SCSI Solutions White Paper – HP-UX

June 15, 2005

Executive Summary 2

1 SCSI Overview 2

2 HP Solutions 5

3 Accessories 13

4 Configuration 15

5 High Availability Configurations 16

6 Storage Management 18

7 References 20

8 For more information 21

Version 1.0

Executive Summary

This paper provides an overview of parallel SCSI (pSCSI) technology and solutions on HP-UX. A basic understanding of mass storage technologies is assumed. Topics discussed include:

- pSCSI history and futures
- HP pSCSI products and solutions
- Configuration rules and guidelines
- Online repairability and High Availability (HA)
- Enclosure management
- Performance and feature comparisons

1 SCSI Overview

1.1 History

SCSI (Small Computer System Interface) is an ANSI standard for interconnecting computers to peripheral devices such as disks, tapes, CD ROMs, and autochangers. In its original form, known as SCSI-1, it was a narrow (8-bit) parallel bus with a 5 MB/second maximum transfer rate which supported only single-ended transmission with passive termination.

SCSI-2 provided a higher-speed option known as Fast SCSI, a wide (16-bit) bus option known as Wide SCSI (which allowed for doubling the number of devices per bus that could be attached, and a doubling of the raw performance, which together with the Fast SCSI option allowed for quadrupling of the raw performance), improved connectors and cabling, more reliable termination known as Active Termination, differential signaling (which later became known as High Voltage Differential – HVD) to allow for longer cable lengths, command queuing for increased performance (pipelining requests to a device) and support for additional types of devices such as removable media, CD-ROMs, and scanners.

SCSI-3, the latest version of the standard, expanded the SCSI horizons beyond the parallel bus to serial technologies, such as Fibre Channel and more recently iSCSI and SAS, separating the underlying transport technologies from the SCSI command protocol. On the parallel bus, SCSI-3 brought significant performance and feature improvements, initially defining Ultra SCSI at double the Fast SCSI transfer rate, and then successively Ultra2, Ultra3 (also known as Ultra160), and Ultra320 SCSI, each doubling the raw speed of the bus.

1.2 Key Features

The basic features evolved with the data transfer rate options, as shown in the table below. In this context MB/s is 1,000,000 bytes per second, and MT/s is 1,000,000 Transfers per second. On a narrow (8-bit) bus 1 MT/s = 1 MB/s; on a wide (16-bit) bus 1 MT/s = 2 MB/s. For all speeds other than Ultra160 and Ultra320, 1 MT/s = 1 Mhz clock speed on the bus. Ultra160 and Ultra320, however, use double-transition clocking (sampling data on both the rising and falling edges) to do two transfers per clock cycle.

There are three types of electrical signaling used in SCSI buses:

- Single-Ended (SE) – an electrical signal protocol that transmits information through changes in voltage. Single-ended SCSI uses standard TTL signal-to-ground pairs to transmit information over the SCSI bus. It is a low cost, low power solution, but the signal quality degrades rapidly which limits cable length to a maximum of 6 meters.
- High-Voltage Differential (HVD) – an electrical signal protocol that transmits information through a current loop rather than by changes in voltage, thereby reducing the susceptibility to electrical interference. As a result, cable lengths can be much longer, up to a maximum of 25 meters.
- Low-Voltage Differential (LVD) – uses a differential scheme like HVD, but at a lower voltage level and hence at a lower cost and power level but at a shorter maximum cable length (12 meters with more than one target and initiator, as shown in Table 1).

HVD signaling is expensive, uses more power, and is electrically incompatible with either SE or LVD signaling. Mixing LVD/SE devices with HVD devices on the same bus can result in actual physical damage because of the higher voltages that might be sent to the SE or LVD devices. LVD signaling is compatible with SE and can be mixed on the same bus, but the maximum speed and cable lengths reduce significantly when SE devices are used, as shown in Table 1.

Note: Use of a Single-Ended (SE) device on a bus will result in the entire SCSI bus running in SE mode at a maximum of 40 MB/s.

The following table outlines the various features that have evolved with SCSI over time.

Table 1: Key Features

SCSI Specification		Bus Speed	Max Data Rate (MB/s)	Bus Width (Bits)	Signaling Method	Max Bus Length (meters)[5]	Max # of SCSI IDs	Term[3]	Cables[4]
SCSI-1		**Slow (5 Mhz)**	5	8	SE	6	8	P/A/F	"A"
SCSI-2		**Fast (10 Mhz)**	10	8	SE	3	8	A/F	"A"
				8	HVD	25	8	HVD	
			20	16	SE	3	16	A/F	"P"
				16	HVD[1]	25	16	HVD	
SCSI-3	**SPI** ‡	**Ultra (20 Mhz)**	20	8	SE	1.5 / 3	8 / 4	A/F	"A"
				8	HVD	25	4	HVD	
			40	16	SE	1.5 / 3	16 / 4	A/F	"P"
				16	HVD	25	4	HVD	
	SPI-2	**Ultra2 (40 Mhz)**	40	8	LVD[2]	12 / 25	8 / 2	LVD	"A"
			80	16	LVD[2]		16 / 2		"P"
	SPI-3	**Ultra160 (40 Mhz)***	160	16†	LVD				
	SPI-4	**Ultra320 (80 Mhz)***	320	16†	LVD				

‡ SPI = SCSI Parallel Interface.

[1] Also known as Fast Wide Differential (FWD).

[2] Only LVD is shown here with Ultra2; HVD was also officially supported for Ultra2 SCSI, but was not generally used.

[3] Termination: P/A/F = Passive/Active/Forced-perfect, HVD=HVD termination, LVD=LVD termination

[4] Cables: "A" cable (50 pin), "P" (68 pin)

[5] Max bus length: 1.5 / 3 = max of 3 meters if no more than 4 SCSI IDs are used, otherwise 1.5 meters
12 / 25 = max of 25m if no more than 2 SCSI IDs (point-to-point), otherwise 12 meters

† Ultra160/320 buses are wide-only.

* Ultra160 and Ultra320 use double-transition (DT) clocking – sampling data on both the rising and falling edges of the clock – to double the MB/s throughput per Mhz.

1.3 Futures

The industry has agreed due to various problems inherent in parallel bus technology that there will be no progression of parallel SCSI beyond Ultra320. The parallel bus has inherent electrical issues with the timing of the various bits being transferred in parallel (skew), electromagnetic interference among the various parallel lines (crosstalk), and related issues which become increasingly difficult to address at higher speeds. As a result the SCSI Trade Association (STA) agreed that Ultra640 technology will not be pursued. Instead, serial technologies such as SAS, SATA, Fibre Channel, and iSCSI will provide the next generation of products beyond Ultra320 SCSI.

Note: SCSI technology continues to expand into the serial realm. pSCSI solutions will continue to be provided and supported for years to come, and will overlap with the solutions that future SCSI technologies provide.

For additional information on the SCSI Specifications and technology development, go to www.t10.org.

2 HP Solutions

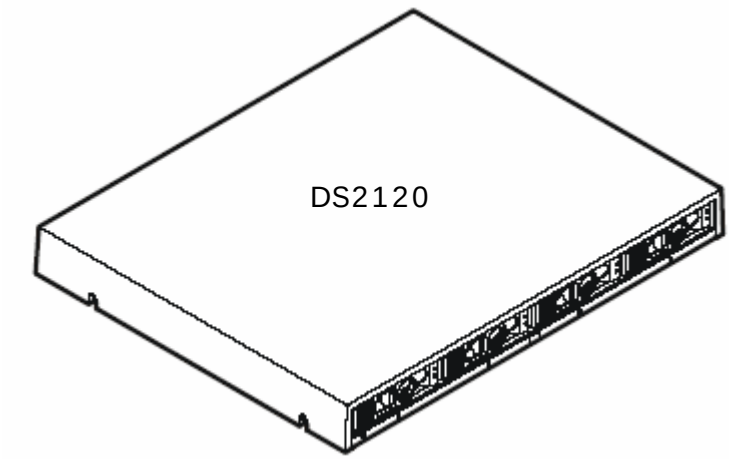
HP supports a range of pSCSI solutions for HP-UX servers, including disk storage enclosures, tape products, optical jukebox devices, and Host Bus Adapters (HBAs). Each of these is discussed below.

2.1 Disk Products

The pSCSI disk products consist of disk drives and enclosures. The enclosures house multiple disk drives, and provide power supplies, fans, hot-swappable drive slots, and other management and availability features. There are two main classes of pSCSI disk enclosures currently available on HP-UX: the Disk System (DS) 21XX Series, and the Modular Smart Array (MSA) 30 Series. The DS21XX products are smaller, lower-cost enclosures, while the MSA30 products are larger, more intelligent enclosures. An overview of each is provided below.

2.1.1 DS2120 Enclosure

The newest DS Series enclosure, the DS2120, is an Ultra160 SCSI 4-disk enclosure. It is intended for customers that need low-cost external non-HA storage without enclosure management or redundant or hot-swappable enclosure components (fans, power supplies, I/O modules). The disk drives are hot-swappable, and supported drives range from 36 to 300 Giga Bytes (GB). It is designed to replace the recently discontinued DS2100 and DS2110 enclosures. The DS2120 comes in a compact 1U package, as seen below from a top-front-left view, with the 4 disk drive module units accessible from the front.



The DS2120 is supported on HP-UX 11.00, 11i v1, and 11i v2 across a range of platforms and HBAs, and is supported as a boot or dump device in most of those configurations. For OS versions, platforms, and HBAs supported with the DS2120, go to www.hp.com/go/serverconnectivity and click on StorageWorks support matrices.

Note: The DS2120 is not supported with the SCSI RAID Controller cards (see section 2.4.2).

The DS2120 is an Ultra160 enclosure, but is supported with Ultra320 HBAs and drives. When used with Ultra320 HBAs, the bus must be speed-limited to Ultra160 speeds[1]. As of AR0505 of

[1] To limit the speed see the sections on "Setting the SCSI Parameters" and "About the Maximum Data Transfer Rate" in the A7173A HBA Support Guide available at http://docs.hp.com/en/J6373-90008/J6373-90008.pdf.

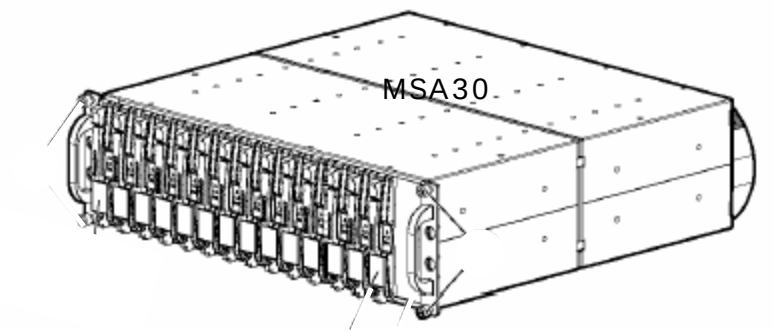
HP-UX 11i v2, HP-UX is delivering self-healing software in the Ultra320 HBA driver which will auto-detect signal integrity issues and downshift to Ultra160 speeds as needed.

Other features of the DS2120 include:

- 3.6TB of total storage per bus: up to three DS2120 enclosures can be daisy-chained in single-initiator configurations (providing a total of up to 3.6TB of storage on a single SCSI bus), and up to two DS2120's can be daisy-chained in multi-initiator configurations (providing up to 2.4 TB of storage per multi-initiator bus).
- Simple, low-cost, easy to install, configure, and support; a good solution for low-end root/boot/swap/dump support and host-based HA solutions such as MirrorDisk/UX.
- Flexible: supports mixed disk capacities.
- Multi-initiator (dual host connect) capability

2.1.2 MSA30 Enclosures

The MSA30 products are Ultra320 SCSI 14-disk enclosures. They are intended for customers that need external HA storage, enclosure management, full Ultra320 speeds, or SCSI RAID Controller support. They provide redundant and hot-swappable fans and power supplies, hot-swappable disk drives, and warm-swappable[2] I/O modules and SCSI cables. The MSA30 enclosures come in a 3U package, as shown below in a top-front-right view, with the 14 disk drive module units accessible in the front.



The MSA30 provides enclosure management support to monitor environmental conditions within the enclosure and components, which can be accessed and controlled using EMS tools[3]. These enclosures are also fully supportive of Ultra320 speeds, drives, and HBAs.

The MSA30 comes in three flavors: the MSA30 SB (Single Bus), MSA30 DB (Dual Bus), and MSA30 MI (Multi-Initiator), based on the type of I/O module in the back of the enclosure. The SB I/O module provides a "full bus" mode of operation, in which all 14 drives are attached to a single bus accessible via a single SCSI port. The DB and MI modules provide a "split bus" mode of operation, splitting the enclosure into two separate SCSI buses, with 7 drive slots on each bus. The DB I/O module contains two SCSI ports, one for each bus; the MI contains four SCSI ports, two for each bus.

The MSA30 SB and DB only provide for the connection of one initiator per SCSI bus, and therefore do not support multiple paths to a disk. If a path goes down and the data isn't mirrored or otherwise replicated on another bus (which in the DB case could be on the other internal bus), the access to the data is lost. The MSA30 MI provides multi-initiator (shared bus) support and the ability to build more complete HA solutions.

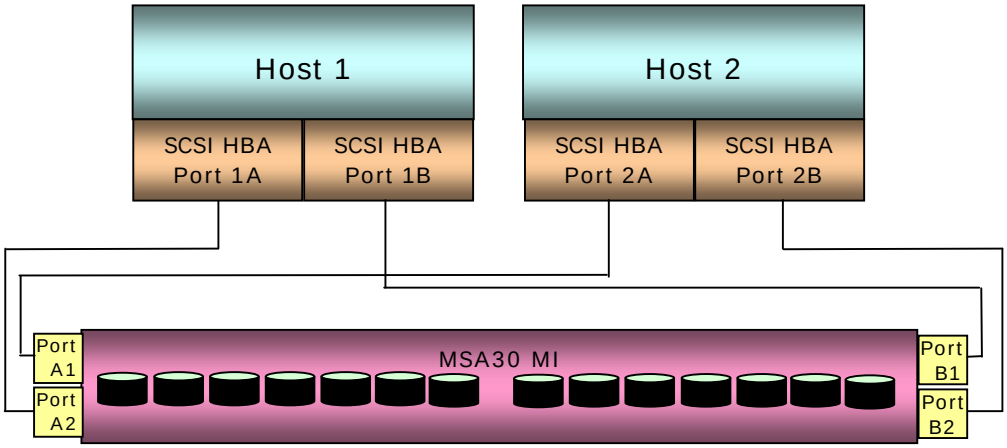
---

[2] Warm-swappable modules allow power to remain on but require all data transfers to the bus be stopped before removing or installing it.
[3] See http://docs.hp.com/en/diag.html. See also Section 6, "Storage Management", below.

2.1.2.1 MSA30 Configuration/Deployment

If shared buses are not required, the choice between SB and DB will depend upon the customer's storage needs and other requirements. For example, a customer with small data storage needs may decide to connect a single MSA30 DB to two hosts, giving each host access to up to 7 disks. Each host can operate independently and not impact the other. On the other hand, a customer with larger data storage needs may decide to connect a single host to the MSA30 SB to give the host access to all 14 disks. In either case, additional data protection can be achieved via SCSI RAID Controllers (see section 2.4.2) or software RAID products.

The MSA30 MI can be used to protect against any single storage component failure (cables, HBAs, disks, etc.), and when used in conjunction with a multi-system HA solution (such as HP Serviceguard) it can also be used to provide data availability across system failures. A common MSA30 MI configuration which can provide many of these characteristics is a 2-host configuration in which both hosts are connected to both MSA30 MI buses, as depicted below:



Mirroring of the drives across the two internal buses (using, for example, MirrorDisk/UX) makes for an even more robust configuration, protecting against the failure of any drive without losing access to the data. See Section 5 for additional information regarding the MSA30 MI in HA configurations.

The MSA30 is also designed to be used as back-end storage behind the MSA1000 and MSA1500 Fibre Channel arrays.

2.1.2.2 MSA30 Product Compatibility

The MSA30 provides a simple upgrade path from the StorageWorks 4300: simply replace the I/O module on the back of the unit with the corresponding MSA30 I/O module. Moreover, since the I/O module is warm-swappable, power to the unit can remain on during the upgrade, but all I/O operations must be stopped during the upgrade to ensure no data is lost.

The MSA30 enclosures are supported on HP-UX 11.00, 11i v1, and 11i v2 across a range of platforms and HBAs, including Smart Array RAID Controller cards. They are supported as a boot or dump device in most configurations. For supported OS versions, platforms, and HBAs go to www.hp.com/go/serverconnectivity and click on StorageWorks support matrices. Note that the Smart Array RAID Controller cards are only supported with the SB and DB products. For supported drives and enclosure components, see the MSA30 Specifications.

The MSA30 MI is designed to be a replacement for the recently discontinued DS2300.

For additional MSA30 product information, go to www.hp.com/go/msa30.

2.1.3 Older supported products

Older supported disk products which are no longer sold by HP include the DS2100, DS2110, DS2300, SC10, HVD10, HASS, and AutoRAID 12H.

The DS2100 and DS2110 are 4-slot Ultra160 SCSI disk enclosures in a 1U form factor with hot-swappable disks. They support Ultra320 disks and HBAs provided that the HBA is configured to operate at Ultra160 speeds (see the footnote regarding this in Section 2.1.2). As with the DS2120, up to three of these enclosures can be daisy-chained on a single bus to provide increased storage.

The DS2300 is a 14-slot Ultra160 SCSI high-availability disk enclosure with hot-swappable drives, power supplies, fans, and bus control cards (BCCs). Redundant power supply/fan modules and BCCs can be removed and replaced without interrupting storage operations. As mentioned above, the MSA30 is designed to be a replacement for the DS2300. As discussed in Sections 6.1.3 and 6.2, standard HP-UX Diagnostics utilities can be used for enclosure management of the DS2300, and centralized management of multiple DS2300s is available via the Command View utility. See Section 5 for information regarding the use of the DS2300 in HA configurations.

The SC10, HP's SureStore E Disk System SC10, is a high-availability10-slot Ultra2 LVD SCSI disk enclosure with hot-swappable drives, power supplies, bus control cards, and cooling fans, and supports 10K and 15K rpm drives. Data throughput is limited to 40 MB/s due to an internal SCSI SE backplane.

The HVD10 is a 10-slot Ultra2 HVD SCSI disk enclosure with hot-swappable components, and supports 10K and 15K rpm drives.

The HASS is an 8-slot Fast/Wide High Availability Storage System (HASS) of hot-swappable disk and/or tape drives, which can contain either HVD or SE devices.

The AutoRAID12H is a Fast/Wide HVD SCSI enclosure which continually fine-tunes itself for maximum performance and data availability, keeping the most frequently accessed data in RAID 0/1 storage, and less frequently accessed data in RAID 5 space. It supports active/active controllers with hot fail-over, and a mirrored cache, keeping a complete copy of the cache on both controllers.

The DS2300, HVD10, and SC10 also support an optional split bus mode which allows the enclosure to be configured into two separate SCSI buses.

All of these older devices are supported on 11.0 and 11i v1. Only the DS Series products and the SC10 are supported on 11i v2.

2.1.4 DVD+RW

The HP StorageWorks DVD+RW drive, the Q1592A product, is an Ultra2 SCSI LVD half-height array module designed for use in the Tape Array 5300 discussed in Section 2.2.1 below. It allows the customer to backup data and to install software to/from CD or DVD disks if the server itself does not have a CD or DVD drive, or if additional DVDs or CDs are desired. A single Tape Array 5300 could, for example, be used to house up to four DVD+RW drives, each connected to a separate server or all four connected to one server. The DVD+RW supports a wide variety of DVD and CD formats with up to 4.7 GB capacity in DVD formats, and up to 700 MB capacity in CD formats.

2.2 Tape Products

HP supports a wide range of SCSI tape solutions for HP-UX servers, including various types of tape drives, autoloaders, and libraries. Currently available tape drives, from lower to higher capacity and speed, include the following shown in Table 2.

Table 2: Tape Drive Products

Tape Drive Products		Capacity[1] (GB)	Speed[1] (MB/s)	Speed[1] (GB/hr)	Reliability[2] (MTBF)	Power[3] (watts)	Form[4] Factor	SCSI Interface
DAT 4mm Cassette	DAT 24	24	2	7.2	37.5	6	half-height	Fast Narrow SE
	DAT 40	40	6	21.6	125	7.3		Ultra Wide
	DAT 72	72	6	21.6		7.1		
DLT Cartridge	DLT VS80	80	6	21	200	15	half-height	Ultra Wide
	DLT VS160	160	16	57.6	250			Ultra160
SDLT	SDLT 110/220	220	22	79	250		full-height	Ultra2 Wide
	SDLT 160/320	320	32	115		27		
	SDLT 300/600	600	72	259		33		Ultra160
Ultrium	Ultrium 215	200	15	54	250	19	half-height	Ultra2 Wide
	Ultrium 230	200	30	108		21	full-height	
	Ultrium 448	400	48	173			half-height	Ultra160
	Ultrium 460	400	60	216		25	full-height	
	Ultrium 960	800	160*	576		30		Ultra320

[1] Compressed (2:1) capacity/speed. Native capacity/speed is one-half these values. Speed=sustained data rate.
[2] Mean Time Before Failure (MTBF) in thousands of hours.
[3] Average power consumption (watts)
[4] 5¼ inch half-height or full-height form factor. DAT 24 and 40 also have 3 1/2 inch options.
* When used with Ultra320 HBAs the Ultrium 960 must currently be speed-limited to Ultra160 (160 MB/s) speeds, which lowers the achievable sustained data rate to below 160 MB/s.

HP also has a variety of autoloaders and libraries which work with a variety of the tape drive products to provide for automated backup, larger storage capacities, and advanced availability and management features. Note: the majority of larger tape libraries are used in Fibre Channel storage area networks, and use FC-to-SCSI routers to convert to the device's pSCSI interface. The StorageWorks Network Storage Routers N1200 and M2402 are two supported FC-to-SCSI routers.

2.2.1 Tape Array 5300

HP offers a compact and flexible tape enclosure, the Tape Array 5300, which can meet many of the availability and flexibility needs of autoloaders and small tape libraries without the higher cost. The HP StorageWorks Tape Array 5300 is a 3U rack enclosure that holds two full-height array modules or four half-height, or one full height and two half-heights. The enclosure slots are offline hot-swappable[4], allowing you to replace drives without taking down the server. It also comes with an option for a redundant hot-swap fan and power supply for even greater

[4] Offline hot-swap drive modules can be added or removed while the array is powered up and the SCSI cables connected to servers, but the SCSI bus must be quiescent, which means no activity on the bus from any host to any drive on the bus.

availability. The Tape Array can be connected to up to four separate HBAs connected to four different servers, or multiple drives in the enclosure can be daisy chained together and connected to a single HBA. The Tape Array 5300 supports a variety of DAT and Ultrium tape drives, as well as a DVD+RW drive. With the DVD+RW, the Q1592A product described in section 2.1.4 above, customers can read and/or write to CD or DVD in addition to tape, offering additional flexibility.

See http://www.hp.com/go/tape for additional details on each of the tape products. For compatibility of the tape products with OS versions, platforms, and HBAs, click on Compatibility in the tape products page.

2.3 Optical Jukebox Devices

An option for archival storage, with faster access than tape or other archival solutions such as microfilm or paper, is magneto-optical disk storage. When large amounts of data require long term storage with faster access than tape, microfilm or paper, the use of optical disks and jukeboxes as an archival storage solution should be considered. Optical media has up to a 100 year archival life, and is available in both rewritable and Write-Once (WORM) formats. Write-once is designed for permanent data storage that can't be altered or erased. Medical archiving of healthcare data and document imaging are examples where optical archiving may be beneficial.

Optical jukeboxes provide automated access to archival storage. They come in two forms: Ultra Density Optical (UDO), which work with the higher density optical media, and Magneto-Optical (MO). HP's optical jukeboxes come with HP-patented robotics which provides exceptional reliability and durability across the optical solution.

For more information on archival storage solutions, go to www.hp.com/go/archival.

2.4 Host Bus Adapters

There are two classes of pSCSI HBA solutions for HP-UX: host-based pSCSI RAID controllers, and pSCSI HBAs. These are discussed below. For more information on the various HBA products, go to www.hp.com/go/serverconnectivity and click on Mass storage interface adapters. See also Section 6.1.5 for information on HBA-specific management utilities available for certain cards.

Some of the pSCSI HBAs and RAID controllers have both internal and external SCSI connectors, for attachment to server-internal buses and external buses. With multi-port HBAs there is still typically only one internal connector, along with multiple external connectors, with the internal connector on the same bus as one of the external connectors.

CAUTION: Simultaneous internal and external connections on the same SCSI bus are generally not supported. Internal and external devices can be used on the same HBA as long as they are on different SCSI buses. See the server connectivity documentation for platform-specific information regarding external and internal device connectivity requirements.

2.4.1 pSCSI HBAs

The following pSCSI HBAs are currently available on HP-UX:

- A7173A – PCI-X 2-port Ultra320 SCSI card
- AB290A – PCI-X 2-port Ultra320 SCSI, 2-port 1000base-t combo card

- A6829A – PCI 2-port Ultra160 SCSI card
- A6828A – PCI 1-port Ultra160 SCSI card
- A5838A – PCI 2-port Ultra2 SCSI, 2-port 100base-t combo card
- A5159B – PCI 2-port FWD SCSI card

All of these except for the FWD HBA work with LVD and SE devices, including SE Asynchronous, Fast, and Ultra devices, and LVD Ultra160, and Ultra320 devices. Only the FWD HBA directly supports HVD devices[5]. All of these support attachment of up to 15 SCSI devices per port at varying device speed capabilities up to the speed of the HBA. For supported OS versions, platforms, boot support and other details see the SCSI HBA Support Matrices at http://docs.cup.hp.com/en/netcom.html#SCSI%20Host%20Bus%20Adapters. For associated device support go to www.hp.com/go/serverconnectivity and click on StorageWorks support matrices.

The newest and highest performing pSCSI HBAs are the A7173A PCI-X dual-port Ultra320 SCSI card and the AB290A combo card. The A7173A is intended to replace the A6829A, and the AB290A is intended to replace the A5838A. The dual SCSI ports on these Ultra320 cards support a combined speed of up to 640 MB/s and capability of up to 100,000 I/Os per second. The Ultra160 cards support approximately ½ the raw performance capability per port of the Ultra320 cards, the Ultra card about ¼, and the FWD cards about 1/16 of the raw performance. For details on the A7173A performance, go to http://docs.hp.com/en/5991-0688/5991-0688.pdf. For details on the AB290A performance see the recently published white paper at http://docs.hp.com/en/6620/Castorwhitepaperbook.pdf.

The A7173A and AB290A support domain validation technology to verify that the physical connection (cables, connectors, targets, etc.) is capable of handling the negotiated transfer speed. Domain validation technology automatically compensates for marginally performing hardware by reducing the communication speed, thus preserving data integrity even when conditions do not permit transfers at the speeds negotiated between the initiator and the target. Then the user is notified to correct the problem.

As of the AR0505 release of HP-UX 11i v2, as mentioned in Section 2.1.1 above, HP-UX plans to deliver self-healing technology in the Ultra320 driver which will detect and compensate for additional data integrity problems using CRC (Cyclic Redundancy Check) error detection technology. The user is notified so the problem can be corrected and the speed adjusted back up again.

2.4.2 RAID Controller Cards

The following RAID Controller cards are currently available on HP-UX:

- A9890A – PCI-X 2-port Ultra320 RAID Controller
- A9891A – PCI-X 4-port Ultra320 RAID Controller
- A7143A –PCI 4-port Ultra160 RAID Controller

These cards are part of the Smart Array (SA) RAID Controller family of products. The A9890A uses the SA6402 Controller, the A9891A uses the SA6404 Controller, and the A7143A uses the SA5304 Controller. Each of these are supported on HP-UX 11i v1 and v2 with the MSA30 SB and DB disks across a variety of HP 9000 and Integrity servers.

[5] HVD devices can be used with the LVD HBAs via an LVD-to-HVD converter, as discussed in Section 3.2.

The A9890A RAID card is a PCI-X 2-port Ultra320 SCSI RAID Controller which provides Advanced Data Guarding (RAID ADG), leading performance, 256MB Double Data Rate (DDR) battery-backed write cache architecture, and a new RAID engine. The A9891A is very similar to the A9890A card, the difference being that the A9891A has two additional SCSI ports for a total of four. The A7143A is a PCI 4-port Ultra160 SCSI RAID Controller with similar RAID and cache features.

For supported configurations of OS versions, platforms, and devices with these RAID controllers see http://docs.hp.com/en/SM-20050510/raidcombinedsupportmatrix.htm.

All three cards support the following RAID levels, from lowest to highest in fault tolerance:

§ RAID 0 – Striping (no fault tolerance)

§ RAID 1 – Mirroring

§ RAID 1+0 – Mirroring and Striping

§ RAID 5 – Distributed Data Guarding (Striping with Parity)

§ RAID ADG – Advanced Data Guarding (Striping with Double Parity)

RAID 0 enhances performance with data striping, but provides no data redundancy to protect against data loss when a physical disk fails. RAID 1 mirrors the data onto two disks without striping.

RAID 1+0 provides both the redundancy of mirroring and the performance benefits of striping, with the disks mirrored in pairs and the data striped across the mirrored pairs. This allows for the failure of multiple disks in the array without incurring data loss, as long as no two failed disks belong to the same mirrored pair. RAID 1+0 has the highest read and write performance of any of the fault-tolerant RAID levels, but is more costly per unit of storage, with only one-half of the disk space being available for non-redundant data (usable capacity).

RAID 5, Distributed Data Guarding, uses striping with parity, with the parity data distributed across the physical disks. When a physical disk fails, the data that was on the failed disk can be calculated from the data blocks on the remaining physical disks. RAID 5 has lower write performance than RAID 1+0, but has more usable disk capacity: in an array composed of n physical disks, it has a usable capacity of (n-1)/n, which calculate to a low of 67% in a 3-disk array up to 93% in a 14-disk array. RAID 5 is recommended for up to 14 physical disks.

HP's RAID ADG, Advanced Data Guarding, provides the highest level of fault tolerance, by generating parity data ala RAID 5, but increasing to two sets of parity data per data stripe. As a result, it can withstand two simultaneous drive failures without downtime or data loss. It also maintains a very good usable capacity, (n-2)/n, which is close to RAID 5 and higher than the other fault-tolerant levels. The only significant disadvantage of RAID ADG is a relatively low write performance (lower than RAID 5), due to the generation of two sets of parity data.

Note: The Smart Array SCSI RAID cards are not supported in multi-initiator configurations.

For additional information on the SCSI RAID products, go to www.hp.com/products/smartarray.

3 Accessories

3.1 SCSI Bus Termination

A SCSI bus is composed of a daisy chain of SCSI devices connected via SCSI cables. SCSI buses need to be electrically terminated at each end of the SCSI bus chain, and only at the ends. This termination is needed for proper operation of the bus. See Table 1 for details on the termination types needed with each SCSI bus type.

Because SCSI buses are connected in a daisy chain fashion, SCSI devices normally have two SCSI connectors to accommodate connections in each direction on the bus chain. If the device is at the end of the chain it must be electrically terminated by either physically plugging a terminator into the unused connector or it may be auto-terminated in the device (circuitry which detects that the connector is not connected to another device).

3.1.1 HBA Auto-Termination

Typically HBAs are designed to connect at the end of a bus with a single connector. When an HBA has two connectors they typically consist of an "internal" connector for connection to a system-internal bus, and an "external" connector to connect to an external bus. If an HBA is located at the end of a bus, it must be terminated internally onboard the HBA, either through auto-termination circuitry or via manual switches or jumpers.

HBA auto-termination generally operates by sensing a pin on the internal and external connectors of the HBA port to see if a device is connected. If both connectors have devices connected[6] the HBA will disable its on-board termination circuits; otherwise the on-board termination is enabled. The auto-termination mechanism can also be affected by jumpers on the card or programmatic[7] controls.

HBA auto-termination works well as long as the HBA is at the end of a bus, or if it's connected between an internal and external bus. If an HBA is connected via a V-cable[8] in the middle of a bus the auto-termination won't work as required. It must be disabled, and the on-board termination forced off either programmatically[7] or via jumpers.

3.1.2 Device Auto-Termination

The MSA30 MI auto-terminates its SCSI bus ports, sensing if a device is connected to it or not. This provides an ease-of-use benefit over previous devices such as the DS2300, which required physically plugging in a terminator in the empty SCSI port of a SCSI port pair.

3.1.3 Inline-Terminated Cables

Inline-terminated (ILT) cables[8] violate the SCSI Specification at Ultra2 speeds and beyond. As a result inline-terminated cables are not supported on Ultra2, Ultra160, and Ultra320 buses. On Ultra or earlier configurations inline-terminated cables were used to allow a host at either end of a shared (multi-initiator) bus to be disconnected without bringing down the other host or stopping the applications accessing that bus. However, online reconnection of a host back to the bus using ILT cables is not supported due to electrical issues that can arise while reconnecting to a live bus.

---

[6] Simultaneous connection of both the internal and external connectors of an HBA is not supported on HP-UX, which means that the auto-termination circuitry will always enable termination unless it's manually disabled. This is the desired behavior for HBAs at the ends of buses.

[7] Software-controlled termination (programmatic control) is not supported on any of the currently available SCSI HBAs.

[8] V-cables and ILT cables are not supported with Ultra2, Ultra160, or Ultra320 HBAs.

3.2 SCSI LVD-to-HVD Converter

HP has recently certified an LVD-to-HVD converter to allow customers to use HVD disks with the newer LVD HBAs. The HP Hardware Provider (HPHP) program has successfully certified Rancho Technology's LVD-to-HVD converter products on 11i v2 servers. This product meets the HP Hardware compliance testing criteria and is now officially "Certified for HP-UX". This will enable customers who wish to continue using their legacy HVD devices on 11i v2 servers[9].

Product Description:

The LVD-to-HVD converter is connected between an LVD HBA port and an HVD peripheral device to allow HVD devices to be used with LVD HBAs (the older HVD HBAs are not supported on 11iv2). The converter comes in two product types:

- RTLVD-PHVD stand alone unit. A small single port converter that runs on 12 volts DC.
- RTLVD-HVDRxE 1 to 4-port enhanced rack mount unit. This is a 1U Rack mounted QUAD SCSI multimode SCSI Converter with up to 4 SCSI ports.

The converter has been tested in single-initiator configurations with Ultra160 and Ultra320 cards only. Multi-initiator configurations are not supported.

Product Features:

- Data transfers up to 40 Mbytes/second (the maximum HVD (Ultra) speed)
- Does not occupy or use a SCSI ID
- Switchable HVD and LVD Termination
- Transparent Operation (no software required)
- TERMPOWER via resetting circuit protector and backflow preventing diodes

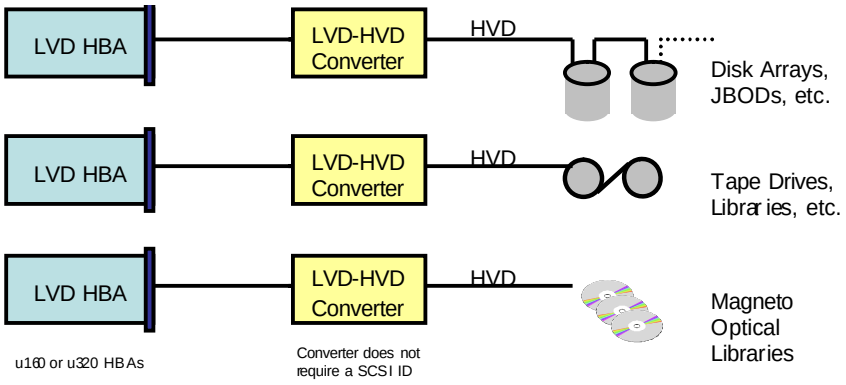
Supported HBA's

- Ultra320 and Ultra160 HBAs: A7173A, AB290A, A6829A, A6828A

Supported HVD target types include

- Disk Arrays and JBODs
- Tape Drives/Libraries
- Magneto-optical Libraries

HP-UX LVD-to-HVD Converter Configurations



For more information on the converter product, go to the "HP Certified Third Party Hardware Products" link on the HPHP web site at www.hp.com/dspp/hphp. For more information on Rancho Technology, go to http://www.rancho.com.

[9] The SCSI HBAs available for sale and supported on 11i v2 are all LVD.

4 Configuration

Setting up parallel SCSI solutions requires careful attention to configuration issues and details. The following basic rules and definitions apply. Additional rules need to be considered in shared SCSI bus environments as discussed in Section 5.

Basic Configuration Rules/Definitions

Note: It is generally not recommended to mix disks with tapes or other non-disk devices on the same SCSI bus. This is due to performance and starvation issues that can result in mixed configurations.

1. Electrical Signaling (do not mix HVD devices with LVD/SE devices on the same bus): As discussed above in Section 1.2, devices and HBAs come in three signaling types: the older SE and HVD, and the newer LVD. When LVD was defined in Ultra2 SCSI and beyond, a multi-mode option was also defined allowing devices to support both LVD and SE, with the restriction that in SE mode the bus is limited to Ultra speeds. HVD devices cannot coexist on the same bus with LVD/SE devices, unless the HVD devices are connected through an HVD-to-LVD converter.

2. Cable length (must not exceed maximum limits for a given type of bus): Exceeding the maximum total cable length for a given type of bus can result in signal integrity issues. It is therefore very important to adhere to the following limits on the total cable length from beginning to end of the bus, including any cable internal to SCSI enclosures, HBAs, disk arrays, etc. The maximum length for different types of buses is shown in Table 1, and the common cases are summarized below:
   - SE Ultra SCSI: 3 meters up to 4 devices; 1.5 meters >4 devices
   - HVD up to Ultra SCSI: 25 meters
   - LVD Ultra2, Ultra160, and Ultra320: 12 meters

   CAUTION: When determining cable length, the internal length within devices, enclosures, or HBAs must be estimated and included. It is often thought, for example, that one can connect 12 meters of external cable on an LVD bus, when in fact the HBAs and enclosures may use 1 to 3 meters of bus length internally.

3. SCSI Addressing Definitions and Priorities: narrow SCSI buses support 8 SCSI addresses, 0..7, with 7 being the highest priority and 0 the lowest, while wide SCSI buses support 16 SCSI addresses, 0..15, with the priority (highest to lowest) going from 7 to 0 followed by 15 down to 8. Ultra160 and Ultra320 buses are wide-only.

4. Unique Address assignment for each device: Each device on a SCSI bus (including SCSI initiators) must have a unique SCSI address. Duplicate SCSI addresses on a SCSI bus can prevent systems from booting or otherwise operating correctly. SCSI enclosures typically guarantee unique addresses among the various internal drives, with mode settings to support higher priority initiator addresses as discussed below.

5. SCSI Initiator Addressing (must be higher priority than any target): Each SCSI initiator must have a unique SCSI address and it must be at a higher priority than any SCSI target on the bus. A SCSI initiator is an HBA port, and a SCSI target is a peripheral device such as a disk or tape drive. SCSI enclosures typically have mode settings that reserve various high priority address assignments for initiators (not allowing them to be assigned to the enclosure's devices).

6. Bus Termination: Each end of a SCSI bus, and only the ends, must be terminated. If improperly terminated at the ends, bus integrity may be compromised. Termination in the middle can result in some devices being inaccessible. See the discussion on termination in Section 3 for additional details.

5 High Availability Configurations

High Availability (HA) refers to the ability to keep a system or component operational (available) for a significant percentage of the time. We refer to HA configurations in pSCSI as configurations which work together within larger HA solutions to allow for failure of a component without bringing down the operation of the system or access to data. Shared bus configurations[10] are a key aspect of HA support in pSCSI. The following general rules and definitions apply to shared pSCSI buses.

Shared Bus Configuration Rules/Definitions

Note: Shared SCSI buses require that both the HBA and the host drivers for the HBA support multiple SCSI initiators. See Section 2.4 and the referenced support matrices for details of which drivers/HBAs support multiple initiators.

1. Shared SCSI bus – termination: two of the initiators must be at each end of the bus with termination enabled at the HBA port.

2. Shared SCSI bus – targets: The only target devices allowed on a shared SCSI bus are disk devices, including disk arrays. No other devices, such as CD-ROMs, Magneto-Optical disks, tape drives, or scanners, are allowed on a shared SCSI bus in HP-UX.

3. Shared SCSI bus – initiators: There can be no more than four SCSI initiators (HBA ports) on any shared SCSI bus: one at each end of the bus, and up to two additional initiators attached via V-cables in the middle of the bus. As noted in section 3.1.1, attachment of HBAs in the middle of the bus requires that the HBA's auto-termination be disabled. Using more than two initiators on a bus will result in online repairability issues as discussed in item 7 below.

4. Cables/Connectors: In-Line-Terminated (ILT) cables and V-cables are not supported at Ultra2 speeds or higher (see section 3.1). The MSA30 MI and DS2300 enclosures provide similar functionality to ILT cables in certain configurations, as discussed below.

5. Electrical Isolation of MSA30 MI and DS2300 SCSI ports: These enclosures have four SCSI ports to which up to four HBA ports (initiators) can be attached (DS2300 supports a maximum of two initiators). Each pair of SCSI ports in these enclosures is electrically isolated from each other, allowing an HBA attached to one of the port pairs to be disconnected without affecting the operation of HBAs attached to the other, as discussed in more detail in Online Repairability below. In the MSA30 MI the electrical isolation is a natural result of its split-bus operation: each pair of SCSI ports is attached to a separate bus. In the DS2300 each pair of SCSI ports is associated with a module called the Bus Control Card (BCC). The DS2300 can be configured in either split-bus or full-bus mode. In split-bus mode, the separate buses provide electrical isolation. In full-bus mode, the electrical isolation is provided with respect to cable disconnection in that the DS2300 detects when a BCC is un-terminated

[10] A shared SCSI bus is a bus attached to multiple SCSI initiators.

and disconnects the BCC from the enclosure's internal bus, thus effectively isolating the BCC whose HBA has been disconnected from the rest of the bus and the other BCC.

6. DS2300 termination: Each BCC in the DS2300, whether in split-bus or full-bus mode, acts like a separate SCSI bus with respect to termination, and thus any unused port of a BCC whose other port is connected to an HBA must be physically terminated. See also section 3.1.2.

7. Online Repairability: Online Replacement (OLR) of an HBA is only supported in configurations in which the physical disconnection of the SCSI cable from the HBA to the bus will not result in instability of the bus (which can lead to data corruption, bus hangs, and other problems). Thus, HBA OLR is supported in single-initiator configurations. It is also supported in dual-initiator configurations on the DS2300 in which each initiator is attached to a separate BCC. Only one port on each DS2300 BCC can be connected to a host because disconnecting an HBA from one port on a BCC will bring the other port down. As a result, the DS2300 must be configured in full-bus mode to allow both hosts to see all the disks and thus allow continuous data availability across the OLR operation.

   Full multi-initiator configurations, with initiators connected to all four SCSI ports, are supported on the MSA30 MI. This is a result of its auto-termination capabilities which auto-terminate the port when the cable is disconnected.

   Use of ILT cables (on Ultra or lower speed buses), as discussed in 3.1.3, allows only for online disconnection and not subsequent reconnection.

8. Configurations without online repairability: Configurations which meet the other shared bus requirements but which don't meet the above requirements for online repairability can be supported without online repairability. In such cases, all initiators connected to the shared bus must be halted and power removed from the bus before any portion of the SCSI bus can be disconnected.

9. Root/Boot disks: root/boot disks can be on a SCSI bus that is shared across multiple hosts as long as the shared bus only contains active root/boot disks for a single host. This is because boot firmware does not support two nodes booting from the same SCSI bus at the same time. Note: mirroring of root/boot to another bus that contains the root/boot disks of another host is allowed.

6 Storage Management

Storage management refers to the management and configuration of storage devices, enclosures, HBAs, and related components, including storage provisioning, backup planning, and the timely detection and recovery from hardware failures or exception conditions. Online management of storage hardware conditions, including drives, enclosures, and HBAs, is provided on HP-UX via the Diagnostics Event Monitoring Service (EMS) tools. Centralized storage management is also provided with certain storage products via the HP Command View tools, as described in section 6.2 below.

6.1 EMS Tools

The EMS tool set is comprised of hardware monitors for different types and classes of devices. These monitors provide automatic detection and notification of hardware events, and offer a variety of notification methods to alert you when a problem occurs. They are integrated into MC/ServiceGuard and other applications, allowing the application to take timely action when hardware problems occur. Among the available SCSI monitors are disk and tape class monitors for the management of disk and tape drives, and enclosure monitors for managing and monitoring the resources within various types of enclosures.

If a problem is detected with a supported hardware module, the monitor immediately sends an event to the Event Monitoring Service (EMS) which alerts the customer using the notification method specified for that instance of the monitor. EMS notification messages identify the problem, what caused it, and what must be done to correct it. In addition to hardware failures, some monitors can provide predictive notifications, alerting the customer to a potential problem and the opportunity to replace a component before it fails.

Brief descriptions of some of the key pSCSI-related hardware monitors are given below. The name in parentheses is the computer-friendly name for the monitor. For more details on the various EMS tools and supported products go to http://docs.hp.com/en/diag.html.

6.1.1 Disk Monitor

The Disk Monitor (disk_em) monitors the operation of disk drives, whether standalone or within an enclosure. It is the default event monitor for SCSI disk resources. It only operates on HP-supported non-removable-media disks. Predictive events monitored include the detection of an "excessive rate of recovered media defects", which indicates that while these errors have all been recovered from with no data loss, the drive may fail due to an unrecoverable error in the near future and should be replaced.

6.1.1.1 diskinfo(1M) command

While the diskinfo command is not an EMS tool, it is a standard command on HP-UX which can be used to obtain information about a disk device. It takes as an argument a character-mode device file for a disk (for example, /dev/rdsk/c1t2d0) and displays information such as the vendor name, product id, disk capacity, block size, and revision number of the disk. See the associated HP-UX man page for additional details.

6.1.2 Tape Monitor

The Tape Monitor (dm_stape) is designed to monitor the operation of individual tape drives, as well as various tape libraries and autoloaders. Predictive events monitored include "severe trouble reading or writing the tape", which indicates that while no data has been lost, there has been a reduction in the performance or capacity of the tape, and it is recommended that the drive head be cleaned.

6.1.3 Enclosure Monitors

The enclosure monitors include a High Availability Storage System Monitor (dm_ses_enclosure), which is supported with the DS2300 and the SC10 enclosures, and the Storage Works Modular Smart Array Monitor (msamon) for the MSA30 and other MSA products. Both of these will monitor the status of the power supplies, fans, temperature, enclosure controllers, internal bus problems, and other events associated with the operation of the enclosure. They will also monitor the general accessibility of the disk drives in the enclosure. In addition, the MSA Monitor will detect and report various hot-plug removal and insertion events.

6.1.4 HBA-Specific Management Utilities

6.1.4.1 Ultra320 SCSI Cards

The Ultra320 SCSI cards, A7173A and AB290A, have online configuration and management utilities included with HP-UX's Ultra320 driver, the MPT driver. These are command-line utilities called mptutil and mptconfig.

The mptutil command can be used to perform online updates of the HBA firmware (known as firmware download). It also allows the user to obtain detailed diagnostics information about the card, firmware, and driver, including the firmware revision, chip revision, Vital Product Data (VPD), hardware termination settings, driver trace buffer, and various statistics. In addition, mptutil provides interfaces to read PCI configuration space and to perform various types of SCSI aborts and resets, including bus reset.

The mptconfig command is used to display and set various SCSI parameter values, such as the HBA port's Initiator Id, target and initiator speed limits (maximum transfer rates), and bus width.

6.1.4.2 SCSI RAID Cards

The SCSI RAID cards have associated command-line management and configuration utilities called saconfig and sautil.

The saconfig command allows the customer to do various RAID configuration tasks such as create a logical drive, add a spare drive, display the current configuration, delete a logical drive. It also can be used to control certain cache functions, including enabling the cache and specifying the percentage of the total cache used for reading versus writing.

The sautil command is a support tool used to retrieve RAID configuration and status information for the controller, logical drive, physical disk, cache, etc. It also provides interfaces to obtain RAID driver state, trace log, and statistics, to update (download) new revisions of the controller or disk firmware, to perform resets and bus scans, and to recreate the controller device files.

6.2 Command View SDM

HP StorageWorks Command View SDM (CVSDM) provides centralized storage management for DS2300 enclosures and various Fibre Channel disk array products. CVSDM allows for the centralized management of an unlimited number of the DS2300's or other supported storage products from a graphical user interface (GUI), command line user interface (CLI), or Web browser. It provides secure device management in both direct-attach and SAN environments. CVSDM is provided in both a CD kit release and on the Web at http://hp.com/support/cvsdm (click "download drivers and software").

7 References

http://www.t10.org/
(SCSI Specification web site)

http://www.hp.com/go/storageworks
(HP StorageWork products)

http://www.hp.com/go/msa30
(MSA30 product information)

http://www.hp.com/go/tape
(Tape product information)

http://www.hp.com/go/archival
(Archival storage information)

http://www.hp.com/products/smartarray
(Smart Array RAID card products)

http://www.hp.com/go/unixserverconnectivity
(HP 9000 and HP Integrity server connectivity)

http://www.hp.com/products1/serverconnectivity/support_matrices.html
(Matrices of supported OS versions, platforms, HBAs, and devices)

http://www.hp.com/dspp/hphp
(HP Hardware Provider Program)

http://docs.hp.com/en/diag.html
(EMS tools for diagnostics and storage management)

8 For more information

http://www.hp.com/go/storage

http://www.hp.com/go/unixserverconnectivity

http://www.t10.org/

XXXX-XXXXEN, 12/2003

invent