

Overview of the HSG80 StorageWorks Fibre Channel RAID Controller



HSG80 Controller Topics

- → Features
- Architecture
- 7 Caching Techniques
- ACS Operating System Software
- Subsystem Components
- Options



StorageWorks RA/ESA MA/EMA Features

- Modular components
- No single point of failure (redundant system level architecture plus storage software)
- Redundant, Hot-swappable components
- → 512MB Battery-Backed-up cache per controller
- → Up to 84 drive configuration scalability
- →Support for Tru64 Unix, OpenVMS, Windows NT, NetWare, IBM AIX, Sun Solaris, HP-UX, SGI IRIX, Linux
- Optical Fibre Channel Interface (FC-AL or switched fabric)
- Controller based Snapshot and Clone
- Controller based Selective Storage Presentation
- Data Replication Manager for NT, OpenVMS, Tru64 and Sun Solaris
- Array Controller Software (ACS)



StorageWorks ACS Functions

- Host Interconnect and Protocol Services
- → Cluster Support
- Support for 84 drives per storage system
- Dual Redundant Controller Operation
- Testing and Diagnosis of the HSG80 array controller
- SCSI device control
- Transparent Controller Failover support
- Multiple Bus Failover support
- Asynchronous Disk Swap (Hot Swap)
- → ACS system management services
- Local program support
- Mirrored Write-Back Cache support
- Read-ahead Cache support
- Disk Mirroring capability (RAID 1)
- Disk Striping capability (RAID 0)
- → RAID capability (RAID3/5)
- Storageset Expansion
- Disk Partition capability



HSG80 Architecture

FibreChannel HSG80 Controller Architecture





Policy Processor

- Coordinates, controls, and executes the controller software
 - ACS (HSG80 Array Controller Software)
- Uses Intel i960 microprocessor
- 16MB Data Memory
- Program Card
 - ↓ 2MB for ACS
- Nonvolatile Memory
 - 128KB to store
 - Controller configuration (SCSI targets, preferred ID's
 - Subsystem configuration (names and locations of disk drives)
 - Last failure entries



- Dual UART
 - Supports serial communication between controllers
- ↗ Timer
 - Controlled by and interrupts the i960
- Diagnostic Registers
 - 2 write and 2 read registers
 - Used by diagnostics to control devices and the controller



- Operator Control Panel (OCP)
 - Contains 6 device port LEDs (1)
 - Contains a "RESET" button (2)
 - Contains 6 port quiesce buttons (3)
 - to quiesce a port push its port button until LED is lit continuously
 - push the button again to resume I/O activity







- ↓ IBUS
 - Transfers code and text from the program card, code and data from the policy processor's I/D cache, and configuration information to the controller's nonvolatile memory
- ↓ CDAL
 - Transfers data to the surviving controller during failover
- ↓ MDAL
 - Conveys data to and from addresses in the local buffer memory
- ↓ NB<mark>US</mark>
 - Transfers data to and from addresses in any of the NBUS components: host ports, device ports, value added functions chip
- ↓ Bus Exchanger
 - Facilitates high speed communication between the local memory and devices on the IBUS, CDAL, MDAL, and NBUS
 - 🗟 Is a 4-way crossover switch



Local Memory

- During controller initialization the policy processor loads and stores the PCMCIA software into local memory
- Consists of dynamic RAM controller, arbitration engine, a gate array controller, and 16MB of dynamic RAM

Value Added Functions Chip

- Generates parity for the RAIDsets when data is written to them
- Also regenerates parity to recover stored data in the event that a member of a RAIDset fails
- Performs XOR parity



Host Port

- 4 2 Fibre Channel host ports per HSG80 controller
- Connection to hub or switch via single mode GLM
- FC-AL or Switched (but not both at same time)
- Device Ports
 - The controller communicates with the devices and modules in the subsystem through 3 backplane connectors
 - Each connector contains 2 ports to give the controller access to 6 Wide Ultra SCSI-2 busses.
 - Each device port contains a Wide Ultra SCSI processor



HSG80 Caching Techniques

Read Caching

- Used to decrease the subsystem's response time for a read request
- When the controller receives a read request from the host, it reads the data from the disk drives and delivers it to the host, and also stores it in cache memory
- By default, enabled for all storage units

Write-Through Caching

- Used to decrease the subsystem's response time to a request
- Called "Write-Through" because data passes through and is stored in - the cache memory on its way to the target disk drives
- If read caching is enabled for a storage unit, writethrough caching is also enabled



HSG80 Caching Techniques (cont.)

Write-Back Caching

- Used to decrease the subsystem's response time for write requests
 - Allows controller to declare write "complete" as soon as the data reaches its cache memory
 - Data is written to the disks later
- Disabled by default for all storage units

Fault Tolerance for Write-Back Cache

- The cache module supports the non-volatile memory, mirrored caching, and dynamic caching policies to protect the availability of its unwritten (write-back) data
- The controller cannot provide write-back caching to a unit unless its cache memory is non-volatile
 - By default the controller expects to use the ECB (External Cache) Battery) as its cache module backup power source



HSG80 Caching Techniques (cont.)

Read-ahead Caching

- Inables controller to keep track of read I/Os
- If controller detects sequential read I/Os from host, it will try to keep "ahead" of the host by reading the next sequential blocks of data (the host has not requested) and put them in cache
- controller can detect multiple sequential I/O requests across multiple units



HSG80 Caching Techniques (normal)





HSG80 Caching Techniques (mirrored)





HSG80 Caching Techniques (cont.)

Cache Policies Resulting from Cache Module Failures

- Way too much detail for this slide
- See HSG80 Maintenance Manual for details
- Cache Policies for ECB Failures
 - Too much detail for slide
 - See HSG80 Maintenance Manual for details



HSG80 External Cache Batteries

- ECB diagnostics are run on every controller restart
 - If batteries are fully charged, the controller reports them as "good" and rechecks them every 24 hours
 - If the batteries are charging, the controller checks them every 4 minutes
 - batteries are reported as being either above or below 50 percent in capacity
 - batteries below 50 percent in capacity are referred as being "low"
 - The four minute polling continues for up to 10 hours
 - the maximum time it should take to recharge the batteries
 - If the batteries have not been charged sufficiently after 10 hours, the controller declares them to be failed



ACS Operating System Software

Array Controller Software for HSG80 controller

- Current version V8.7
- See SPD in "Supplemental Handouts"
- All versions support Clone and Selective Storage Presentations (SSP)
- ↓ V8.7*G*
 - FC-AL only
- ↓ V8.7F
 - Switched or FC-AL
- ↓ V8.7<mark>5</mark>
 - Switched only, snapshot support
- ↓ V8.7P
 - Data Replication Manager, switched only, snapshot
- ↓ V<mark>8.</mark>7L
 - HSG60 only, Switched or FC-AL



RA/ESA Subsystem Components

DS-BA370-AA enclosure

- Supports up to 24 UltraSCSI devices
 - ↓ 6 Ultra Wide SCSI channels
- Packaging
 - HSx controller (1)
 - Environmental monitoring unit (2)
 - AC power distribution (3)
 - 180-watt power supply (4)
 - N+1 or 2N power
 - ECB (external cache battery (not shown))⁽³⁾
 - Storage building block (SBB) (5)
 - Cooling fans (7)
 - \downarrow I/O module (8)
 - Power Verification and Addressing (9)
 - Cache module (10)





RA/ESA Options

- Dual redundant controllers
- Dual redundant cache
 - up to 512MB/controller HSx80
- ↗ N+1 power
 - Requires an additional AC power distribution
 - Requires 3 additional 180-watt power supplies
 - May require additional AC power controller (depending on cabinet)
- 2x external cache batteries



ESA Options

- In Up to 2 additional BA370s can be daisy-chained behind a controller (or dual-redundant pair)
- This configuration supports up to 72 SBBs







Modula

- Jast architecture for MA/EMA product set
- → Uses new "Universal" drives
- Common package for all platforms





StorageWorks MA8000 and EMA12000

- Modular components: Separate controller and drive enclosures
- > 10 or 14 drives per SCSI bus (higher capacity/spindle count)

No single point of failure (redundant system level architecture plus storage software)

- Redundant, Hot-swappable components
- → Supports Universal Drive
- → 512MB Battery-Backed-up Cache Standard per Controller
- → Up to 84 drive configuration scalability. Additional subsystems can be attached for limitless capacity

Support for NT, NetWare, Unix, and OPS clusters; simultaneous access by non-clustered dissimilar/similar hosts

→ Windows NT, NetWare, IBM AIX, Sun Solaris, SGI IRIX, Tru64 Unix, OpenVMS, HP-UX support

→ Optical Fibre Channel Interface

 Data Replication Manager for NT, OpenVMS, Tru64 and Sun Solaris, and Netware

Configured to order



M2200 Controller Shelf



M2 Front View





14 (1") Disk Drive Enclosure



* Can be configured in "Single Bus"(4314) or "Split Bus" (4354) mode

4314 rear view



10 (1.6") Disk Drive Enclosure (obsolete)







Universal Drive Carrier

Carriers both for 3.5" disk drives

1.0" carrier

1.6" carrier (obsolete)

- » EMI containment
- Activity, fault and in use LEDS on front
- SCA II direct connect interface
- Cam mechanism locks carrier in place
- Designed to absorb higher vibrations and dissipate heat





Environmental Monitoring

- SWCC/Insight Manager Reporting
- Physical to logical mapping
- FRU identification
- Three temperature sensor
- Shelf status LED
- In-band reporting
- EMU hot swap



- Monitor fan and power supply presence and status
- Sets multiple fan speeds on high temperature and power or fan fault



Configuring Modula - Best Practices

- Jse all six buses for best performance
- → Use all six drive shelves in "Single Bus" mode connected to one M2200 controller shelf for highest capacity (84 disks)
- → Use 3 drive shelves in "Split Bus" mode for highest performance, medium capacity (42 disks), medium availability
 - → shelf failure could take out 2 channels
 - → low probability with dual redundant power
- → Use all six shelves in "split bus" mode between two M2200 controller shelves for highest availability, highest performance, medium capacity (84 disks)
- Spread stripes across channels





Configuring Modula - Best Practices

→ 2 controller pairs, 6 Split Bus shelves (84 Devices)



Controller A												2			
Bus #1	B	B		:	•	•		•		8			8	8	
Bus #2	:	8			•			•							
Bus #3	:	8		•	•			•						•	
Bus #4	8	8	8	8	•			:				8	8	8	
Bus #5	8	8	•	:	:	•	8	:			•	•	B	8	
Bus #6	8	8		•	:			:						•	
	0	1 1	2 3 Di	3 4 sk	D	5 8 ev	3 9 vice	9 1 2 S	0 5C	11 [5]	12 []	13 D:	} 14 S	4 1	5
														_	_

→1 controller pair, 6 Standard Bus shelves (84 Devices)

Model 2200 (M2)	<u>MA8</u>	000/EM	[A12000 F	<u>`ibre</u>	Cha	nnel Pi	roduct	Set ^{May}	7 12, 2000	EMA12000 S14
Controller Shelf - AU	Platform ki	ts 1 required per p	olatform <u>, adapters</u> 1 pe	r host con	nection.				1	
Controller Shell - 40	Platform kits 1 re	equired per platform	, adapters 1 per host con	nection.						
Front	O.S. Versions	G80 SFT Kit	FC Adapters(AKA);Bu	<u>15</u>	ACS	<u>Multi Path</u>	HA	FC	MA 8000	
	Support WNT X86 NT4.0/W Loop	V2K 380551-001	176479-B21(DS-KGPSA-CH	3);PCI	8.5G,F,S	Secure Path V2.2.2	MSCSV1.0	Switch,		
	N 1/1 4.0 Tru64 4 0F	128696-B21 380553-001	3805/4-001 (KGPSA-BC);F 168794-B21(DS-KGPSA-C	YCI A)·PCI	8.5P 8.5G F S	secure Path V2.2.2 in Tru64 5 0A	MSCSV1.0 TruCluster 1.6	Switch Switch		
Back	Tru64 UNIX 4.0F	(Configured and	d ordered through CSS only)	,	8.5P		Tructuster 1.0	Switch	A COMPANY OF THE REAL PROPERTY OF	
Duck	Tru64 UNIX 5.0A	128693-B21	168794-B21(KGPSA-CA);	PCI	8.5P	in OS (NOT YET	AVAILABLE)	Switch	A CONTRACTOR OF A DESCRIPTION	TOTORI SECONDARY
	OpenVMS 7.2-1 OpenVMS 7.2-1	380555-001 128694-B21	380574-001(KGPSA-BC):P	CA);PCI CI	8.5G,F,S 8.5P	in OS	OpenVMS Cluster	rs Switch Switch		
8	Solaris 2.6, 7.0	380554-001	123503-001(DS-SWSA4-SC	C); SBU	8.5G,F,S	Secure Path V2.1	FirstWatch onV2.6	Switch,		
	Loop			~				7.0	175992-B21/60Hz	175991-B21/60Hz
125920 D21	NetWare 4250	380559-001	380576-001(SWSA4-PC);P0 120186-B21	CI	8 5G F S		VCS V1.01 on V2.6 NCSV1.0	,/.0 Loon	(1) 22U Modular	(1) 36U Modular
133820-B21 Includes: (2) 180W Power Supplies (3)	SGI IRIX 6.5.3	380556-001	Origin 2000 SGI XT-FC-2P	:XIO	8.5G,F		1105 11.0	Loop	Storage Cab. (1) M2200 Shelf	(1) M2200 Shelf
Fans, (6) UltraSCSI I/O Modules, EMU,			Origin 200 SGI PCI-FC-IP;I	PCI (requires	X-F-OE-KI	(MIA))			(3) 4254 Shelves	(6) 4214 Shelves
(2) IEC C13-C14 Power Cords, RETMA	<u>HP-UX-10.2, 11.0</u> Switch Loop	380557-001	HP9000/X00 K Class A 3404	<u>A;HS(</u>	<u> </u>	$\frac{PVlinks in HP-UX}{Q}$	1.2 MC Service Guar	d	(6) SCSI VHD-VHD	(6) SCSI VHD-VHD
mounting kit		<u>C10 or</u>	HP9000/800 D Class A3391	A;GSC	10102-00	8 on the first fi	ne		Cables Controllers, Batteries	Cables Controllers, Batterie
Model 42xx Device Shelf - 3U	HSG80X4: Laty Co	ontrollensestationere ((14984 CB2taller) or RS/600	0	8.5G,F <u>M</u>	ulti-path Software	HACMP V4.3	Loop	and Disks ordered	and Disks ordered
	128697-B21 ACS V	8.5F FC-AL/SW Soft	ware		16	5989-B21 Secure P	ath V3.1 for WNT		separately	separately
	160091-B21 ACS V	/8.5S FC/SW Software	w/CLONE+SNAPSHOT		16	5991-B21 Secure P	ath v2.1 for SUN			
	128698-B21 ACS V	/8.5P FC/SW Data Rep	olication Software		н	A/F500 Clusters				
and provide the state of the st					В	asic 103250-B21			EMA12000 D14	EMA12000 Blue
103381-001 4214B Single Due 14 Dee Dueine Shelf										
RETMA mounting kit, (2) Blower				Sunno	rted Hard	Drives			ALLER STREET, T	
(1) Power Supply, (1) IEC-C13 to	Storage contro	ollers, cache, batt	eries:	328939-	-B22	9.1GB Ultra2 SC	CSI 10K RPM 1"			
NEMA 5-15P Power Cord, (1) 3.7 VHD VHD SCSI Cable	M176622-B21 HSG	80 RAID Array Contro	ller w/ 256MB Cache	142671-	-B22	9.1GB Ultra3 SC	CSI 10K RPM 1"			The second s
VIID - VIID SESI Cable	176623-B21 HSG	80 Cache bulkhead upg	grade	128418- 142673-	-B22 -B22	18.2GB Ultra2 SC 18.2GB Ultra3 SC	CSI 10K RPM 1" CSI 10K RPM 1"			
	135823-B21 Singl	le external cache batter	ý	123065-	-B22	9.1GB Ultra2 SC	CSI 7200 RPM 1"			A NEW YORK WARNING TO
	Shelf options			388144-	-B22	18.2GB Ultra2 SC	CSI 7200 RPM 1"			
119826-B21 4214R Redundant powe		4R Redundant power su	ipply	SCSI C	ables - VH	DCI to VHDCI		ALTERIAL CAL	an ag san si san san san si sa si sa sa	
129151 001	119829-B21 4214	4R Dual Bus I/O Modul	le	168256-B2	21 1Meter, 168	257-B21 2M, 189505-B2	1 3M, 400983-005 5M,	400985-010 10M		
4254R - Dual Bus 14 Bay Device Shelf				FC Cal	bles 50 Mic	ron Short Wave M	ulti mode		ter of sector any subsection and	
RETMA mounting kit, (2) Blowers,	Со	mpaq 9142 Rack, o	ptions:	234457	-B21/B22/E	323/B24/B25 Optica	al cables $(2/5/15/3)$	J/50m)		175003 B21/60Hz
(2) Power Supplies, (2) IEC-C13 to NEMA 5 15P Power Cords (2) IEC	120	663 B22 0142 42U C	ah (anal) No PDU Shock Pal	of	Modula	r Storage Cabinet			(1) 42U Modular	(1) 41U Modular
C13 to IEC-C14 Power Cords.	120	664-B22 9136 -36U Ca	ab (opal) No PDU- Shock Pal	e e	180311-1	321 42U (opal) 60Hz	w/ (26) IEC320-C13	- C14 Power	Storage Cab.	Storage Cab.
(2) 3.7M VHD - VHD SCSI Cables	Compaq 120	665-B22 9122 -22U C	ab (opal) No PDU- Shock Pal	Module	Cords	322 42U (opal) 50Hz	w/ (26) IEC320-C13	- C14 Power	(3) M2200 Shelves	(1) M2200 Shelf (3) 4254 Shelves
FC Storage Area Network Component	9142 1200	670-B21 Side Panel 673-B21 Rack Stabilize	ər	Storage	e Cords	522 420 (opai) 50112	w/ (20) IEC520-C15	- 01410wei	(9) 4254 Sherves (18) SCSI VHD-VHD	(6) SCSI VHD-VHD
FC SAN Switches	Rack 1200	669-B21 Multi-Bay Kit	t	Cabine	et 180313-1	321 36U (opal) 60Hz	w/ (22) IEC320-C13	- C14 Power	Cables	Cables
158223-B21 FC SAN Switch 16 ports	OPAL only 120	677/678-B21 Fan Kit 1	10/220V		Cords 180314-1	322 36U (opal) 50Hz	w/ (22) IEC320-C13	- C14 Power	Controllers, Batteries	Controllers, Batterio
158222-B21 FC SAN Switch 8 port	295	940-B21 Blank Panel K 363-002 PDU DOM 16	Kit (Qty 4) Amn - Zero II	CTO ava	ail. Cords	522 500 (opar) 50112	w/ (22) IEC520-C15	- 01410wei	and Disks ordered	separately
160407-B21 8/16 SAN Switch Blacket	2)5.	6-20P Pdu su	pports 6 p/s		180315-1	321 22U (opal) 60Hz	w/ (12) IEC320-C13	- C14 Power		
GBICs, Connection Kits	2953 CTO	363-003 PDU DOM 32	Amp - Zero U		Cords 180316-1	322 22U (opal) 50Hz	w/ (12) IEC320-C13	- C14 Power		
380561-B21 Optical Short Wave GBIC		12 ft L6-30P 1 Order 91xx with 1 or	Pau supports 6 p/s 2 PDU		Cords	dular Storage Cabs i	nclude: els (2) $2 \amalg 24 \text{ Amp P}$	DUs each with	Power Cord. IEC22	0 C13 to JEC220
380579-B21 Kit 3 SW GBIC, 2-2M cables	42U Rack	Note: Not all options s	hown	– '		121 41U (blue) 60Hz L6-30P input Power C	Cord, (16) TEC320-C1	- C14 Power 3 receptacles	C14	0-C13 10 IEC320-
127508-B21 Optical Long Wave GBIC	\bullet $\overline{\bullet}$				180318-I	322 41U (blue) 50Hz	w/ (26) IEC320-C13	3 - C14 Power	BN35S-02 2.5M IEC	cable, black



Setting Up the HSG80 Using the Command Line Interface



Setting Up the HSG80 Using the CLI

- Initial Controller Setup (HSG80)
- Creating Storagesets
- Creating Spares
- Creating Partitions
- All CLI Commands Listing



HSG80 Setup

- Connect console cable from controller maintenance port to VT terminal or emulator
 - Maintenance port accepts standard RS-232- jack from any EIA-423 compatible terminal
 - Port support serial communications with default values of 9600 baud using 8 data bits, 1 stop bit, no parity
 - If possible use "good" emulator (ie Keaterm, Powerterm, etc..)
 - Hyperterm has limitations
 - Press "Enter" or "Return" key CLI prompt appears
 - When entering commands in a dual-redundant controller configuration, remember that the controller to which you are connected is "this_controller" and the remaining controller is the "other_controller"


- If configured for dual-redundant controllers, set the controllers CLI prompts. This may eliminate any confusion later
 - - commands may be abbreviated
 - HSGtop> set other prompt="HSGbottom> "



At command prompt type

HSG> show this controller

- Note condition of cache and batteries both should be good. If cache is invalid - CLI will be reduced
- Note that controller is configured for dual redundancy
- HSG> show other_controller
- If error accessing other controller then move the CLI cable to the other controller

(This could occur if the controllers are not configured for dual redundancy or cache is invalid)







7 HSG> show this controller full

Controller:

```
HSG80 (C) DEC ZG7400120 Software R052G-0, Hardware 0000
   Node ID = 5000-1FE1-FF00-00B0 (set this node id = 5000-1fe1-ff00-00b0 fc)
   ALLOCATION CLASS = 0 (set this allocation lass=1)
   SCSI VERSION = SCSI-2 (set this scsi version=scsi-2)
   Not Configured for dual-redundancy (set failover copy=this)
   Device Port SCSI address 7
   TIME: NOT SET (set this time= dd-mmm-yyyy:hh:mm:ss)
   Command Console LUN is 0 (set this command console lun)
Host PORT 1:
   Reported PORT ID = 5000-1FE1-FF00-00B1
   Port 1 PROFILE = PLDA
   PORT 1 TOPOLOGY = LOOP HARD (loop up) (set this port 1 topology = loop hard)
   PORT 1 AL PA = 71 (set this port 1 alpa=71)
Host PORT 2:
   Reported PORT ID = 5000-1FE1-FF00-00B2
   Port 2 PROFILE = PLDA
   PORT 2 TOPOLOGY = LOOP SOFT (loop up)
   PORT 2 AL PA = 02 (negotiated)
```



7 HSG> show this_controller full (continued)

Cache:

256 megabyte write cache, version 0012 Cache is GOOD No unflushed data in cache CACHE_FLUSH_TIMER = DEFAULT (10 seconds) (set this cache_flush_timer=65535) NOCACHE_UPS (set this nocache_ups)

Mirrored Cache:

Not Enabled (set this mirrored_cache)

Battery:

MORE THAN 50% CHARGED Standby Capacity: Less than one hour Time to full charge: 31 hours Expires: 23-AUG-1957 WARNING: BATTERY AT END OF LIFE WITHIN ONE WEEK, REPLACE BATTERY SOON!

Extended Information:

Terminal speed 19200 baud, eight bit, no parity, 1 stop bit Operation control: 00000001 Security state code 33506 Configuration backup disabled Other controller not responding - RESET signal NOT asserted -Temperature within optimum limit



HSG80 Setup - Host Name

Host domain for fibre channel uses World Wide Names

- Use SHOW THIS command to see the controller World Wide Name
- ✓ If the controller reports a node WWN of all zeros (0000-0000-0000-0000) set the subsystem WWN to the IEE WWN that came with the controller located on a sticker on the frame
- ↓ If replacing a failed controller in a dual-redundant configuration, the remaining controller remembers the subsystem WWN

✓ All ports on controllers have unique World Wide Names

Save the configuration to a storageset or disk (especially in a single controller configuration)



HSG80 Setup Host Ports

The HSG80 has 2 host ports

Set the port topology for each port SET THIS CONTROLLER PORT 1 TOPOLOGY = "topology" Topology settings: $\sqrt{Loop}_hard - allows$ you to pick the ALPA Specify the ALPA for the host ports set this_controller port_1 ALPA="address" Loop_soft - the controller will pick the ALPA Ψ Fabric - for switched fabric environments $\mathbf{VOffline}$ - specify when the port is not to be used → Examples: HSG80> set this controller port 1 topology =

fab<mark>r</mark>ic



HSG80 Setup - Finding Devices

Run the "config" program which will will add all the disks to the configuration

HSG> run config

CONFIG LOCAL PROGRAM INVOKED

CONFIG IS BUILDING ITS TABLES AND DETERMINING WHAT DEVICES EXIST ON THE SUBSYSTEM. PLEASE BE PATIENT

ADD DISK10000 1 0 0

- ADD DISK10100 1 1 0
- ADD DISK20000 2 0 0
- ADD DISK20100 2 1 0
- ADD DISK30000 3 0 0
- ADD DISK30100 3 1 0
- ADD DISK40000 4 0 0
- ADD DISK<mark>401</mark>00 4 1 0
- ADD DISK50000 5 0 0
- ADD DISK50100 5 1 0
- ADD DIS<mark>K6</mark>0000 6 0 0
- ADD DISK<mark>6</mark>0100 6 1 0

CONFIG NORMAL TERMINATION



BA370 SCSI Device Domain



Device address 3/11/15

Device address 2/10/14

Device address 1/9/13

Device address 0/8/12

PVA address 0/2/3 Device address 6 Device address 7



Target numbers

BA370 Device Port Target LUN (PTL)

Device port numbers



Figure 3–10

PTL Addressing in an Extended Configuration

PTL location = Device port number = 3



BA370 - Setting the PVA Switch



PVA switch setting 1 must <u>not</u> be used as this would assign target ids 4, 5, 6, & 7 to the device shelves. This would cause a conflict with the HSG80 controller target Ids of 6 and 7.



Creating Storagesets

- Decide how many and what type of storage sets to create based on customer/application requirements.
- What are storagesets?
 - Raid 0 = Stripeset
 - Raid 1 = Mirrorset
 - Raid 0+1 = Striped Mirrorset
 - Raid 3/5 = Raidset
 - JBOD = Just a Bunch of Disks
- All stripes should (but don't have to) go across SCSI channels.



Creating Storagesets - Stripeset

7 You can have up to 24 disks in a stripeset 7 There are no stripeset switches 7 Syntax ADD STRIPESET stripeset-name containername1 container-name2 [container-nameN] 7 Examples: HSG> add stripeset stripe1 disk10000 disk20000 HSG> add stripe s2 disk11200 disk21200 disk31200



Creating Storagesets - Mirrorset

7 You can have up to 6 disks in a mirrorset

- Mirrorset switches:
 - \$\$\psystample copy=fast, normal (default)
 - vpolicy=best_fit, best_performance (default)
 - ↓nopolicy

vread_source=disk-name, least_busy (default), round_robin

↗ Syntax

ADD MIRRORSET mirrorset-name disk-name1 [disk-nameN]

→ Examples:

HSG> add mirrorset mirror1 disk10000 disk20000



Creating Storagesets - RAIDset

You can have up to 14 disks in a RAIDset

RAIDset switches:

vreconstruct=fast, normal (default)

- vpolicy=best_fit, best_performance (default)
- ↓nopolicy
- ↓reduced

↓noreduced (default)

- ↗ Syntax
- ADD RAIDSET raidset-name disk-name1 [disknameN]
- → Examples:
- HSG> add raidset raid1 disk10000 disk20000 disk30000



Creating Storagesets - Initializing

IVSE INITIALIZE command when:

 Creating a unit from a newly installed disk
 Creating a unit from a newly created RAIDset, stripeset, or mirrorset

Initializing the data structure of a previously partitioned container

Do not use the INITIALIZE command when:

Creating a unit from the same disks previously initialized, such as when a RAIDset is moved
 Creating a storageset from deleted members
 Adding a RAIDset with the REDUCED switch



Creating Storagesets - Initializing

Switches:

↓ chunksize=default,n

- specifies the block chunk size to be used for stripesets and RAIDsets
- the default chunksize for storagesets with less than 9 members is 256 blocks, or 128 kbytes.
- the default chunksize for storagesets with more than 9 members is 128 blocks, or 64 kbytes
- a chunksize less than 64k is not recommended
- accept the default for most applications
- see HSx Configuration Manual for information regarding recommended chunk size settings
- does not apply to mirrorset
- √des<mark>tro</mark>y (default)

✓ nodestroy

- to preserve forced error metadata during init
- use only when a unit is to be created from disks reduced from mirrorsets
- valid only with stripesets and mirrorsets



Creating Storagesets - Initializing

Switches:

↓save_configuration

- instructs the controller to save the controller's configuration information to the container being initialized
- intended for use with a single controller
- if specified for multi-device storageset, the complete controller configuration is stored on each member of the storageset
- a disk initialized with the save_configuration will have slightly less storage available for data

✓Syntax:

↓INITIALIZE container name

 Ψ Examples:

HSG> initialize disk30300 save_configuration



Creating Storagesets - LUNS

Logical Units (LUNs) are created from storagesets. Storagesets are created from disks.



CXO5564A



Creating Storagesets - LUNS

Use the ADD UNIT command to create the LUN

Switches for Units

enable_access_path/disable_access_path (WWN of host that can see this LUN)
identifier/noidentifier
maximum_cached_transfer (default = 32, max=1024)
preferred_path/nopreferred_path (this_controller or other_controller, used in mulitibus failover mode)

√read_cache/noreadcache

√run/norun (availability to host)

write_protect/nowrite_protect

vreadahead_cache/noreadahead_cache

writeback_cache/nowriteback_cache



Creating Storagesets - LUNS

→ Examples:

HSG> add unit d102 disk10000
 preferred_path=this_controller
HSG> add unit d0 raid3 writeback_cache
 maximum_cached_transfer=1024
HSG> add unit d100 stripe3
 enable_access_path=(!newcon11,!newcon12)



Creating a RAIDset and LUN

HSG> add raidset raid1 disk10500 disk20500 disk30500 HSG> init raid1 HSG> add unit d105 raid1 writeback cache





Creating a Stripe Mirrorset and LUN

Create the mirrorsets HSG> add mirror m1 disk10100 disk20000 HSG> add mirror m2 disk20100 disk30000 HSG> add mirror m3 disk30100 disk40000 Create the stripeset HSG> add stripe s0 m1 m2 m3 Save configuration data (to at least one storageset) HSG> init s0 save configuration Create the LUN HSG> add unit s0 d0 writeback cache



Creating Spares

Create the spare from available disks ↗ Use ADD SPARESET command \downarrow no switches **↗** Syntax: ADD SPARESET disk-name → Examples: HSG> add spareset disk20300 HSG> add spareset disk40300 Sparing policies are specified by storagesets Specify "best fit" or "best performance" HSG> add raidset r1 disk40200 disk50200 disk60200 policy=best performance



Creating Partitions

The CREATE_PARTITION command

- divides a non_transportable disk drive or multiple disk storageset into several, separately addressable storage units
- Imarks a specified percentage of a disk drive or storageset to be used as separate units
- ↓you can create up to 8 partitons per storageset
- the individual units can be presented back to different hosts
- ✓ Supported in "multi-bus failover" mode (sometimes).



Creating Partitions

7 Switches: size=percent, largest ↗ Syntax: CREATE_PARTITION container-name SIZE=percent → Examples: HSG> add raidset raid9 disk10000 disk20000 disk30000 init raid9 HSG> create partiton raid9 size=25 HSG> create partiton raid9 size=25 HSG> create partiton raid9 size=25 HSG> create partiton raid9 size= largest HSG> add unit d0 raid9 partiton=1 HSG> add unit d1 raid9 partiton=2 HSG> add unit d2 raid9 partiton=3 HSG> add unit d3 raid9 partiton=4



Failover Modes

Dual-redundant HSG80 pairs must operate in one of two failover modes.
 Transparent
 Multiple Bus
 Does not apply to single controller configurations



Failover Modes - Transparent

7 Transparent

LUNs (transparently to the host) failover to the surviving controller in the event that a controller is taken offline or has failed.

HSG80> set failover copy=this_controller

	EMU	PVA				
HSG80 Controller A Por	t 1 = Active 🕅	Port 2 = Stand	Ьу			
HSG80 Controller B Por	t 1 = Standby 📩	Port 2 = Active				
	A Cache	B Cache				



Transparent Failover

- If using "Transparent Failover"
 - Set your controllers for transparent failover <u>before</u> configuring devices
 - When creating LUNS, DO-D99 "appear" on port 1 and D100-D199 "appear" on port 2





Failover Modes - Multiple Bus

Multiple Bus bus

↓Also known as Host Assisted failover

- The host controls failover of the LUNs
- \checkmark The host must have 2 host bus adapters
- ↓ All 4 host ports are "active"

HSG80> set multibus_failover copy=this_controller







Load Balancing LUNS

- If using "Multi-bus" or host assisted failover
 - Set your controllers for multi-bus failover before configuring devices
 - ↓ All LUNs appear on all ports
 - The host distributes the I/O between controllers and available ports
 - Load balancing may be "Static" or "Dynamic"
 - The host should have at least 2 fibre channel adapters as well as operating system software support
 - Secure Path for Windows NT
 - Secure Path for Sun Solaris
 - Secure Path for AIX
 - Secure Path for Netware
 - Native support built in to Tru64 UNIX and OpenVMS



- SSP is provided as part of the ACS
- Allows LUNs to be "selectively" presented to hosts based on the WWN of the host bus adapter
- Uses connection table in the HSG80
- Connection table holds 96 entries
- Table provides a map for WWN to user specified name
- Connection names appear dynamically as the hosts logs into the fabric unless you've locked the connection table (8.7 new feature).
- Default format for name in connection table is INEWCONXX (where XX gets incremented for every new connection)
- Re-name the connection to something more logical



Displaying the Connection Table and re-naming connections

HSG80> show c	onnections	5				
Host Name	Operating	Controller	Port	Address	Status	O <mark>ff</mark> set
	System					
!NEWCON00	WINNT	THIS	1	000000	OFFLINE	0
HOST_	ID=1000-00	<mark>)00-</mark> C920-273E	C	PORT_ID	=1000-0 <mark>00</mark>	<mark>0</mark> -C920-273E
!NEWCON01	WINNT	OTHER	2	000000	OFFL <mark>INE</mark>	0
HOST_	ID=10 <mark>00-0</mark> 0	00-C920-345E	ŗ	PORT_ID	=10 <mark>00-0</mark> 00	0-C920-345F
!NEWCON02	VMS	THIS	1	000000	ONLINE	10
HOST_	ID= <mark>100</mark> 0-00	000-C920-0591	L	PORT_ID	<mark>=10</mark> 00-000	0-C920-0591
!NEWCON03	VM <mark>S</mark>	OTHER	2	000000	ONLINE	10
HOST_ID=1000-0000-C920-0655 PORT_ID=1000-0000-C920-065					0-C920-0655	
HSG80> rename	!newcon00) NTSQL1_A				
HSG80> rename	newcon01	L NTSQL1_B				
HSG80> rename	!newcon02	2 VMSCLS1_A				
HSC80> rename	Inewcon03	NMSCLSI B				



7Assign access to the LUN by using the connection
name
HSG80> set D0 disable_access=all
HSG80> set D0 enable access=(ntsql1 a,ntsql1 b)

¬Assign the operating system type to the
connection
HSG80> set connection vmscls1_a
operating system = vms



">>Use "Offsets" with the connections for logical grouping of LUNS and hosts

7 i.e.. NTSQL1 offset = 0
7 use LUNs D0-D9
7 D0 becomes Logical Unit 0 to NT
7 VMSCLS1_A offset = 10
7 use D10-D19

↗D10 becomes Logical Unit 0 to VMS

Relationship between LUN number, unit number, and offset is as follows:

7LUN number = unit number - offset

7Logical unit number or LUN number = the logical unit number
presented to the host connection

7Unit number = the number assigned to the unit in the ADD UNIT command

HSG80> set connection vmscls1_a unit_offset=10


HSG80 Storagesets

- Created from individual disks
- Define underlying RAID level for logical units





HSG80 Logical Units

- Created from storagesets
- Presented to host to be mapped as a SCSI disk device
- Presented as 128 bit WWID to hosts
- Use Logical Unit Numbers D0-D199





HSG80 Selective Storage Presentation

 Selective Storage Presentation (SSP) is used to select which host(s) can access (map) the LUN

 Access_path is based on the connection ID from the WWID of host bus adapter (HBA)





HSG80 Partitioned Storagesets

 Multiple LUNs can be created from one storageset (up to 8 partitions per storageset)

- Each LUN can be presented to a different host
- Minimizes RAID5 parity cost





Presenting Logical Units to Hosts



Host WWIDs from HBA become connections to controller
 Controller uses connections for Selective Storage Presentation





All CLI Commands

7 ADD CDROM **ADD DISK ADD MIRRORSET 7 ADD PASSTHROUGH ADD RAIDSET ADD SPARESET ADD STRIPESET ADD UNIT** 7 CLEAR ERRORS CLI **7 CLEAR ERRORS UNKNOWN 7 CREATE PARTITION DELETE FAILEDSET**



All CLI Commands

↗ DELETE SPARESET **DESTROY PARTITION 7** DIRECTORY **→** HELP **7** INITIALIZE *ILOCATE* **MIRROR REDUCE RENAME 7 RETRY ERRORS UNWRITEABLE DATA RUN →** SET EMU



All CLI Commands

SET FAILEDSET
SET FAILOVER
SET NOFAILOVER
SHOW
UNMIRROR



HSG Controller Utilities



Topics:

7Overview **7**Utilities **⊅DILX ∧**VTDPY *¬CONFIG* **⊅**CLCP **FRUTIL HSUTIL** *TCLONE* **⊅FM**U



Hsx Utilities: Overview

⁷Used to perform maintenance, diagnostic, and performance monitoring on the HS series controllers

Part of HSOF/ACS controller software

Invoked from CLI using "RUN" command

¬Use "DIR" at CLI prompt to see what utilities are available

Not supported by remote CLI (SWCC)

Some utilities may work though

7Usage much too detailed for this seminar 7See HS Controller Service Guide for more information



Utilities: DILX (Disk Inline Exerciser)

Ised to Check the data transfer capability of disk drives

Used to initialize physical disk

¬Good way to wipe out NT signature

Generates intensive read/write loads to the disk drive while monitoring the drives performance and status

Four tests available (read, write, data xfer, seek)

Ability to test as many drives as available

HP recommends stopping host based I/O during tests

> Writes WILL destroy user data



Use to monitor controller performance
 Communication between controller and host
 Communication between controller and devices
 State and I/O activity of the logical units, devices, and device ports in the subsystem

Detailed descriptions on how to interpret the display output can be found in the HSG80 Service Manual



To initiate VTDPY from a maintenance terminal at the CLI> prompt, enter the following command:

HSZ70> RUN VTDPY

Use the following control key sequences with the VTDPY display:

VTDPY control keys	
Control Key Sequence Function	
Ctrl/C	Prompts for commands.
Ctrl/G	Updates the screen (same as Ctrl/Z).
Ctrl/O	Pauses or resumes screen updates.
Ctrl/R	Refreshes current screen display (same as Ctrl/W).
Ctrl/W	Refreshes current screen display (same as Ctrl/R).
Ctrl/Y	Terminates VTDPY and resets screen characteristics.
Ctrl/Z	Updates the screen (same as Ctrl/G).

While VTDPY and a maintenance terminal interface support passing all of the listed control characters, some host-based terminal interfaces restrict passing some of the characters. Using the VTDPY Command Line

VTDPY contains a command line interpreter that you can invoke by entering Ctrl/C any time after starting the program. The command line interpreter is used to modify the characteristics of the VTDPY display.



VTDPY commands Command String DISPLAY CACHE

DISPLAY DEFAULT

DISPLAY DEVICE

DISPLAY STATUS EXIT INTERVAL <seconds> HELP REFRESH QUIT UPDATE Function

Use 132 column unit caching statistics display.

Use default 132 column system performance display.

Use 132 column device performance display.

Use 80 column controller status display. Terminates program (same as QUIT).

s> Changes update interval.

Displays help message text.

Refreshes the current display.

Terminates program (same as EXIT). Updates screen display.



VTDPY NOTES:

The keywords in the command strings can be abbreviated to the minimum number of characters necessary to uniquely identify the keyword. Entering a guestion mark (?) after a keyword causes the parser to provide a list of keywords or values that can follow the supplied keyword. The command line interpreter is not case sensitive, so keywords can be entered in uppercase, lowercase, or mixed case. Upon successful execution of a command other than HELP, the command line interpreter is exited and the display is resumed. Entering a carriage return without a command also exits the command line interpreter and resumes the display. If an error occurs in the command, the user prompts for command expansion help, or the HELP command is entered, the command line interpreter prompts for an additional command instead of returning to the display.



Utilities: CONFIG (Configuration Utility)

Ise to add one or more storage devices to the subsystem

Utility checks the device ports for new disk drives then adds them to the controller's configuration automatically names them
If you're adding new disk drives to a BA370 or a Modula shelf, just run CONFIG and it will automatically add the devices to the subsystem

correctly



Utilities: CLCP (Code Load Code Patch)

Use CLCP to download new software to the program card while it's installed in the controller
Uses Kermit software
If too complex for you, get a newer version
PCMCIA card

Obtain image from customer services representative

Zoad the image onto a PC or workstation

From the host, quiesce all port activity and dismount the storage units in the subsystem

Disable write-protect switch on PCMCI card

Run CLCP



Utilities: FRUTIL (Field Replacement Utility)

Use the field replacement utility to replace a failed controller, cache module, cache battery, EMU, or PVA module without shutting down the subsystem
 Dual-redundant controller configurations are assumed

Allows quiescing of device SCSI buses during swap of failed module



Utilities: HSUTIL (Hierarchical Storage Utility)

¬Use HSUTIL to upgrade the firmware on disk and tape drives

Drives must not be members of units

Drives not available to the host during upgrade

>Use to perform low level format to disk drives



Utilities: CLONE

- Use to duplicate the data on any unpartitioned single-disk unit, stripeset, mirrorset, or striped mirrorset
- When the cloning operation is done, you can back up the clones rather than the storageset or single disk unit, which can continue to service its I/O load

When cloning a mirrorset, CLONE does not need to create a temporary mirrorset. Instead, it adds a temporary member to the mirrorset and copies the data onto this new member



Utilities: CLONE

- CLONE creates a temporary, two member mirrorset for each member in a single-disk or stripeset
- After the CLONE utility copies the data from the members to the clones, it restores the unit to its original configuration and creates a clone unit you can backup



Utilities: FMU Fault Management Utility

Provides a limited interface to the controller's fault management software

⊅Use FMU to:

Display the last-failure and memory-system-failure entries that the fault management software stores in the controller's non-volatile memory

Translate many of the code values contained in event messages

For example, entries may contain code values that indicate the cause of the event, the software component that reported the event, the repair action, etc..

Control the display characteristics of significant events and failures that the fault management system displays on the maintenance terminal



Utilities: FMU Fault Management Utility

Using FMU to describe event log codes

The Fault Management Utility (FMU) has a DESCRIBE function you can use to interpret event codes produced by the controller. Use this function to understand events that have occurred in the subsystem (instance, codes) and to find the recommended repair action (repair action codes), as well as to interpret other codes. The types of codes FMU can describe are:

- · INSTANCE,_CODE
- · REPAIR_ACTION_CODE
- · LAST_FAILURE_CODE
- · ASC_ASCQ_CODE
- · COMPONENT_CODE
- · CONTR<mark>OLL</mark>ER_UNIQUE_ASC_ASCQ_CODE
- · DEVICE_TYPE_CODE
- · EVENT_THRESHOLD_CODE
- · RESTART_TYPE
- · SCSI_COMMAND_OPERATION_CODE
- · SENS<mark>E_</mark>DATA_QUALIFIERS
- · SENSE_KEY_CODE
- · TEMPLATE_CODE



Utilities: FMU Fault Management Utility

FMU Output Example

HSZ50> RUN FMU Fault Management Utility FMU> SHOW LAST_FAILURE MOST_RECENT Last Failure Entry: 1. Flags: 000FF301 Template: 1.(01) Description: Last Failure Event Occurred on 07-DFC-1995 at 09:21:44 Controller Model: HSZ50 Serial Number: ZG51909900 Hardware Version:0000(00) Controller Identifier: Unique Device Number: 01519090 Model: 40.(28) Class: 1.(01) Firmware Version: W18J(FF) Node Name: "HSZA3" CI Node Number: 12.(OC) Informational Report Instance, Code: 01010302 Description: An unrecoverable hardware detected fault occurred. Reporting Component: 1.(01) Description: Executive Services Reporting component's event number: 1.(01) Event Threshold: 2.(02) Classification: HARD. Failure of a component that affects controller performance or precludes access to a device connected to the controller is indicated. Last Failure Code: 018800A0 (No Last Failure Parameters) Last Failure Code: 018800A0 Description: A processor interrupt was generated with an indication that the program card was removed. Reporting Component: 1.(01) Description: Executive Services