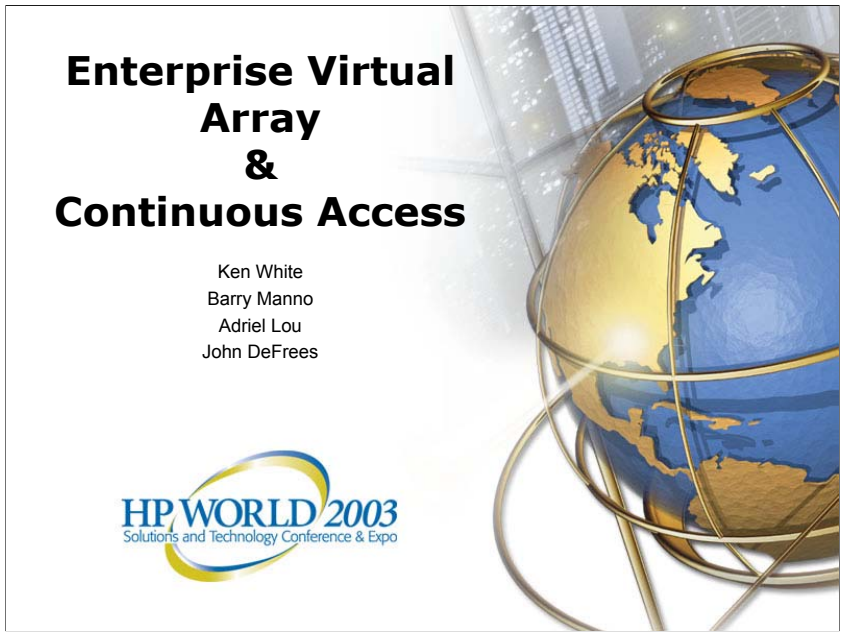
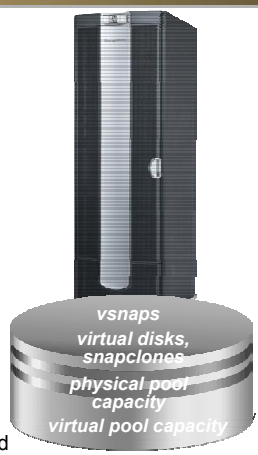

Enterprise Virtual Array
&
Continuous Access
Ken White
Barry Manno
Adriel Lou
John DeFrees
HP WORLD 2003
Solutions and Technology Conference & Expo

hp StorageWorks enterprise virtual array

HP WORLD 2003 Solutions and Technology Conference & Expo

- Design features
  - modularity, scalability and density
  - full end to end 2Gbps fibre channel performance
    - host, infrastructure, controllers and disks
  - high availability with vRAID and component redundancy
  - centralized, lite-touch management
  - storage array based virtualization architecture
    - lower TCO and increase ROI
      save time, space and money
- Target deployments
  - enterprise open system data center operations
  - VLDB, OLTP, ERP, CRM
  - HP-UX, OpenVMS, Tru64, Windows NT/ 2000, Red Hat/SUSe Linux, Solaris, IBM AIX, Netware



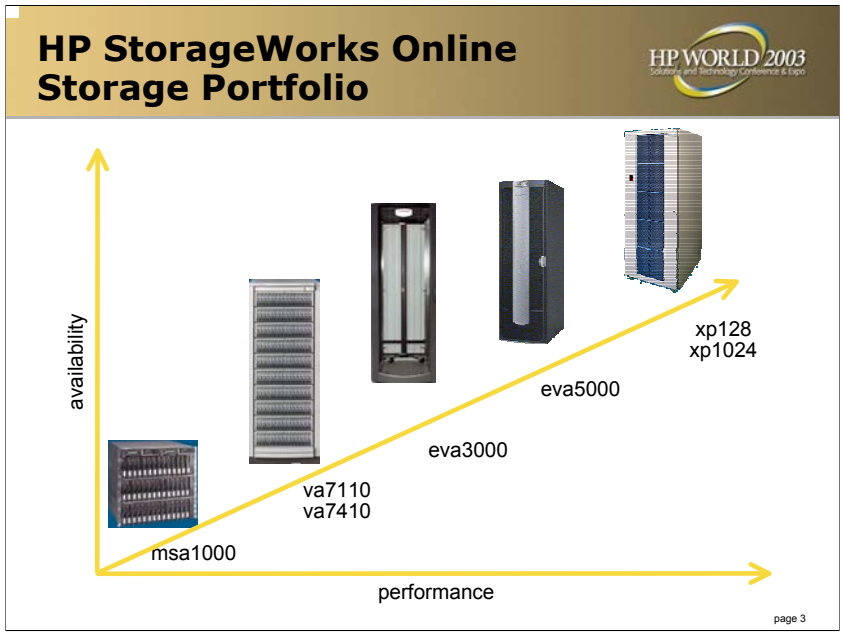
page 1

The HP StorageWorks enterprise virtual array is executing on the advantages of virtualization at the storage system level enabled by SANworks VersaStor technology by hp.

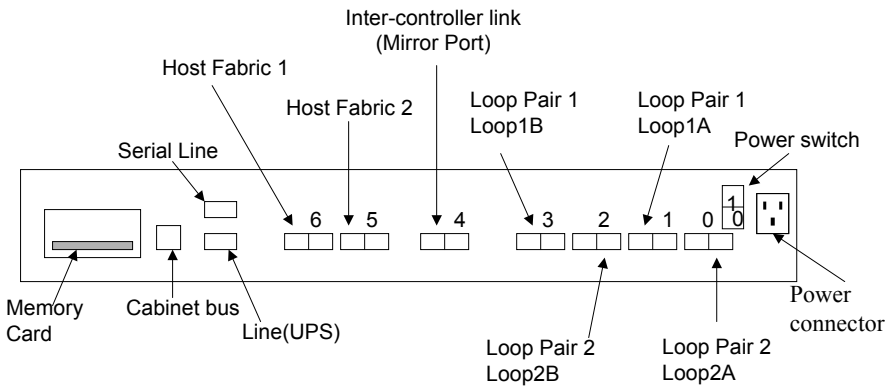
While virtualization is being addressed by hp at all levels of the SAN, the most powerful application of this technology today is at the storage system level. With the HP StorageWorks enterprise virtual array, organizations are able to take advantage of the many benefits of virtualization with a solution designed specifically for open systems.

hp enterprise virtual array

- Features
  - scaleable modular networked storage
  - higher storage capacity utilization
  - high capacity in dense packaging footprint
  - centralized and simplified management
  - auto load balance & self-tuning performance
  - hot-swap redundant components, vRAID 0, 1, 5
  - 2Gbps fibre channel array controller
  - up to 35TB, up to 240 fibre channel 1” HDDs behind single pair of array controllers
  - 1GB cache per controller
  - centralized management via SAN Appliance


HP StorageWorks Online Storage Portfolio
HP WORLD 2003
availability
performance
msa1000
va7110
va7410
eva3000
eva5000
xp128
xp1024
page 3

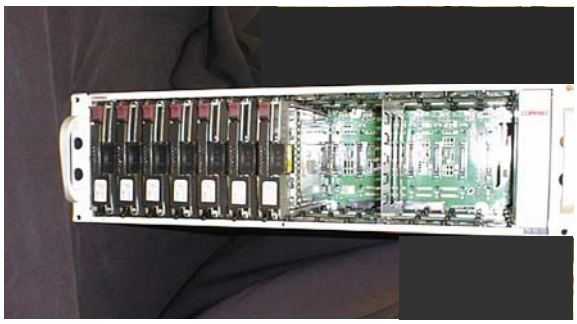
HSV array controller connectors—rear view



page 4

M5214—drive enclosures—front view

- 3U disk enclosure
- Dual redundant active-active 2Gb/s FC busses
- Fourteen 1-in. FC disks per enclosure



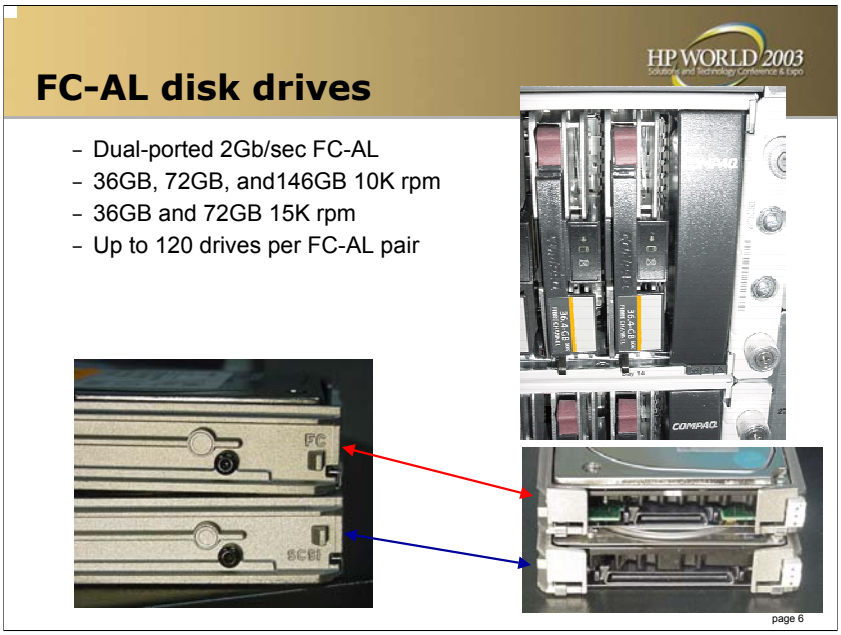
page 5

From the front, with disks installed, it is hard to distinguish a FC drive enclosure from a SCSI drive enclosure.

This view shows the smaller FC drive mating connector on the FC drive enclosure backplane, as opposed to the larger SCSI connector on the SCSI drive enclosure backplane.

Only 1” drives, 14 of them, are supported.

The FC drive enclosure supports dual FC busses to the drives. These are redundant and allow for load balancing between the busses. Both busses are active simultaneously.



SCSI drives only have one port, because of the number of signals/pins required for the interface, with no redundancy or load balancing option of a second port. FC drives have two ports, providing redundancy and load balancing.

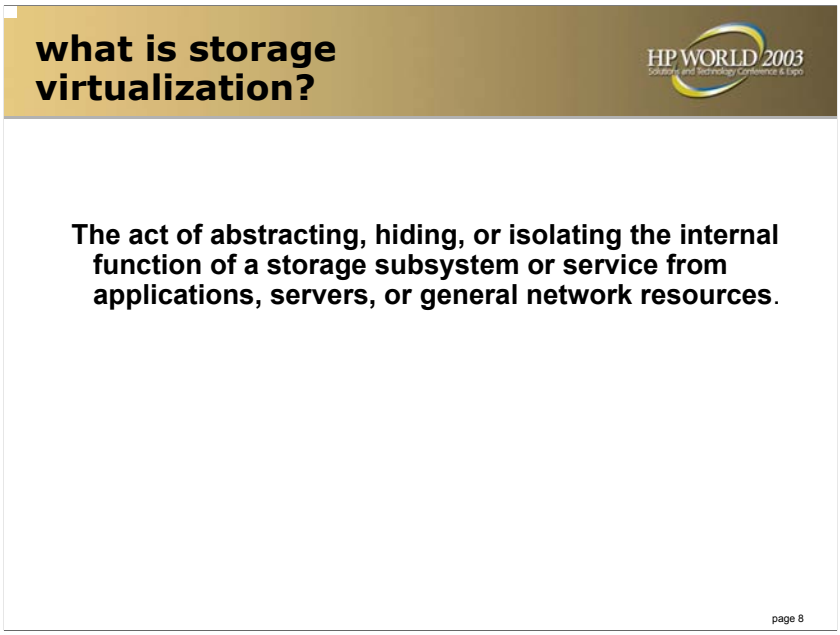
The SCSI interface has a limited distance that it can support and as a result a limited number of disks per bus. The FC interface allows, in Compaq's implementation up to 120 drives on redundant pairs of FC busses. The redundant pair of busses also allows load balancing, so that if one buss or port is busy, any drive can be accessed on the other buss or port.

In some cases, SCSI and FC drives share the same mechanics, but they are beginning to diverge. Some SCSI drive vendors are designing new drives with lower cost in mind, trading off performance for cost, as the drives have a wide range of applications including PCs and low-end servers. FC drives, on the other hand, are being designed for high performance and high capacity storage applications.

Seagate is our current FC drive vendor.

Terminology

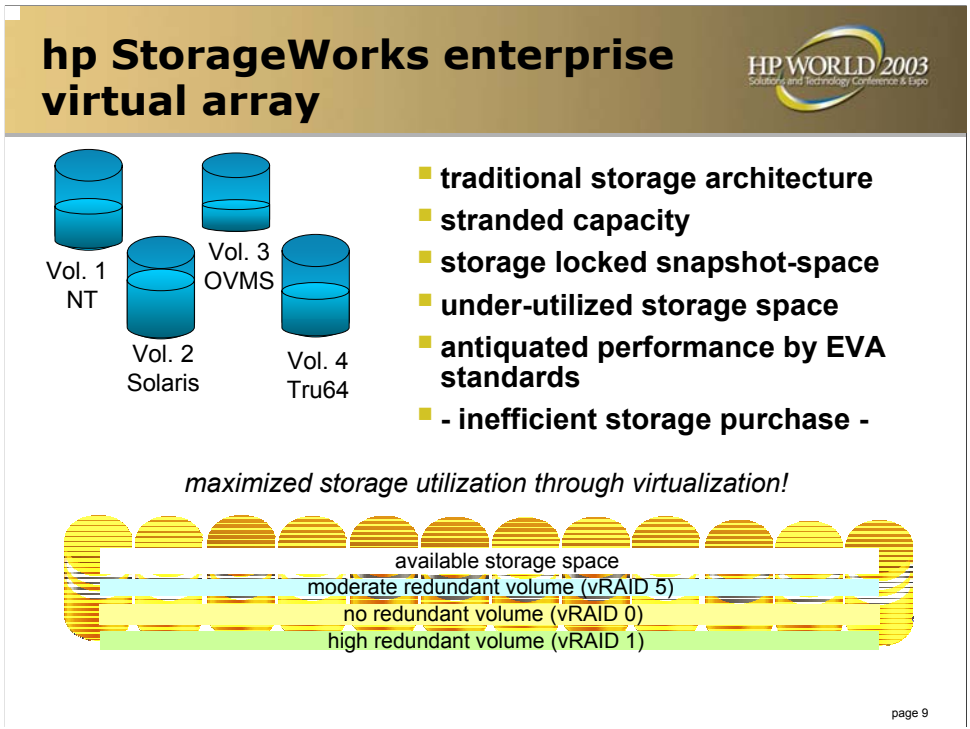
- Define the term:
  - Virtualization
  - VRAID
  - Disk Group
  - Virtual Disk
  - LUN
  - Host
  - Presentation
  - Virtual Disk Leveling
  - Distributed Sparing
  - Virtually Capacity Free Snapshots
  - Snap Clones
  - Command View EVA
  - SAN Script
  - Virtual Controller Software



Isn't new.

In 1974, IBM introduced Multiple Virtual Storage (MVS), the primary operating system used on mainframes.

MVS evolved over time and was prepackaged with an extensive set of utilities and renamed OS/390. MVS now refers to the base control program in OS/390.
(Source: Techencyclopedia)

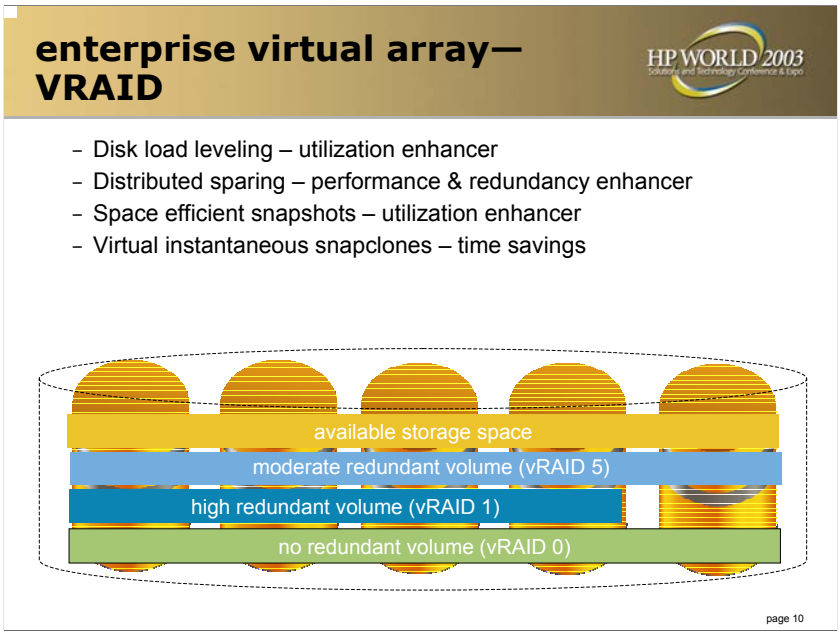


In a traditional storage environment you would purchase as many drives as the server could hold to make sure you did not have

To perform an upgrade until the disk is completely full. This lead to your server using only 40-50% of your storage because you

Were planning on growth that may never happen.

With Virtualization you can carve a LUN (Vdisk) out of the pool of storage starting at 1GB and increasing to 2TB in 1GB increments, this allows you to grow your volumes as needed. No Wasted Space!



Increased utilization of available capacity

- no need to allocate large files for databases **to insure you don't exceed the size of the allocated database. With enterprise you can easily expand the database dynamically interrupting the application**

Eliminate stranded capacity

- No longer need to create a volume large enough to handle the anticipated level of activity for the volume and then fine you've over-allocated and locked out the disk space

Reduce capacity requirements with "virtually capacity-free" snapshots

- No need to allocate a block of disk space equivalent in size to the disk to get the snapshot.

Manage overall utilization at higher levels

- Utilize more of your disk by reducing the amount of stranded capacity and space needed for 'virtually capacity-free' snapshots

Capacity-on-demand virtual disks

- Easily expand the size of your disk pool by adding physical disk drives without interrupting your applications. Enterprise will automatically evenly distribute the files across all physical disks. When a virtual disk is reaching it's capacity, simply enter the additional space needed and virtual disk will automatically expand to the new size.

Virtual Disk redundancy:

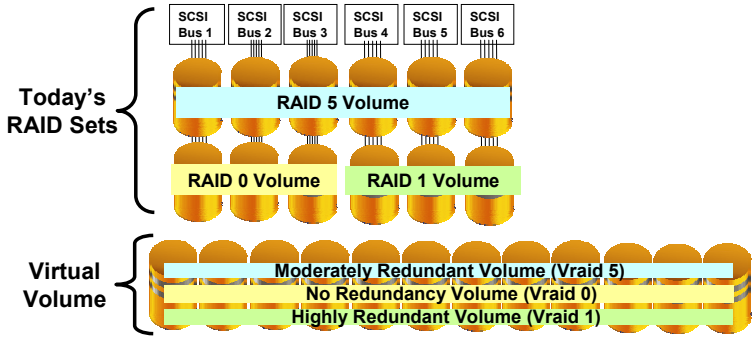
None (VRAID0): Data is striped across all physical disks in the Disk Group.

Moderate (VRAID5): Data is striped with parity across all physical disks in the Disk Group. Always 5 (4+1) physical disks per stripe are used.

High (VRAID1): Data is striped mirrored across all physical disks (even number of them) in the Disk Group. Established pairs of physical disks mirror each other.

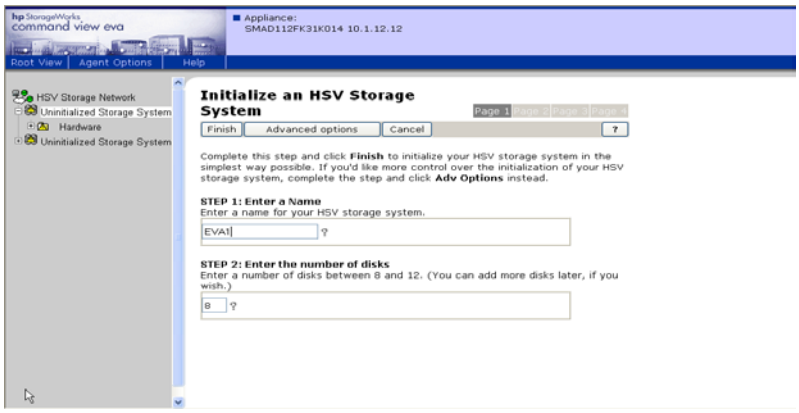
VRAID versus RAID

- Eliminate throughput bottlenecks
- Workload is evenly distributed across all spindles in group
- Eliminate load balancing procedures for application and database



terminology—disk group

- Group of physical disks, from which virtual disks will be created
- Minimum of 8 disks



page 12

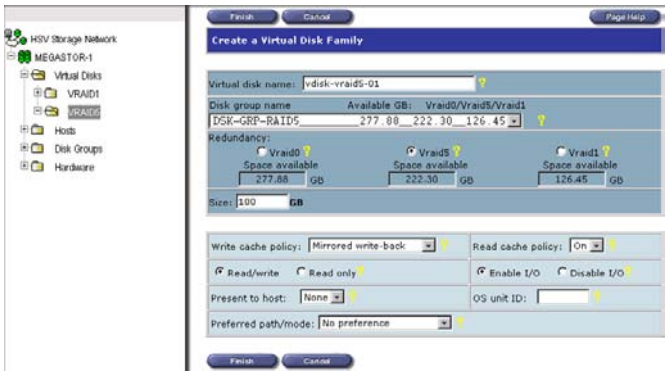
sizing HSV disk groups— considerations disk count formula

- Hardware versus software capacities
  - Physical 1000 000 000Bytes = 1GB
  - Software 1073 741 824Bytes = 1.07GB Physical (230)
  - ~ 7% Variance → 1GB Physical = 0.93GB Software
- System metadata overhead — 0.2%
  - System metadata
  - MLD—HSV Element Manager metadata
  - Virtual Disk metadata
- Vraid overhead
  - Vraid0 — 0% (1 block for every 1 block usable)
  - Vraid1 — 50% (2 blocks for every 1 block usable)
  - Vraid5 — 20% (1.25 blocks for every 1 block usable)
- Snapshot working space
  - Snap — capacity of original is “allocated” *or* dependent on rate of change of original data
  - Snapclone — same physical capacity as virtual disk
- Spare capacity
  - 2 X physical capacity of the largest physical disk X protection selected

page 13

terminology—virtual disk

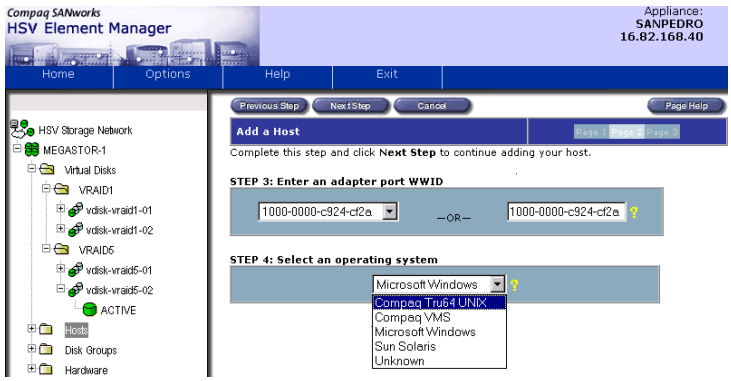
- Logical disk with certain attributes
- Resides in a disk group



page 14

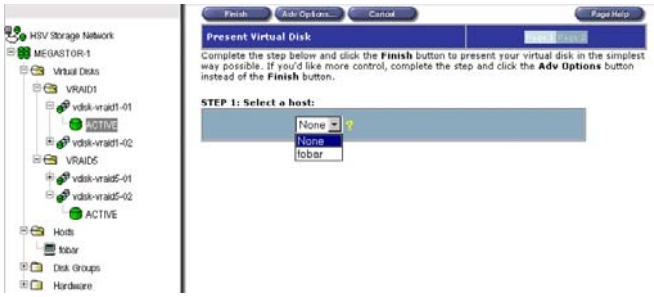
terminology—host

- Collection of HBAs, that reside in the same (virtual) server



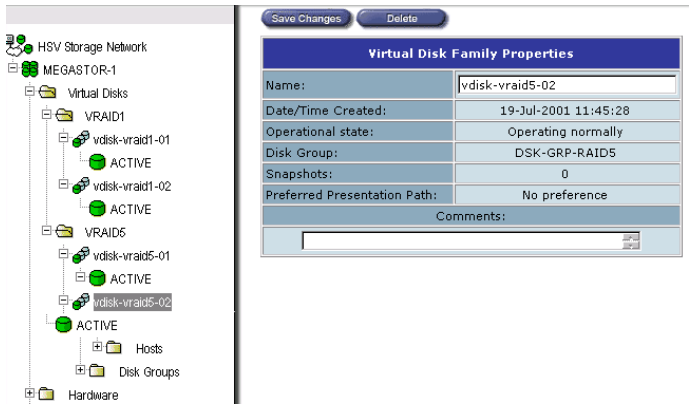
terminology—presentation

- A persistent instruction to make a specified virtual disk available to all adapters on the specified host as the specified SCSI LUN

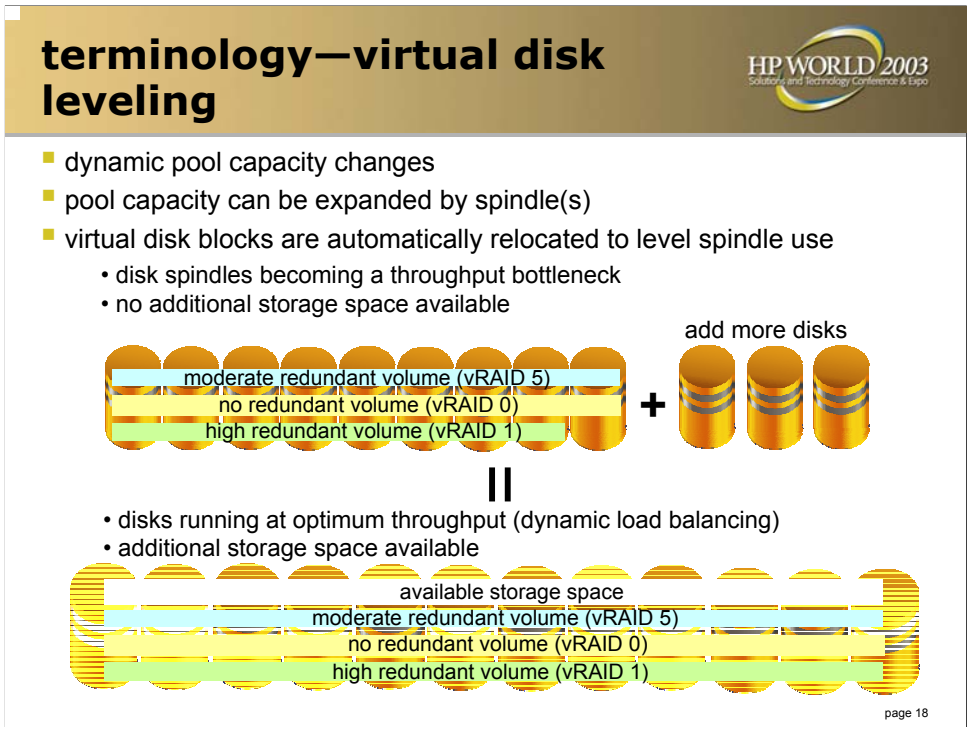


terminology—LUN

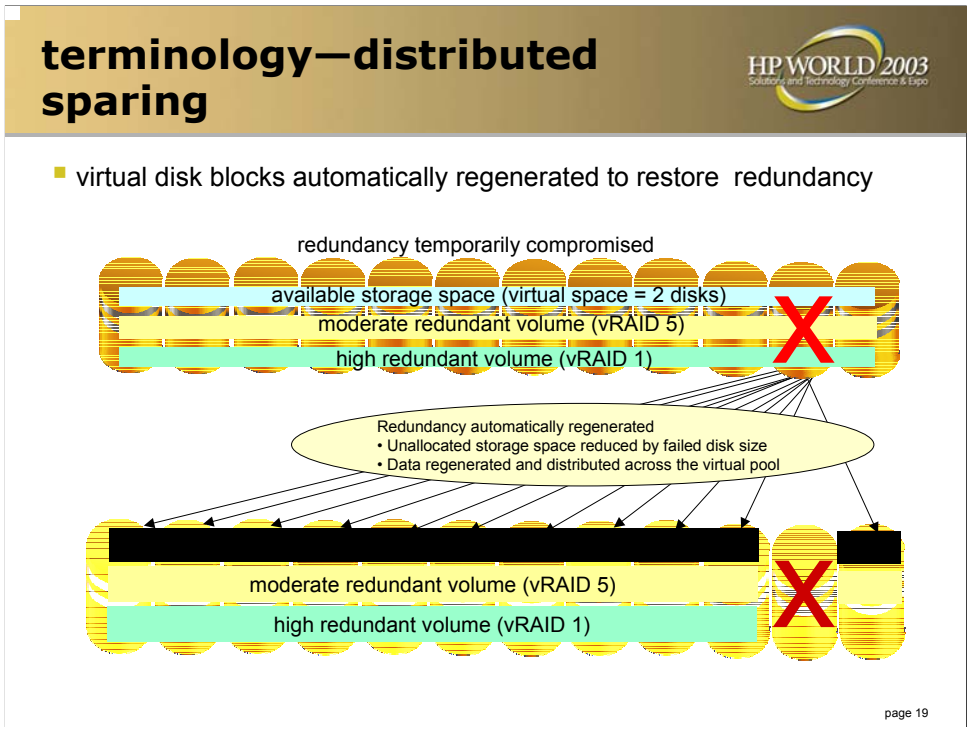
- Virtual disk presented to one or more hosts



page 17



With virtualization you can now manage by attribute, think of only blocks not disks… In the upper diagram let's assume that there are 100 72GB drives in the pool. You also have a 100GB database on a server attached to the EVA, therefore there would be 1GB of blocks on each drive. Now add the 3 drives for the diagram at the bottom, this will now span 103 drives but the database will still only be 100GB, this allows performance to increase because the data is automatically distributed to the other 3 drives.



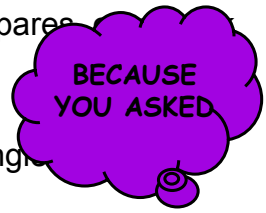
Think about all the servers that you have…. Now think about each array having a separate spare drive in case a drive fails in the middle of the night… That is probably a large number of drives…

In the EVA we use a process called distributed sparing, this process allows us to add the spare drives into the disk pool for performance and we take the capacity of those drives out of the pool to allow for a failed drives VD data to be moved into that space.

With the EVA we change the process of sparing and give you extra performance as well…..

GO TO NEXT SLIDE for better example

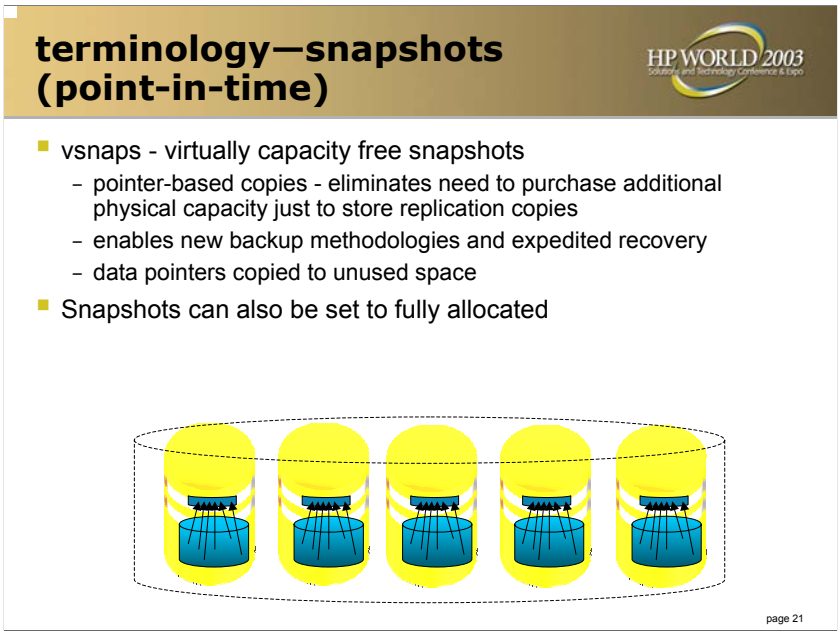
terminology—distributed sparing sizing



- Chunks allocated, but not dedicated as spares [illegible] drives to survive 1 or 2 disk drive failures

- Example:
  - Disk Group: 8 * 36G, Protection level: Sing[illegible]
    - Total Disk Group size: 8 * 36 = 288 GB
    - Maximum size for Virtual Disks (8-(1*2)) * 36 = 216 GB
  - Disk Group: 120 * 36G, Protection level: Double
    - Total Disk Group size: 120 * 36 = 4.32 TB
    - Maximum size for Virtual Disks (120-(2*2)) * 36 = 4.18 TB

- Note: This does not account for difference between physical and software GB, more later



The enterprise virtual array controller provides three types of data replication tools:

Traditional snapshot

Virtually Capacity-Free Snapshot

Virtually Instantaneous Snapclone

These data migration and replication tools have been refined in the enterprise virtual array to provide tremendous flexibility and data protection.

What are Virtual snapshots?

Virtual disk snapshots, or Vsnaps, are disk-based copies that reduce backup windows to seconds. This is achieved by replicating pointers to data — instead of the data itself.

Benefits and value of Vsnaps:

Snapshots of virtual disks are virtually capacity-free

Ad hoc Vsnaps can be taken in seconds

Vsnaps require minimal physical capacity;only modified blocks take space

Vsnaps can be mounted in seconds for disk-based recovery

Vsnaps can be mounted and copied to tape anytime for archive

There are two Snapshot options:

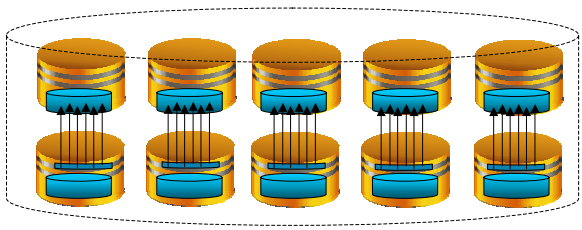
Traditional Snapshot is the standard way of doing snapshots where an exact amount of space is reserved, equal to the original Virtual Disk or LUN.

Virtually Capacity Free Snapshot, which uses space for the snapshot only as required as the original copy changes. This is a feature most competitors do not have.

terminology—snapclone (point-in-time)

HP WORLD 2003 Solutions and Technology Conference & Expo

- snapclone - virtually instantaneous replication
  - combines the best of snapshot and clone (physical copy)
  - provides virtually instantaneous Snapclone data replication
  - similar to snapshot
  - data cloned to second set of disks in background
  - data available after snap



page 22

What are Snapclones?

A Snapclone integrates the best features of Vsnaps and clones with automation to provide the ultimate data mining solution. Independent physical clone copies of data can be accessed instantaneously — rather than waiting for the complete physical movement of the data.

Benefits and value of Snapclone:

Enables application and live testing of data without disrupting business operations

Can be mounted and copied to tape for archive

Enables data to be relocated from main data center in case of site disaster

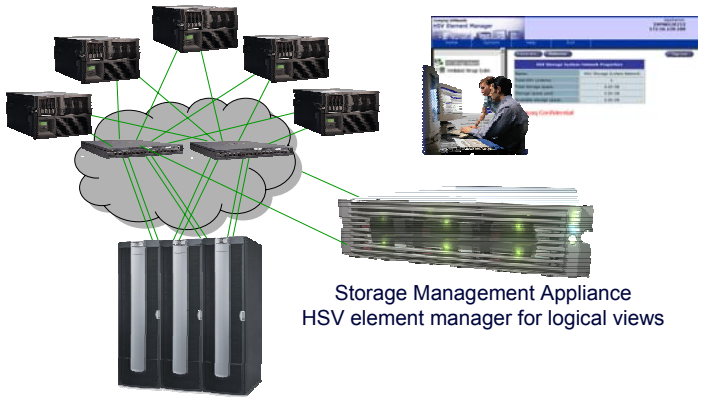
Virtually Instantaneous Snapclone is a new implementation of clone or copy, which can save customers time and money.

Virtually Instantaneous Snapclone is an improved type of data cloning similar to a traditional clone because duplicate space is reserved. A complete copy of the original virtual disk is made as quickly as data transfer rates permit, resulting in two identical independent copies of the data in the shortest time possible.

There is an important difference between Virtually Instantaneous Snapclone and the traditional clone. With traditional controllers, the clone copy is not available until the copy is complete. With Virtually Instantaneous Snapclone, the snapclone data can be accessed virtually immediately.

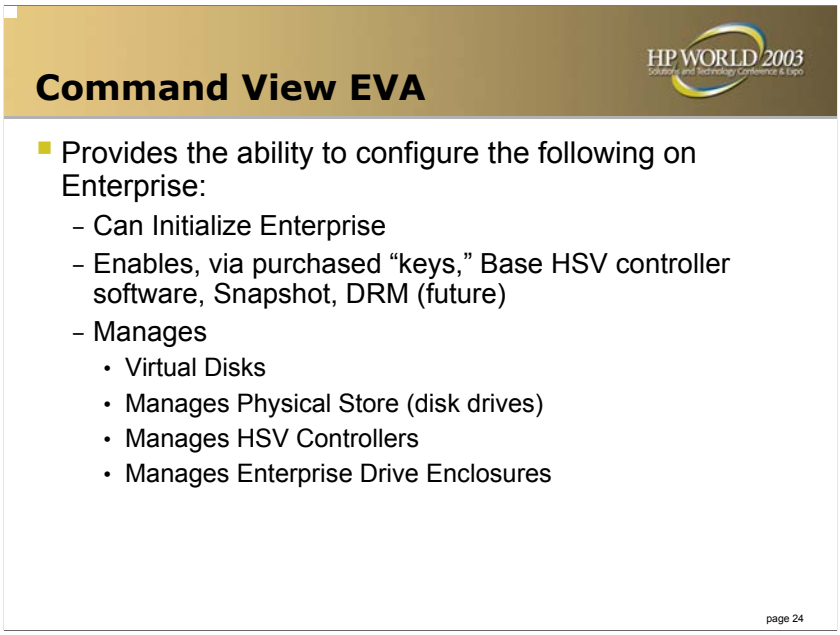
terminology—Command View EVA

- Application to manage Enterprise Virtual Array
- Runs on OpenView Storage Management Appliance



Storage Management Appliance
HSV element manager for logical views

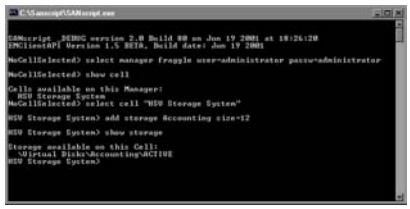
page 23



This slide outlines the basic Configuration features of the HSV Element Manager.

terminology—Storage System Scripting Utility (SSSU) aka SANscript

- Tool to issue commands through a shell UI and execute scripts
- Scripts to create/modify a configuration
- Interactive mode as a command prompt
- Configuration commands to add, set, and delete
- Capture of a configuration into a script file
- Show command to display configuration
- Establishes a session with HSV agent on SAN Management Appliance



virtualization controller software v3.0

- non-disruptive download firmware updates
  - controllers, EMU and disk drives
- controller code
  - distributed with controller hardware from the factory
  - updates distributed on CD-ROMs and updated from the SAN Appliance
  - enabled (base, snapshot and, later, DRM) by software license keys which are managed by SAN Appliance / HSV element manager
- support for:
  - dual redundant controllers
  - mirrored cache, 1GB per controller, running across a dedicated 2Gb/s FC port
  - support for multi-bus failover (native or with Secure Path)
  - support for up to 512 hosts
  - support for up to 512 LUNs / virtual disks (a max of 256 LUNs per HBA*)
  - support for up to 1024 HBAs
  - support for SMART environmental monitoring unit (EMU)
  - Continuous Access EVA Disaster Tolerant configurations

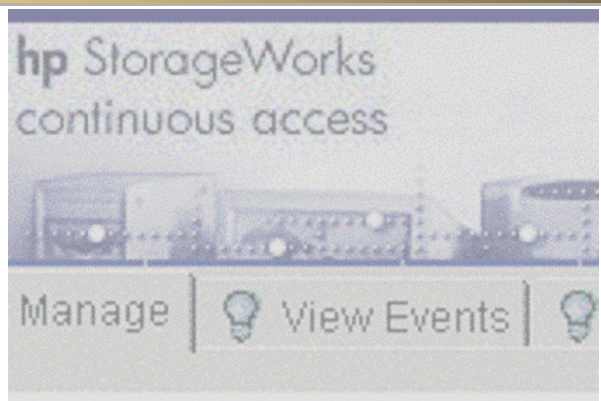
*Some HBAs won't support a max of 256 LUNs. The HBA used with IBM AIX will only support 16 LUNs. The HBAs used with HP-UX will support 128 LUNs.

page 26

•Another important feature is non-disruptive download of controller software, EMU firmware and drive firmware.

• The controller code is no longer distributed on PCMCIA cards, like it is on the HSG80. It is stored in flash memory on the controllers when they leave the factory. New code can be distributed on CD-ROMs (later electronically over the web, but not at first release). The code is enabled, including Base and Snapshot (and later DRM) with Software License keys which are managed by the HSV Element Manager running on the SAN Appliance.

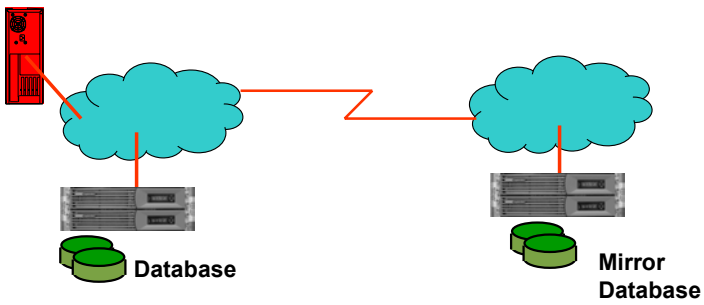
CONTINUOUS ACCESS



page 27

Continuous Access concepts

- What is Continuous Access EVA ?
  - CA copies data online and in real time via synchronous replication to remote locations through a local or extended SAN



page 28

Continuous Access for EVA
Terminology

- Source
  - EVA with Data Replication functionality, which acts as the master for a Data Replication (DR) Group
- Destination
  - EVA with Data Replication functionality, which acts as the slave for a given DR Group
- Copy Set (CS)
  - Relation between a virtual disk on a source EVA and a virtual disk on a destination EVA. Data is copied from the source to the destination. A copy set is uni-directional
- Data Replication (DR) Group
  - A collection of 1 or more copy sets, logically grouped together, usually for data consistency reasons

page 29

XP Terminology:

Remote Copy Set equals to CA-pair (Continues Access Pair)

Association Set equals to Consistency Group

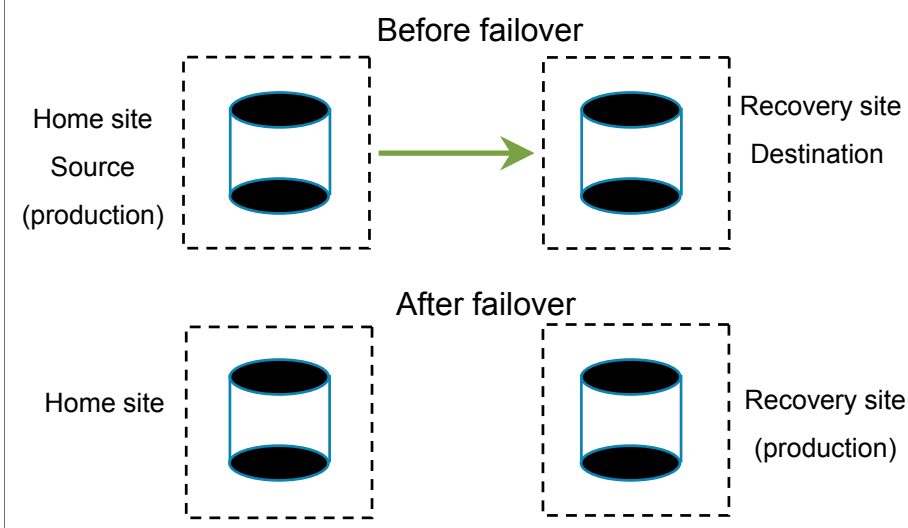
Continuous Access for EVA
Data Replication (DR) Groups and Managed Sets

HP WORLD 2003 Solutions and Technology Conference & Expo

- DR Group characteristics
  - A DR Group is a collection of 1 or more copy sets which share the following common attributes
    - exist in the same EVA
    - they share a log
    - if one copy set fails, host access is removed from all other copy sets as well (for failsafe_mode = enabled)
    - I/O ordering is kept across all copy sets
    - all copy sets failover together
  - In other words: DR Groups are used by a host to keep multiple copy sets consistent with one another

page 30

Continuous Access for EVA
Terminology



page 31

Continuous Access for EVA
Terminology

- Log space
  - Space taken from a disk group that will be used when a DR Group starts logging
  - Log space is taken from the disk group with the most available free space, by default, when more than one disk group exists
  - Log space is "per DR Group" and can grow to either 2TB, 2xDR Group membership size, or all available disk group free space (minus protection level), whichever comes first
- Command View EVA
  - Web-based application for managing/configuring EVAs
- CA-UI
  - Abbreviation for Continuous Access User Interface. Web-based application. Server part resides on the SAN Appliance
- Managed Set
  - Collection of DR Groups that can be failed over all at once by CA-UI. May span multiple EVA CA configurations

page 32

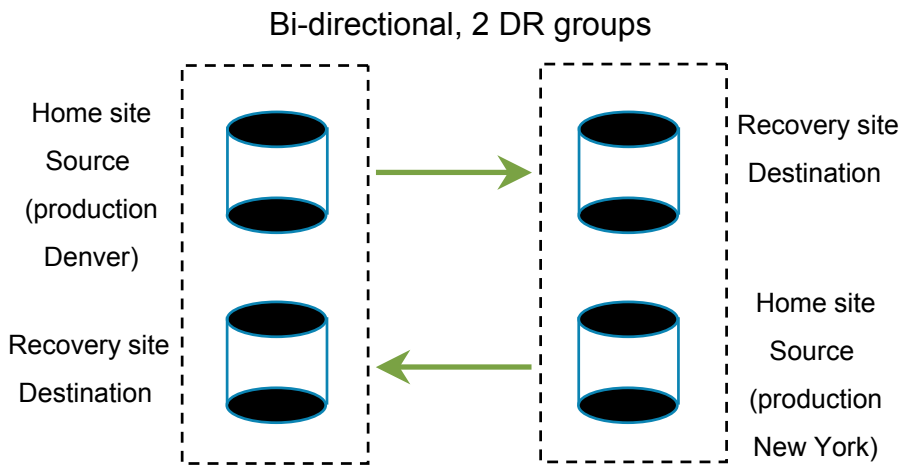
Managed Sets are only available via DRM-UI (aka Doctor MUI)

We always use logging and cannot be switches off.

Continuous Access for EVA
Data Replication (DR) Groups and Managed Sets

- Managed Sets
  - A Managed Set is a collection of DR Groups which share common attributes, say:
    - are used by the same line of business
    - all DR Groups within an EVA
  - A Managed Set may span multiple EVA CA setups
  - Only available in CA-UI, not in Command View EVA
  - Managed sets are used by an end-user to group DR groups into a single management entity

Continuous Access for EVA
Terminology



page 34

Continuous Access for EVA
Feature Set

- 64 Copy Sets
- Up to 8 Copy Sets per DR Group
- 256 Vdisks (this includes snapshots & snapclones, and 7 snap or snapclone per Vdisk on either source or destination, but not both)
- Synchronous Replication
- Bi-directional Replication
  - On DR Group level
- No restriction on virtual disk characteristics
  - Size: 1GB ←→ 2 TB
  - Vraid levels : 0*, 1 and 5

page 35

So no asynchronous in final version.

Try to avoid Vraid0 as long as possible.

Continuous Access for EVA
Feature Set

- **Dynamic expansion of virtual disks** (for OS's that support this feature)
- **ISLs can use:**
  - DWDM
  - FC-IP
  - Multi Mode Fiber/Single Mode Fiber

- Up to 16 Disk Groups per storage system
- Members of a DR Group can reside in different disk groups

page 36

Continuous Access for EVA
Feature set

- **Management:**
  - Continuous Access User Interface (CA-UI)
  - Command View EVA
  - Scripting (SSSU)

page 37

Continuous Access for EVA
Synchronous and Asynchronous Copy Modes

HP WORLD 2003 Solutions and Technology Conference & Expo

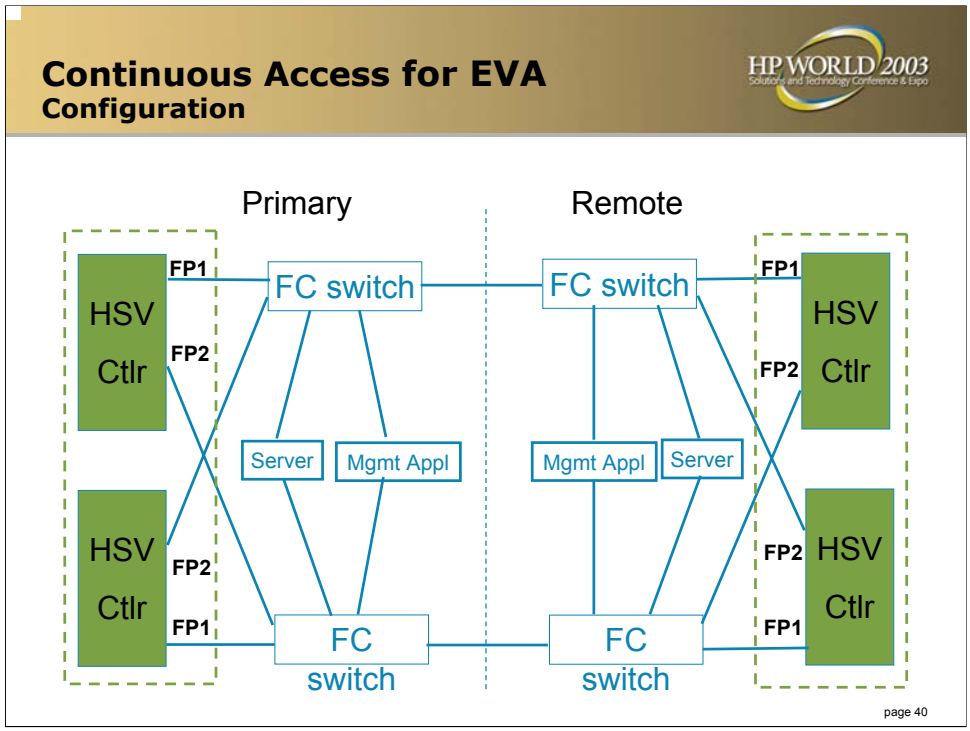
- Synchronous Mode
  - Remote site data is constantly kept consistent with local site data
  - I/O completion status is not returned to the host until both the local and remote writes (to cache) are complete
  - Provides real-time mirroring of data
  - In-order delivery is guaranteed (using Group Sequence Numbers (GSNs)
  - Appropriate when exact data consistency is crucial to the application

page 38

Continuous Access for EVA
Synchronous and Asynchronous Copy Modes

HP WORLD 2003 Solutions and Technology Conference & Expo

- Synchronous Mode
  - Synchronous mode can affect overall performance by increasing response time on writes
  - The higher the latency of the link, the more impact on performance
  - Synchronous mode is what VCS V3.0 uses



2 SANAppliances, being 1 per site. Advised is to have the SWMA on the primary site shutdown and manage everything from the SWMA on the remote site. On the SWMA, 2 components are important: HSV Element Manager and DRM-UI. The SWMA must be able to 'see' both the Source and Destination Storage System.

HSV controllers are cross-wired, thus allowing fabric failover on the same controller and load balancing over both ISL's

DRM traffic is preferred to FP1. This is true for V3 and may change in V4 (so that it will use FP2 as well)

FP means Fabric Port

DP means Device Port

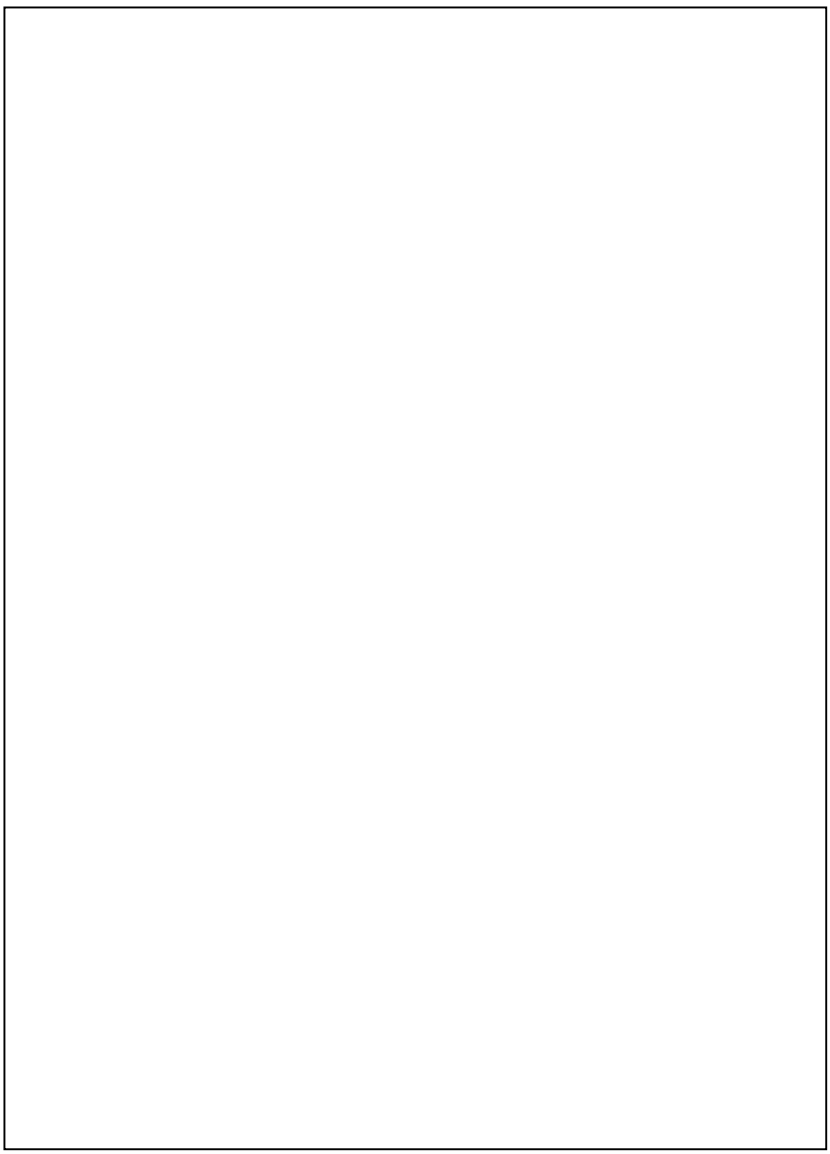
MP means Mirror Port

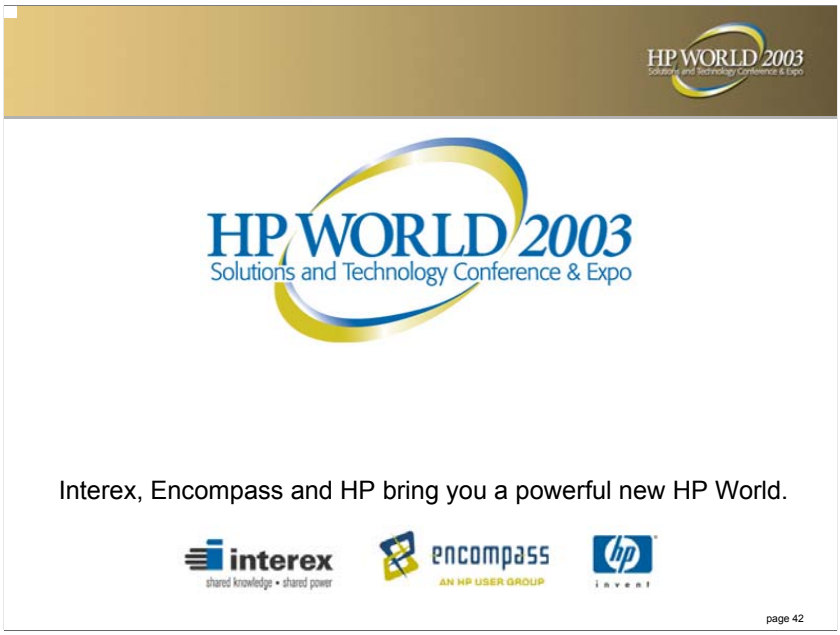
DRM-UI does NOT have access to the MLD, hence the

multi operating system support

- Windows NT/2000
  - secure path v4.0A
- HP-UX 11.0 and 11.i
  - MC Service Guard
    - 11.13 and 11.14
  - secure path v3.0A
- Tru64 UNIX 5.1, 5.1a and 5.1b
- OpenVMS
  - 7.2-2 with Fibre_SCSI-V0200
  - 7.3
- Netware 5.1, 6.0
  - secure path v3.0C
- SUN Solaris 2.6, v7 and v8
  - VERITAS Clustering v1.3
  - SUN Cluster v2.2
  - PCI, cPCI & S-Bus HBAs
  - secure path v3.0A/3.0B
- IBM AIX
  - HACMP certification
  - 4.3.3 and 5.1
  - secure path v2.0C
- Linux (x86)
  - Red Hat 7.2
  - Advanced Server 2.1
  - SLES 7
  - secure path v3.0

page 41




HP WORLD 2003
HP WORLD 2003
Solutions and Technology Conference & Expo
Interex, Encompass and HP bring you a powerful new HP World.
interex
shared knowledge • shared power
encompass
AN HP USER GROUP
hp
invent
page 42

Reference Information

Continuous Access for EVA
Configuration Rules

- Cabling method of EVA host ports to allow fabric failover per controller
- SAN Appliance
  - Minimum of 2, so 1 per site
  - Both HBAs of each OV Storage Mgmt Appliance must be connected to a different fabric
  - Must be able to “see” both source and destination EVA
  - Only one OV Storage Mgmt Appliance may be actively monitoring the EVAs
  - Hops between the local OV Storage Mgmt Appliance and the local EVA are not allowed
- Destination virtual disk can not already exist
- Maximum of 7 hops between source and destination EVA
- Minimum of 4 FC Switches in 2 Fabrics
- 2 ISLs, 1 per Fabric

page 44

Continuous Access for EVA
Comparison with DRM on HSG80

EVA

- Nr of CS's: 64
- Nr of DR Groups: 64
- Copy methods:
  - Synchronous
- Nr of CS's per DR Group: 8
- Not applicable due to architectural difference
- Always uses write history log
  - Always ORDER_ALL
- Max CS size: 2 TB

HSG80

- Nr of RCS's: 12
- Nr of association sets: 12
- Copy methods:
  - Synchronous
  - Asynchronous
- Nr of RCS's per association set: 12
- No controller based partitions of RCS's
- Write history log *optional*
  - NOORDER_ALL is default
- Max RCS size: 1.1 TB

page 45

Continuous Access for EVA
Comparison with DRM on HSG80

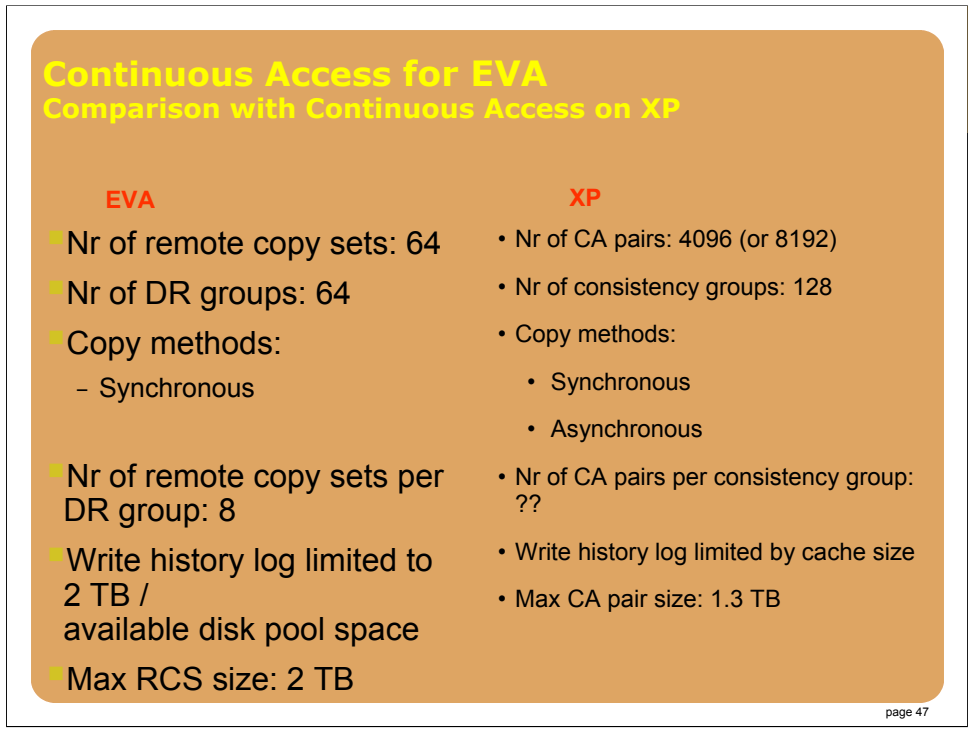
HP WORLD 2003 Solutions and Technology Conference & Expo

EVA	HSG80
▪ Automatic selection of HSV host port for peer to peer communication – Host access is possible too ▪ Options on DR Group level: – Synchronous – Failsafe mode: enabled / disabled ▪ Fabric Failover per controller – Both controller ports connected to different fabrics ▪ Max “local” Virtual Disks: 256 ▪ Vraid levels supported: – Vraid 5, Vraid 1 ▪ Functionality enabled via license key	• Dedicated HSG host port for peer to peer communication • No host access • Options on RCS level: • Synchronous / asynchronous • Failsafe / normal • Fabric Failover by controller failover • Both controller ports connected to the same fabric • Max “local” LUNs: 12 • Raid levels supported: • 0, 1, 0+1, 5, JBOD • Functionality enabled with special firmware card (ACS V8.xP)

page 46

XP: Fence level=Data equals to Failsafe Mode=Enabled

Still to be determined whether DRM is a capacity or a feature license.



XP: Consistency Groups only used for async (which is used for > 10km)

XP: Site File equals to Write History Log. Max is limited to a certain (settable) amount of shared memory. Typical 30%.

additional resources

- EVA 5000 Documentation (External Resource)
- http://h18006.www1.hp.com/products/storageworks/enterprise/documentation.html
- EVA 3000 Documentation
- http://h20000.www2.hp.com/bizsupport/TechSupport/DocumentIndex.jsp?contentType=SupportManual&locale=en_US&prodTypeId=12169&prodSeriesId=315127&taskId=115&prodSeriesId=315127&docIndexId=763
- Continuous Access Information
- http://h18006.www1.hp.com/products/storage/software/conaccesseva/index.html
- Education and Training
- http://h18014.www1.hp.com/training/

page 48


invent