Certified Professional – Modular certification, candidates will need to pass the HH0-120 exam. Hitachi offers an official course on the material (Course #THI0515) but there are no publicly available materials to review or purchase.

Hitachi's description of the exam can be found here.

<http://www.hds.com/services/education/certification/exam-description-hh0-120.html>

Essential Terminology

Out of band – Outside the AMS

Inband – Within the AMS

Mapping – logical volume is mapped to a front-end port

Masking – restricting a LUN to a specific host

Recovery Point Objective (RPO) – Time of the last backup. How much data can you afford to lose

Recovery Time Objective (RTO) – The length of time it takes to locate and restore the data. How long can you afford to be offline

History

Let me give you a brief history of the Hitachi storage array product line. Hitachi used to call their enterprise storage the Hitachi Lightning and their modular storage was the Hitachi Thunder. The product names were changed several years ago. Now the enterprise line is called the Universal Storage Platform (USP) and modular storage is broken into three groups:

- Adaptable Modular Storage (AMS)
- Workgroup Modular Storage (WMS)
- Simple Modular Storage (SMS) platforms.

The AMS product line began with three models, the AMS200, AMS500, and AMS1000. In December of 2008 the AMS200, 500, and 1000, collectively called the AMS1000's were replaced with the AMS2000's. The replacement for the AMS200, AMS500, and AMS1000 are the AMS2100, AMS2300, and AMS2500 respectively.

Finally, Hitachi also has a product called the Network Storage Controller (NSC) which is a cross between enterprise and modular storage. Below is a chart with specifications from the different modular lines. Specifications for the product line is given below:

Table 1.1: HDS Product Line Specifications

Model	Max Drives	Max Trays	Cache	Front end ports	Back end ports	Max RAID Groups	Max LUNs
SMS100	12	1	2GB	iSCSI	N/A		128
WMS100	105	6	2GB	4 FC	FC	25	512
AMS200	105	6	2GB	4 FC	FC	25	512
AMS2100	120	7	4GB	4 FC	16 (4x4) SAS	50	2048
AMS2300	240	15	8GB	8 FC	16 (4x4) SAS	75	2048
AMS2500	480	32	16GB	16 FC	32 (4x8) SAS	100	4096

AMS 2000 Series

In December of 2008 the AMS200, 500, and 1000, collectively called the AMS1000's were replaced with the AMS2000's. The table below shows how the AMS 2000 line maps to the AMS 1000 line.

Table 1.2: AMS 1000 and AMS 2000 product line comparison

AMS1000 Series	AMS2000 Series
AMS 200	AMS 2100
AMS 500	AMS 2300
AMS 1000	AMS 2500

Features

Here is a list of the features of the AMS2000 product line

- FC was replaced with SAS, all are SAS shelves
- NAS is through NAS gateway
- LUNs do not have an owning controller
- Online firmware upgrades are possible
- Double the cache (4GB, 8GB, 16GB) – more capability for replication, additional snapshots
- 60TB LUN support (not used in the real world)
- Support for Microsoft VDS and VSS

SMS (Simple Modular Storage) 100 or 110 – iSCSI storage

Modular has 15 drives per shelf and enterprise has 16 drives per shelf so the drives cannot be changed between them because they are a slightly different size to fit in the same rack.

AMS500 often sold as a secondary storage for enterprise products

AMS1000 has 4 fiber ports per controller – very solid for replication because the extra ports can be dedicated to Truecopy.

Front end board types – NAS, iSCSI, or FC (AMS1000 can have different front-end board types per controller)

Service Oriented Storage Solutions

Service Oriented Storage Solutions (SOSS) – Business centric approach to aligning IT storage resources to changing business requirements. Looks at cost (tiered storage), risk (BC/DR - replication), and efficiency.

Put the application on a platform that makes sense for the application. Hiachi does this by looking at the type of data

- High-end Enterprise Apps Structured Data/RDB / Apps – Tiered storage virtualization using USP and NSC
- Midrange Application DB Enablement – NSC, AMS WMS
- Unstructured Data – archiving and object level awareness
- NAS

Integrated Strategy – Common Storage and Data management – when you learn how to manage an AMS that knowledge works on a USP/V box too because storage navigator is similar, LUN masking is similar, replication management is similar.

Tiered Storage – The ability to pull a number of different types of storage and virtualize them together. Pull these technologies together with Universal Volume Manager

Mainframes cannot be connected to modular storage – Must use NSC or USPV

NSC55 marketing position – high-end modular, low-end enterprise customer

Can virtualize other storage platforms

Modular storage upgrades are “data-in-place” upgrades that require downtime but preserve the data.

Shared Memory – stores the host group mappings and directories. This is written to a 48MB LUN called the differential manager LUN.

Active/Active configuration – no LUN ownership, microcode upgrades are non-disruptive

Host Groups

The smaller the RAID group the faster the rebuild time.

The more we distribute across the storage array, the better the performance of our logical units.

Mapping and masking are done in the same process by creating a host-group map.

Host group mapping process:

1. Pick a port and create a host group for each node. (Mapping)
2. Place LUNs into appropriate host groups. (Masking)
3. The host will not be able to see the storage unless there are mappings on the switches between the array and the HBA.

How does a storage array recognize a host? WWN of the HBA.

The settings in Storage navigator are all front-end configurations.

Software

Storage Navigator Modular 2 – works like device manager with SNM2 server.

Linux is supported but only certain kernel levels.

Controller batteries will keep the data in cache for 48 hours and drives will still spin for another 15 minutes.

Hitachi Dynamic Link Manager

Virtualizes the paths. The host does not care which HBA is providing the I/O.

The new AMS will load balance better because it is active/active.

Random I/O such as SQL server can be distributed between ports on HBAs and then make use of both controllers on the 2000 AMS and better use of the cache. However, sequential I/O will not be as easily distributed because data is dependent upon previous transactions. HDLM will burst data to one controller for a while before sending data to another one with Extended Round Robin. It is not as good on random in ExRR. These changes have to be made manually or through HDML scripts. HGLAM can manage this and specify which LUNs will be random (RR) and which will be sequential (ExRR).

HDLM daemon acts like a router to determine the best path to send the data down.

VMWare is going to need HDML version 5.9 or better.

HDev (Hitachi Device) – Modular LUN or Enterprise LDEV.

HGLAM – uses HiRDB. Specific LUNs on an HBA can be configured for specific round-robin modes.

HiRDB (Hitachi Relational Database) – Run-time database for the management suite. This is sold as a full product that competes with Oracle and SQL Server in Japan.

Storage Navigator

Enterprise boxes use a dedicated management machine that runs a web server and communicates with the SAN.

DAMP was the old modular storage management application. Written in Java.

Storage Navigator Modular 2 requires Java 1.6

SNM2 is more like device manager. It is server based and it can pick up information from more than one modular device at a time. The SNM2 server can now track performance data. (It works like the Essential NAS management software)

Device Manager will allow you to manage all Hitachi products

Default username and password is root and storage.

WWN for the AMS is made up of a combination of serial number and port.

Licenses can be permanent, temporary, or emergency. Permanent licenses are the ones purchased from Hitachi. These do not expire. Temporary keys are used for testing and proof of concept. They last 120 days. Emergency keys last for 5 days.

License files are .plk

A Hitachi wizard will walk you through the process of creating LUNs, RAID groups, and host group mappings.

Host Groups can be given a name rather than a number.

Raidscan can read the metadata coming from a LUN on a host machine and can be used to verify the host detection of the WWN and host settings such as middleware.

The middleware setting for the host group / host mode is used for setting cluster settings.

Host group settings can be changed without impact to the system.

PSUE Read Reject Mode shuts down the LUN if replication fails.

LUN Expansion can be used to combine LUNs on the AMS so that the AMS performs the concatenation instead of the host. The LUNs must be removed from the host group first. This is a destructive process that causes all data on the LUNs to be lost. The first LUN number added to the group is used as the new LUN number. Expanding to the concatenated LUN is not destructive to the concatenated LUN but it will destroy data on the new LUN (sub-LUN) that is

added to the group. Upt to 128 LUNs can be combined but there is the possibility of metadata corruption when you combine this many LUNs.

LUSE (“loosey”) Logical Unit Size Expansion – another name for this process on older boxes.

Storage Navigator Commands

CLI tools must be separately installed from the CD.

Startsnmen.bat will start the CLI services. This file is in program files\Storage Navigator Modular 2 CLI\

Run set and then STONAVM_HOME = . needs to be there.

Adding an array – auunitadd unit unit_name –ctl0 device...

Create a RAID group – aurgadd

Looking at raid configuration aurgref

Delete a RAID group – aurgdel

Adding a LU – auluadd

Viewing a LU – auluref

All the commands that have “ref” in them are read only report commands

Software Features

If you have an application that is using a lot of random I/O you can give it less cache and one that uses sequential I/O could have more cache. Applications that hog resources could be restricted with a cache partition.

Segmentation is per cache partition so that it can be customized for the applications.

MRU (Most Recently Used)/ LRU (Least Recently Used) queue – used to manage cache so that there are more cache hits.

RAID Group stripe size – default 64KB

AMS500 Max partitions was 8. AMS2300 has a max partition of 16.

Cache partitions should have the same stripe size as the RAID groups they service.

Cache residency – Reserves some cache for read hits to a specific LUN. It operates like a RAM drive. A separate license is needed for this. This is removed automatically when the LUN is removed.

Copy Pace settings are used to set the amount of write cache replication activities utilize.

Replication configurations are stored in the differential LUN so that when the machine boots it will copy that information back into cache.

Device Manager

Allows you to manage all of your HDS storage from a single interface. The devices are managed from “file system to spindle”.

Device Manager can be purchased for modular customers who have multiple arrays they need to manage but it has a large price tag. Device manager ships with the Enterprise products, however, each device must be licensed to be discoverable by device manager.

Other products can be used within device manager. Device manager serves as the framework for these applications so they all have a similar look and feel.

- Global link availability manager
- Tuning Manager
- Dynamic Link Manager
- Replication Manager – GUI replication functions. It allows you to configure all the replication pairs and then replication manager outputs the HORCM files.
- Storage Capacity Reporter
- Storage service Manager
- Tiered Storage Manager

Resource groups can be configured in device manager that allows resources to be grouped depending on your own needs such as by facility, department, type, etc... Group members are added manually. Reports can be generated upon groups and access control can be assigned to groups. Reports are exported in CSV.

- Physical configuration of storage systems
- Storage utilization by host
- Storage utilization by logical group
- Detail array reports
- User and group reports

By default an all resources group is created with everything in it.

Host-oriented storage – a device manager agent will allow you to have access to (Device manager has the permissions assigned to a peer account that you set up)

Device manager will provide the SMI-S API so that other heterogeneous storage management solutions can talk to Hitachi devices.

Software provisioning – Allocate storage until a logical pool so that storage usage can be tracked per group. For example, one department purchases storage and their storage can then be tracked on its own. Storage parameters such as host group, LUN size, concatenation, drive letters or mount points, formatting, and cluster size can be stored so that future storage is configured in the same way. This requires a host agent on the machine.

The device manager machine can manage inband devices if it has an HBA in it.

Device manager operations can be scripted with CLI.

Single management of all of the storage (Hitachi, Hitachi RSD, and Sun StorageTek).

Tuning Manager

Device Manager is needed to run tuning manager ever since version 5 because we are pulling capacity and configuration information from device manager. The recommendation was to run these on separate machines but it could be on the same machine if it is powerful enough and in this case they will both use the same database. This can be a virtual machine or a physical machine (Windows or Solaris). Tuning manager needs to have an HBA into each fabric. It also talks via IP to the host agents on the hosts.

All the data is obtained by the collection manager and stored in a HiRDB database.

Hitachi documentation does not give a decision tree on what might be the problem when counters reach certain numbers but the reports from tuning manager can greatly assist support in finding problems.

The reporting available in tuning manager can save a huge amount of time for management activities.

Performance monitoring to determine where trouble points are on the entire SAN. It can monitor the storage array, switch fabric, fiber directors and HBAs. Can show historic trends to see IOPS on the host and the storage array.

Email DSS, and rich media have a higher % of I/O content

Allows you to differentiate between perception and reality.

35-40 reports are predefined in the main console but you can use performance reporter to design your own reports.

Performance monitor in device manager only tracks short-range (15 days) and long-range data (up to 93 days) so data must be consistently exported to be useful for trend analysis and planning. Tuning manager can store much more and it will be in a database and it can track many metrics that performance monitor cannot.

Licensed per monitoring server so it can monitor many HDS devices.

Architecture

Agents

- Hardware – can run on the tuning manager machine
 - RAID agent (performance from storage arrays in band) – cache, IOPS, controller utilization
 - SAN agent (talks in band to the switch)
 - NAS agent (talks to NAS blades, essential NAS, and other vendor's NAS)
- Operating System - CPU and memory stats
- Application agents (Oracle, SQL Server, DB2, and SyBase)

Capacity and Performance management tools – used for generating historical reports, forecasting, and alerting.

Tiered Storage Manager

Tiered Storage Manager needs an enterprise box in order to work because the USP or NSC55 operates as the universal domain controller. All storage has to be virtualized using universal volume manager.

Migration tool that makes migrations simple and transparent. It allows you to migrate data from one level of storage to another without the application knowing about it.

Components:

- Tiered Storage manager server
- Management Client – laptop
- Domain Control Storage System – USP (could be diskless), NSC55 requires 5 drives (3+1 + Spare)

Storage domain – where our disks live

Migration Group – the LUNs that will be moved.

Storage Tier – This can be defined as you want. It could be drive types, RAID types, or system types.

The USP will need one fiber port configured as external. If connecting to one other array you can direct connect the array to the USP but if you are connecting multiple arrays you can use a switch to only allocate one external fiber port.

NAS

NAS Node Controller (NNC) – Blade that goes into the AMS

- Gives you 4 Gigabit Ethernet ports for each controller instead of two fiber channel ports.
- Debian with SAMBA for CIFS
- Web based management utility
- File system for the blade is HiXFS (Hitachi Tuned XFS)
- LUNs are created and then assigned to a host group associated with the blade. A LUN has to be created for the Debian OS.
- A NAS blade can join a Kerberos domain in both mixed and native mode AD.
- Additional Products: ShadowImage, Sync Image, and AntiVirus
- Backup: ShadowImage or Sync Image, or NDMP
- This option did not sell well because you lose fiber ports.

iSCSI

- 2 ports per controller, Gigabit Ethernet
- Will work with any iSCSI HBA NIC
- Which of the modular boxes can do a protocol intermix? AMS1000 or AMS2500

Content Archive Platform

Sarbanes Oxley requires this.

Hitachi keeps everything on an online drive so that it is always there.

Data is sent to the archive server or array when it is written to the production server or array.

Retention policies are put in place and the data is not modified. Metadata is retained for the data.

Components:

- Nodes (Linux Boxes)
- Software – Content Archive Platform File System
 - Metadata manager
 - Policy manager – constantly checks objects to see if they comply
 - Storage manager
- Policies
- Gateways – NFS, CIFS, HTTP
- Metadata – standard metadata for all supported file types

The file system is very basic so that it can be easily migrated as technology changes.

Data is straight ASCII and raw data.

HCAP sold as an appliance consisting of a WMS 100 with a brocade switch along with two 2-node clusters.

SAN plus Array of Independent Nodes (SAIN) – linux cluster that expands out to 64/128 nodes.

Replication requires help from HDS.

The system will allow another application to interface with it to find information using the HCAP API.

ShadowImage Replication

In system – local replication

Shadowimage create a point in time full copy of data on a LUN.

Replicate the data and then split the pair so that the replicated data sits on its own for backup purposes. The data could be presented to another host if needed.

Asynchronous Cache Destaging – New data that comes to the P-VOL is destaged first to the P-VOL and then to the S-VOLs that are in a paired stage. The track table is flagged for tracks that changed for non-paired S-VOLs.

RAID Manager – (Use to be called RAID manager Command Control Interface CCI) – This is the way to schedule jobs.

P-VOL (Production Volume) – Primary

S-VOL (Secondary Volume) - Copy

3 ShadowImage backups can run concurrently. (8 on the 2000 series)

P-VOL and S-VOL must be same sizes and on the same controller. (2000 series can operate on either controller)

First steps for replication: create a command device and a differential management LUN (DMLU).

Command Device – LUN that is 48MB (min) – This LUN is the place where commands are written to so that the array can pick it up.

Differential management LUN - Replication configurations are stored in the differential LUN so that when the machine boots it will copy that information back into cache. Need one per controller.

Process:

Want to back up a 500GB LUN at each hour in an ABC rotation starting at 10:00 AM.

1. Create three 500GB LUN
2. Create the initial copy
 - a. Issue a paircreate command to join them together at 9:15 AM expecting a 30-minute backup. The pair is in a state called "simplex".
 - b. A full volume copy starts between the two (1TB takes about 45 minutes) – the pair is in a state called "copy". Once the copy completes the pair is in a state called "pair".
 - c. The pair is treated as a mirror until 10:00 AM.
 - d. Issue a pairsplit command to separate the two. New data is only written to the P-VOL. The pair is now labeled as "suspend" PSUS and SSUS.
3. Updating the copy
4. Repeat steps 2a-d for the B backup.
5. Repeat steps 2a-d for the C backup.
6. ShadowImage keeps track of every track that has been changed in a table containing all tracks and either a 0 (no change) or a 1 (changed).
7. Issue a pairresync command on the A backup. The tracks that have changed since the split are transferred to the S-VOL.
8. Do the same thing on the B and C backups on schedule.

PC using ShadowImage has RAID manager on it. It talks to the differential LUN and the P-VOL.

If a second server accesses the S-VOL, the S-VOL will be unavailable when it is re-synced.

If 80% of the tracks have changed the entire volume will be replicated.

Reverse Resync – pairresync –r

Reverting to an S-VOL process

1. Take the application offline and unmount the P-VOL
2. Do a reverse resync from the S-VOL to the P-VOL.
3. Mount the P-VOL again.

Steady Split – flushes the last few things in cache for the P-VOL and S-VOL when the pairsplit command is issued.

Paireventwait – can pause until the pair reaches a specific state such as suspend.

See also class CSI0157 Modular Replication Fundamentals

Copy on Write Snapshot

Copy on Write is a virtual copy of data. This option does not require the entire storage volume to be fully replicated.

V-VOL – Virtual Volume – series of pointer tables back to the P-VOL. Points to a save pool where original tracks from the P-VOL that change are stored.

1000 series AMS can have 15 V-VOLs per P-VOL

2000 series AMS can have 32 V-VOLs per P-VOL

How much space do you need to allocate in the save pool for a copy on write snapshot? It depends is the answer.

If you do not provide enough space in the same pool the snapshots will fail and become unusable.

You need less space for V-VOL save pools because it only preserves the changes.

P-VOL and V-VOLs must be on the same controller in the 1000 series AMS.

Can you restore from a V-VOL to a P-VOL? Yes but you can only restore what has changed.

These are not business continuity snapshots.

You can monitor the size of a pool and add additional storage to it as needed. If the pool does fill all the V-VOLs will change to a status of PSUE (suspend with error) and all snaps in the pool will not work.

Sizes, RAID types, and drive types can be mixed within the save pool.

The only difference between setting up a shadowimage or copy on write is whether you specify a pool. If you specify a pool then it is a copy on write.

Paircreates and pairsplits are instantaneous because the only updates are the pointers.

TrueCopy Remote Replication

Moves from BC to DR and insystem to remote.

TrueCopy alone implies TrueCopy Synchronous

TrueCopy does not work on the AMS200 but it does work on the 500, 1000 and 2000 models.

Two front-end fiber ports are used for the replication. In the AMS500 this would mean that 50% of the front-end ports are unavailable.

This is a 1:1 relationship. No other S-VOLs can be mapped to the P-VOL so you cannot replicate to two different locations. Drive types cannot be mixed but you can mix RAID levels.

TrueCopy Synchronous – Licensed from IBM many years ago. Hitachi Open Remote Copy (HORC) was the old name for it. Guarantee of an RPO of 0. Uses a P-VOL and S-VOL relationship like in system replication. I/O comes in and is placed in cache. We do not send the node a confirmation until the data is placed in cache at the remote site. Synchronous replication has a distance limitation of 120KM (length of a dark fiber connection). It is not recommended that you use anything other than a direct connection or dark fiber because of latency concerns with telco offerings. The first time a replication pair is created the entire LU has to be replicated. This will require downtime to the application until this completes. The replication will time out after 10 seconds and then the pair will be labeled as PSUE. When the connection is restored the P-VOL and S-VOL will need to be resynced. The replication can be single direction or bi-directional.

Failover for TrueCopy Synchronous – Use the RAID manager host at the RCU to issue the horctakeover command which makes the S-VOL available on the RCU side. Horctakeover can be set to monitor the communications line so that it will fail back and replicate the changes back to the P-VOL when communications resume.

TrueCopy Extended – Asynchronous replication. This solution can go beyond 120KM. Often implemented with some telco WAN such as SONET or OC. I/O comes in and the array sends a confirmation to the host. The I/O is destaged to the P-VOL and queued up on a save pool on the array and the data is replicated from that save pool to the other save pool at the RCU. The data at the RCU save pool is destaged to the S-VOL as data is consistent. The application does not have to be offline when the initial copy is made. Best RTO is about 1-2 minutes on modular storage.

Consistency Group (CTG) – contains the P-VOLs, S-VOLs, and save pools. The consistency group does not destage data from the save pool to the S-VOL until the data is consistent (in other words, the data is all in the pool).

MCU (Main Control Unit) – Source array

RCU (Remote Control Unit) – Remote array

HORCM (Hitachi Open Remote Copy Manager)

TrueCopy Synchronous and Extended cannot be on the same box.

TrueCopy – creates a real-time copy of data

RAID Manager and CCI

Host talks to the array inband by sending data to the command LUN. Commands are written to the LUN and the daemon checks the LUN to see if there are responses on the LUN.

2 instances of the service are started and each has a HORCM command file. (HORCM0.conf and HORCM1.conf). If the configuration files are on two separate machines then they will need to talk via ports 11000 and 11001 on an IP network.

The last line in each of the HORCM files points to the other file. One file is for the P-VOL and the other is for the S-VOL.

TrueCopy will require a RAID manager CCI setup at both sites.

P-VOL HORCM files should be on even numbered files and S-VOLs on odd numbered files.

Start and stop the HORCM instances with the following commands:

`Horcmshutdown`

`Horcmstart`

Command devices process commands serially so if there are multiple replication actions taking place on one command device it will process them one at a time. This can be resolved by creating more command devices.