VIOS Shared Storage Pools

Phase 2 – December 2011

Nigel Griffiths
IBM Power Systems
Advanced Technology Support, Europe (version 12b)

© 2011 IBM Corporation

VIO Shared Storage Pool phase 2

- Announced   14th Oct 2012
  http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=AN&subtype=CA&htmlfid=897/ENUS211-354&appname=USN

- Available   16th Dec 2012

- **Please read the Release notes:**
  http://www-01.ibm.com/support/docview.wss?rs=0&uid=isg400000876

  VIOS 2.2.1.3 FixPack 25 SP01 Virtual I/O Server 2.2.1.0 VIOS 2.2.1.3 FixPack 25 SP 01 Readme

  Readme file for: VIOS 2.2.1.3 FixPack 25 SP01
  Product/Component Release: 2.2.1.0
  Update Name: VIOS 2.2.1.3 FixPack 25 SP 01
  Fix ID: VIOS_2.2.1.3-FP25-SP01
  Publication Date: 14 Dec 2011
  Last modified date: 14 Dec 2011

  Contents
  Download location
  Known issues

  Installation information
  Prior to installation
  Installing
  Performing the necessary tasks after installation

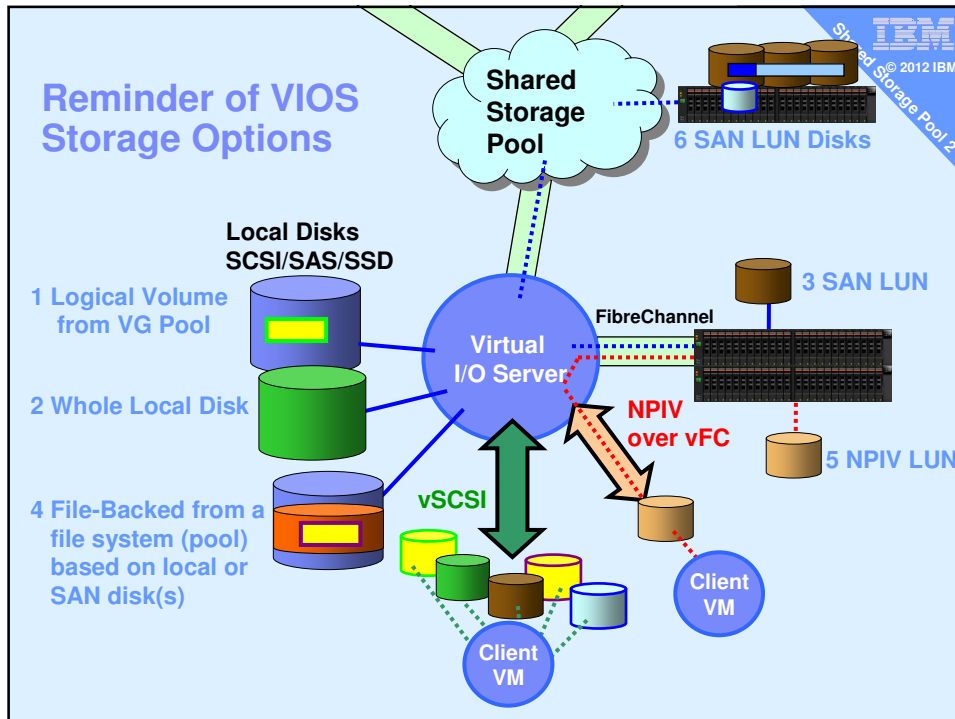 Rate this page:
  Average rating 0 user(s)

  Add comments

  Document information
  Virtual I/O Server

  Reference #:
  00000876

  Modified date:

  Power Systems

  Virtual I/O Server

- **Then read VIOS 2.2 User Guide:**
  http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/topic/p7hb1/p7hb1.pdf

Reminder of VIOS Storage Options

Shared Storage Pool

6 SAN LUN Disks

Local Disks SCSI/SAS/SSD

1 Logical Volume from VG Pool

2 Whole Local Disk

Virtual I/O Server

FibreChannel

3 SAN LUN

NPIV over vFC

5 NPIV LUN

4 File-Backed from a file system (pool) based on local or SAN disk(s)

vSCSI

Client VM

Client VM

© 2012 IBM

Shared Storage Pool 2

Is vSCSI LUN or NPIV dead?

No, absolutely not
Customers continue to use all 6 options

Some people have the idea NPIV is somehow a
   strategic direction with IBM - this is not true
IBM equally supports all 6 disk options.

Are the SAN guys spreading this rumour?
- they like the extra control of NPIV
- don't care if it means 10 times the server setup work

© 2012 IBM

Shared Storage Pool 2

How is it paid for?

Shared Storage Pools is a feature
of PowerVM Standard & Enterprise

How is it installed?

Shared Storage Pool is a VIOS feature
so just upgrade to VIOS 2.2.1.3 = FP25+sp1
December 2011 service pack

Note: This VIOS is AIX 6.1 TL7 based
NIM server needs to be AIX 6.1 TL7 or AIX 7.1 TL1

Why SSP?   Nigel's Opinion here

- Fibre-Channel LUN & NPIV is complex
    1. SAN switch, SAN disk subsystem = hard work & weird GUI !!
    2. Typical LUN lead time: 4 minutes, 4 hours, 4 days, 4 weeks?
    3. With rapidly changing needs with mandatory responsiveness it is simply not good enough!
    4. Many smaller computer rooms have no dedicated SAN guy
    5. LPM hard work as most people don't pre-Zone the target so have to Zone before the move = complexity, slow, error prone
    6. LPM = zero outage for Hardware & Firmware upgrades
- Shared Storage Pool
    1. Allocate LUNs to the Virtual I/O Servers once
    2. One VIOS command to allocate space to a VM
        - Or use: cfgassist (VIOS's smitty)
        - Or use: HMC Virtual Storage Management GUI
    3. LPM any time you like

Shared Storage Pool phase 2 Requirements

© 2012 IBM

Read the Release Notes
http://www-01.ibm.com/support/docview.wss?rs=0&uid=isg400000876

- Platforms: **POWER6** & **POWER7** only (includes **Power Blades**)
- VIOS Storage Pool (minimums):
    - Direct fibre-channel attached LUNs:
    - **1 for repository ~10 GB (NEW INFO: 1GB is enough) &**
    - **1 or more for data, 10 GB → in practice lots more [like 1TB+]**
- Pool Storage Redundancy: Repository & pool storage must be **RAIDed**
- VIOS **name resolution** to resolve hostnames
- Virtual I/O Server(s):
    - **Minimum CPU: 1 (shared, uncapped is good)**
    - **Minimum Memory: 4 GB**
    - Nigel's recommendation: Please, no skinny VIOS
- To further VIOS upgrades
    - Client VM's must be stopped, plus cluster down (use clstartstop)

Shared Storage Pool phase 2 Limits

© 2012 IBM

- Max nodes: **4 VIOS nodes**

- Max physical disks in a pool: 256
- Max virtual disks (LUs) in a cluster: 1024
- Number of Client LPARs per VIOS   1   to 40
  (that is, 40 clients per VIOS, or 40 clients per VIOS pair)
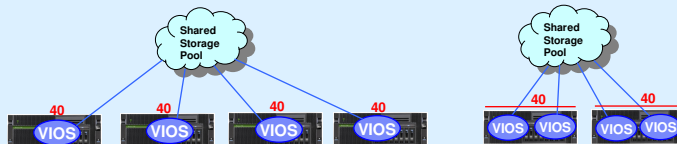
- Capacity of Physical Disks in Pool (each)   5GB to 4TB
- Storage Capacity of Storage Pool (total)   10GB to 128TB
- Capacity of each Virtual Disk (LU) in Pool   1GB to 4TB
- Number of Repository Disks   1   to 1 (CAA limit)

Read the Release Notes & README

Shared Storage Pool phase 2   3 of 3
Limits

Network
- Reliable & only IPv4
- No changes to hostname or IP address
- VIOS can't use VLAN tagging
- DNS should use local /etc/hosts first
- Hostnames must be fully qualified
- Forward & reverse lookup must work
- Recommended to synchronise clocks
- Restoring VIOS from viosbr – get the networks configured 1st

- SEA must use default threaded mode
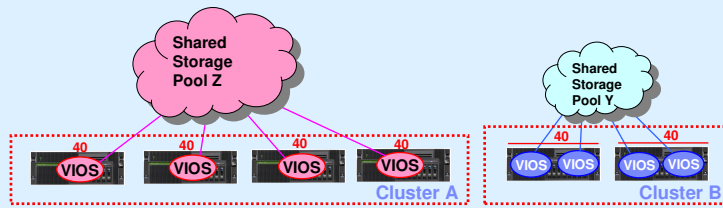- If cluster or pool name >16 characters add APAR IV11852m13

Storage
- No interruption
- Can't resize a LUN
- SSP may take more CPU
- No SCSI reservations (Reserve/Release)
- HA SAN solutions used to mitigate outages
- SANCOM not supported

- Don't use vSCSI adapter "Any client partition can connect"
- AMS or Suspend/Resume can't use SSP for Paging Space

If you used phase 1 then many limits removed

- Now OK to
  – Use Live Partition Mobility
  – SSP VIOS can be a LPM Data Mover
  – Can use VIOS which is a AMS Pager
  – Can do Non-disruptive cluster upgrade
  – Can use 3rd party multi-pathing software support

- Live Partition Mobility across VIOS SSP cluster
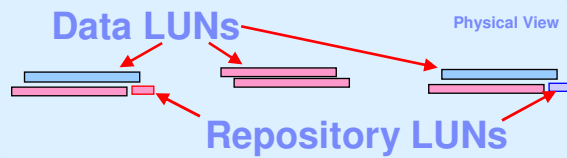 – They all see the disks so available by default
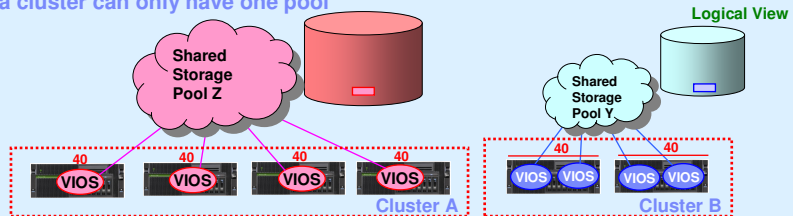
Terms   **Shared Storage Pool phase 2 = SSP2**

© 2012 IBM

Shared Storage Pool 2

Shared Storage Pool Z

Shared Storage Pool Y

40 40 40 40
VIOS VIOS VIOS VIOS
Cluster A

40 40
VIOS VIOS VIOS VIOS
Cluster B

1 SSP2 cluster = set of co-operating Virtual I/O Servers
Currently a VIOS can only be in one cluster.

Here we show two clusters

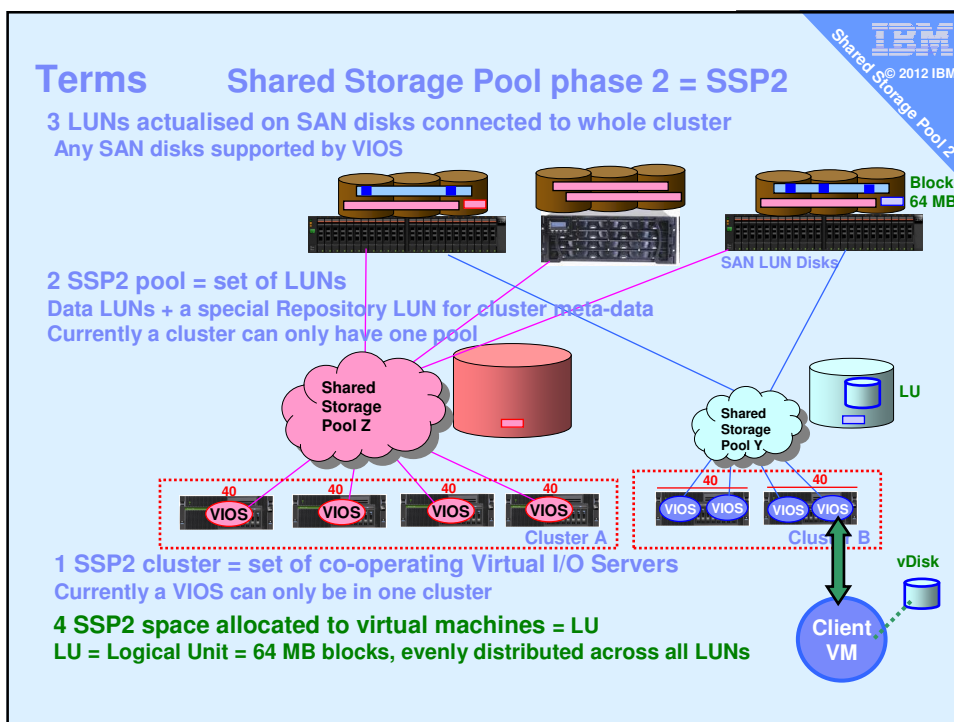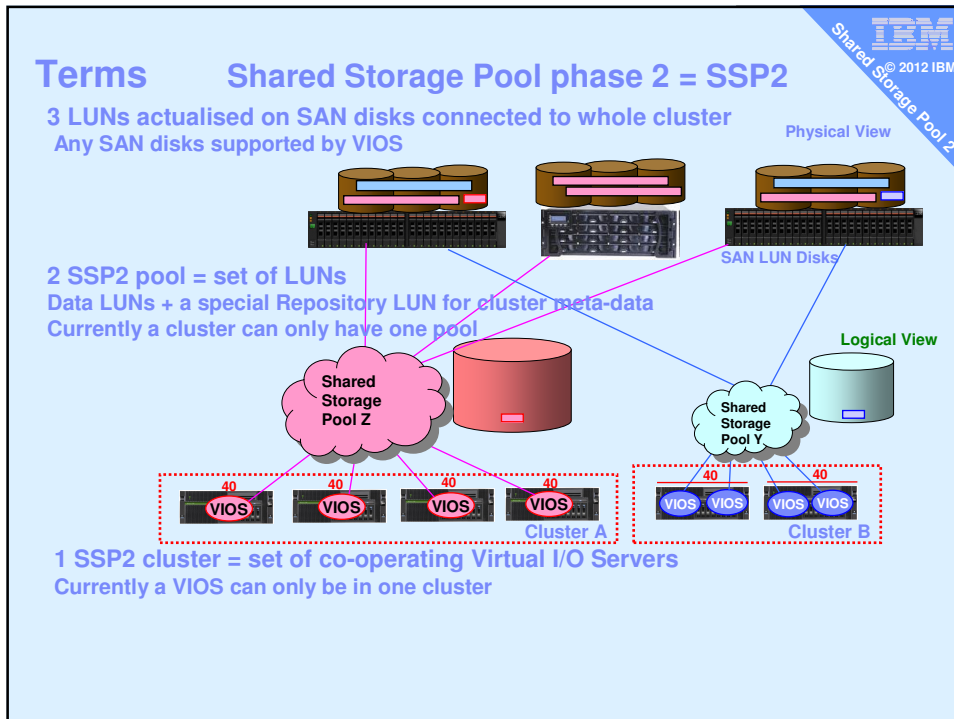**Terms**   **Shared Storage Pool phase 2 = SSP2**

© 2012 IBM

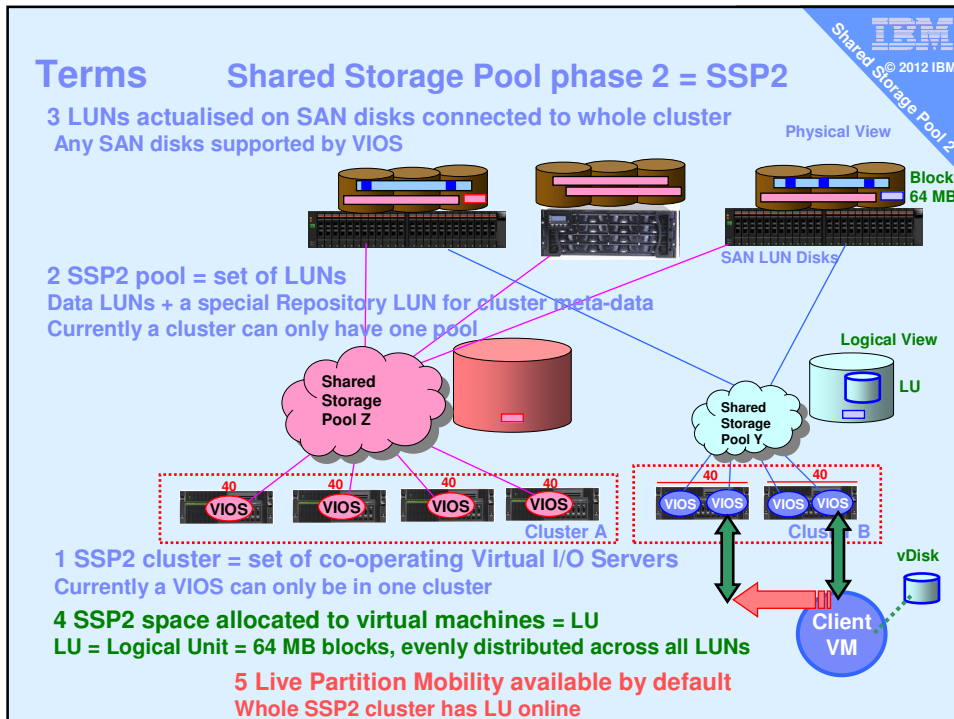Shared Storage Pool 2

Data LUNs   Physical View

Repository LUNs

2 SSP2 pool = set of LUNs
Data LUNs + a special Repository LUN = cluster meta-data
Currently a cluster can only have one pool

Logical View

Shared Storage Pool Z

Shared Storage Pool Y

40 40 40 40
VIOS VIOS VIOS VIOS
Cluster A

40 40
VIOS VIOS VIOS VIOS
Cluster B

1 SSP2 cluster = set of co-operating Virtual I/O Servers
Currently a VIOS can only be in one cluster

Terms   Shared Storage Pool phase 2 = SSP2

© 2012 IBM

Shared Storage Pool 2

3 LUNs actualised on SAN disks connected to whole cluster
Any SAN disks supported by VIOS

Physical View

SAN LUN Disks

2 SSP2 pool = set of LUNs
Data LUNs + a special Repository LUN for cluster meta-data
Currently a cluster can only have one pool

Logical View

Shared Storage Pool Z

Shared Storage Pool Y

40   40   40   40
VIOS VIOS VIOS VIOS
Cluster A

40   40
VIOS VIOS VIOS VIOS
Cluster B

1 SSP2 cluster = set of co-operating Virtual I/O Servers
Currently a VIOS can only be in one cluster

Terms   Shared Storage Pool phase 2 = SSP2

© 2012 IBM

Shared Storage Pool 2

3 LUNs actualised on SAN disks connected to whole cluster
Any SAN disks supported by VIOS

Blocks 64 MB

SAN LUN Disks

2 SSP2 pool = set of LUNs
Data LUNs + a special Repository LUN for cluster meta-data
Currently a cluster can only have one pool

LU

Shared Storage Pool Z

Shared Storage Pool Y

40   40   40   40
VIOS VIOS VIOS VIOS
Cluster A

40   40
VIOS VIOS VIOS VIOS
Cluster B

1 SSP2 cluster = set of co-operating Virtual I/O Servers
Currently a VIOS can only be in one cluster

vDisk

Client VM

4 SSP2 space allocated to virtual machines = LU
LU = Logical Unit = 64 MB blocks, evenly distributed across all LUNs

Terms   Shared Storage Pool phase 2 = SSP2

© 2012 IBM

Shared Storage Pool 2

3 LUNs actualised on SAN disks connected to whole cluster
Any SAN disks supported by VIOS

Physical View

Blocks 64 MB

2 SSP2 pool = set of LUNs
Data LUNs + a special Repository LUN for cluster meta-data
Currently a cluster can only have one pool

SAN LUN Disks

Logical View

LU

Shared Storage Pool Z

Shared Storage Pool Y

40   40   40   40
VIOS   VIOS   VIOS   VIOS
Cluster A

40   40
VIOS VIOS   VIOS VIOS
Cluster B

vDisk

1 SSP2 cluster = set of co-operating Virtual I/O Servers
Currently a VIOS can only be in one cluster

4 SSP2 space allocated to virtual machines = LU
LU = Logical Unit = 64 MB blocks, evenly distributed across all LUNs

Client VM

5 Live Partition Mobility available by default
Whole SSP2 cluster has LU online

Preparation

© 2012 IBM

Shared Storage Pool 2

- All the Cluster VIOSs need the LUNs online
  Make sure they are available = Zoned

- **BEFORE** you start the cluster for all LUNs
  chdev -dev <device name> -attr reserve_policy=no_reserve

- Don't forget this for extra disks that you add later

- Forgetting this = a real mess as you have to stop
  the cluster to make low level disk attributes changes

Cluster Management

Section Contents
- Create Cluster and Pool
- Find out information about the SSP
- Add node
- House Keeping

Cluster full hostnames mandatory + DNS

1. Need full DNS working or /etc/hosts
2. $ hostname
   redvios1.ibm.com
   $ host redvios1.ibm.com – must work
   $ host 9.123.456.99   – must work

Not just "redvios1" here

VIOS
oem_setup_env
smitty tcpip
Example →

```
                    Minimum Configuration & Startup

  To Delete existing configuration data, please use Further Configuration menus

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

[TOP]                                                  [Entry Fields]
* HOSTNAME                                           [redvios1.aixncc.uk.ib>
* Internet ADDRESS (dotted decimal)                  [9.69.44.50]
  Network MASK (dotted decimal)                      [255.255.255.0]
* Network INTERFACE                                   en2
  NAMESERVER
         Internet ADDRESS (dotted decimal)           [9.137.62.2]
         DOMAIN Name                                  [aixncc.uk.ibm.com]
  Default Gateway
       Address (dotted decimal or symbolic name)      [9.69.44.98]
       Cost                                           [0]                    #
       Do Active Dead Gateway Detection?              no                     +
[MORE...2]

F1=Help              F2=Refresh           F3=Cancel          F4=List
F5=Reset             F6=Command           F7=Edit            F8=Image
F9=Shell             F10=Exit             Enter=Do
```

Cluster create on 1st node

Create cluster on one VIOS (here called bluevios1)

```
$ cluster –create –clustername galaxy \
  –repopvs hdisk2 \
  –spname atlantic –sppvs hdisk3 hdisk5 \
  –hostname bluevios1.ibm.com
...
Cluster galaxy has been created successfully.
```

It will take a minute or two, then output Cluster created
You will find a bunch of new daemons running.

If it complains the disks are "in use" check.
If certain they are correct, wipe the disk content with:
   # cleandisk –r hdiskX

   # cleandisk –s hdiskX
It may ask you to confirm y/n ?

Then on that 1st node - add other nodes

On the first VIOS running the cluster

```
$ cluster –addnode –clustername galaxy \
    –hostname redvios1.ibm.com
```

Partition redvios1.aixncc.uk.ibm.com has been added to the galaxy cluster
$

Add other node(s) as necessary.

List cluster & cluster nodes

```
$ cluster –list
Cluster Name     Cluster ID
galaxy           68c06102fc5311e093c8f6027171fc64
$


$ cluster –status –clustername galaxy
Cluster Name        State
galaxy              OK

Node Name         MTM                 Partition Num  State  Pool
                                                            State
diamondvios1      8233-E8B02100271P              2   OK     OK
diamondvios2      8233-E8B02100271P              1   OK     OK
redvios1          8203-E4A0310E0A41              1   OK     OK
$
```

Example of a 3 node cluster

House keeping

You can remove a node from the cluster
- LPM any important client Virtual machines elsewhere
- Stop remaining VMs
- Remove the client VMs
- Remove their allocated virtual disks
- then

```
$ cluster –rmnode –clustername galaxy \
  –hostname redvios1.ibm.com
```

You can also remove the cluster completely
- Once all disk space unassigned & nodes removed

```
$ cluster –delete –clustername galaxy
```

Cluster Aware AIX (CAA) commands

- SSP is built on top of Cluster Aware AIX
- So lscluster command provides more info

  - `lscluster -c`   ← `Configuration`

  - `lscluster -d`   ← `Lists all the hdisks`

  - `lscluster -I`   ← `Network Interfaces`

  - `lscluster -s`   ← `Network Stats`

Cluster Aware AIX (CAA) commands

- Cluster configuration

```
$ lscluster -c
Cluster query for cluster galaxy returns:
Cluster uuid: 68c06102-fc53-11e0-93c8-f6027171fc64
Number of nodes in cluster = 3
      Cluster id for node diamondvios1.aixncc.uk.ibm.com is 1
      Primary IP address for node diamondvios1.aixncc.uk.ibm.com is 9.69.44.221
      Cluster id for node diamondvios2.aixncc.uk.ibm.com is 2
      Primary IP address for node diamondvios2.aixncc.uk.ibm.com is 9.69.44.222
      Cluster id for node redvios1.aixncc.uk.ibm.com is 3
      Primary IP address for node redvios1.aixncc.uk.ibm.com is 9.69.44.50
Number of disks in cluster = 3
       for disk hdisk9 UUID = 5cd2400… cluster_major = 0 cluster_minor =3
       for disk hdisk6 UUID = 6ef71f2d… cluster_major = 0 cluster_minor =2
       for disk hdisk7 UUID = 957a8286… cluster_major = 0 cluster_minor =1
Multicast address for cluster is 228.69.44.221
```

```
$ lscluster -d
Storage Interface Query
Cluster Name: galaxy
Cluster uuid: 68c06102-fc53-11e0-93c8-f6027171fc64
Number of nodes reporting = 3
Number of nodes expected = 3
Node redvios1.aixncc.uk.ibm.com
Node uuid = 85eebf9e-0671-11e1-861c-f60271718d0d
Number of disk discovered = 4
     hdisk9
        state : UP
        uDid  : 3E213600A0B8000294FF8000007DE4E6F18DB0F1814    FAStT03IBMfcp
        uUid  : 5cd24000-5c18-74b5-e873-49841d016e22
        type  : CLUSDISK
     hdisk6
        state : UP
        uDid  : 3E213600A0B800029492E00001A084E6F15DA0F1814    FAStT03IBMfcp
        uUid  : 6ef71f2d-467d-732f-3aee-f6dc865dde53
        type  : CLUSDISK
     hdisk7
        state : UP
        uDid  : 3E213600A0B8000294FF8000007E04E6F192F0F1814    FAStT03IBMfcp
        uUid  : 957a8286-c93d-e46e-84a8-151aed13c5f3
        type  : CLUSDISK
     hdisk8
        state : UP
        uDid  :
        uUid  : 187b5b66-6df2-ed90-e91b-0839aed7cda4
        type  : REPDISK                        ← REPOSTIORY DISK
Node diamondvios1.aixncc.uk.ibm.com
Node uuid = 68aab88e-fc53-11e0-93c8-f6027171fc64
Number of disk discovered = 4
     hdisk5
        state : UP
        uDid  : 3E213600A0B8000294FF8000007DE4E6F18DB0F1814    FAStT03IBMfcp
        uUid  : 5cd24000-5c18-74b5-e873-49841d016e22
        type  : CLUSDISK
     hdisk2
        state : UP
        uDid  : 3E213600A0B800029492E00001A084E6F15DA0F1814    FAStT03IBMfcp
        uUid  : 6ef71f2d-467d-732f-3aee-f6dc865dde53
        type  : CLUSDISK
     hdisk3
        state : UP
        uDid  : 3E213600A0B8000294FF8000007E04E6F192F0F1814    FAStT03IBMfcp
        uUid  : 957a8286-c93d-e46e-84a8-151aed13c5f3
        type  : CLUSDISK
     hdisk4
        state : UP
        uDid  :
        uUid  : 187b5b66-6df2-ed90-e91b-0839aed7cda4
        type  : REPDISK   …
```

```
Hostname redvois1
$ lspv
NAME     PVID             VG               STATUS
hdisk0   000e0a41ff0ec86c None
hdisk1   000e0a41a06ed683 rootvg           active
hdisk2   000e0a41a06ed737 None
hdisk3   000e0a41d4654e89 None
hdisk4   000e0a41d4654f64 None
hdisk5   000e0a41ba665a09 None
hdisk6   00f6027187d44895 None
hdisk7   00f6027187d51e64 None
hdisk8   00f6027187d5f029 caavg_private active
hdisk9   00f6027187d6c664 None

$ lspv -size
NAME     PVID             SIZE(megabytes)
hdisk0   000e0a41ff0ec86c 140013
hdisk1   000e0a41a06ed683 140013
hdisk2   000e0a41a06ed737 140013
hdisk3   000e0a41d4654e89 140013
hdisk4   00e0a41d4654f64  140013
hdisk5   000e0a41ba665a09 140013
hdisk6   00f6027187d44895 16384
hdisk7   00f6027187d51e64 16384
hdisk8   00f6027187d5f029 15158
hdisk9   00f6027187d6c664 20480
```

Pool Disk Space Management

Content

- Allocate pool disk space and give to a VM
  - Ditto as two commands
- Removing the disk space

Allocate disk space & assign to client VM

```
$ mkbdsp –clustername galaxy \
  –sp atlantic 16G –bd vdisk_diamond6a \
  –vadapter vhost2

Logical Unit vdisk_diamond6a has been created with udid:
 615af85de5acad39a8827e9cd01d6b36.
Assigning file "vdisk_diamond6a" as a backing device.
Vtscsi3 Available.
$
```

Notes:
- 16 GB is not actually allocated until written too
- vdisk_diamond6a is just a name = reminder of the VM using it
- vhost2 is the virtual SCSI adapter for client VM diamond6

Same but two steps

```
1) Create Logical Unit            Note: no –vadapter option
$ mkbdsp -clustername galaxy -sp atlantic 10G -bd LU42
Lu Name:LU42
Lu Udid:374a609cb072e4015d558ff290b9f0bd
```

```
List the pool contents
$ lssp -clustername galaxy -sp atlantic -bd
Lu Name       Size(mb) ProvisionType   Lu Udid
LU42          10240     THIN           374a609cb072e4015d558ff290b9f0bd
…
```

```
2) Example of two ways using "-bd LU42" or  "–luudid <hexidecimal>"
```
- -bd only works if LU42 is unique
- Note: below **no Size argument** (or it creates another LU with same name!)

```
$ mkbdsp -clustername galaxy -sp atlantic -bd  LU42 -vadapter vhost2
Assigning file "vdisk_diamond6a" as a backing device.
VTD:vtscsi1
- or - -
$ mkbdsp -clustername galaxy -sp atlantic \
   –luudid  374a609cb072e4015d558ff290b9f0bd -vadapter vhost2
Assigning file "374a609cb072e4015d558ff290b9f0bd" as a backing device.
VTD:vtscsi1
```

Dual path via Two VIOSs

1 Setup virtual SCSI adapter pairs as normal
- **client VM virtual SCSI adapter A ←→ VIOS C**
- **client VM virtual SCSI adapter B ←→ VIOS D**

2 on VIOS C: use "lsmap –all" to map slot to vhostN
- **mkbdsp -clustername galaxy -sp atlantic 16G**
      -bd vdisk_red6a -vadapter vhostN

3 on VIOS D: use lsmap –all to map slot to vhostM
- **mkbdsp -clustername galaxy -sp atlantic**
        -bd vdisk_red6a -vadapter vhostM

Note: No size (16G) 2nd time

4 On the client VM
```
$ lspath
Enabled hdisk0 vscsi0
Enabled hdisk0 vscsi1
```

5 LPM still available – dual VIOS to dual VIOS

Logical View

LU

Shared Storage Pool Y

40   40

VIOS VIOS   VIOS VIOS
          C   D

A   B

Client VM
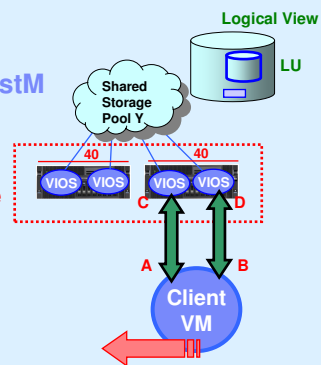
Removing an LU (Logical Unit)

Assuming it is NOT used !!

On the VIOS remove disk space
rmbdsp = remove backing device from storage pool

```
$ rmbdsp –clustername galaxy \
  –sp atlantic –bd vdisk_diamond6a
```

or via the LU hexadecimal name
```
$ rmbdsp –clustername galaxy –sp atlantic
 –luudid 858152297879adfe0d75b05f586d36ee
```

House keeping

Add more physical LUNs to the Pool

```
$ chsp –add –clustername galaxy \
  -sp atlantic hdisk8
```

Remove a physical LUN from the Pool
- You can not …. with this release.
- We can replace a disk but not remove one
  - Replacement disk - equal or larger size

Thin provisioning
= Allocating disk blocks only when they are used i.e. written

- Conceptual model
- Monitoring the pool space

Thin Provisioning Blocks can be missing

Block 0 1 2 3 4 5 6 7 8 9
- - - - - -

Unused blocks

List storage pool "lssp" output shows block size = 64MB

Thin Provisioning Added on 1st Write

Block 0 1 2 3 4 5 6 7 8 9
- - - - -

1 Client VM writes to block 5
2 VIOS spots it is not there & allocates the block from free list
3 then VIOS completes the write
4 so the client is unaware of this

Unused blocks
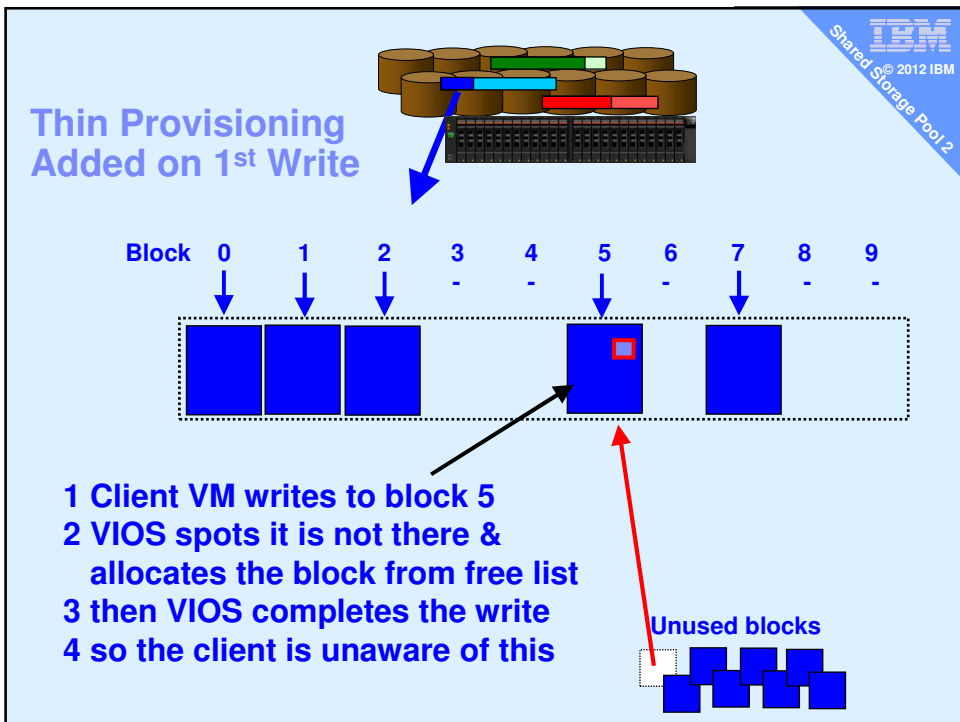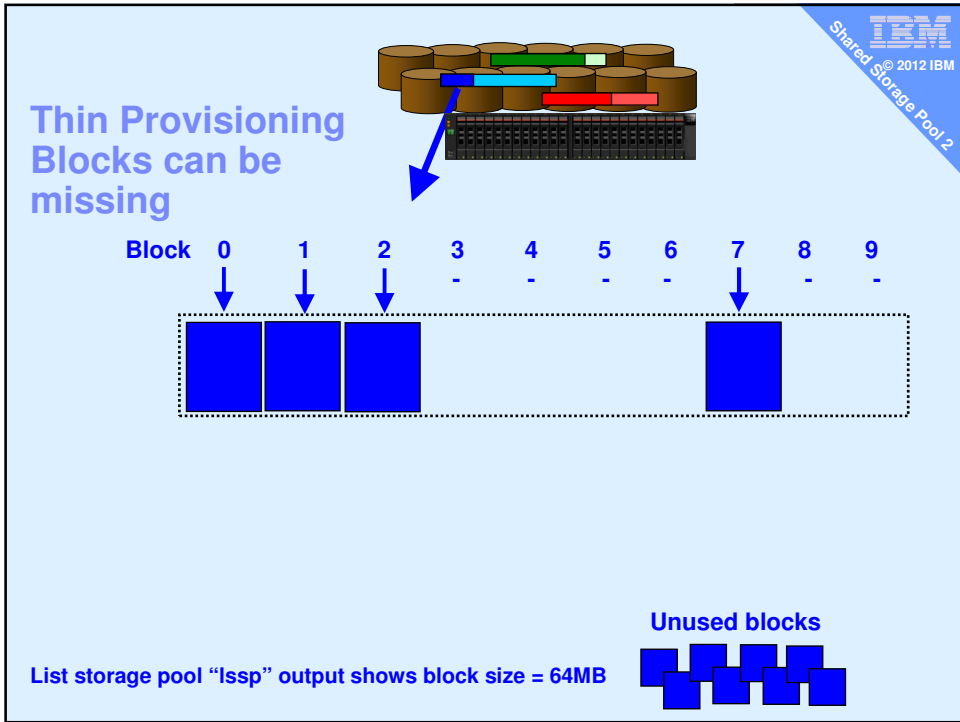
© 2012 IBM

Shared Storage Pool 2

Thin Provisioning

Size 16 GB is actually the max.

Only 3 GB Reduction of free space

FibreChannel

Virtual I/O Server

vSCSI

lspv hdisk0 Disk 16GB

lsvg rootvg Free = 11GB Used = 5GB

Client VM

- mkbdsp states the "LU" size

- Blocks assigned only when written
- After installing AIX 7 (could be any supported OS)
- AIX sees 16 GB disk
- AIX has allocated 5 GB in rootvg
- But not actually written to all 5 GB
  - Paging space not used
  - Free space in filesystems not used
  - Sparse files have "holes"
- Brand new pool & AIX 7
  only 3 GB used from the pool

- Instead of unused disk space in every VM, now it is SSP "pooled"

Complete guesswork: 20,000 machines * 20 VMs* 16 GB unused = 6 PetaBytes

---

Thick Provisioning

- Doh! A no-brainer!
- Like Thin but actually allocate all the disk space
- New option: `mkbdsp … –thick`

The point is
- Avoids problems, if the free list empties
- Good for more important work/production
  or you prefer not to dynamically add blocks

Monitoring: topas on VIOS then "D"

```
Topas Monitor for host:     diamondvios1Interval:    2     Fri Jan 14 14:46:00 2011
====================================================================================
Disk       Busy%  KBPS    TPS    KB-R   ART   MRT   KB-W    AWT   MWT    AQW    AQD
cldisk2    41.0   17.6K   493.0  0.0   0.0   174.6  17.6K   1.1   14.6   0.0    0.0
cldisk3    34.0   20.0K   160.0  0.0   0.0   186.4  20.0K   2.9   13.1   0.0    0.0
cldisk1     3.0   24.0    6.0    0.0   0.0   112.0  24.0    0.6   158.8  0.0    0.0
hdisk0      0.0   8.0     2.0    0.0   0.0   10.2   8.0     4.1   64.2   0.0    0.0
caa_priva   0.0   17.0    5.0    9.0   0.1   2.1    8.0     0.5   6.9    0.0    0.0
hdisk1      0.0   0.0     0.0    0.0   0.0   0.0    0.0     0.0   7.2    0.0    0.0
cd0         0.0   0.0     0.0    0.0   0.0   0.0    0.0     0.0   0.0    0.0    0.0
```

One client VM running: yes >/tmp/x

Disk I/O spread across disks
 Allocation unit is 64MB (see lssp output)

Monitoring Disk use with lssp

```
$ lssp –clustername galaxy –sp atlantic –bd
Lu(Disk) Name       Size(MB) ProvisionType Lu Udid
vdisk_diamond6a     16384    THIN          615af . . .
vdisk_diamond8a     16384    THIN          917c0 . . .
vdisk_diamond5a      8192    THICK         f1442 . . .
vdisk_diamond5b      8192    THICK         ebecd . . .
vdisk_diamond3a     10240    THIN          afcec . . .
$ lssp –clustername galaxy
POOL_NAME:        atlantic
POOL_SIZE:          47552
FREE_SPACE:         17945
TOTAL_LU_SIZE:      59392
TOTAL_LUS:        5
POOL_TYPE:        CLPOOL
POOL_ID:          0000000009893EDD000000004F174D22
```

47522 Pool Physical Size
17945 Pool Physical Free
29607 Pool Physical Used
Pool use 29607/47522x100=62%

59392 Allocated
Pool Over commit 59392/47522= 1.25
allocated 25% more than I have!
= Thin provisioning

Thin provisioning risks pool free space = zero Ek!
Next write needing a new SSP block, gets a disk error!
Just don't go there – you need to be warned!

Thin Provisioning Alerts

- Set alerts to warn on <u>free pool space</u> getting too low
  alert -set -clustername *galaxy* **-spname** *atlantic* **-value 10**

- To list the alert threshold:
  - **alert -list -clustername** *galaxy* **-spname** *atlantic*
    ```
    $ alert –list –clustername galaxy –spname atlantic
    PoolName:        atlantic
    ```
  - PoolID: 0000000009893EDD000000004F174D22
  - ThresholdPercent: 35

- To remove the alert:
  - **alert -unset -clustername** *galaxy* **-spname** *atlantic*
  - Threshold is set to 0 (zero) – it will not happen!!
- The default alert is <u>free pool space</u> below 35%

House keeping – Alert Reporting

- Reported on **any one of the VIOS cluster**
- padmin user: errlog
  - Like AIX errpt
  - $ errlog | more
    ```
    IDENTIFIER TIMESTAMP  T C RESOURCE_NAME   DESCRIPTION
    0FD4CF1A   0215112612 I O VIOD_POOL       Informational Message
    ...
    ```
    ②
  - $ errlog –ls | more
    See example on the next page

- Can also be reported to high levels SM like Systems Director etc.

!
CAUTION
BE ALERT

```
$ errlog –ls ...
LABEL:      VIO_ALERT_EVENT  ①
IDENTIFIER: 0FD4CF1A

Date/Time:      Wed Feb 15 11:26:32 CST 2012
Sequence Number: 86
Machine Id:     00F602714C00
Node Id:        diamondvios2
Class:          O
Type:           INFO
WPAR:           Global
Resource Name:  VIOD_POOL  ②

Description
Informational Message  ③

Probable Causes
Asynchronous Event Occurred

Failure Causes
PROCESSOR

     Recommended Actions
     Check Detail Data

Detail Data
Alert Event Message
25b8001                      ④
A Storage Pool Threshold alert event occurred on pool D_E_F_A_U_L_T_061310 pool id 92d2fd5f2ec45382 in cluster
   galaxy cluster id 00841e2a422711e194cbf60271715fc2 The alert event received is:   Threshold Exceeded.
                                                                                          ⑤
Diagnostic Analysis
Diagnostic Log sequence number: 250
Resource tested:    sysplanar0
Menu Number:        25B8001
Description:
A Storage Pool Threshold alert event occurred on pool D_E_F_A_U_L_T_061310 pool id 92d2fd5f2ec45382 in cluster
   galaxy cluster id 00841e2a422711e194cbf60271715fc2 The alert event received is:    Threshold Exceeded.
```

CAUTION
BE ALERT

House keeping - Thin Provisioning Alerts

- It is vital that you get these warning messages

- Suggest on ALL VIOS
  1. Email the Pool stats every night to the admin guys (cron as root)

     ```
     . /home/padmin/.profile
     lssp –clustername galaxy | /usr/bin/mailx –s "SSP stats" ps@acme.com
     ```

  2. Script to check and if free space is low
     then email or send phone TEXT message or escalate

- Possible reactions are:
  - Add a new LUN to the pool,
  - Delete allocated space = unused LU or entire VM & space
  - Drop a Snapshot

House keeping - Thin Provisioning Alerts

Snapshot

Snapshots on VM disks and Cloning

Snapshot available using
– Advanced SAN disks or SAN Volume Controller (SVC)

but now VIOS admin can do this too!

Currently: no way to save the snapshot off-line

Snapshot + Drop
– Very quick
– Allows point in time backup
– Later delete the original to reclaim the space

Examples:
▪ Backup VM stopped, quiesce, live

Snapshot + Roll-back
– Very quick
– Useful for lots of reasons →
– Stop the client VM
– Restart on original copy
– Discard newer copy

Examples:
▪ Practice OS or App update
▪ Training & reset
▪ Benchmark & reset
▪ Failure & avoid recovery from tape
▪ Save points for batch runs

Supports single disk or a consistent set of disks

Snapshot – create, list, delete or rollback

Snapshot Usage:

snapshot -create <filename>   -clustername galaxy -spname atlantic -lu LUs

snapshot -delete <filename>   -clustername galaxy -spname atlantic -lu LUs

snapshot -rollback <filename> -clustername galaxy -spname atlantic -lu LUs

snapshot -list   -clustername galaxy -spname atlantic

Notes:
- Alternatively swap "-lu LU_name(s)" for "-luudid Hexadecimal"
- LUs means a space separated list disk names

Snapshot – create and list

Create
```
$ snapshot –create diamond5s.snap  –clustername galaxy
  -spname atlantic –lu vdisk_diamond5a
```

List
```
$ snapshot –list -clustername galaxy -spname atlantic
Lu Name          Size(mb)      ProvisionType    Lu Udid
vdisk_diamond5a  16384         THIN             b3f3a . . .
Snapshot
diamond5s.snap
```

Also snap shots appear in the lssp output
```
$ lssp –clustername galaxy -sp atlantic –bd
Lu Name          Size(mb)      ProvisionType    Lu Udid
vdisk_diamond5a  16384         THIN             b3f3a . . .
Snapshot
diamond5s.snap

vdisk_diamond6a  16384         THIN             4c9e9 . . .
```

Watch those options!

Two different options for the storage pool name

Example:

        snapshot … -spname <name>

        lssp …   -sp   <name>

It can easily catch you out

Snapshot – delete or rollback

When sure you never want to rollback
Delete original & continue on the current blocks
$ **snapshot –clustername galaxy –delete** diamond5t.snap
  -spname atlantic **–lu** vdisk_diamond5a

Rollback to a snapshot
Stop the virtual machine/LPAR then
$ **snapshot –clustername galaxy –rollback** diamond5t.snap
  -spname atlantic **–lu** vdisk_diamond5a

You loose any updated you made since the last snapshot

Snapshot Model

Original Set

Block	0	1	2	3	4	5	6	7	8	9
				-	-		-		-	-

Block	0	1	2	3	4	5	6	7	8	9
				-	-		-		-	-
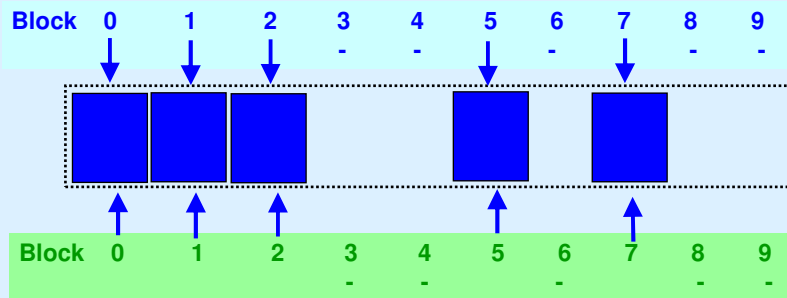
New Snapshot

**Creating a snap shot only involved copying the meta data
i.e. list of the blocks within the LU (not the block themselves)**

Snapshot + Update

Original Set

Block	0	1	2	3	4	5	6	7	8	9
				-	-		-		-	-

Block	0	1	2	3	4	5	6	7	8	9
				-	-		-		-	-
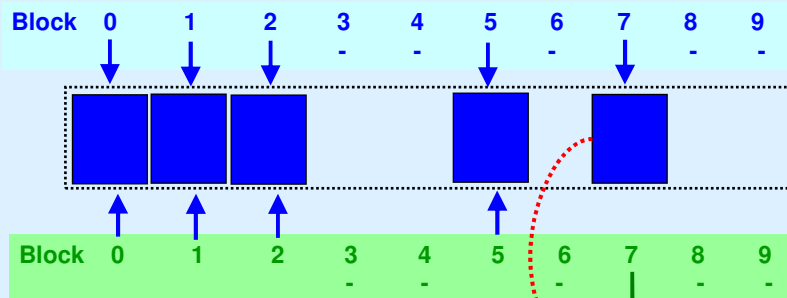
New Snapshot

**1 Client VM update to block 7
2 VIOS allocates a new block
3 Copies original 64MB
4 VIOS completes the write**

Snapshot delete or Roll-back

Original Set

Block	0	1	2	3	4	5	6	7	8	9
				-	-		-		-	-

Block

	0	1	2	3	4	5	6	7	8	9
				-	-		-		-	-

New Snapshot

Delete = remove original just release blue block 7 to free list

Roll-back = remove new set just release green block 7 to free list

---
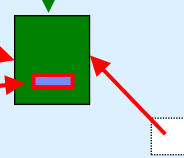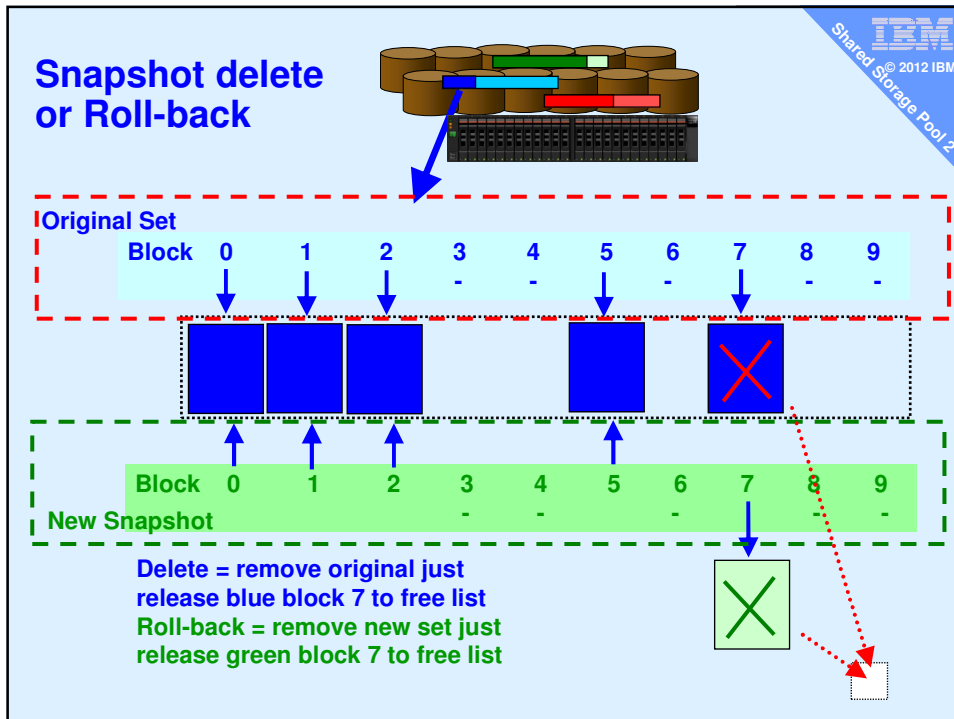
Storage Management

- Reminder currently,
  - One pool of large LUNs - syntax suggests multiple pools later
  - Pool can be on a mix of brands or generations of disk sub-systems
  - 64 MB chunks are spread as evenly as possible across LUNs

Live Storage Mobility

Server Admin would like to :

A. Replace a faulty LUN

B. Move all blocks off one disk subsystem (retiring a disk subsystem)

C. Recover from repository failure

D. Select which disk subsystems a particular VM uses

E. Ensure mirrors are on different subsystems (even different sites)

- A and B → via replace physical disk
  - chsp -replace -clustername galaxy -sp atlantic -oldpv hdisk4 -newpv hdisk24

- C   → see viosbr command (later)

- D and E → in a later SSP release
  - Multiple pools is an obvious solution here (not in the current release)
  - Could use SVC now for lower level mirror (E)

Replace Disk

New LUN

FibreChannel

Network

Virtual I/O Server

Virtual I/O Server

Virtual I/O Server

vSCSI

vSCSI

vSCSI

Client VM

Client VM

Client VM

Live Storage Mobility phase 1 replace faulty disk or move all from old sub-system

---
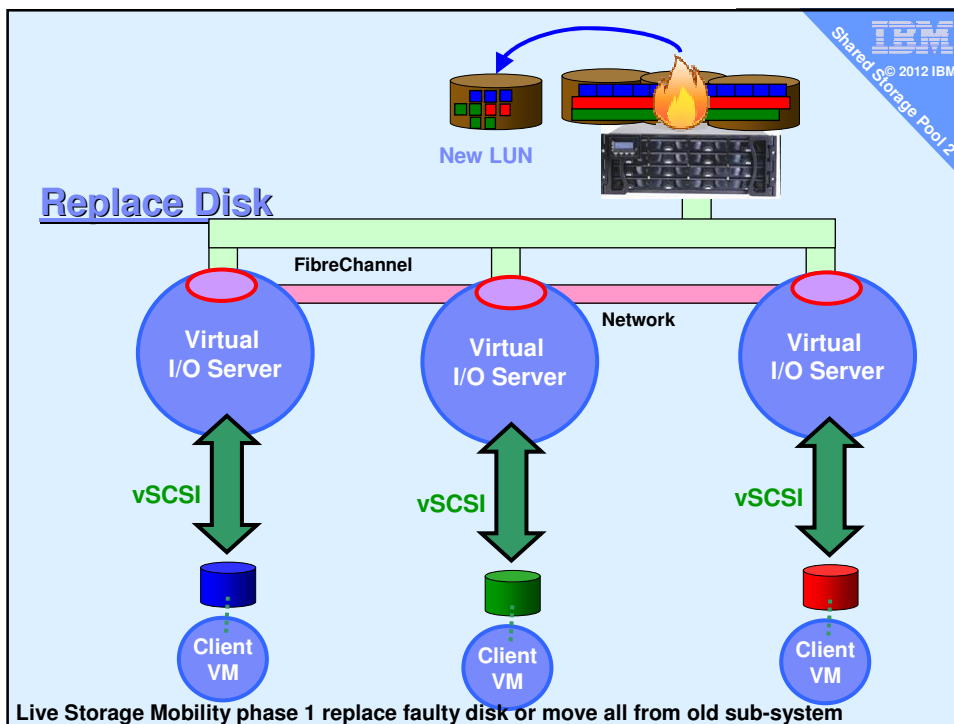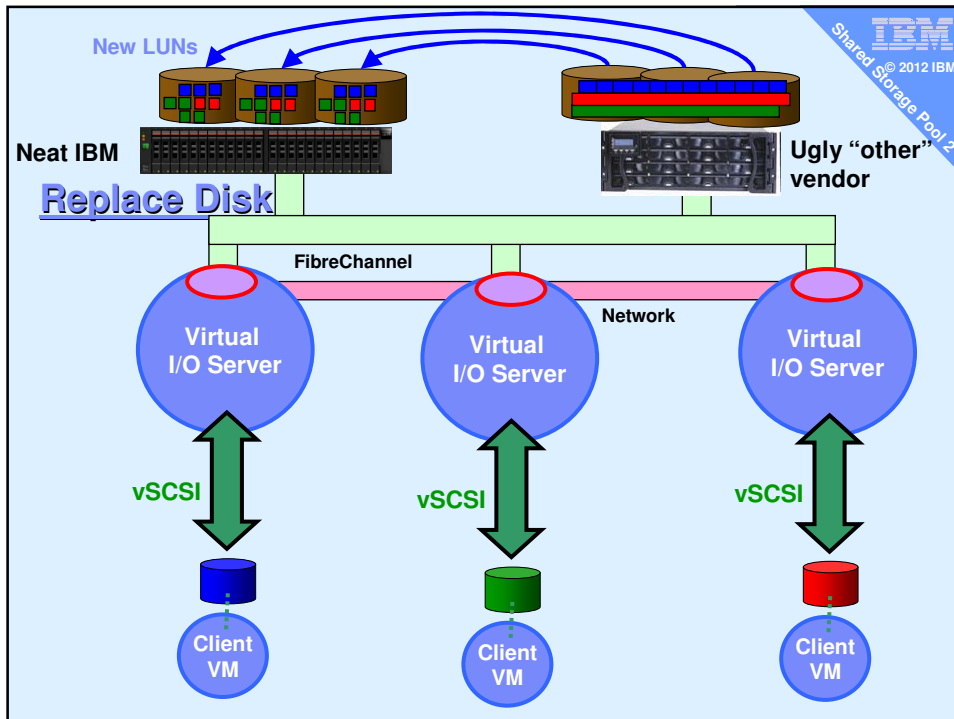
User Interface

- Command line
  - Already shown in this presentation by example
  - Some feature will remain command line only → like: cluster -create
- cfgassist
  - This is the VIOS version of smitty
  - Menu driven interface for CLI
  - Fully covers SSP functions
- Graphical User Interface
  - HMC – now (note the don't have any VIOS vSCSI slots in "Any mode"
  - SDMC - planned for next major release in 2012

- System Director - Future release
  - Already has Storage Pools concept and features
  - SSP is just another storage pool type
  - Then may adds new unique items – like
    VMControl appliance deploy to SSP disk space or cloning

Hardware Management Console

- HMC from October 2011
  - for SSP support & LPM of SSP LPARs
  - Addition feature to Virtual Storage Management
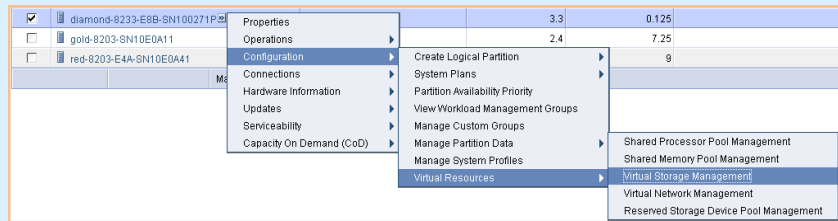
- Shipped with HMC upgrade V7 R7.4 SP0+
  - Not part of the VIOS package

HMC

SSP Pool

HMC

Virtual Storage Management - diamond-8233-E8B-SN100271P

Shared Storage Pool phase 2 – Call to Action

As a result of this presentation: I want you to
Do
1. Start negotiating with SAN team to hand-over a few TB
2. Get to VIOS 2.2.1.3 on all POWER6/7 … ASAP

Feel
– Excited with easy SAN disk management & LPM

Think
– About how this technology could save you time,
boost efficiency & increase responsiveness to users