



2E01
OpenVMS
SAN/Storage
Technical Update

Manfred Kaser
Hewlett Packard GmbH
Manfred.Kaser@hp.com

© 2006 Hewlett-Packard Development Company, L.P.
The information contained herein is subject to change without notice



OpenVMS V8.3 Storage Features

4Gb Fibre Channel



- VMS supports 4Gb FC HBAs with V8.3
 - AB378A/B – 1 port 4gb PCI-X 266Mhz (qlogic)
 - AB379A/B – 2 port 4gb PCI-X 266Mhz (qlogic)
 - “B” version works in all Integrity platforms
 - “A” version doesn’t work in 16xx, 26xx and 46xx
- Support is limited to Integrity only
 - Boot setup is identical to existing 2gb
 - One way read performance of 400MB/sec has been observed to the new 4gb EVAs
- 4gb FC switches have been shipping for 12+ months
 - 4gb EVA started shipping on mid 2006
 - 4gb (full bandwidth) XP will ship soon
- Fabric does all speed matching for 1/2/4gb devices
 - Cables are the same as 2gb

4Gb Fibre Channel



- New FC cards in process...
 - PCI-e dual 4gb FC
 - AD300A
 - Qlogic Based
 - Mezz format for blades
 - PCI-e single 4gb FC
 - Emulex based
 - PCI-X 1/2 port 4gb fc + 1/2 port Gbe (Vega)
 - AD193A / AD194A
 - Qlogic based
 - Merlion / RuSA support only
 - PCI-e 1/2 port 4gb fc + 1/2 port Gbe (Deneb)
 - Emulex based
 - AD221A / AD222A
 - Merlion / RuSA support only

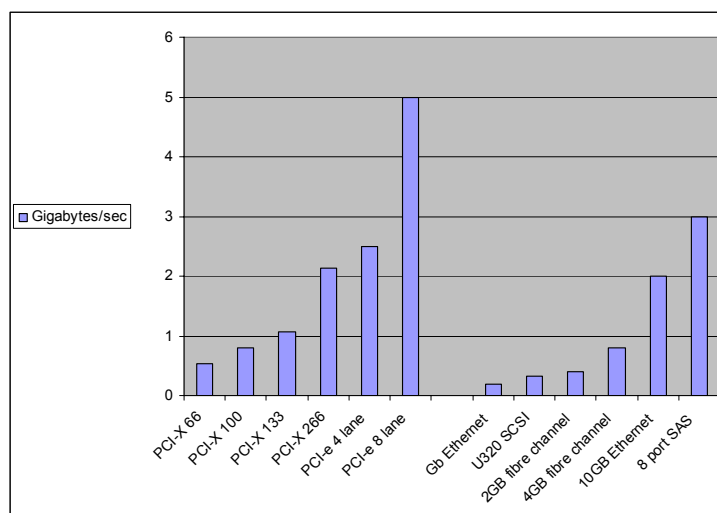


PCI-e

- PCI-x
 - 133Mhz = 1.05 GB/sec (most current Integrity cards)
 - 266Mhz = 2.1GB/sec (4gb FC cards)
- PCI-e
 - X1 = 5gb/sec = 625MB/sec
 - X4 = 20gb/sec = 2.5GB/sec (initial cards)
 - X8 = 40gb/sec = 5GB/sec
 - X16 = 80gb/sec = 10GB/sec



PCI and IO Interface Bandwidths





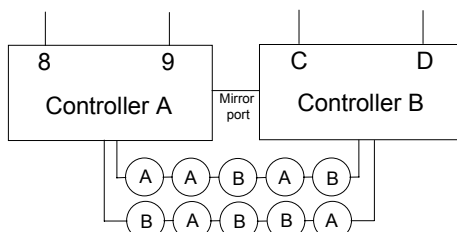
PCI-e

- PCI-e slots will be incorporated into Integrity servers shipping in late CY06
 - RX2660
 - RX36xx/66xx
 - Initial systems have both PCI-x and PCI-e slots
- Initial PCI-e cards will be
 - 2 port 4gb Fibre Channel
 - SAS Raid (Vostok)
 - SAS SmartArray (Spawn – Saturn V)
 - FC / Gbe Combo cards *(after initial ship)*
 - 10Gbe *(after initial ship)*



Active-Active Storage Host Ports

- EVA / MSA active-active host ports
 - Original EVA/HSG/MSA designs force access to any given lun to be through a single controller



I/O paths to device 4

Path PGA0.5000-1FE1-0011-B158 (CLETA), primary path.		
Error count	0	Operations completed 8
Path PGB0.5000-1FE1-0011-B15D (CLETA), current path.		
Error count	0	Operations completed 1557
Path PGA0.5000-1FE1-0011-B15C (CLETA).		
Error count	0	Operations completed 0
Path PGB0.5000-1FE1-0011-B159 (CLETA).		
Error count	0	Operations completed 0



Active-Active Storage Host Ports

- EVA / MSA active-active storage host ports
 - With active-active host ports a lun can be accessed via all host ports
 - The lun is still bound to a single controller, but the reference comes into the other controller, the request will be forwarded via the mirror port
 - little write impact since the data needed to be mirrored anyway
 - there will be a small latency impact on reads via the “non-optimized” controller
 - Pre 8.3 VMS will not distinguish between optimal and non-optimal ports for path selection
 - The EVA will move a lun across controllers if the workload is sufficiently unbalanced



Active-Active Storage Host Ports

- EVA / MSA active-active storage host ports
 - Active-active host ports requires no changes to VMS and will be backwards compatible with current releases
 - XP already works this way...
 - Active-active host ports is primarily a feature to make our storage arrays more Windows / HP-UX friendly
 - In V8.3, the VMS multipath algorithm will choose an optimized path if one is available
 - Set device /switch/path=xxxx will force the specified path to become optimized
 - Other hosts will receive notification and will migrate to the new optimized path(s)
 - Active-active host ports will be part of the EVA-XL (4000/6000/8000) and MSA V7 firmware base
 - Active-active host ports is released on the EVA-GL (3000/5000) as part of a VCS V4 firmware base level

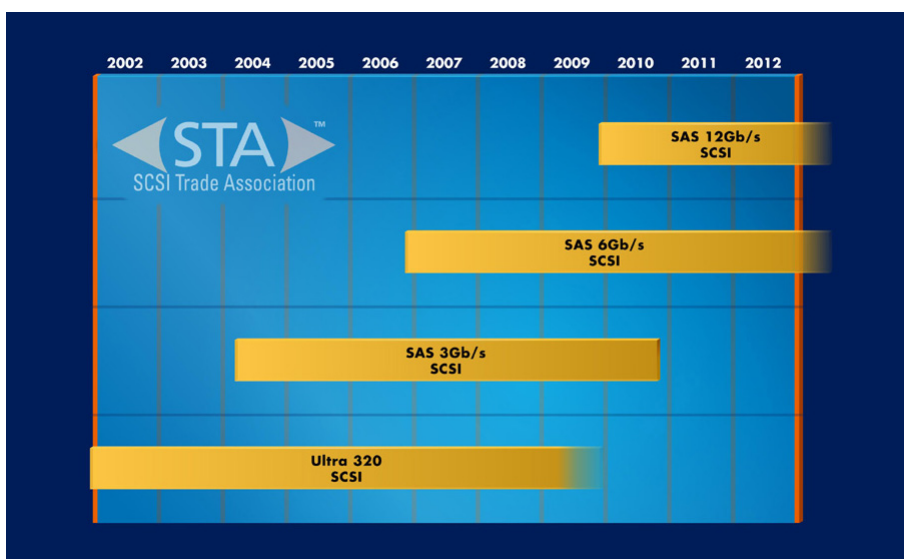


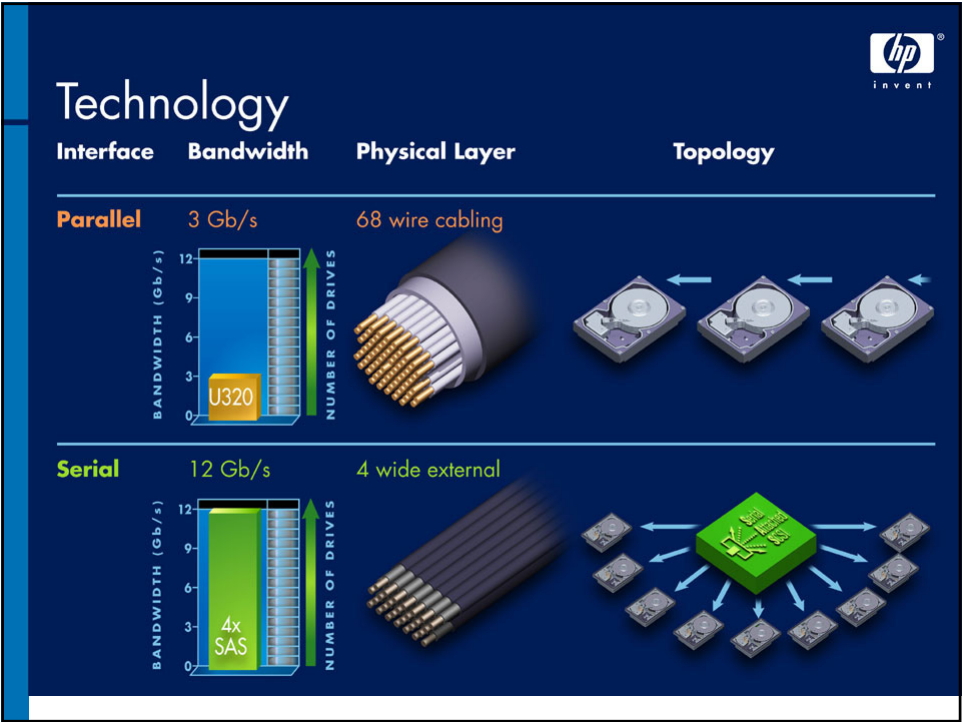
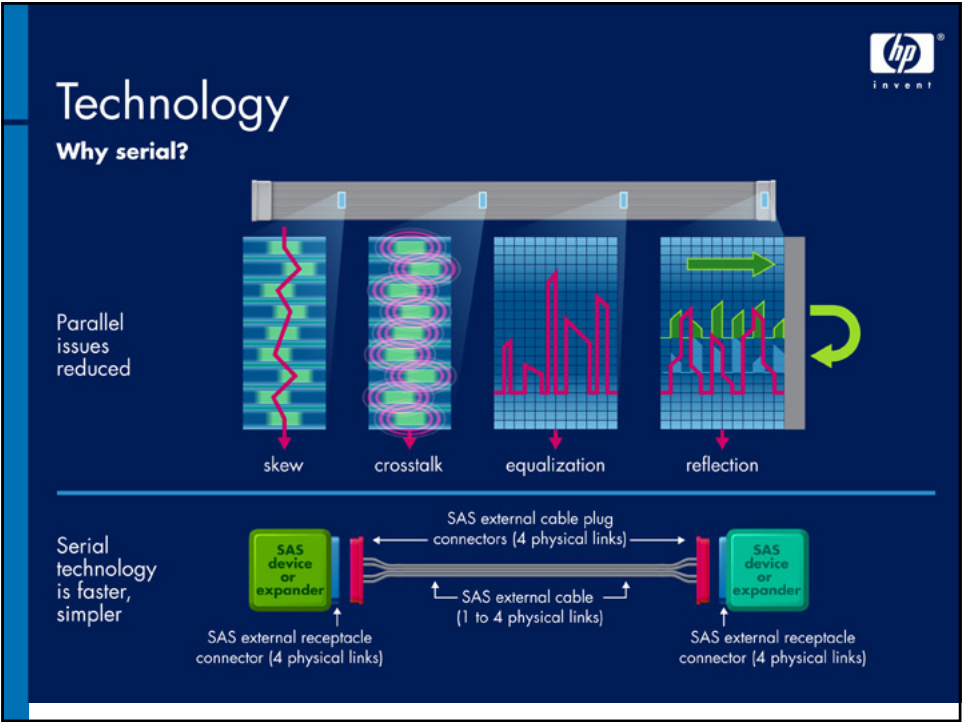
SCSI to SAS/SATA Transition

- Starting in mid CY2006, Integrity servers will start to transition host based storage from SCSI to SAS/SATA
 - RX3300 / RX6600 house 8-16 2.5" internal SAS drives
 - Over the following 6-12 months we expect to see
 - Mixture of SAS/SATA devices in the same enclosures
 - Multihost configurations
 - SAS based SmartArray technology
 - External SAS/SATA shelves (MSA60 / MSA70)
 - SAS tape drives
 - External SAS Storage Arrays (think MSA with SAS instead of FC)



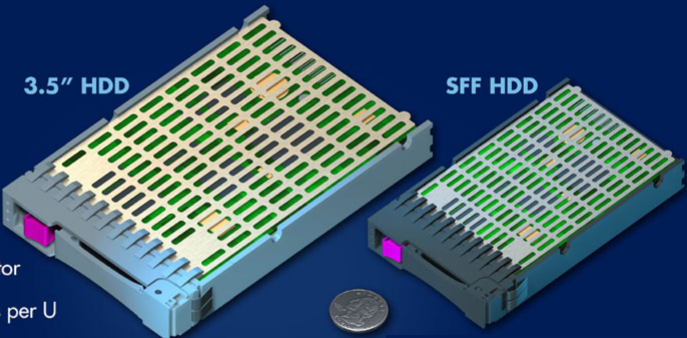
SCSI to SAS/SATA Transition





Technology

What is SFF?



- 2.5" form factor
- More spindles per U
- Higher I/Os per U
- Higher capacity per U
- Lower accoustics
- Same enterprise characteristics

- SFF enables more disk spindles / U yielding higher IOPS / U
- Each SFF drive utilizes only 50% the power of one 3.5" drive
- Improved airflow and more drives per server
- MTBF is ~17% improved with SAS/SFF (1.75 Million hrs. MTBF vs. 1.5 Million hrs. MTBF)

SCSI to SAS/SATA Transition

- Initial SAS/SATA controller is LSI Logic based and will be core IO on RX36xx/RX66xx/RX2660 platforms
 - pkmdriver controls both U320 and SAS/SATA disks
 - controller based raid 1 available
 - Max 2 raid sets
 - no sata devices allowed
- 2nd SAS/SATA controller is PCI-e SmartArray based
 - PKRdriver
 - Initially just supports internal SAS drives on Merlion/RuSA
- New universal 2.5" carrier for both internal and external sas/sata drives

Common packaging between Integrity and ProLiant



SCSI to SAS/SATA Transition

- SAS/SATA device naming
 - SAS/SATA has a complex naming scheme driven by the need for simple internal device naming, but cluster compatible external device naming
 - Internal SAS/SATA devices will never be have shared direct access across multiple hosts
 - External SAS/SATA devices may someday be shared across multiple hosts
 - Naming scheme is necessary for
 - SAS
 - SATA
 - Tapes
 - Optical Media
 - Host based Raid Arrays



SAS/SATA Device Naming

	Internal Devices	External Devices
SAS Disk	SCSI Based name based on slot * 100 dka100	User created name stored in nvrom \$3\$dga329
SATA Disk	SCSI Based name based on slot * 100 dkb600	User created name stored in hidden blocks on the disk \$3\$dga165 (no boot support)
SAS/SATA Tape	SCSI Based name based on slot * 100 mka400	User created name stored in sys\$devices.dat \$2\$mga12 (same as fibre)
SAS/SATA Optical	SCSI Based name based on slot * 100 dka200	SCSI Based name based on slot * 100 dka200 (not multihost capable)
SAS/SATA Host Based Raid Array	SCSI Based name based on lowest slot name of any member dkc2	SCSI Based name based on lowest slot name of any member dkc7 (not multihost capable)

SAS/SATA Device Naming



sas\$util is used to manage sas/sata devices

```
* THORA:[SYSMGR]>run sys$system:sas$util
* SAS> set adapter pka

* Selected controller _PKA0: is a LSI SAS10xx device.
* SAS> show disk
* Port _PKA0:
-----
* ID  Enc  PHY  Slt  Device  Type  Size(MB)  SAS Address  Product Name
-----
* 14  02  12  12      -  ATAPI      0  500508B3-00A1396C  HP MSA50.. 1.20F
* 15  02  02  02  DGA6417  SAS      70007  5000C500-0022DC0D  HP DG072.. HPD43
* 16  02  03  03  DGA6418  SAS      70007  5000C500-002802C9  HP DG072.. HPD43
* 17  02  04  04  DGA6419  SAS      70007  5000C500-0023240D  HP DG072.. HPD43
* 18  02  05  05  DGA6420  SAS      70007  5000C500-002772E9  HP DG072.. HPD43
* 19  02  06  06  DGA6421  SAS      34732  5000C500-0033AA49  HP DG036.. HPD43
* 20  02  07  07  DGA6422  SAS      34732  5000C500-0030B295  HP DG036.. HPD43
* 21  02  08  08  DGA6423  SAS      70007  5000C500-00277EDD  HP DG072.. HPD43
* 22  02  09  09  DGA6424  SATA     57231  500508B3-00A13979  ATA FUJIT.. 034EN
* 23  03  12  13      -  ATAPI      0  500508B3-00A139CC  HP MSA50.. 1.20F
* 24  03  02  02  DGA6427  SAS      34732  5000C500-00330871  HP DG036.. HPD43
* 25  03  03  03  DGA6428  SAS      34732  5000C500-00339B89  HP DG036.. HPD43
* 26  03  04  04  DGA6429  SAS      34732  5000C500-0030B785  HP DG036.. HPD43
* 27  03  05  05  DGA6430  SAS      34732  5000C500-0034163D  HP DG036.. HPD43
* 28  03  06  06  DGA6431  SAS      34732  5000C500-0030B2E1  HP DG036.. HPD43
* 29  03  07  07  DGA6432  SAS      34732  5000C500-00339ABD  HP DG036.. HPD43
* 30  03  08  08  DGA6433  SAS      34732  5000C500-003340C5  HP DG036.. HPD43
* 31  03  09  09  DGA6434  SATA     57231  500508B3-00A139D9  ATA FUJIT.. 034EN
* 05  -  -  -      DKA5  RAID 0    68664  33A6D625-0AF85AD9  LSIL Logic.. 3000L
* 12  -  -  -      DKA12 RAID 1    69607  B2086E81-0D57C37D  LSIL Logic.. 3000L
* SAS>
```

SAS/SATA Device Naming



```
* THORA:[SYSMGR]>run sys$system:sas$util
* SAS> set adapter pkb

* Selected controller _PKB0: is a LSI SAS10xx device.
* SAS> show disk
* Port _PKB0:
-----
* ID  Enc  PHY  Slt  Device  Type  Size(MB)  SAS Address  Product Name
-----
* 02  DA  02  02  DKB200  SAS      34732  5000C500-00341621  HP DG036.. HPD43
* 03  DA  01  03  DKB300  SAS      70007  5000C500-002CC419  HP DG072.. HPD43
* 04  DA  00  04  DKB400  SAS      34732  5000C500-00340185  HP DG036.. HPD43
* 05  DA  07  05  DKB500  SAS      34732  5000C500-0033FAF1  HP DG036.. HPD43
* 07  DA  05  07  DKB700  SATA     57231  D83F3C15-C9A9809B  ATA FUJIT.. 034EN
* 07  -  -  -      DKB7  RAID 1    69607  F5F4AA91-04B8276D  LSIL Logic.. 3000L
* SAS>
```



Tape Performance Monitoring

- Tape performance arrays similar to existing fibre channel disk arrays are now kept in V8.3
 - Performance data collected for FC attach tapes (anything named \$2\$mgaxxx)
 - we'll eventually extend this to SAS tapes as well
 - sda> fc perf /tape [/clear]



Tape Performance Monitoring

```

• SDA> fc perf /tape

• FibreChannel Device Performance Data
• -----

• $2$mgall2 (read)
• Using EXE$GQ_SYSTIME to calculate the I/O time
• accumulated read time = 1357282706us
• reads = 1349951
• total blocks = 127569106

• LBC    <2ms    <4ms    <8ms    <16ms   <32ms   <64ms   <128ms  <2s    <4s
• ===    =====
• 1      278      23      159      117      39      -        8        3        2      629
• 2        2        -        -        -        -        -        -        -        -        2
• 4       727      11      45      -        5        -       17        -        -      805
• 16     2354      10      78      -        3        3       17        -        -     2465
• 32        -        3      45      -        -        -       16        -        -        64
• 64    678222     138      -        -        -        -        2        2        2     678366
• 128   667392     222      -        -        5        -        -        1        -     667620

•      1348975     407     327     117     52      3      60      6      4  1349951

```



Tape Performance Monitoring

• FibreChannel Device Performance Data

```

• $2$mgall2 (write)
• Using EXE$GQ_SYSTIME to calculate the I/O time
• accumulated write time = 865674716us
• writes = 1140812
• total blocks = 114517211

```

LBC	<2ms	<4ms	<8ms	<32ms	<64ms	<128ms	<2s	<4s	<8s	
1	68	22	2	43	3	132	10	6	1	287
2	1	-	-	-	-	-	-	-	-	1
4	296	4	-	16	-	48	-	-	-	364
16	745	2	-	24	-	72	-	-	-	843
32	-	-	-	16	-	48	-	-	-	64
64	471562	66	3	-	-	-	-	2	-	471633
128	667454	160	2	2	-	-	1	1	-	667620
	1140126	254	7	101	3	300	11	9	1	1140812



Fibre Channel Queue Depth Management

- Fibre Channel storage arrays are very sensitive to the depth of their host queues
 - Since V7.3-2 we have had various host queue depth limit algorithms based on qfull signals from the storage array
 - HSG has very shallow host port queues (~240), and can't stop host traffic soon enough
 - XP12K doesn't use qfull and slows down with host port queues at very modest levels (32-64 commands)
 - EVA and MSA work pretty well with the existing qfull scheme
 - We've had SDA mechanisms to limit storage port queues, but its not convent to use

Fibre Channel Queue Depth Management



- SANCP is a new utility to control storage port queue depths
 - `sancp set wwid /port=xxxx-xxxx-xxxx-xxxx/cap=value`
 - Wildcards supported in wwid to make it easy to set multiple ports with one command
 - The cap is the max number of concurrent commands that the host will allow to a given array host port (remember to consider the number of hosts accessing a given storage port)
 - If possible its nice to allow a queue depth of (~4 commands * the number of active luns) per storage port per host
 - Its best not to use hard port caps unless necessary
 - Check `hsg qfull` responses (tune for none) (`sda> fc show wtid`)
 - Check XP12000 port processor utilization (tune for under 75%)
 - Its possible that the XP requires very careful balancing of luns across host ports and limits on luns per port to achieve best performance

Fibre Channel Queue Depth Management



- Backup is the worst VMS abuser of storage resources
 - Backup will queue 100+ commands to a single source lun if given enough working set
 - Experiments show that backup using modern FC disk and tapes work fine with a queue depth of 8
 - There is a new backup switch “`/io_limit=x`” which will control the number of concurrent commands that can be issued to an input lun
 - The default is 8
 - This should greatly help the numerous disk array performance problems we see during backup windows
 - `vms732_backup-v0600` backports `/io_limit` to `vms v7.3-2`



MSA\$UTIL

- MSA\$UTIL is a command line utility to manage MSA and SmartArray based storage
 - Available for field test now
 - Will ship in Alpha / Integrity TIMA kits in early 2007
 - Replaces sys\$etc:msa_util which was previously used only for MSA firmware update
 - Has similar functionality to MSA serial line interface
 - Supports configuration of new SAS based SmartArray cards (Spawn / Saturn V)



iSCSI

- iSCSI is a means to tunnel storage traffic through TCP/IP based LANs
 - Low cost \$\$\$
 - Can use low cost LAN infrastructure
 - Dedicated infrastructure recommended
 - Can use vanilla NICs in workstations or servers
 - Long distance
 - If you can ping it, you can put your storage there...
 - High Host CPU cost
 - Using software based initiators burns CPU cycles to parse IP packets and to copy data
 - Hardware based initiators are \$\$\$ and not yet mature
 - Modest Performance
 - Uses Gbe LANs today
 - High CPU cost limits high MB/sec applications
 - Long distance and potential routers will limit performance
 - Low cost storage arrays will limit performance

iSCSI



- VMS V8.3 includes latent support for a software based iSCSI initiator
 - Qualification will happen after V8.3 ships
 - Limited initial target support
 - Most likely EVA / MPX100
 - No boot support
 - Requires VMS TCP/IP stack (no 3rd party stack support)
 - Both Alpha and Integrity supported

iSCSI



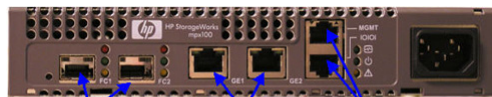
- HP iSCSI solutions
 - EVA / MPX100
 - Tested and works well... Easiest setup via CommandView EVA
 - MSA1510i
 - Basic testing so far... Currently only qualified for windows
 - We may wait for next gen MSA products
 - Cisco iSCSI Routers (9216i)
 - Pretty extensive testing... Worked well... Hard to setup
 - XP iSCSI Port
 - Not tested yet
 - Brocade Multi Protocol Router
 - Limited testing... Didn't work well... Windows only per Brocade
 - StorageWorks 2122 Router
 - Worked OK... Pretty old hardware at this point

MPX-100

(aka HP StorageWorks EVA iSCSI Connectivity Option)



- Direct connect to EVA 4000/6000/8000
- Fabric connect supported (to a single EVA)
- Supports EVA 3000/5000 in fabric mode
- Highly integrated with CommandView EVA
- Dual FC ports
- Dual GBE ports
- Redundant config possible with 2 units
- 1U ½ rack package



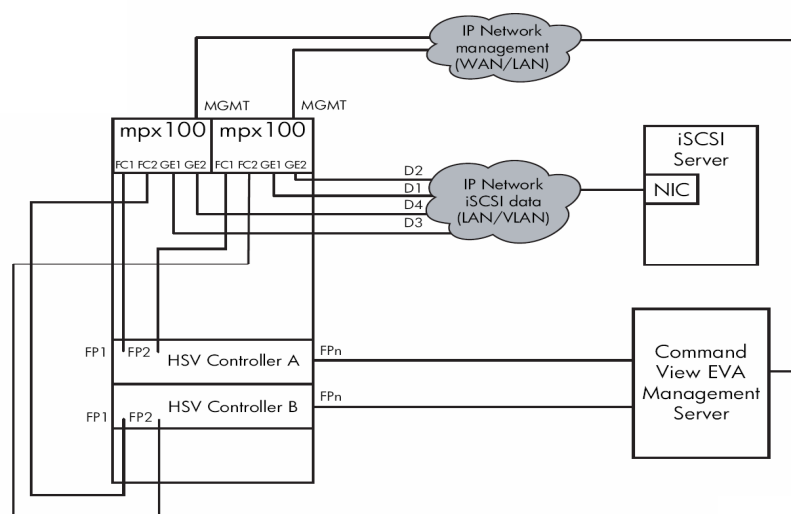
(2) 2Gb FC ports for high performance connectivity to FC switches and storage

(2) 1GbE ports for connectivity to servers and Ethernet networks

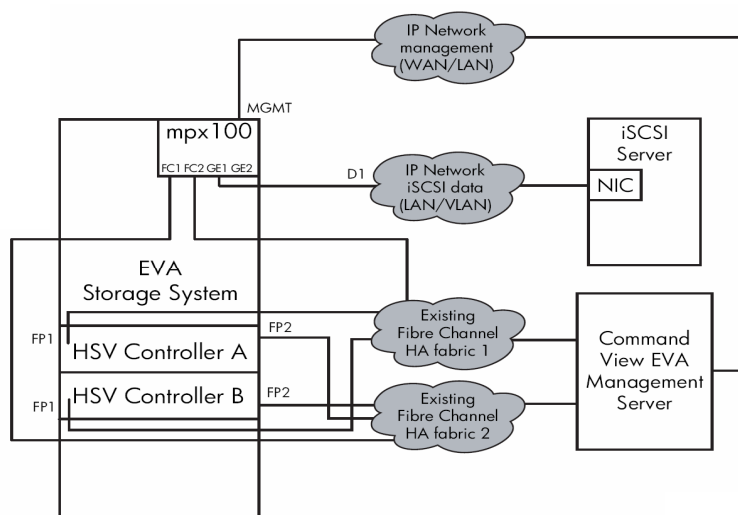
(2) Management ports Ethernet & RS-232

MPX 100 "Crystal Cove"

HP StorageWorks EVA iSCSI Connectivity Option – Direct Connect



HP StorageWorks EVA iSCSI Connectivity Option – Fabric Connect



MPX 100 CommandView EVA



The screenshot displays the HP StorageWorks Command View EVA interface. The left sidebar shows a tree view of Storage Systems, including Hosts, Disk Groups, Hardware, and iSCSI Devices. The main panel shows the iSCSI Controller Properties for iSCSI Controller 1.

Storage Systems

- Hosts
 - Jim Host Folder
 - JimTestHost
 - Host1
 - Host10
 - Host2
 - Host3
 - Host4
 - Host5
 - Rg2S2901
 - iSCSI Host
 - rg2S2902
- Disk Groups
 - Data Replication
- Hardware
 - iSCSI Devices
 - iSCSI Controller 1
 - Rack 1
 - Unmappable Hardware
 - EVA5000-C0

iSCSI Controller Properties

Save changes | Delete | Refresh | ?

General | IP Ports | FC Ports

Identification

Name:	iSCSI Controller 1
Manufacturer:	Hewlett-Packard
Model:	iSR-6140
Software revision:	1.0.0.10
Hardware revision:	0 a
Serial number:	0539A01061

Management Port

Connection status:	☒ On line
MAC address:	00-C0-DD-0C-60-22
IP Address:	☒ Dynamic IP address: 16.112.33.131 ☐ Static IP address: N/A
Subnet mask:	255.255.192.0
Gateway:	16.112.0.1

Comments

iSCSI



```

Mrvp41> sh dev /full $1$dga18994

Disk $1$dga18994: (MRVP41), device type HSV210, is online, file-oriented device,
shareable, device has multiple I/O paths, served to cluster via MSCF Server,
error logging is enabled.

Error count          0      Operations completed          0
Owner process        **      Owner UIC                    [SYSTEM]
Owner process ID      00000000      Dev Prot            S:RWPL,O:RWPL,G:R,W
Reference count       0      Default buffer size    512
Current preferred CPU Id 0      Fastpath                1
WWID 01000010:6005-08B4-0010-2EA5-0000-7000-0A6A-0000
Allocation class      1

I/O paths to device          4
Path DEA0.446F-6821-6F0B-A223 (MRVP41), primary.
Error count          0      Operations completed          0
Last switched to time: Never      Count          0
Last switched from time: 23-MAY-2006 16:21:37.38
Path DEA0.446F-6821-FB6D-1041 (MRVP41), current.
Error count          0      Operations completed          0
Last switched to time: 23-MAY-2006 16:21:37.38      Count          1
Last switched from time: Never
Path DEA0.446F-6821-22DD-5DC1 (MRVP41).
Error count          0      Operations completed          0
Last switched to time: Never      Count          0
Last switched from time: Never
Path DEA0.446F-6821-B6BB-EFA3 (MRVP41).
Error count          0      Operations completed          0
Last switched to time: Never      Count          0
Last switched from time: Never
Mrvp41>

```



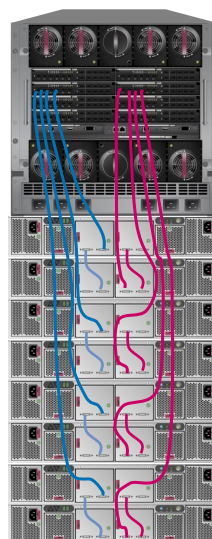
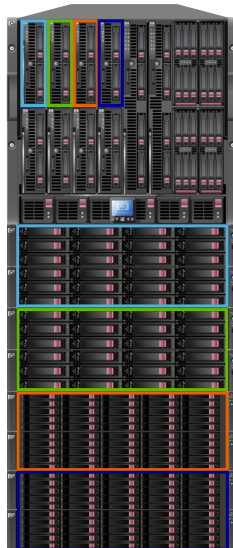
Post V8.3 Work



Blades

- Fibre Channel Mezz Cards
 - 4gb Qlogic based card
 - 4gb Emulex based card
- SAS Mezz cards
 - Dumb SAS
 - SmartArray SAS
- Virtual FC switch
- SAS switch
- Smart SAS switch (MSA / Spitfire based)

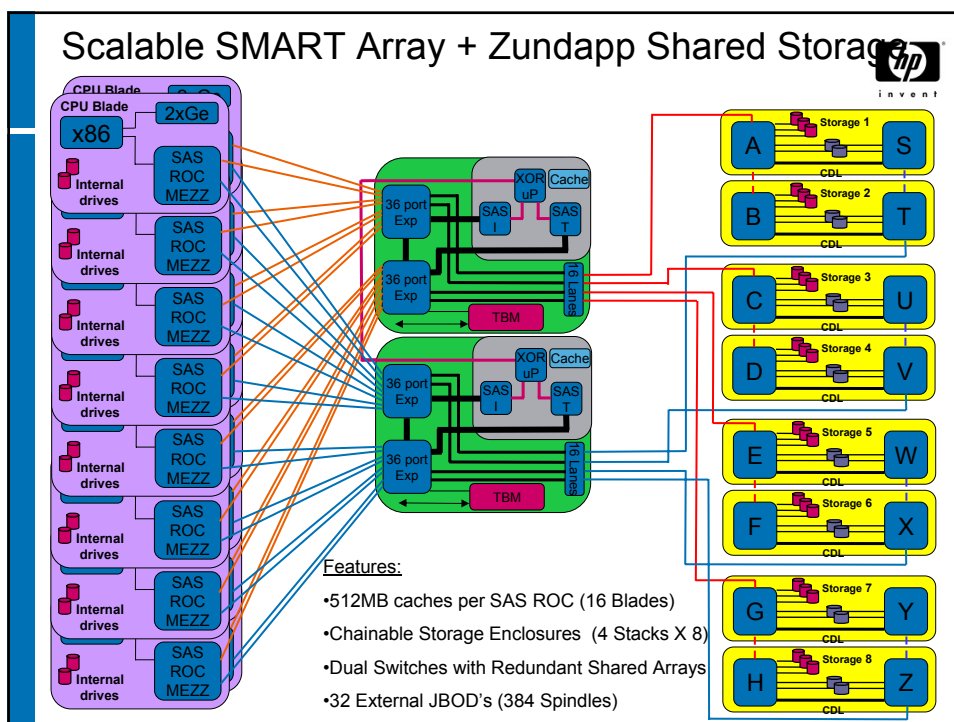
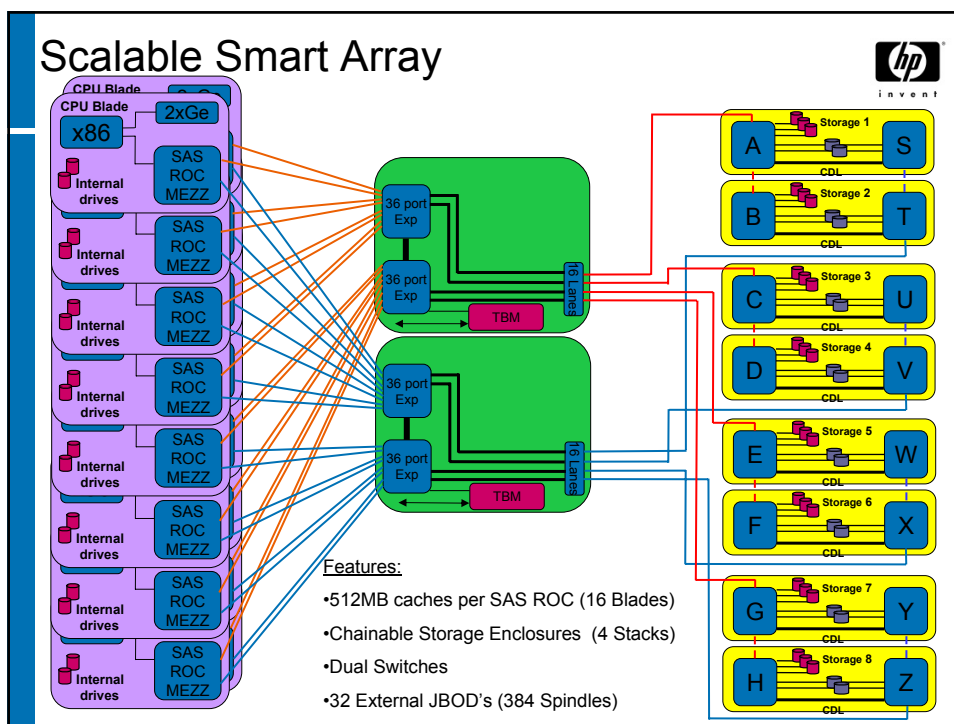
Scalable Smart Array concept



16 Blade
Servers with
SuperSabres

32/64 Storage
Enclosures

2/4 Belle





iSCSI

- Next generation 10gbe cards contain iSCSI assists
 - Onboard tcpip stack
 - DMA engine
 - iSCSI protocol offload
- Performance and CPU load could be similar to FibreChannel HBAs
- One card could do storage, cluster and LAN
- Late 2008 best case deployment for iSCSI
 - Sooner for LAN



New StorageWorks Hardware



HP StorageWorks MSA1500

HP StorageWorks MSA1500



- 2-Gb Fibre Channel SAN 2U controller shelf—Smart Array Technology
- Connects up to 8 MSA 20 SATA enclosures—max 96 drives equaling 24 Terabytes raw storage in a space of only 18U
- Connects up to 4 StorageWorks MSA 30 SCSI enclosures—max 56 drives equaling 16.8 Terabytes raw storage
- Mix StorageWorks MSA 20 SATA and StorageWorks MSA 30 SCSI with one MSA1500 for tiered storage
- Dual or Single I/O cards available



MSA1500 configuration maximums

Version SATA +MSA20



**24-TB Configuration
96 SATA Drives**

Version SCSI +MSA30



**16.8-TB Configuration
56 SCSI Drives**

MIXED version SATA and SCSI +MSA20 & MSA30



**20.4-TB with 71
SCSI/SATA
hard drives**

EVA4000/6000/8000 New controllers



2nd generation EVA controller architecture

Feature

- Faster microprocessor
- Faster internal busses
- New cache management ASIC
- Dual FC cache mirror ports
- Larger cache capability
- Double the FC host ports (on EVA8000)
- Larger policy memory
- New modular design

Benefit

- Faster command execution
- Improved data bandwidth
- Improved data bandwidth
- Improved write data bandwidth
- Improved performance for larger configs
- Improved data bandwidth
- More table space for larger capacities
- High redundancy, easy serviceability

EVA 4000/6000/8000 Three capacity and performance ranges



EVA 4000
2C4D



EVA 6000
2C8D



EVA 8000 **EVA 8000**
2C12D 0C6D

Maximum of

4 drive enclosures
56 drives
16.8 TB
13,000 random read I/Os
(8 KB blocks)
360 MB/s read bandwidth
(sequential 64 KB blocks)

Maximum of

8 drive enclosures
112 drives
33.6 TB
26,000 random read I/Os
(8 KB blocks)
720 MB/s read bandwidth
(sequential 64 KB blocks)

Maximum of

18 drive enclosures
240 drives
72.0 TB
54,500 random read I/Os
(8 KB blocks)
1300 MB/s read bandwidth
(sequential 64 KB blocks)

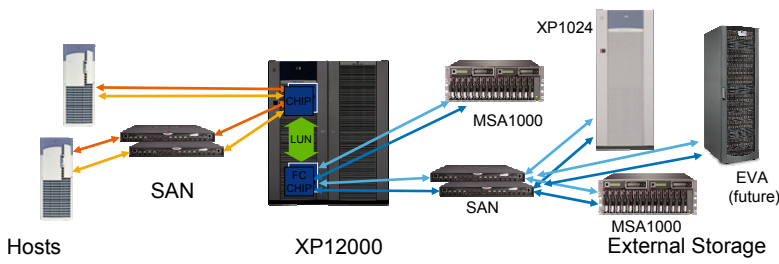


XP 12000 External Storage

Configure both internal and external storage to match cost, performance, and availability objectives

- Connect lower-cost MSA or legacy XP storage through XP12000
- Accessed as internal XP12000 storage
- Supported by all XP solutions (BC, CA*, AutoLUN)
- Up to 14 Petabytes of low cost storage

** Business Intelligence, low-cost archiving, on-line recovery images **



* 2nd release



HP StorageWorks EML E-Series Tape Libraries

Dense, **scalable solutions** that make data protection and recovery simple for your SAN environment

Product Features

- Easy to manage single or multiple libraries through a single pane of glass with Command View ESL Tape Library Software
- Scalable from 103 slots to over 440 slots and up to 16 Ultrium tape drives



EML 103e
EML 245e
EML fully configured

HP StorageWorks 6000 Virtual Library System



Integrating seamlessly into the existing environment, the HP StorageWorks 6000 Virtual Library System **improves the performance and reliability** of your data protection process.

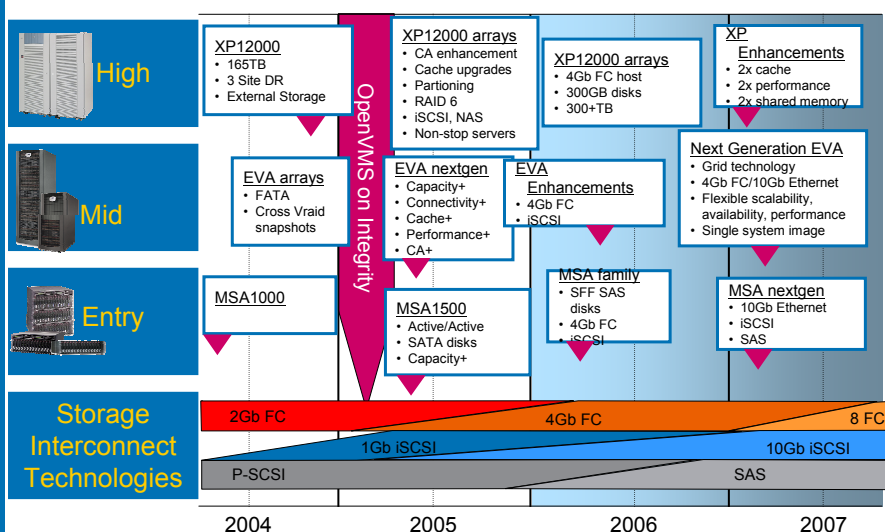
Product Features

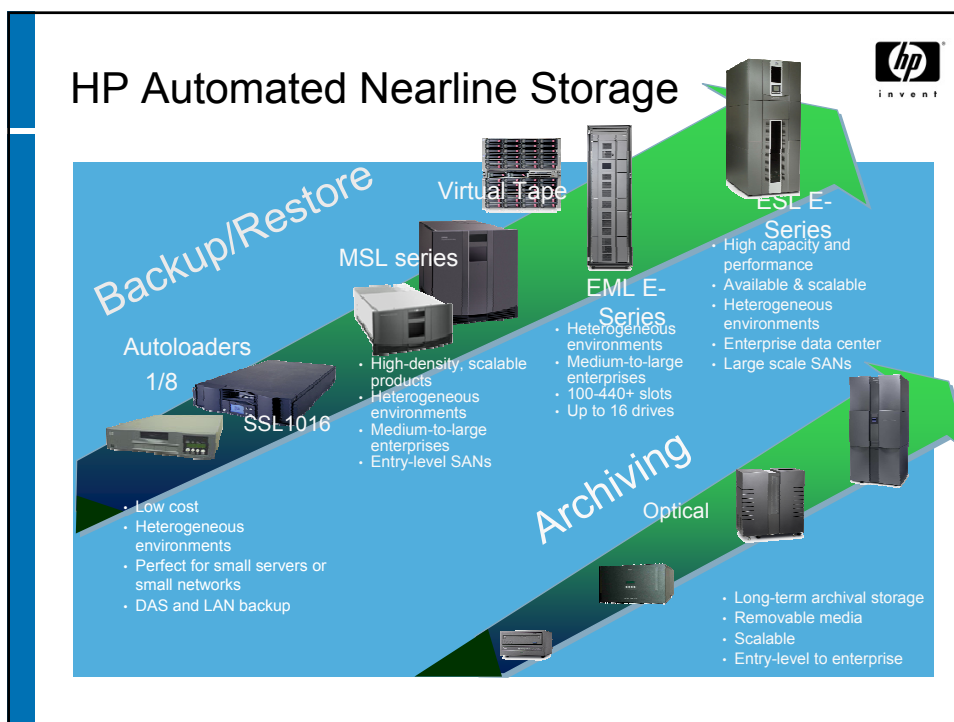
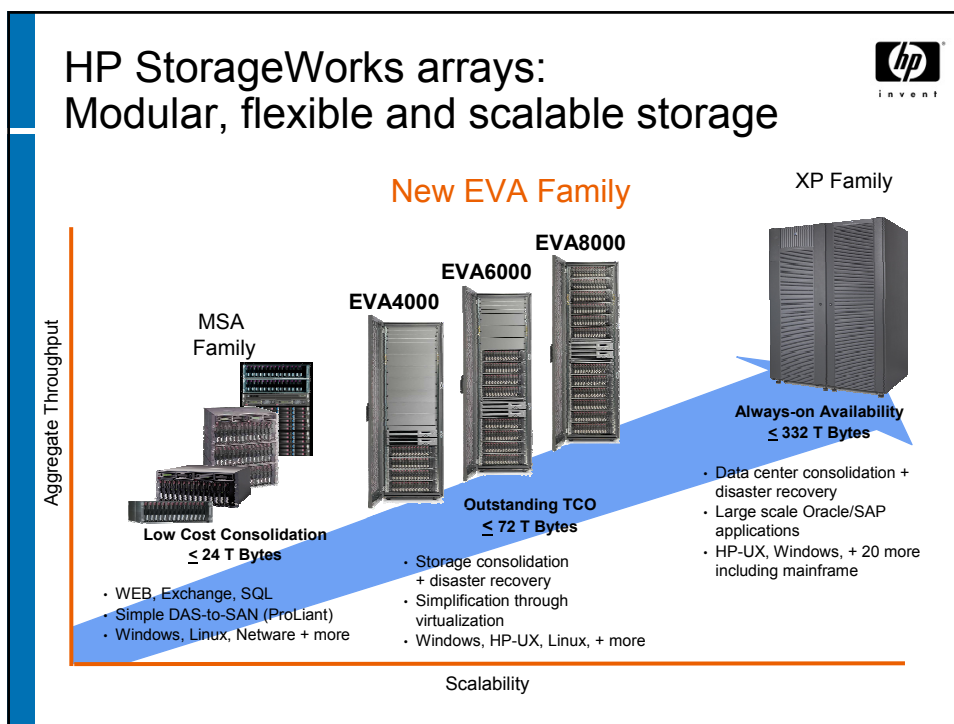
- Emulates popular HP tape drives and libraries – integrating seamlessly into your current backup and recovery application and processes.
- Scales capacity and performance as your requirements change
- Utilizes LUN masking and mapping to protect backup and recovery jobs from disruptions caused by SAN events



StorageWorks array feature roadmap

OpenVMS





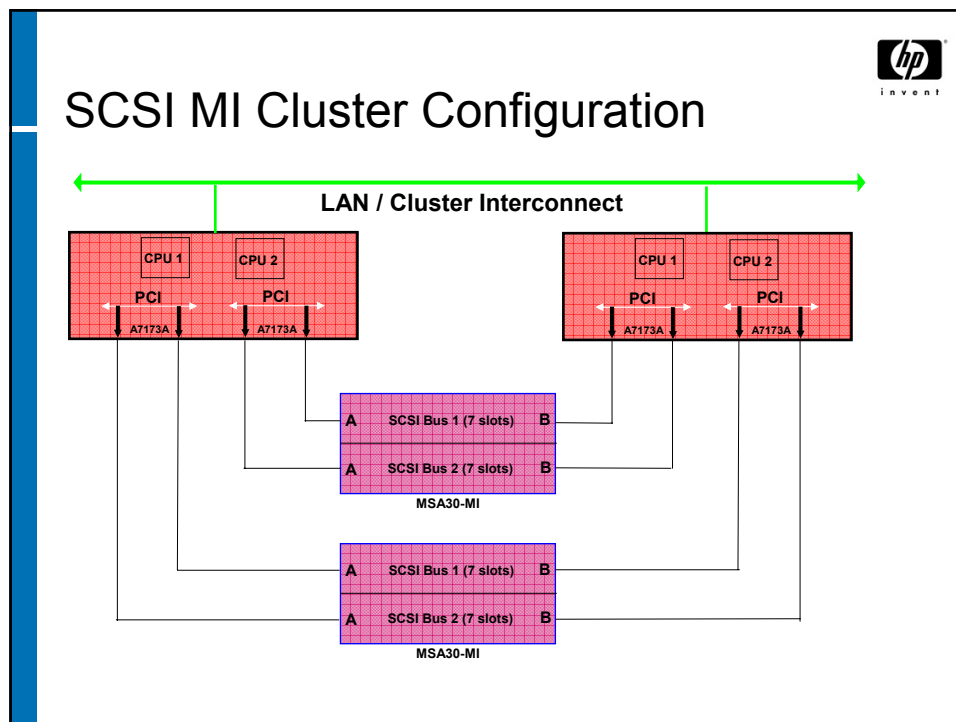


OpenVMS V8.2 / V8.2-1 Storage Features Worth Talking About Again



Multi-Host SCSI (V8.2-1)

- Our desire to drive FC to multi-host SCSI prices never happened
 - VMS FC-AL code / MSA interaction problems
 - Cheap switches never really got that cheap
- Multi-Host SCSI in V8.2-1
 - Integrity only (1600 / 2600 / 4600)
 - A7173A only
 - MSA30-MI only (no funny “Y” cables needed)
 - 2 host only
 - 4 shared busses max



FC Performance Monitoring

- Alpha and IPF FC drivers now keep a performance array for every mounted disk
 - Keeps count of completed IO across latency, IO size and read/write
 - Data can be displayed with SDA (or with a \$QIO code)
 - SDA> fc performance
 - [device-name] (device to display)
 - [/rsc | /systime] (time resolution cc or systime)
 - [/compress] (compress white space)
 - [/csv] (writes data to csv file(s))
 - [/clear] (clear array)
 - Data collection always on in systime mode in V8.2 and V7.3-2 fibre_scsi_v400



FC Performance Monitoring

```
SDA> fc perf $1$dga3107 /comp
```

```
FibreChannel Disk Performance Data
```

```
$1$dga3107 (write)
```

```
Using EXESGQ_SYSTIME to calculate the I/O time  
accumulated write time = 2907297312us  
writes = 266709  
total blocks = 1432966
```

LBC	<2us	<2ms	<4ms	<8ms	<16ms	<32ms	<64ms	<128ms	<256ms	<512ms	<1s
1	46106	20630	12396	13605	13856	15334	14675	8101	777	8	145488
2	52	21	8	9	5	5	6	1	2	-	109
4	40310	13166	3241	3545	3423	3116	2351	977	88	-	70217
8	2213	1355	360	264	205	225	164	82	5	-	4873
16	16202	6897	3283	3553	3184	2863	2323	1012	108	-	39426
32	678	310	36	39	47	44	33	27	6	-	1220
64	105	97	18	26	41	43	42	24	7	-	403
128	592	3642	555	60	43	31	23	9	2	-	4957
256	-	9	7	-	-	-	-	-	-	-	16
	106258	46127	19904	21101	20804	21661	19617	10233	995	8	1 266709



FC Performance Monitoring

```
FibreChannel Disk Performance Data
```

```
$1$dga3107 (read)
```

```
Using EXESGQ_SYSTIME to calculate the I/O time  
accumulated read time = 1241806687us  
reads = 358490  
total blocks = 1110830
```

```
I/O rate is less than 1 mb/sec
```

LBC	<2us	<2ms	<4ms	<8ms	<16ms	<32ms	<64ms	<128ms	<256ms	<512ms	<2s
1	46620	12755	6587	7767	3758	2643	1133	198	5	-	81466
2	574	134	66	158	82	20	21	4	1	-	1060
4	162060	35896	20059	18677	15851	11298	5527	1300	25	2	270696
8	355	79	46	97	59	36	28	10	-	-	710
16	241	103	32	150	77	24	13	1	-	-	641
32	916	355	76	302	316	61	25	10	-	-	2061
64	725	380	64	248	140	17	10	3	-	-	1587
128	13	22	13	36	21	6	-	-	-	-	111
256	10	41	28	15	49	13	2	-	-	-	158
	211514	49765	26971	27450	20353	14118	6759	1526	31	2	1 358490

```
SDA>
```



Questions and Answers???



Integrity Fibre Channel Bootling



IPF Fibre Channel Booting

- IPF FC booting is very different from Alpha and can be pretty confusing to setup
 - Alpha uses WWIDMGR to make bootable devices known to the console
 - wwidmgr is part of the Alpha console sub-system
 - Alpha uses the bootdef_dev console variable to hold default boot devices
 - IPF uses “boot_options.com” to make bootable devices known to EFI
 - boot_options is a VMS runtime command procedure
 - we had to implement a VMS runtime solution because EFI doesn't know how to find “offline” paths to HSG/EVA
 - IPF uses the EFI boot option list to hold default boot devices



IPF Fibre Channel Booting

- In order to boot FC devices the A6826A card **MUST** have the following minimum firmware
 - EFI driver version V1.40 (**V1.42 is current**)
 - RISC firmware version 3.03.001 (**3.03.150 is current**)
 - Below rev firmware can be updated with the latest IPF firmware CD
 - See power-up log for firmware rev levels

```

Loading 'FPSWA'...
Loading 'lsi1030'...
Loading 'gigundi'...
2 0 0x00020B 0x0000000000000006 EFI Launching Boot Manager
Scsi (Pun0,Lun0) HP 36.4GMAS3367NC HPC3 (320 MBytes/sec)
Scsi (Pun1,Lun0) HP 36.4GMAS3367NC HPC3 (320 MBytes/sec)
Scsi (Pun2,Lun0) HP 36.4GST336753LC HPC4 (320 MBytes/sec)
Broadcom NetXtreme Gigabit Ethernet Adapter is detected (PCI)
HP 2Port 2Gb Fibre Channel Adapter (driver 1.41 Test1, firmware 3.03.001)
HP 2Port 2Gb Fibre Channel Adapter (driver 1.41 Test1, firmware 3.03.001)
HP 2Port 2Gb Fibre Channel Adapter (driver 1.41 Test1, firmware 3.03.001)
HP 2Port 2Gb Fibre Channel Adapter (driver 1.41 Test1, firmware 3.03.001)

```

IPF Fibre Channel Booting



- Basic install/boot steps
 - Verify A6826A firmware revs
 - Install VMS to target FC device using install CD (don't reboot yet)
 - Select Option 7 (execute VMS DCL commands)
 - Verify that you have all expected paths to the system disk
 - \$ show dev /mul \$1\$dga3207:
 - \$ @sys\$manager:boot_options.com (run boot_options from the CD)
 - Follow the prompts to add the new system disk to boot options list (add before EFI Shell if you want to auto boot)
 - Follow the prompts to set up the dump disk
 - Boot new system disk
 - Use cluster_config to build additional roots
 - Boot distribution CD on all other cluster members and run boot_options to setup boot and dump disks

IPF Fibre Channel Booting



- Your boot option list should look something like this when you're done
- If you get confused, boot the distribution CD, use boot_options to clear the boot options menu and re-populate it

```
EFI Boot Manager ver 1.10 [14.61]  Firmware ver 2.31
[4411]
```

```
Please select a boot option
```

```
EFI Shell [Built-in]
$1$dga3206 FGA0.5000-1FE1-0011-B15C
$1$dga3206 FGA0.5000-1FE1-0011-B158
$1$dga3206 FGB0.5000-1FE1-0011-B15D
$1$dga3206 FGB0.5000-1FE1-0011-B159
```



VEVAMON



VEVAMON

- VEVAMON = VMS EVA Monitor Tool
 - Provides a way to get VMS host access to performance metrics kept by the EVA
 - Collects the same data as PC based EVAperf
 - Requires that each EVA has a unique non-zero Console LUN ID
 - Set this up using the EVA GUI... Its on the 1st page
 - VMS should configure \$1\$GGAXxx
 - VEVAMON should work on all versions of VMS (and yes we'll do an IPF version too)
 - Distributed as part of the T4 Package
 - CommandView EVA V4, plans to include EVAperf



```
Cleta> vevamon $lsgga4: /vd/psta/hc/pd
VEVAMON Version 1.0 (Compiled on Dec 8 2004 14:48:37, and will expire on Feb 06 2005).
Startup time is Wed Dec 8 16:06:50 2004
Using $lsgga4: data file
Found 1 hsv device
HSV 5000-1FE1-0015-2C50 ($lsgga4:) Wed Dec 8 16:06:50 2004:
```

Disk	Read Miss RPS	KBS	Latency	Read Hit RPS	KBS	Latency	Write RPS	KBS	Latency	Flush RPS	KBS	Mirror RPS	Prefetch KBS
\$lsgga4001	48	1368	19259	2606	67957	11776	1464	36640	37744	9	4608	36397	9782
\$lsgga4002	28	510	19886	317	6931	15264	218	6199	43470	5	2299	6157	2021
\$lsgga4003	20	826	20267	2656	65260	12131	1533	41142	38697	2	1024	41211	9190
\$lsgga4004	26	800	20952	338	8017	16868	213	6080	44784	4	2171	5978	3193
\$lsgga4101	0	0	0	26	761	8604	71	2379	1086875	30	1532	2225	0
\$lsgga4102	0	0	0	29	451	9115	76	2492	1064659	32	1851	2525	0
\$lsgga4103	0	0	0	28	700	11059	71	2053	1086572	22	1135	3385	0
\$lsgga4104	0	0	0	34	765	7840	40	1397	1091926	18	894	2596	0
\$lsgga6101	0	0	0	0	0	0	80	1684	15641	2	383	1684	0
\$lsgga6102	0	0	0	0	0	0	74	1580	15505	4	638	1580	0
\$lsgga6103	0	0	0	0	0	0	76	2094	14289	7	510	2605	0
\$lsgga6104	0	0	0	0	0	0	91	2525	15318	14	766	3291	0
\$lsgga4/vd	122	3504	20091	6034	150842	11582	4007	106265	379623	149	17811	109634	24186

Found 44 logical DGA disks; printed data for 12 disks.

Date	WWN	Name	Read RPS	Read KBS	Read Latency (us)	Write RPS	Write KBS	Write Latency (us)	Ave. Queue Depth
08-Dec-04 16:06:50	5000-1FE1-0015-2C50	P0-this: 2C5C	2438	71737	13545	1516	40530	39560	91
08-Dec-04 16:06:50	5000-1FE1-0015-2C50	P1-this: 2C5D	2896	63809	10795	1483	37330	36875	86
08-Dec-04 16:06:50	5000-1FE1-0015-2C50	P0-that: 2C58	780	18013	16404	593	17714	327222	191
08-Dec-04 16:06:50	5000-1FE1-0015-2C50	P1-that: 2C59	54	964	361	102	2929	1061734	95
08-Dec-04 16:06:50	5000-1FE1-0015-2C50	\$lsgga4/psta	6168	154523	10276	3694	98503	366347	463



```
HSV 5000-1FE1-0015-2C50 ($lsgga4:) Wed Dec 8 16:06:50 2004:
```

Disk	Volume Queue Depth	Volume Latency (us)	Drive Queue Depth	Drive Latency (us)	Read RPS	Read KBS	Write RPS	Write KBS	Shelf	Bay
0	3	5233	3	5230	97	2501	19	1468	0	1
1	2	3408	2	3403	61	1677	6	829	1	10
2	2	3904	2	3900	68	1952	11	638	0	2
3	2	5433	2	5430	68	1804	12	1405	1	2
4	3	8856	3	8852	61	1780	18	1405	0	3
5	2	3208	2	3204	53	1179	13	1277	1	3
6	2	4216	2	4211	43	1055	21	1661	0	4
7	3	7027	3	7024	55	1110	28	2429	1	4
8	3	9838	3	9835	62	1664	37	2812	0	5
9	2	5700	2	5697	56	1361	25	2173	1	5
10	3	5620	3	5617	62	1614	26	2301	0	6
11	2	3889	2	3886	79	1858	11	1021	1	6
12	3	7019	3	7016	63	2206	18	1532	0	7
13	3	5921	3	5918	69	1908	22	1405	1	7
14	2	3427	2	3422	84	1766	14	894	0	8
15	2	2579	2	2576	60	1273	2	383	1	8
16	3	7838	3	7835	69	1828	16	1135	0	9
17	3	5772	3	5769	105	2796	13	1021	1	9
18	2	5123	2	5120	114	2853	13	1007	0	10
\$lsgga4/pd	2	5474	2	5470	1329	34185	325	26796		

Found 19 physical disks; displayed data for 19 disks.



HSV 5000-1FE1-0015-2C50 (\$1\$GGA4:) Wed Dec 8 16:06:50 2004:

Name	Source ID	Queue Depth	Busies
-----	-----	-----	-----
CLETA-P0	0x110700	1	0
CLETA-P1	0x033E00	30	0
VMSROC-P0	0x120300	0	0
VMSROL-P0	0x120400	0	0
NOME-P0	0x120500	30	0
WFGLX0-P0	0x111700	0	0
CEAGLE-P0	0x110800	17	0
PAVER-P0	0x150100	15	0
C930COA9-P0	0x150800	1	0
MTDIB7-P0	0x111600	0	0
FALM25-P0	0x113200	0	0
MRVP1-P0	0x110C00	0	0
MRVP1-P0	0x110D00	32	0
VMSMO-P1	0x030B00	0	0
MRVP1-P0	0x140100	0	0
VMSROC-P1	0x033300	30	0
C9374056-P0	0x140200	0	0
VMSMO-P0	0x110900	0	0
VMSJO-P1	0x030A00	0	0
VMSJO-P0	0x110A00	0	0
VMSROL-P1	0x033500	0	0
CEAGLE-P1	0x033900	16	0
Q2020DD6-P0	0x112E00	0	0
Q22286D6-P0	0x121A00	0	0
Q22286ED-P0	0x120A00	0	0
IPFIO9-P0	0x120600	0	0
IPFIO8-P0	0x110500	0	0
MTDIB7-P1	0x033400	0	0
IPFIO1-P0	0x112200	0	0
IPFIO5-P0	0x121200	0	0
MTDIB7-P0	0x110600	0	0
IPFIO7-P0	0x110400	0	0
IPFIOA-P0	0x113800	0	0
VMSTIC-P0	0x113100	0	0
-----	-----	-----	-----
\$1\$GGA4/hc (sum)		172	0
Cleta>			:34 connections