VIOS Shared Storage Pools

Phase 2 – December 2011

Nigel Griffiths
IBM Power Systems
Advanced Technology Support, Europe (version 11)

Announcement 14th Oct 2011 covering VIO Shared Storage Pool phase 2

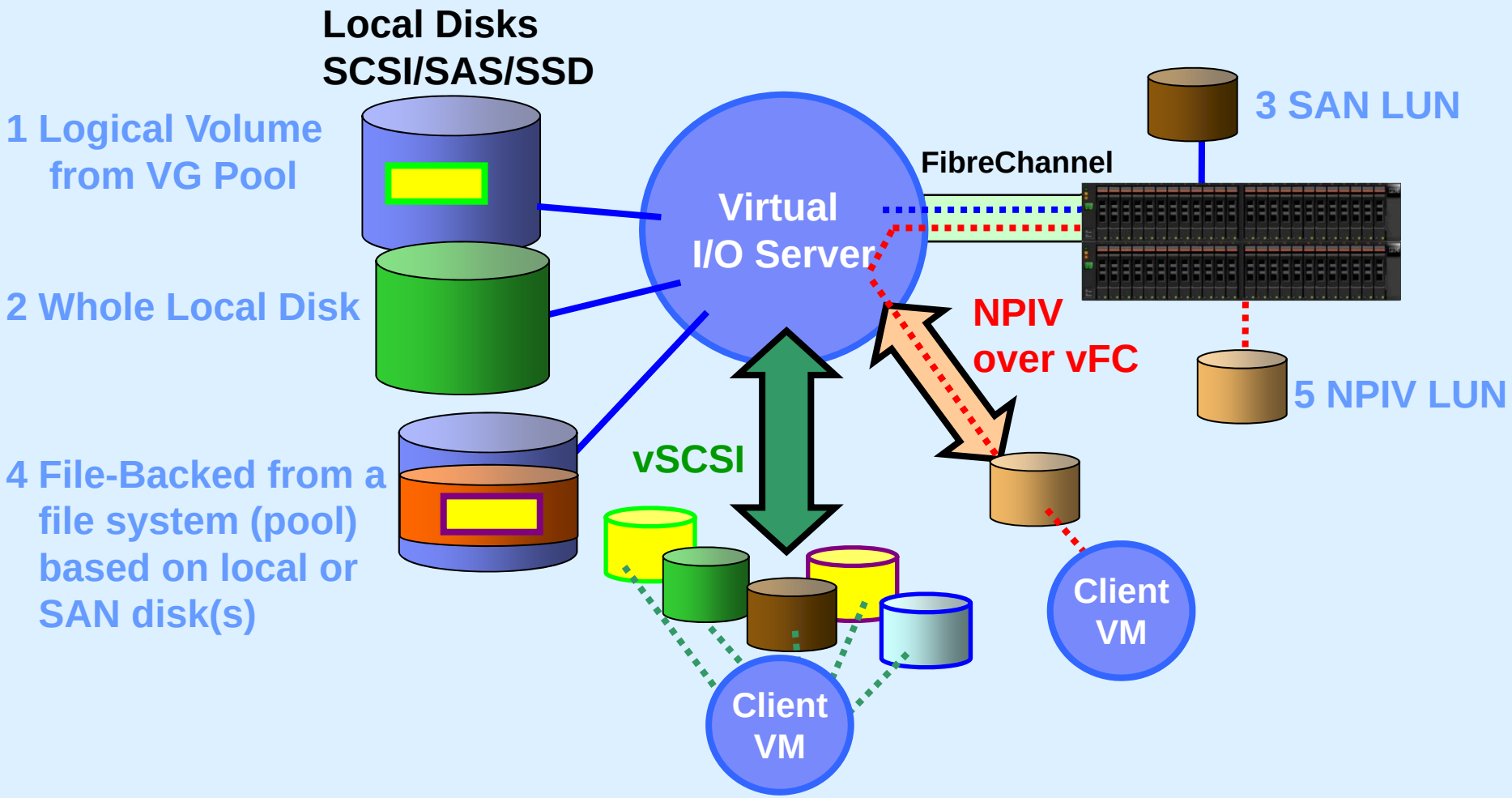
http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=AN&subtype=CA&htmlfid=897/ENUS211-354&appname=USN

GA: December 16, 2011: PowerVM 2.2 Service Pack

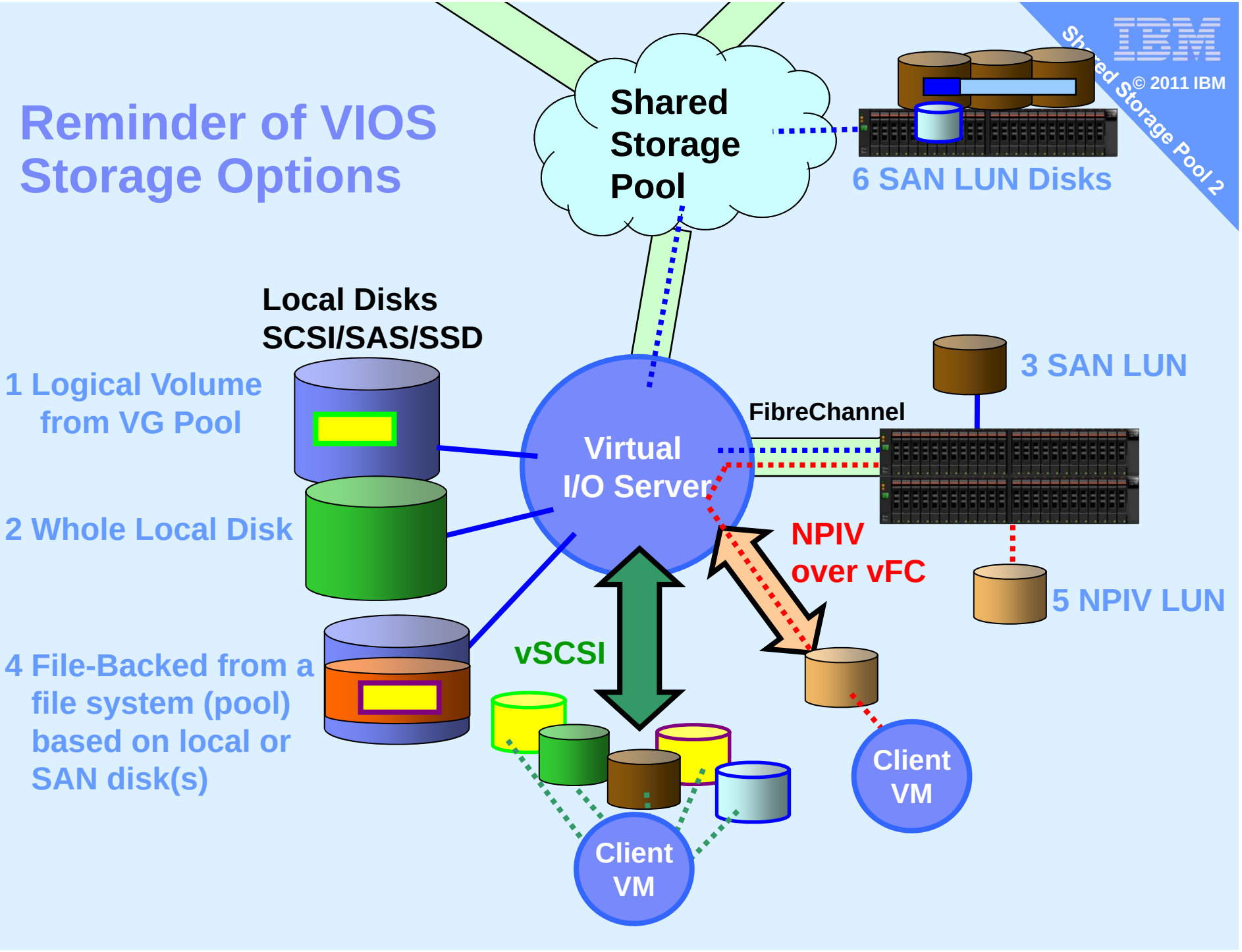
Please check with the Release notes delivered with the product for fine detail.
These slides were prepared during the beta tests.

All statements regarding IBM's future direction & intent are subject to change or withdrawal without notice, & represent goals & objectives only.

Reminder of VIOS Storage Options



Reminder of VIOS Storage Options



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.

Are the SAN guys spreading this rumour!

- they like the extra control
- don’t care if it means 10 times the server setup work

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 your VIOS 2.2.1 to the
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

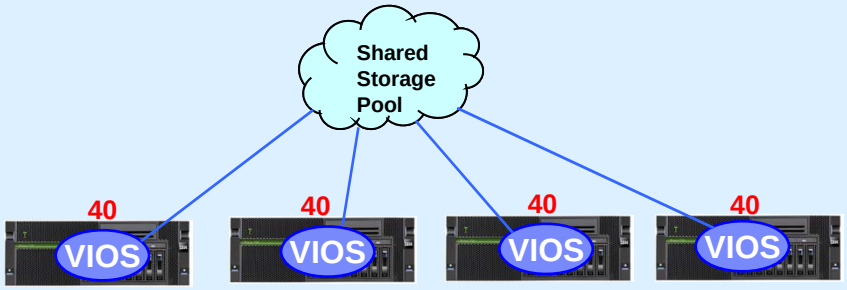
Shared Storage Pool phase 2 Requirements

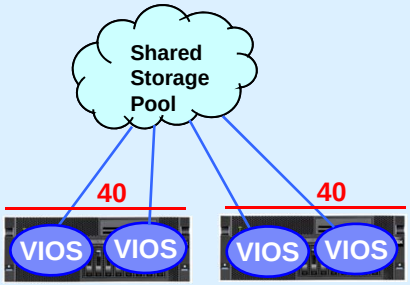
- Platforms: **POWER6** & **POWER7** only (includes **Power Blades**)
- VIOS Storage Pool (minimums):
  - Direct fibre-channel attached LUNs:
  - **1 for repository ~1 GB &**
  - **1 or more for data, 1 GB → in practice lots more [like 1TB+]**
- Pool Storage Redundancy: Repository & pool storage must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

Shared Storage Pool phase 2 Limits

- Max nodes: **4 VIOS nodes**

- Max physical disks in a pool: 256
- Max virtual disks 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) 20GB 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

If you used phase 1 then many limits removed

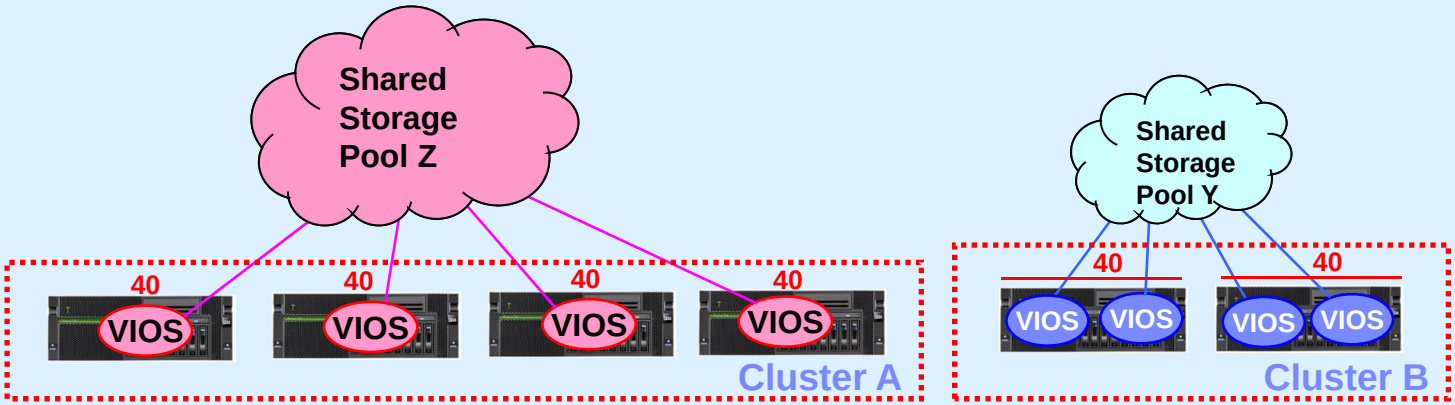
- Now OK to
  - Use Live Partition Mobility
  - SSP VIOS can be a LPM Data Mover
  - Can use VIOS which is a AMS Pager
  - Can do Non-disruptive cluster upgrade
  - Can use 3rd party multi-pathing software support

- Live Partition Mobility across VIOS SSP cluster
  - They all see the disks so available by default

Note: BANNED AMS paging space on a SSP disk!

Terms

Shared Storage Pool phase 2 = SSP2



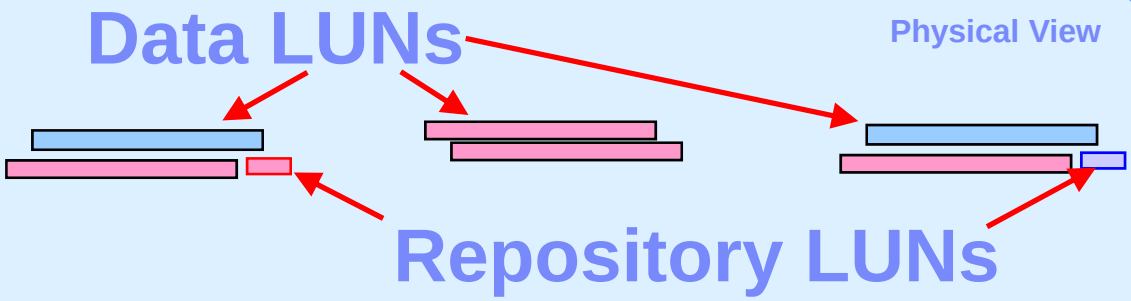
1 SSP2 cluster = set of co-operating Virtual I/O Servers

Currently a VIOS can only be in one cluster.

Here we show two clusters

Terms

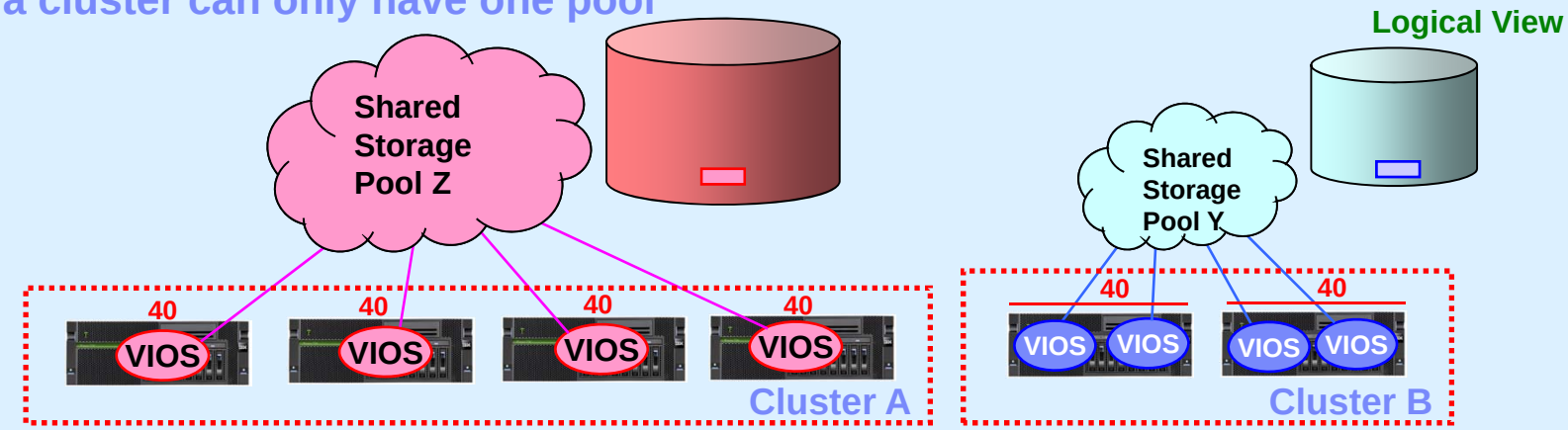
Shared Storage Pool phase 2 = SSP2



2 SSP2 pool = set of LUNs

Data LUNs + a special Repository LUN = cluster meta-data

Currently a cluster can only have one pool



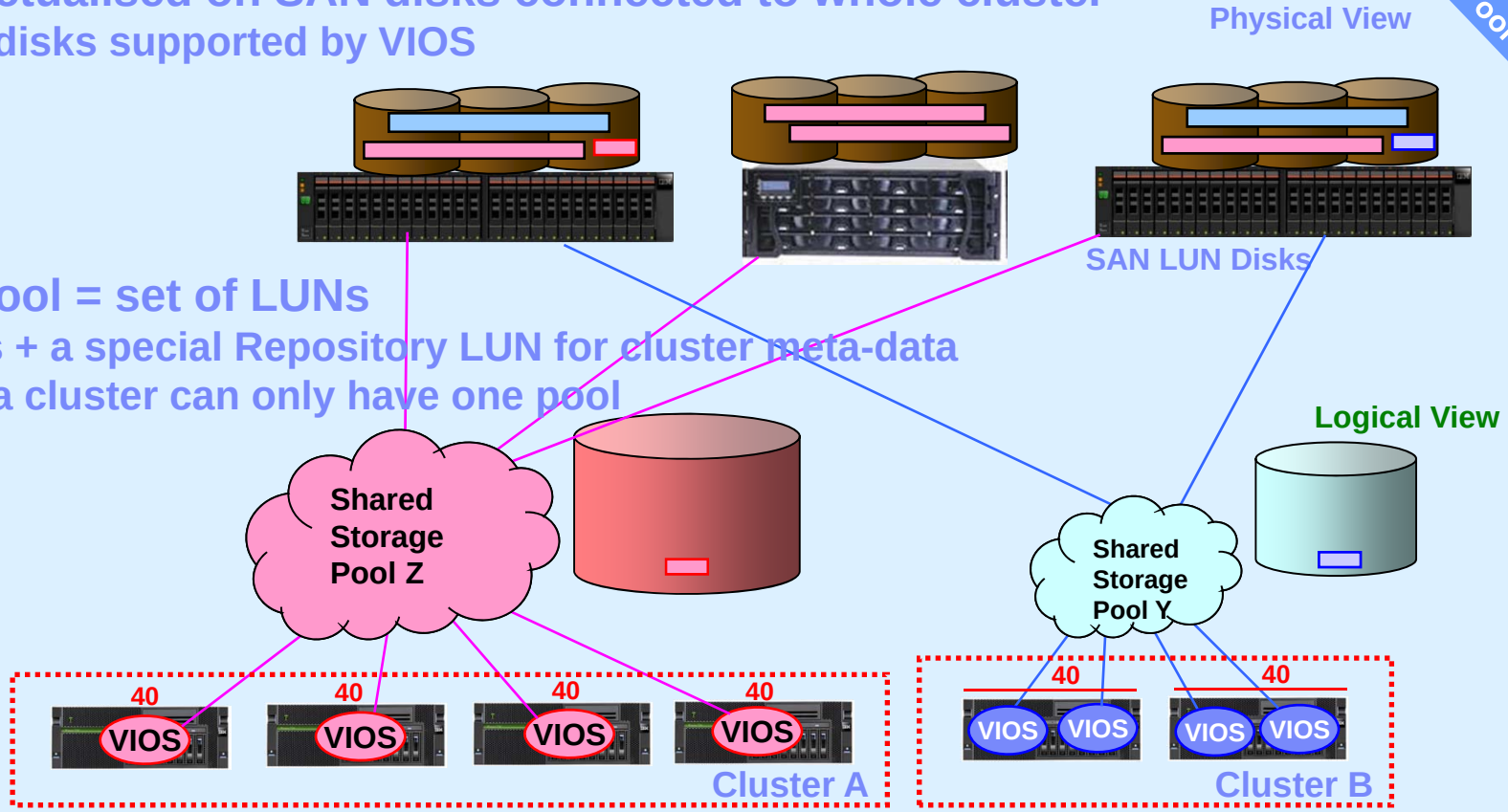
1 SSP2 cluster = set of co-operating Virtual I/O Servers

Currently a VIOS can only be in one cluster

Terms Shared Storage Pool phase 2 = SSP2

3 LUNs actualised on SAN disks connected to whole cluster

Any SAN disks supported by VIOS



2 SSP2 pool = set of LUNs

Data LUNs + a special Repository LUN for cluster meta-data

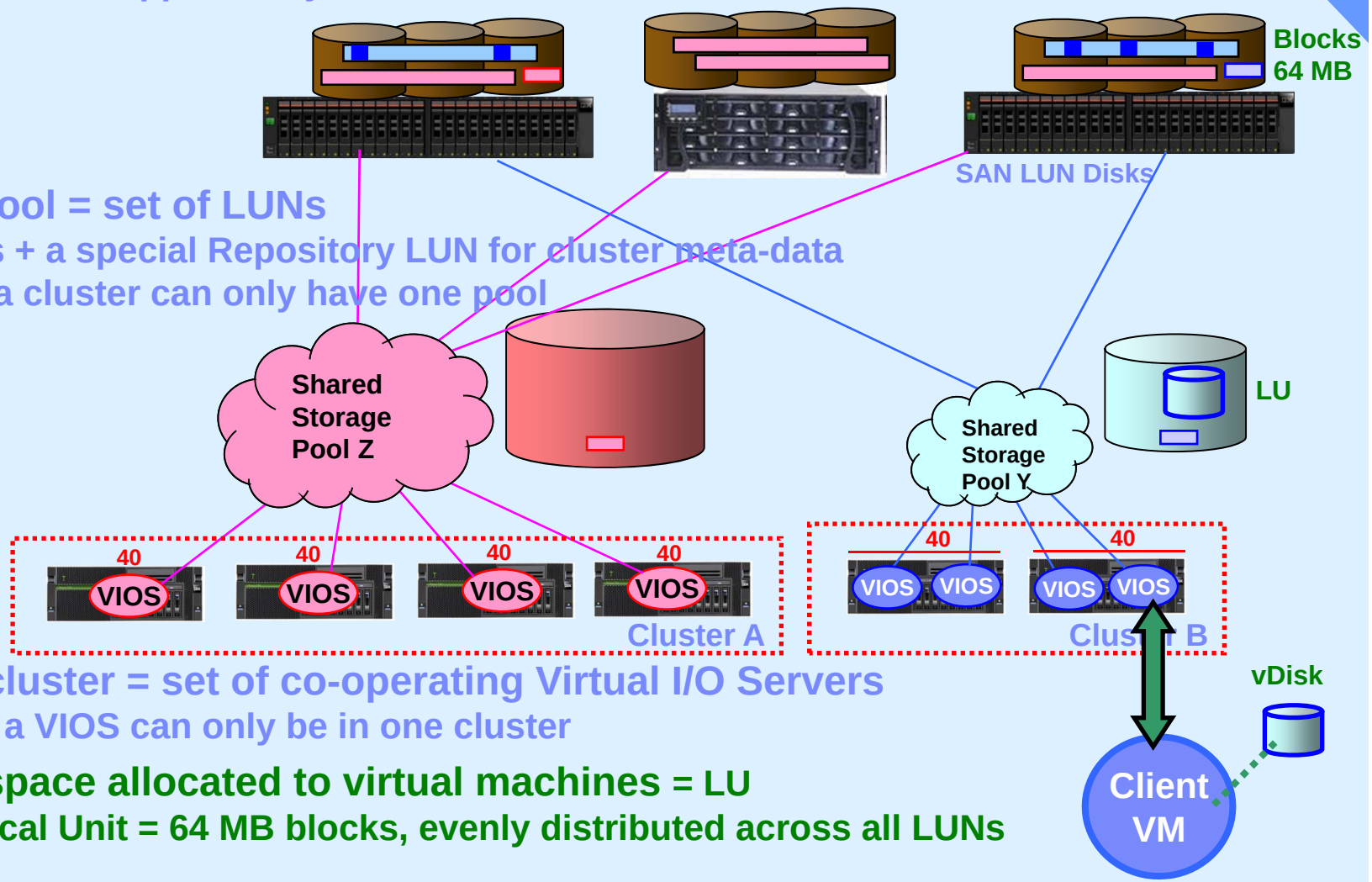
Currently a cluster can only have one pool

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

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



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

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

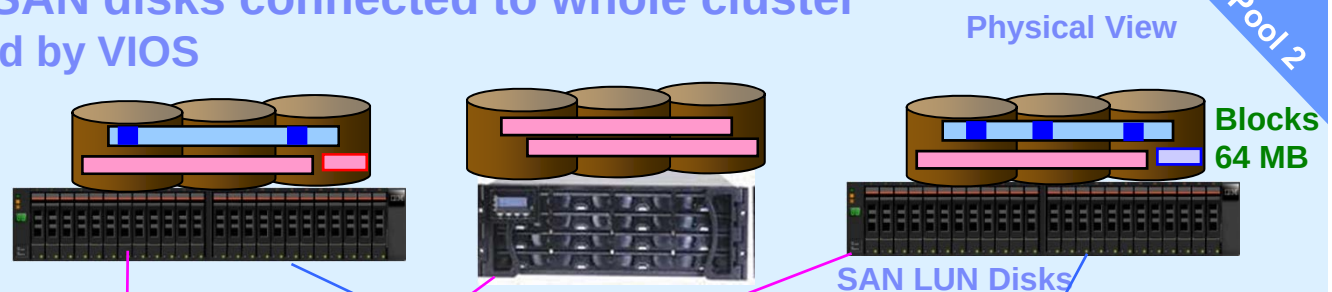
4 SSP2 space allocated to virtual machines = LU
LU = Logical Unit = 64 MB blocks, evenly distributed across all LUNs

Shared Storage Pool 2

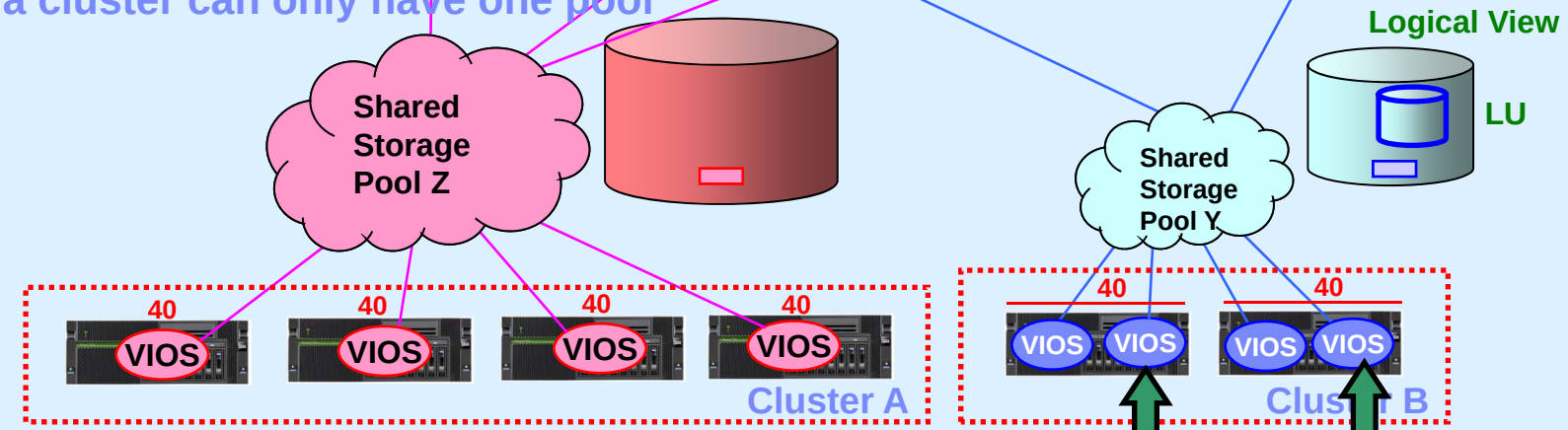
Terms

Shared Storage Pool phase 2 = SSP2

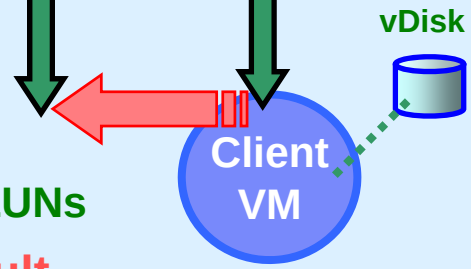
3 LUNs actualised on SAN disks connected to whole cluster
Any SAN disks supported by VIOS



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



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

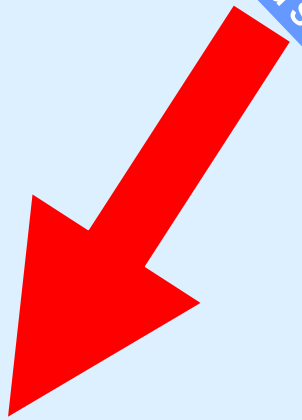


4 SSP2 space allocated to virtual machines = LU
LU = Logical Unit = 64 MB blocks, evenly distributed across all LUNs

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

Preparation

- 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 can’t simply stop the cluster to make low level disk attributes changes

Cluster Management

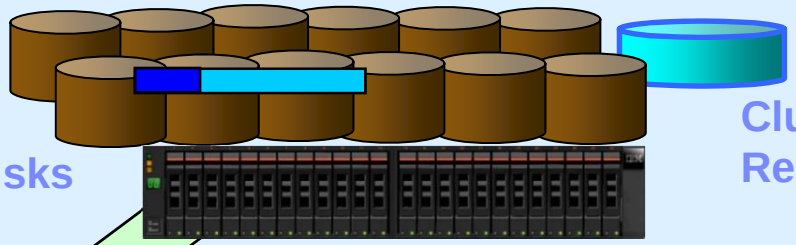
- Create Cluster and Pool
- Add node
- House Keeping

1. `cluster -create …`
2. `cluster –list`
3. `cluster -status …`

4. `cluster –addnode …`

5. `cluster -rmnode …`
6. `cluster –delete ..`

Architecture



Cluster Repository

LUNs = Pools Disks

FibreChannel

Network

Virtual I/O Server

If using FCoCEE this could be just one FCoCEE adapter (or one pair)

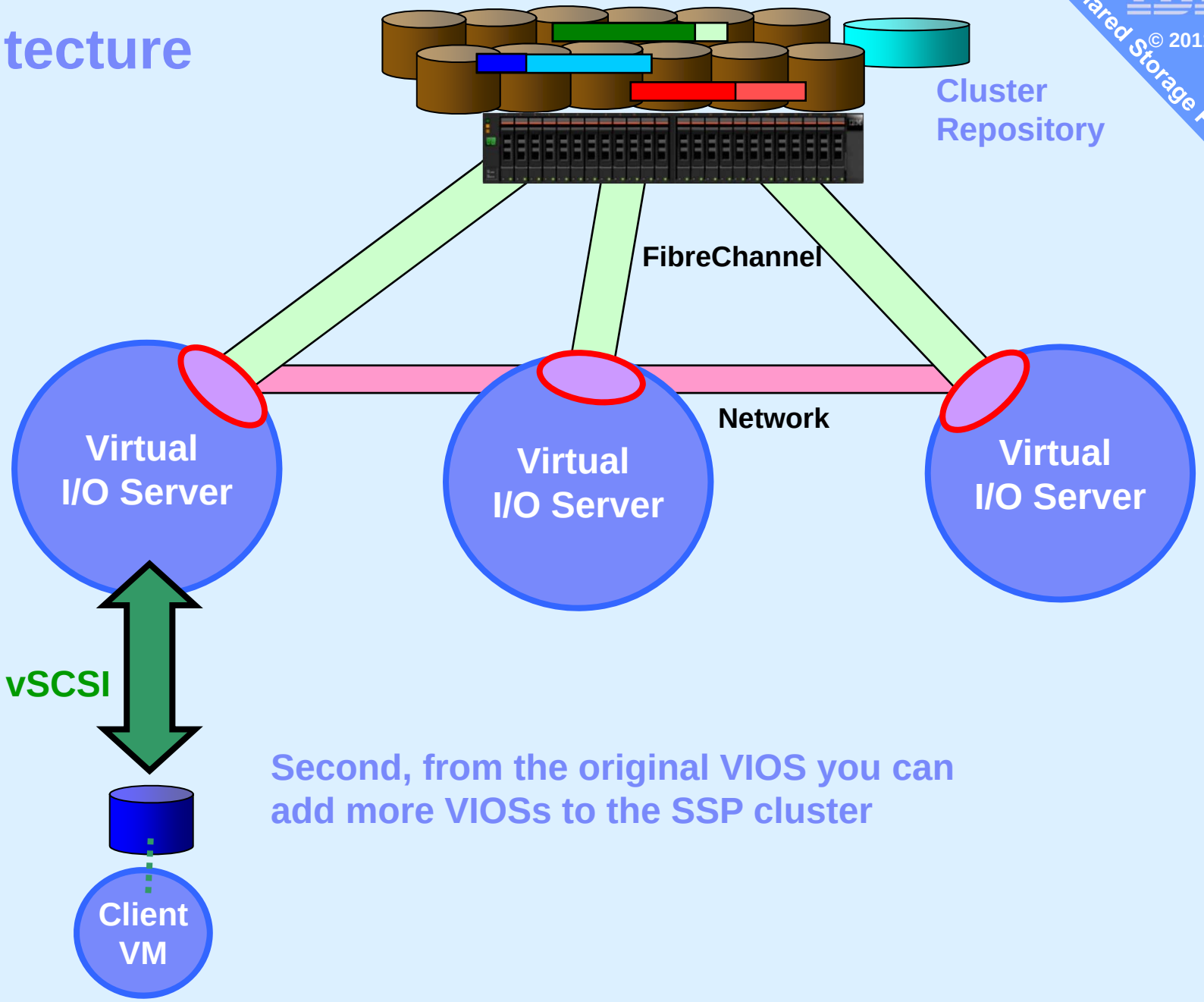
Requires good network
Different subnets OK,
if multi-cast allowed,
Multi-cast = CAA requirement

vSCSI

Client VM

First Create the cluster of one VIOS
-Builds the repository
-Checks the LUNs are available
-You can then allocate space to its VM clients

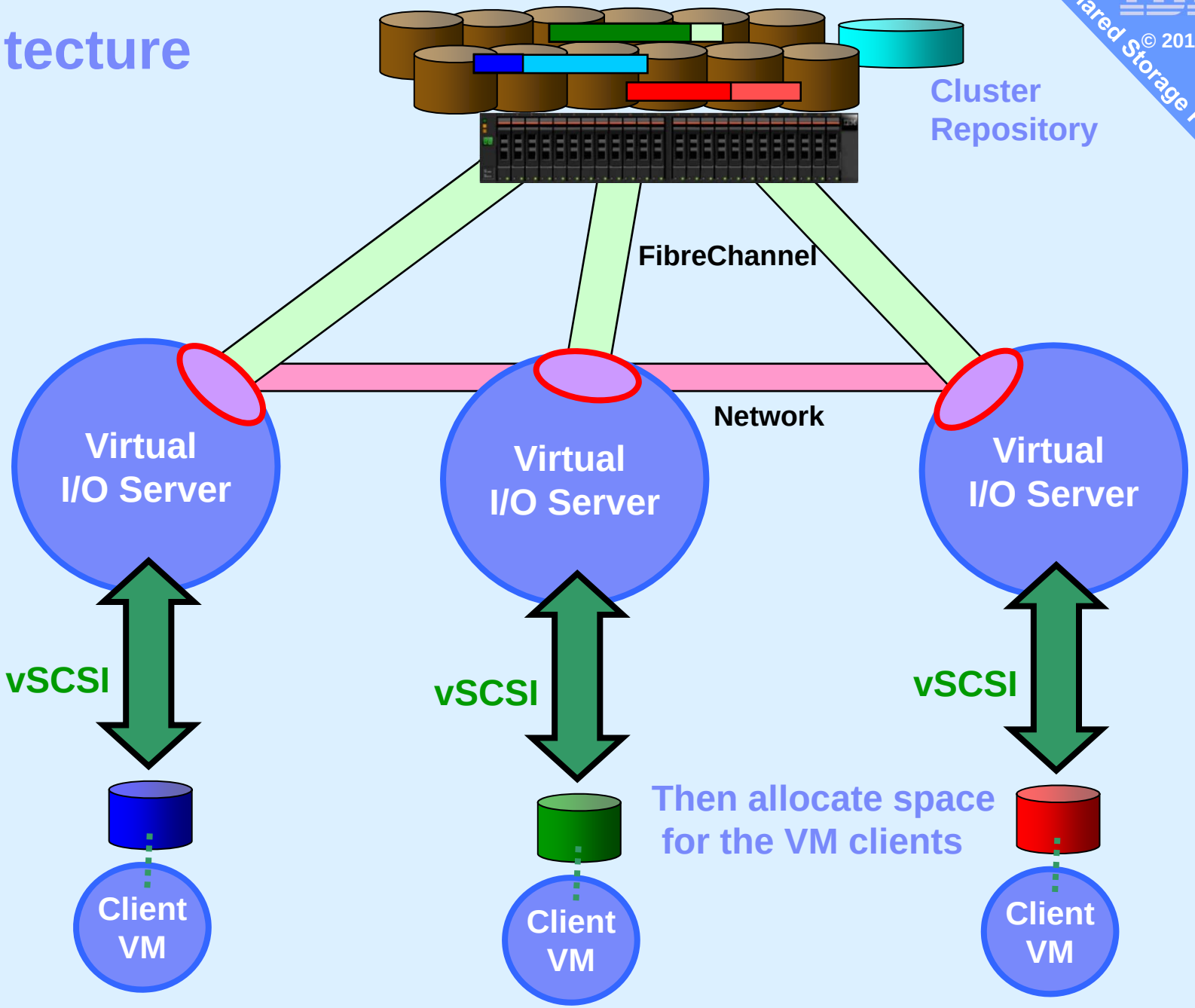
Architecture



Second, from the original VIOS you can add more VIOSs to the SSP cluster

Architecture

Shared Storage Pool 2



Cluster hostnames

1. Need full DNS working or /etc/hosts
2. Full hostname is mandatory
3. hostname command must show the full hostname

   $ hostname

   `redvios1.ibm.com`

Not just “redvios1” here

```
                        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
```

VIOS
oem_setup_env
smitty tcpip
Example

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

On 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
```

Space Management

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

1. `mkbdsp -clustername galaxy -sp atlantic 16G -bd vdisk_red6a -vadapter vhost2 [-thick]`

2. `rmbdsp -clustername galaxy -sp atlantic -bd vdisk_red6a`

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

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
…
```

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 one 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
```

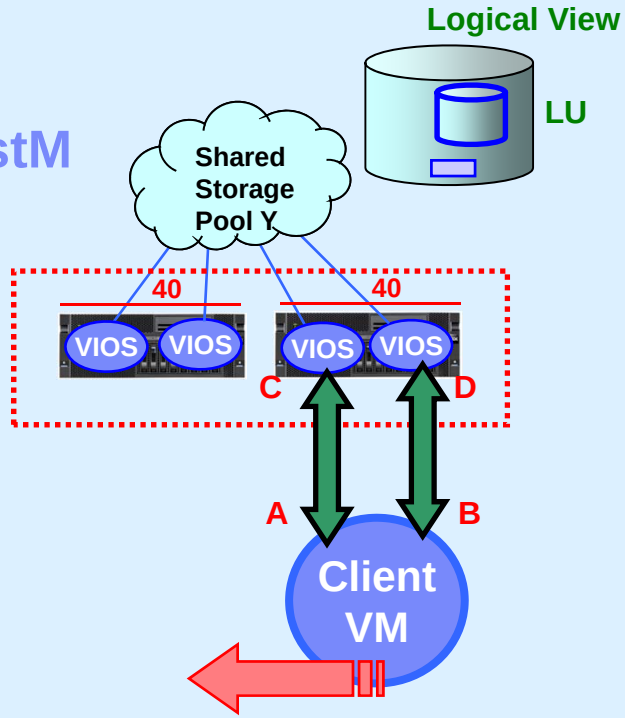
Dual path via Two VIOSs

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

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

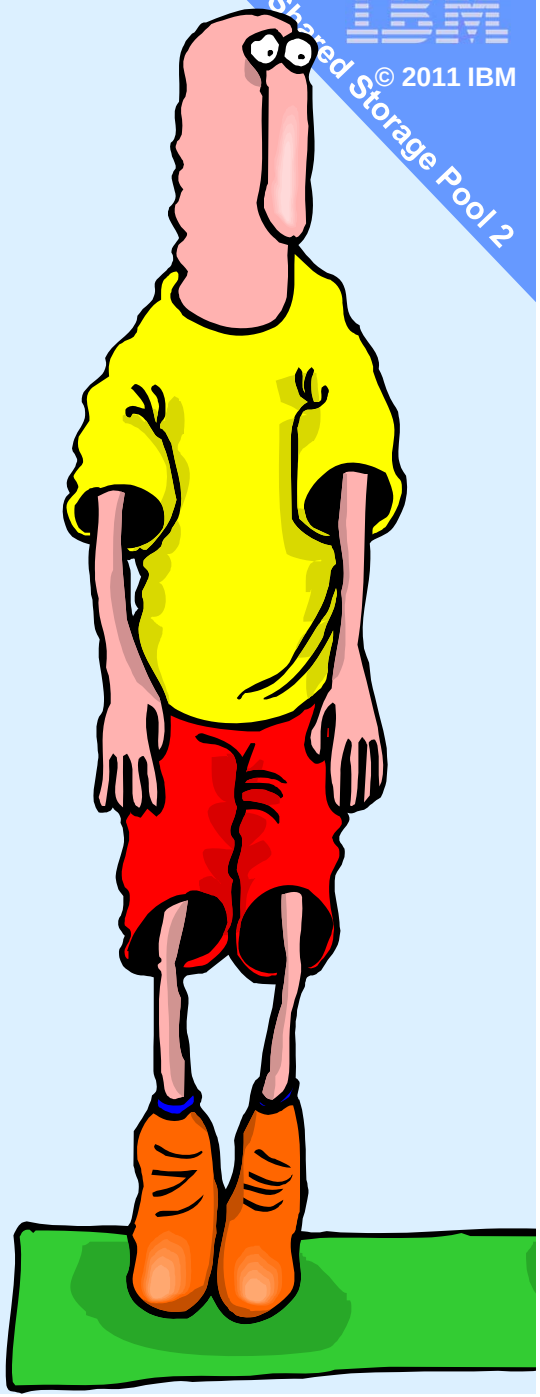
```
$ chsp –add –clustername galaxy \
  –sp atlantic hdisk8
```

Remove a LUN from the Pool

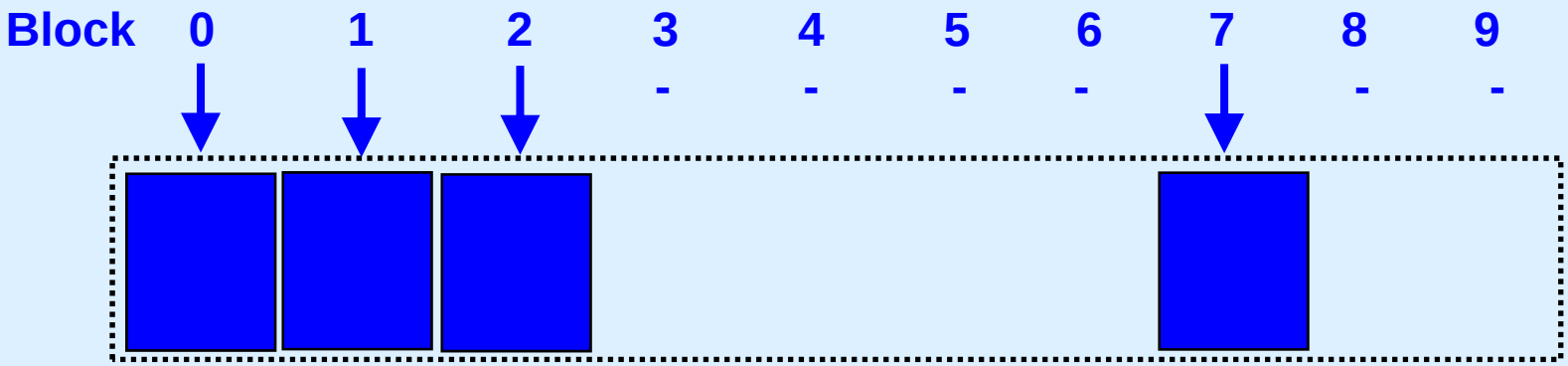
- You can’t
- We can replace a disk but not remove one

Experiments in Thin provisioning

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

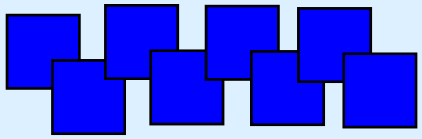


Thin Provisioning Blocks can be missing

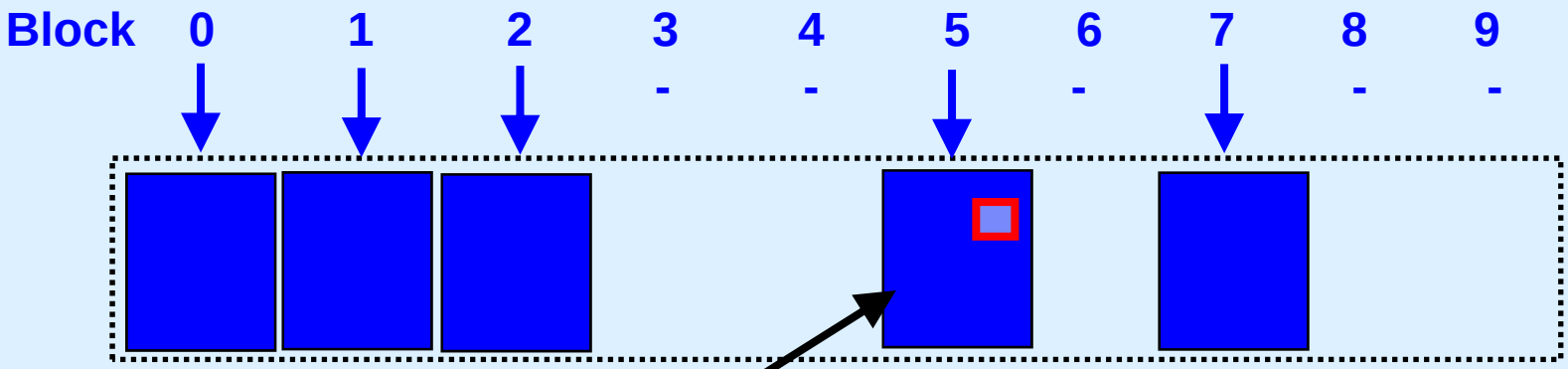


List storage pool “lssp” output shows block size = 64MB

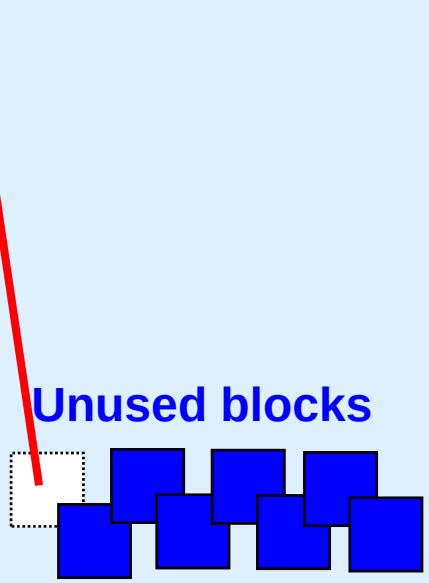
Unused blocks



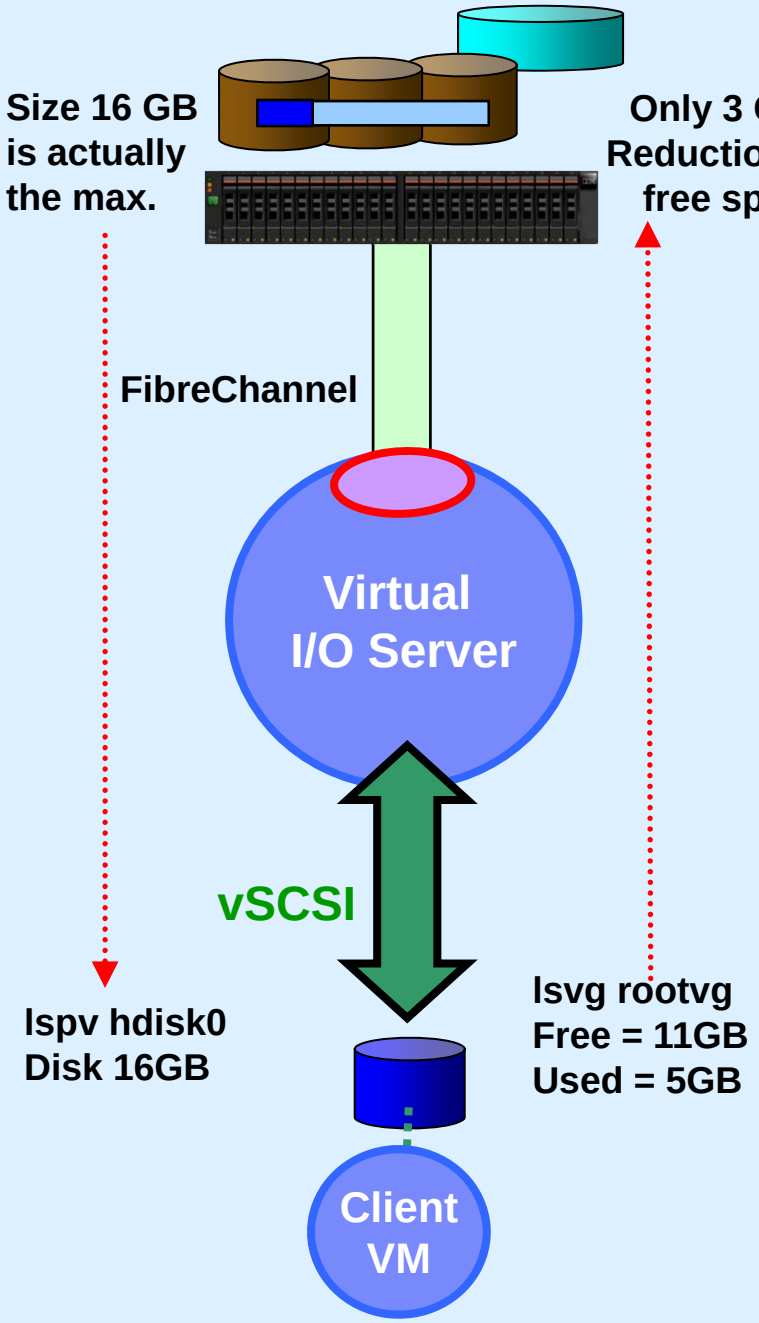
Thin Provisioning
Added on 1st Write



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 then client is unaware of this



Thin Provisioning



- 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”

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        Size(mb)  Free(mb)  TotalLUSize(mb)  LUs   Type PoolID
atlantic    47552       17945   59392            5     CLPOOL …
```

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

Pool space all used up? Ek!

Next write needing a new SSP block, gets a disk error!

Just don’t go there

House keeping - Thin Provisioning Alerts

- Set alerts to warn on the pool filling beyond 80%
  - **alert -set -clustername** *galaxy* **–spname** *atlantic* **-value** *80*

- To list the alert threshold:
  - **alert -list -clustername** *galaxy* **-spname** *atlantic*

```
$ alert -list -clustername galaxy -spname atlantic
PoolName      PoolID                Threshold%
atlantic      0009452CDD04EA226DF   35
```

- To unset the alert:
  - **alert -unset -clustername** *galaxy* **-spname** *atlantic*

- Alert reporting in AIX errpt = VIOS errlog command
  - Also reported to high levels SM → Systems Director etc.

House keeping - Thin Provisioning Alerts

- It is vital that you get these warning messages
- Suggest
  - Email the Pool stats every night to the admin guys
    - lssp -clustername galaxy | mailx ops@acme.com –s “SSP status”
  - Script to check and if low - email or send Mobile text msg
  - Set off the fire alarm bell or Claxton horns
- Possible reactions are:
  - Add a new LUN to the pool,
  - Delete allocated space = unused LU or entire VM & space
  - Drop a Snapshot or two

My own script for better pool stats for all clusters and all pools

Example:

```
$ lspool
Cluster list:  galaxy
Pools in galaxy are: atlantic
atlantic Pool-Size: 52864 MB
atlantic Pool-Free: 45346 MB Percent Free 85
atlantic Pool-Used:  7518 MB Percent Used 14
atlantic Allocated: 62768 MB for 2 Logical Units
atlantic Alert-Percent: 35
atlantic OverCommitted: yes by 9904 MB
```

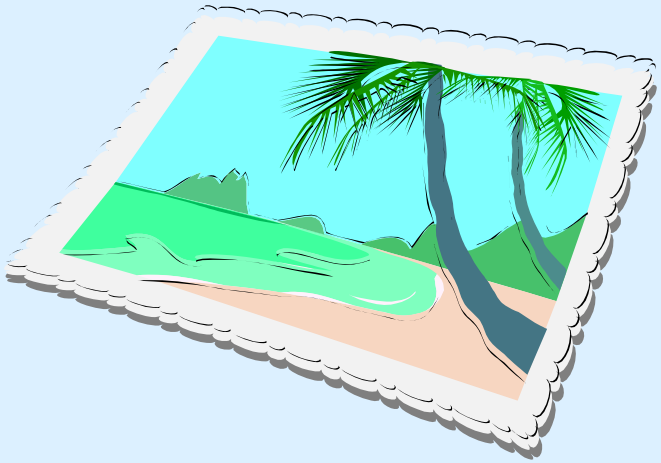
If you are not over committed

```
atlantic OverCommitted:no
```

My lspool script

```
# lspool list each cluster and for each list its pools and pool details
. ~/.profile

clusters=`cluster -list | sed '1d' | awk -F " " '{ printf $1 " " }'`
echo "Cluster list: " $clusters

for clust in $clusters
do
        pools=`lssp -clustername $clust | sed '1d' | awk -F " " '{ printf $1 " " }'`
        echo Pools in $clust are: $pools

        for pool in $pools
        do
                lssp -clustername $clust | sed '1d' | grep $pool | read p size free totalLU numLUs junk

                let freepc=100*$free/$size
                let used=$size-$free
                let usedpc=100*$used/$size

                echo $pool Pool-Size: $size MB
                echo $pool Pool-Free: $free MB Percent Free $freepc
                echo $pool Pool-Used: $used MB Percent Used $usedpc
                echo $pool Allocated: $totalLU MB for $numLUs Logical Units

                alert -list -clustername $clust -spname $pool | sed '1d' | grep $pool | read p poolid percent
                echo $pool Alert-Percent: $percent

                if [[ $totalLU > $size ]]
                then
                let over=$totalLU-$size
                echo $pool OverCommitted: yes by $over MB
                else
                echo $pool OverCommitted: no
                fi
        done
done
```

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 names for the storage pool option

Example:

snapshot … -spname <name>

lssp … -sp <name>

It can easily catch you out.

Snapshot – delete or rollback

After your backup or when sure you never want to rollback
Delete original and continue on currently blocks

```
$ snapshot -clustername galaxy -delete diamond5t.snap
  -spname atlantic -lu vdisk_diamond5a
```

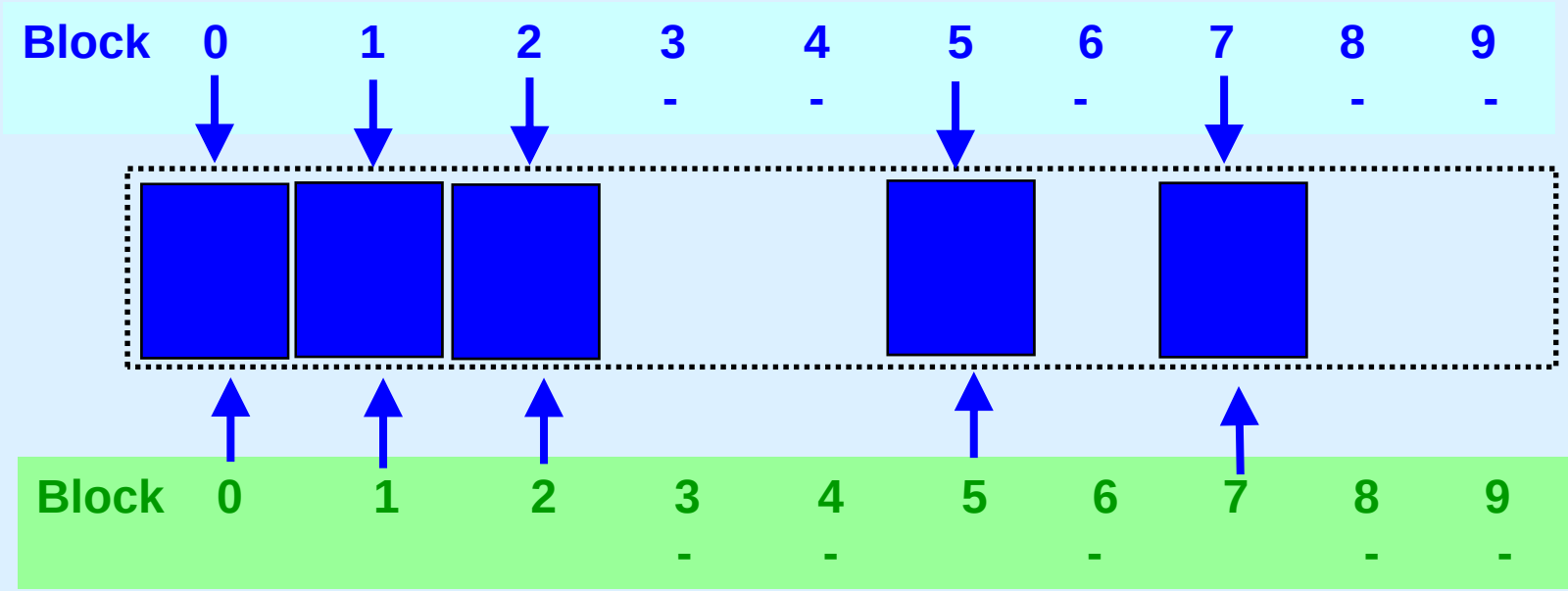
Rollback to a snapshot
Stop the virtual machine/LPAR then

```
$ snapshot -clustername galaxy -delete diamond5t.snap
  -spname atlantic -lu vdisk_diamond5a
```

You loose any updated you made since the last snapshot

Snapshot Model

Original Set



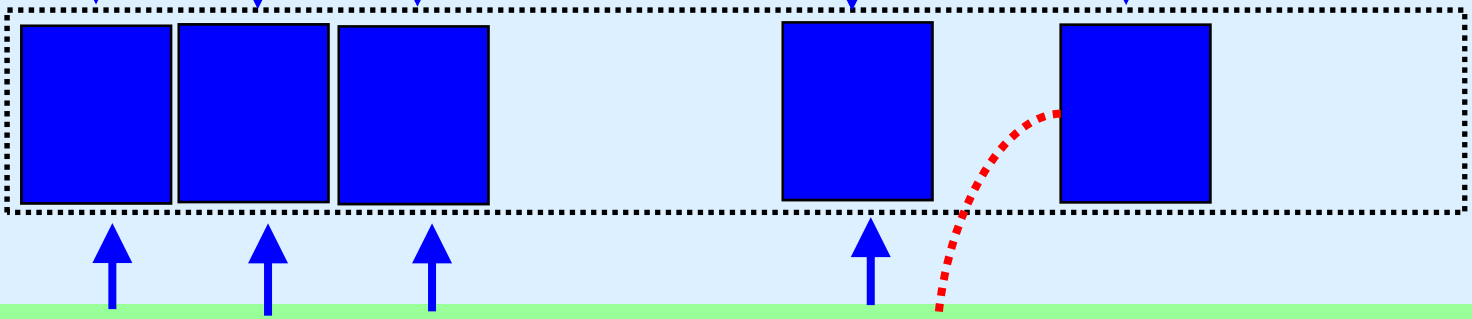
New Snapshot

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

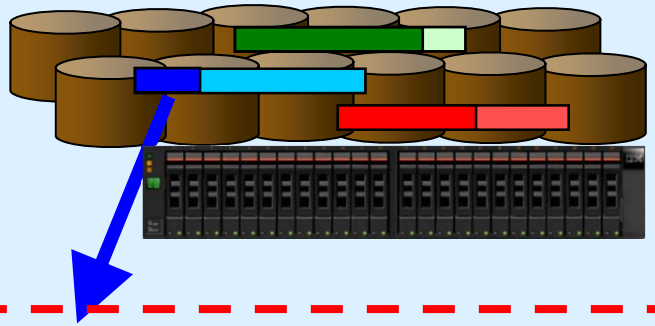


Block 0 1 2 3 4 5 6 7 8 9

New Snapshot

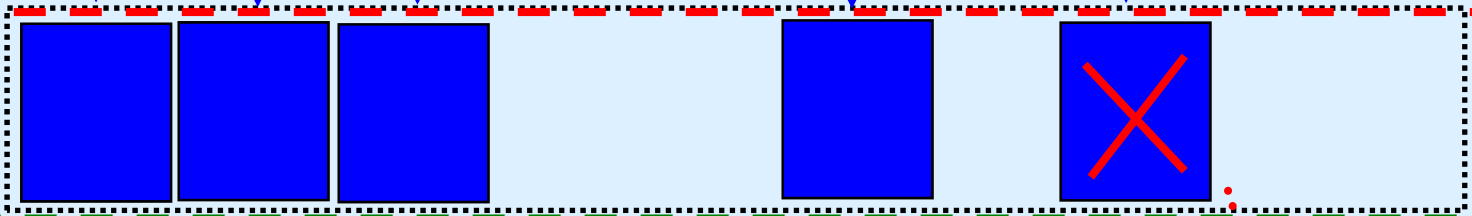
1 Client VM update to block 7
2 VIOS allocates a new block
3 Copies original 64MB
4 VIOS completes the write

Snapshot delete or Roll-back



Original Set

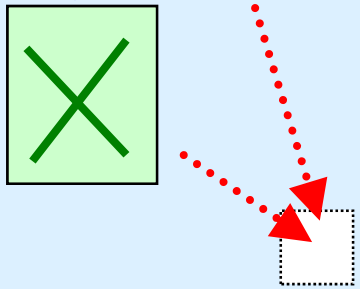
Block	0	1	2	3	4	5	6	7	8	9
				-	-		-		-	-



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

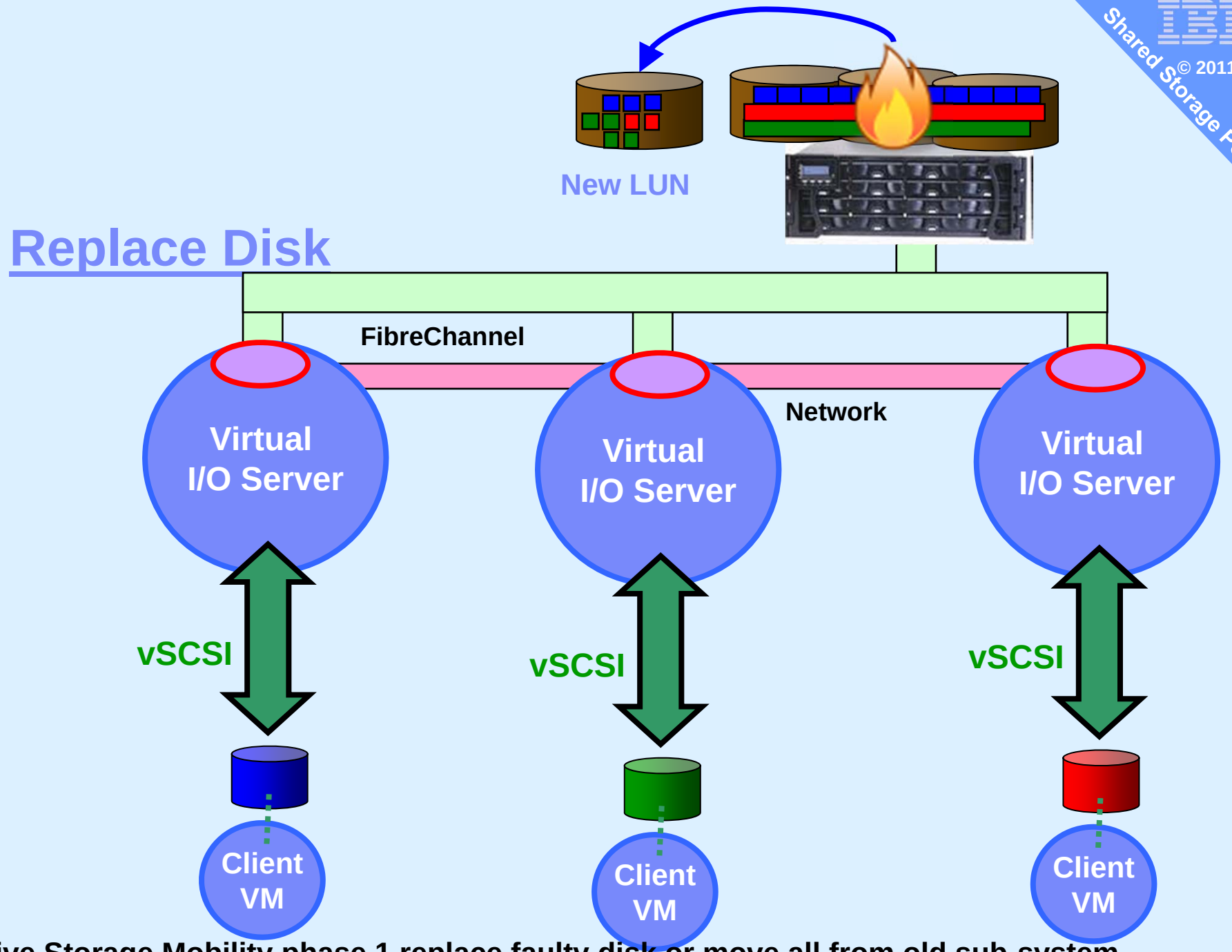
Live Storage Mobility

Server Admin would like to :

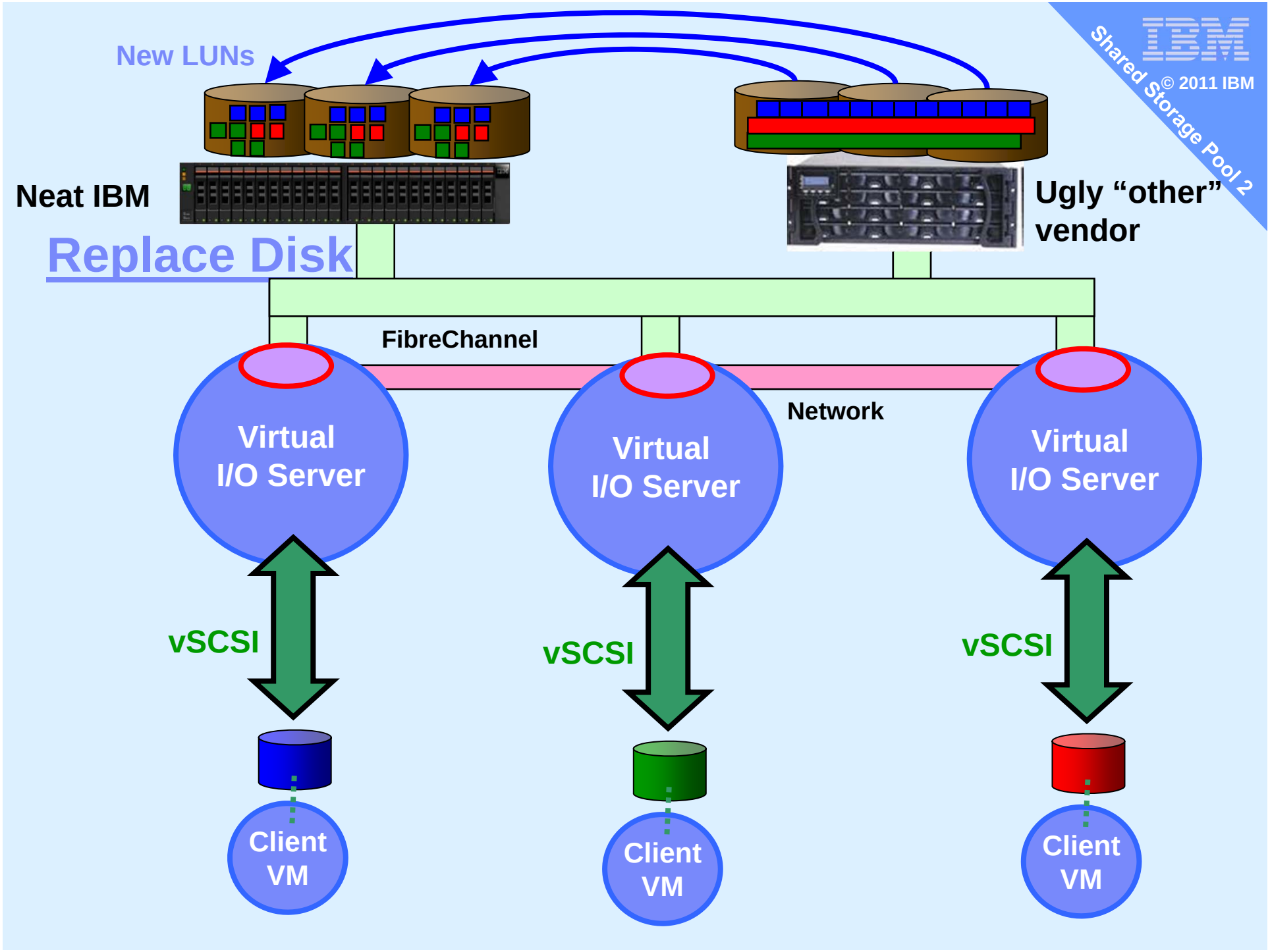
A. Replace a faulty LUN
B. Move all blocks off one disk subsystem (retiring a disk subsystem)
C. Recover from repository failure
D. Select which disk subsystems a particular VM uses
E. Ensure mirrors are on different subsystems (even different sites)

- 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
  - Could use SVC now for lower level mirror (E)

Replace Disk

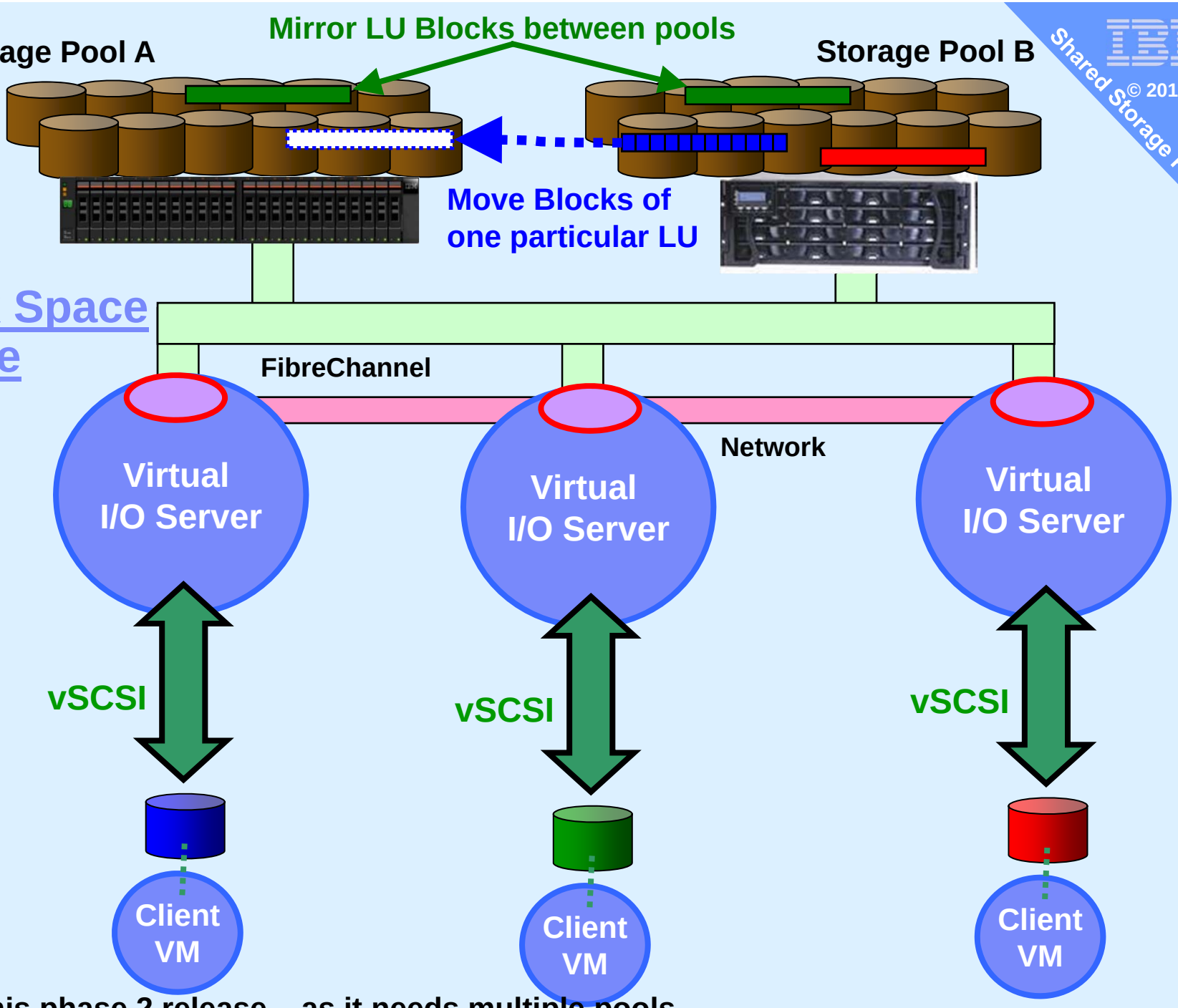


Live Storage Mobility phase 1 replace faulty disk or move all from old sub-system


New LUNs
Neat IBM
Ugly "other" vendor
Replace Disk
FibreChannel
Network
Virtual I/O Server
Virtual I/O Server
Virtual I/O Server
vSCSI
vSCSI
vSCSI
Client VM
Client VM
Client VM
Shared Storage Pool 2
© 2011 IBM

Shared Storage Pool 2

Disk Space Move



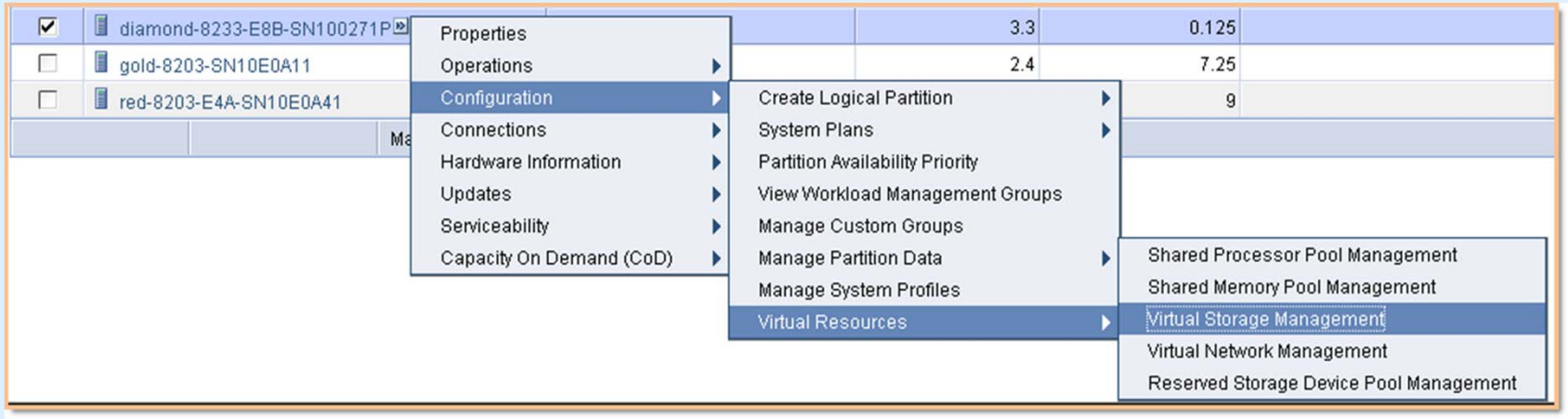
Not in this phase 2 release – as it needs multiple pools

User Interface

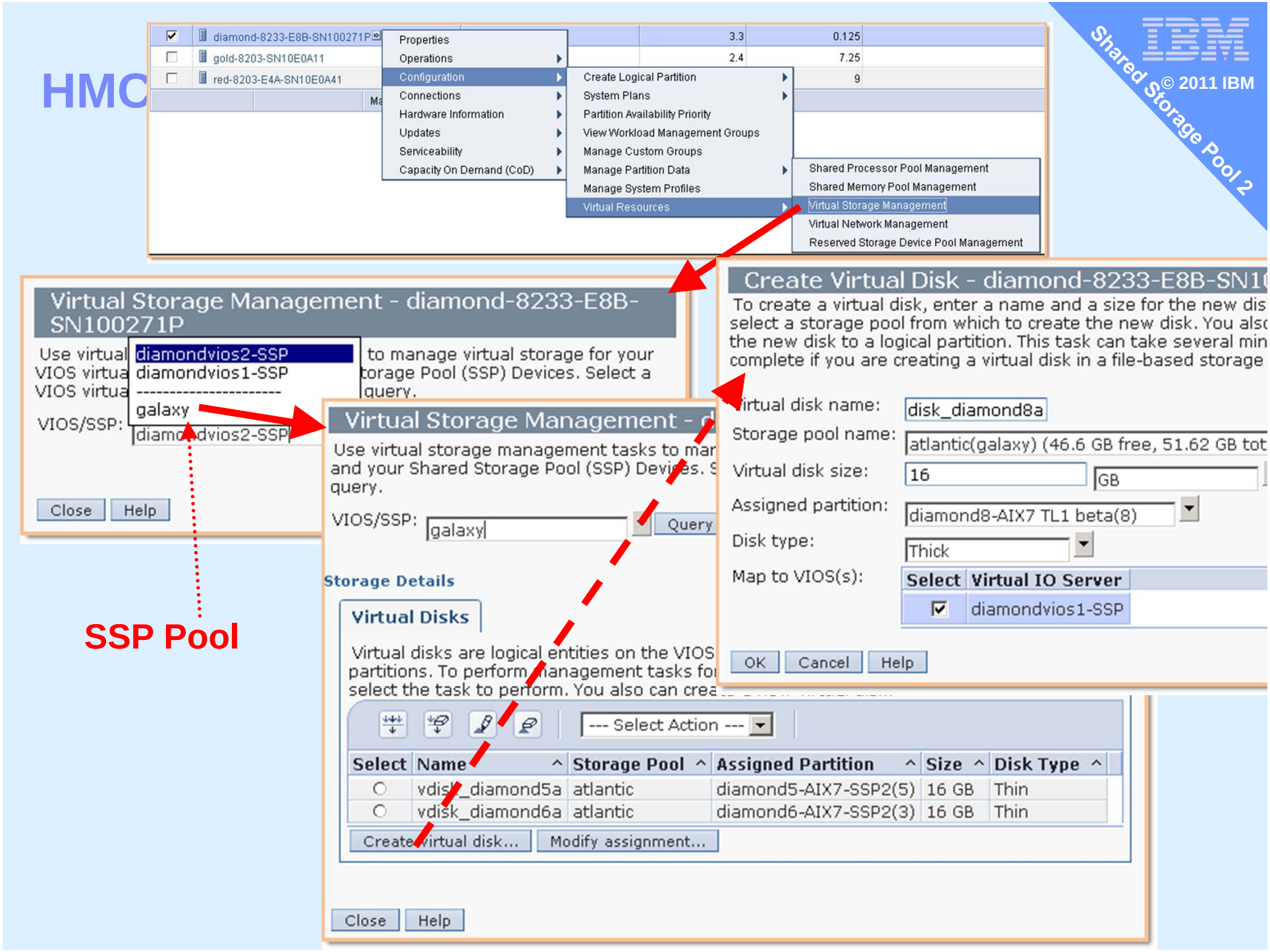
- Command line
  - Already shown in this presentation by example
  - Some feature will remain command line only → like: cluster -create
- cfgassist
  - This is the VIOS version of smitty
  - Menu driven interface for CLI
  - Fully covers SSP functions
- Graphical User Interface
  - HMC - now
  - Planned SDMC at next major release
- 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

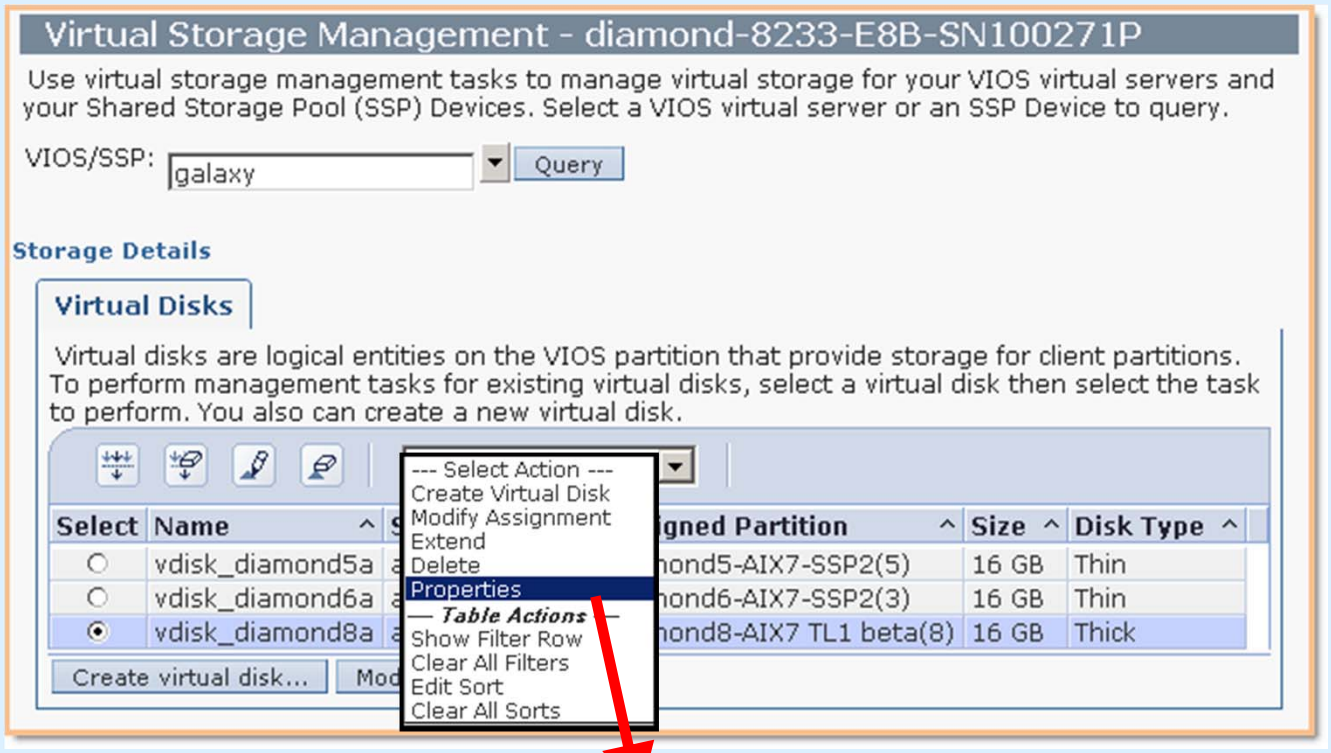
- HMC from October 2011
  - for SSP support & LPM of SSP LPARs
  - Addition feature to Virtual Storage Management

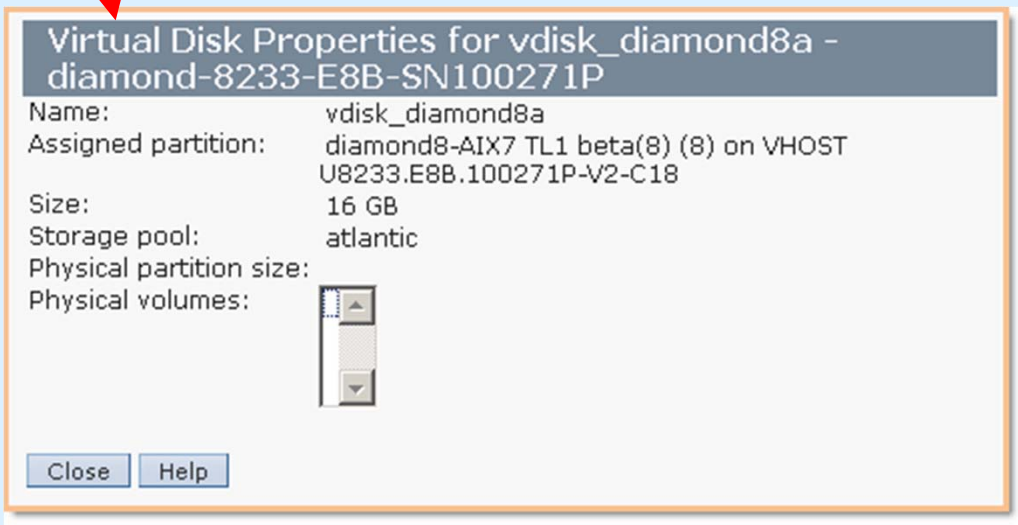


- Shipped with HMC upgrade V7 R7.4 SP0+
  - Not part of the VIOS package


HMC
Shared Storage Pool 2
© 2011 IBM
diamond-8233-E8B-SN100271P
gold-8203-SN10E0A11
red-8203-E4A-SN10E0A41
3.3
0.125
2.4
7.25
9
Properties
Operations
Configuration
Connections
Hardware Information
Updates
Serviceability
Capacity On Demand (CoD)
Create Logical Partition
System Plans
Partition Availability Priority
View Workload Management Groups
Manage Custom Groups
Manage Partition Data
Manage System Profiles
Virtual Resources
Shared Processor Pool Management
Shared Memory Pool Management
Virtual Storage Management
Virtual Network Management
Reserved Storage Device Pool Management
Virtual Storage Management - diamond-8233-E8B-SN100271P
Use virtual
VIOS virtua
VIOS virtua
to manage virtual storage for your
torage Pool (SSP) Devices. Select a
query.
diamondvios2-SSP
diamondvios1-SSP
galaxy
VIOS/SSP:
Close
Help
Virtual Storage Management - d
Use virtual storage management tasks to mar
and your Shared Storage Pool (SSP) Devices. S
query.
VIOS/SSP:
galaxy
Query
Storage Details
Virtual Disks
Virtual disks are logical entities on the VIOS
partitions. To perform management tasks fo
select the task to perform. You also can crea
--- Select Action ---
Select
Name
Storage Pool
Assigned Partition
Size
Disk Type
vdisk_diamond5a
atlantic
diamond5-AIX7-SSP2(5)
16 GB
Thin
vdisk_diamond6a
atlantic
diamond6-AIX7-SSP2(3)
16 GB
Thin
Create virtual disk...
Modify assignment...
Close
Help
Create Virtual Disk - diamond-8233-E8B-SN1
To create a virtual disk, enter a name and a size for the new dis
select a storage pool from which to create the new disk. You also
the new disk to a logical partition. This task can take several min
complete if you are creating a virtual disk in a file-based storage
Virtual disk name:
disk_diamond8a
Storage pool name:
atlantic(galaxy) (46.6 GB free, 51.62 GB tot
Virtual disk size:
16
GB
Assigned partition:
diamond8-AIX7 TL1 beta(8)
Disk type:
Thick
Map to VIOS(s):
Select
Virtual IO Server
diamondvios1-SSP
OK
Cancel
Help
SSP Pool

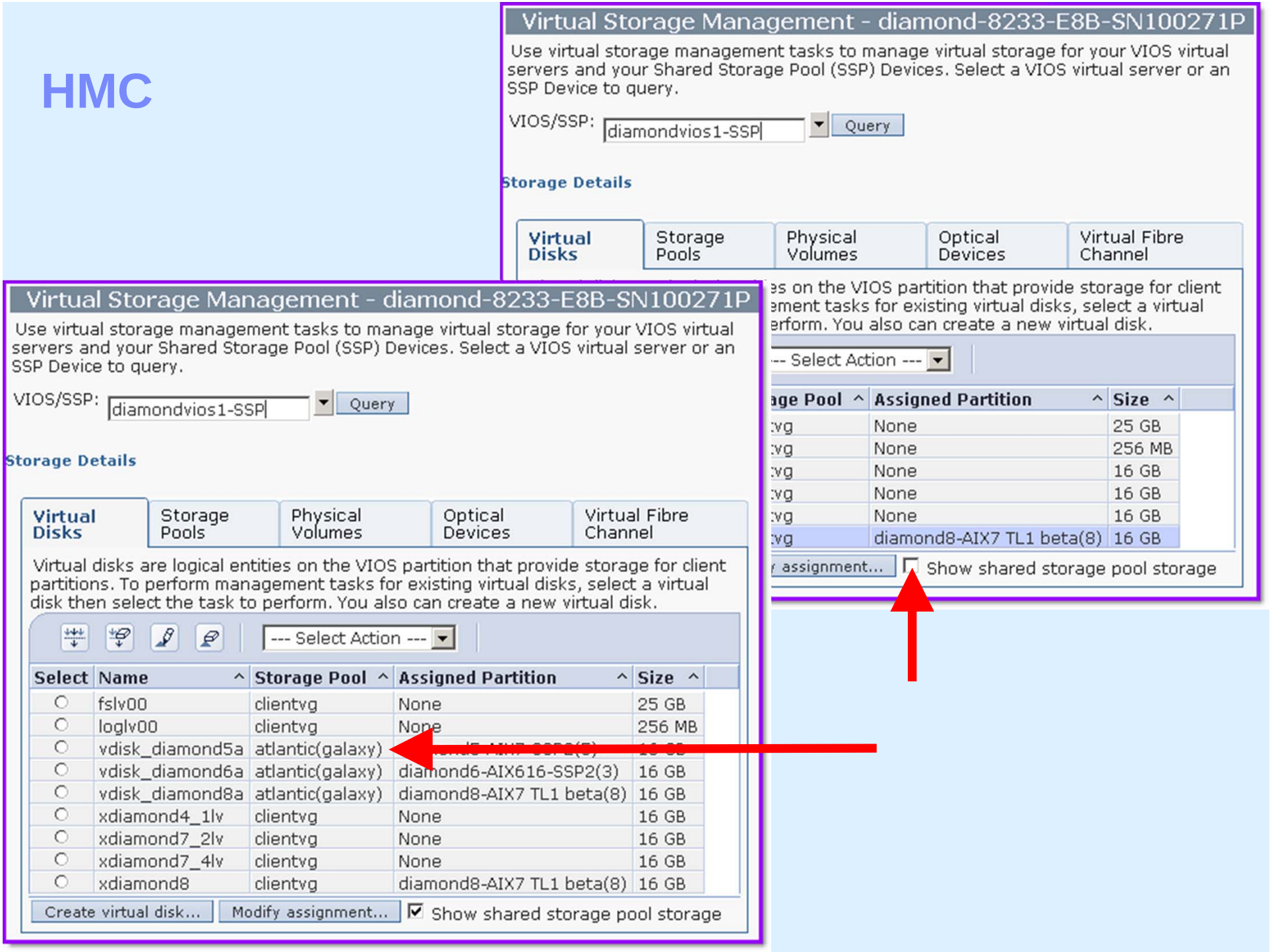
HMC





Looks like a "work in progress" with more to come

HMC


Virtual Storage Management - diamond-8233-E8B-SN100271P
Use virtual storage management tasks to manage virtual storage for your VIOS virtual servers and your Shared Storage Pool (SSP) Devices. Select a VIOS virtual server or an SSP Device to query.
VIOS/SSP: diamondvios1-SSP
Query
Storage Details
Virtual Disks
Storage Pools
Physical Volumes
Optical Devices
Virtual Fibre Channel
Virtual disks are logical entities on the VIOS partition that provide storage for client partitions. To perform management tasks for existing virtual disks, select a virtual disk then select the task to perform. You also can create a new virtual disk.
--- Select Action ---
Select
Name
Storage Pool
Assigned Partition
Size
fslv00 clientvg None 25 GB
loglv00 clientvg None 256 MB
vdisk_diamond5a atlantic(galaxy)
vdisk_diamond6a atlantic(galaxy) diamond6-AIX616-SSP2(3) 16 GB
vdisk_diamond8a atlantic(galaxy) diamond8-AIX7 TL1 beta(8) 16 GB
xdiamond4_1lv clientvg None 16 GB
xdiamond7_2lv clientvg None 16 GB
xdiamond7_4lv clientvg None 16 GB
xdiamond8 clientvg diamond8-AIX7 TL1 beta(8) 16 GB
Create virtual disk...
Modify assignment...
Show shared storage pool storage
Assigned Partition
Size
None 25 GB
None 256 MB
None 16 GB
None 16 GB
None 16 GB
diamond8-AIX7 TL1 beta(8) 16 GB
Show shared storage pool storage

What if you loose the VIOS?

- Updated **viosbr** supports backup / restore of SSP config
  - Warning: this saves the config but not the data
- Backup – will perform regular backups for you

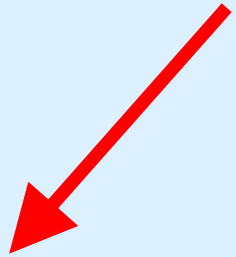
  viosbr -backup -clustername Name -file File \
  [-frequency daily|weekly|monthly [-numfiles fileCount]]
- View

  viosbr -view -file File -clustername Name [-type devType][-detail | -mapping]
- Restore

  viosbr -restore -clustername N -file F -subfile NodeFile [-validate | -inter | -force][-type devType]
  viosbr -restore -clustername N -file F -repopvs disks [-validate | -inter | -force][-type devType][-currentdb]
  viosbr -restore -clustername N -file F -subfile NodeFile –xmlvtds
  viosbr -recoverdb -clustername N [-file F]
  viosbr -migrate -file F
- Can recover from
  1. Repository Disk is corrupted (see -repopvs)
  2. One SSP VIOS is reinstalled
  3. SSP Database is corrupted
  4. Restore to old configuration on the VIOS node
     - Changes done to SSP mappings on the node after a backup

viosbr by example

As padmin
$ viosbr -backup -clustername galaxy -file thursday
Backup of node diamondvios1.aixncc.uk.ibm.com failed
Backup of node diamondvios2.aixncc.uk.ibm.com successful
Backup of this node redvios1.aixncc.uk.ibm.com successful
$
$ viosbr -backup -clustername galaxy -file Daily -frequency daily -numfiles 10
$
$ ls –l /home/padmin/cfgbackups
$ ls -l cfg*
total 1288
-rw-r--r-- 1 root staff 341647 Dec 01 16:54 Daily.01.galaxy.tar.gz
-rw-r--r-- 1 root staff 352132 Dec 02 09:00 Daily.02.galaxy.tar.gz
-rw-r--r-- 1 root staff 310478 Dec 01 16:01 thursday.galaxy.tar.gz
-rw-r--r-- 1 root staff 358646 Dec 01 17:48 Weekly.01.galaxy.tar.gz



Cluster lsmap → MachineType.Serial Number


```
$ lsmap -clustername galaxy -all
Physloc                                                  Client Partition ID
-------------------------------------------------------- ------------------
U8203.E4A.10E0B41-V1-C18                                 0x00000005

VTD                  vtscsi0
LUN                  0x8100000000000000
Backing device       vdisk_diamond5a.b3f3a049c067ada140bc9f53f0a92b12

Physloc                                                  Client Partition ID
-------------------------------------------------------- ------------------
U8233.E8B.100C71P-V1-C32                                 0x00000003

VTD                  vtscsi0
LUN                  0x8100000000000000
Backing device       vdisk_diamond6a.4c9e9eb95ae518e7567dcc063a47f719

Physloc                                                  Client Partition ID
-------------------------------------------------------- ------------------
U8233.E8B.100C71P-V2-C13                                 0x00000003

VTD                  vtscsi1
LUN                  0x8100000000000000
Backing device       vdisk_diamond6a.4c9e9eb95ae518e7567dcc063a47f719

$
```

POWER6 520

POWER7 750

POWER7 750

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 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

SSP2 command cheat sheet

1. chdev -dev <device name> -attr reserve_policy=no_reserve
2. cluster -create -clustername galaxy -repopvs hdisk2
   -spname atlantic -sppvs hdisk3 hdisk5 -hostname bluevios1.ibm.com
3. cluster –list
4. cluster -status -clustername galaxy
5. cluster –addnode –clustername galaxy –hostname redvios1.ibm.com
6. cluster -rmnode [-f] -clustername galaxy -hostname redvios1.ibm.com
7. cluster –delete –clustername galaxy
8. lscluster –s or –d or –c or –i = CAA command
9. chsp –add –clustername galaxy -sp atlantic hdisk8 hdisk9
10. chsp -replace -clustername galaxy -sp atlantic -oldpv hdisk4 -newpv hdisk24
11. mkbdsp -clustername galaxy -sp atlantic 16G
    -bd vdisk_red6a -vadapter vhost2 [-thick]
12. rmbdsp -clustername galaxy -sp atlantic -bd vdisk_red6a
13. lssp -clustername galaxy -sp atlantic -bd
14. lssp -clustername galaxy
15. alert -set -clustername galaxy –spname atlantic -value 80
16. alert -list -clustername galaxy -spname atlantic
17. errlog –ls
18. snapshot -create name -clustername galaxy -spname atlantic -lu LUs
19. snapshot -delete name -clustername galaxy -spname atlantic -lu LUs
20. snapshot -rollback name -clustername galaxy -spname atlantic -lu LUs
21. snapshot –list -clustername galaxy -spname atlantic
22. viosbr -backup -clustername galaxy -file Daily -frequency daily -numfiles 10
23. viosbr -view -file File -clustername Name …
24. viosbr -restore -clustername Name …
25. lsmap -clustername galaxy -all

SSP Demo Commands – as padmin

1. `cluster -list`
2. `cluster -status -clustername galaxy`
3. `lscluster -c`

4. `lssp -clustername galaxy`
5. `lssp -clustername galaxy -sp atlantic -bd`

6. `mkbdsp -clustername galaxy -sp atlantic 32G -bd my_disk_name -vadapter vhost2 -thick`
   then see the new LU with
   `lssp -clustername galaxy -sp atlantic -bd`

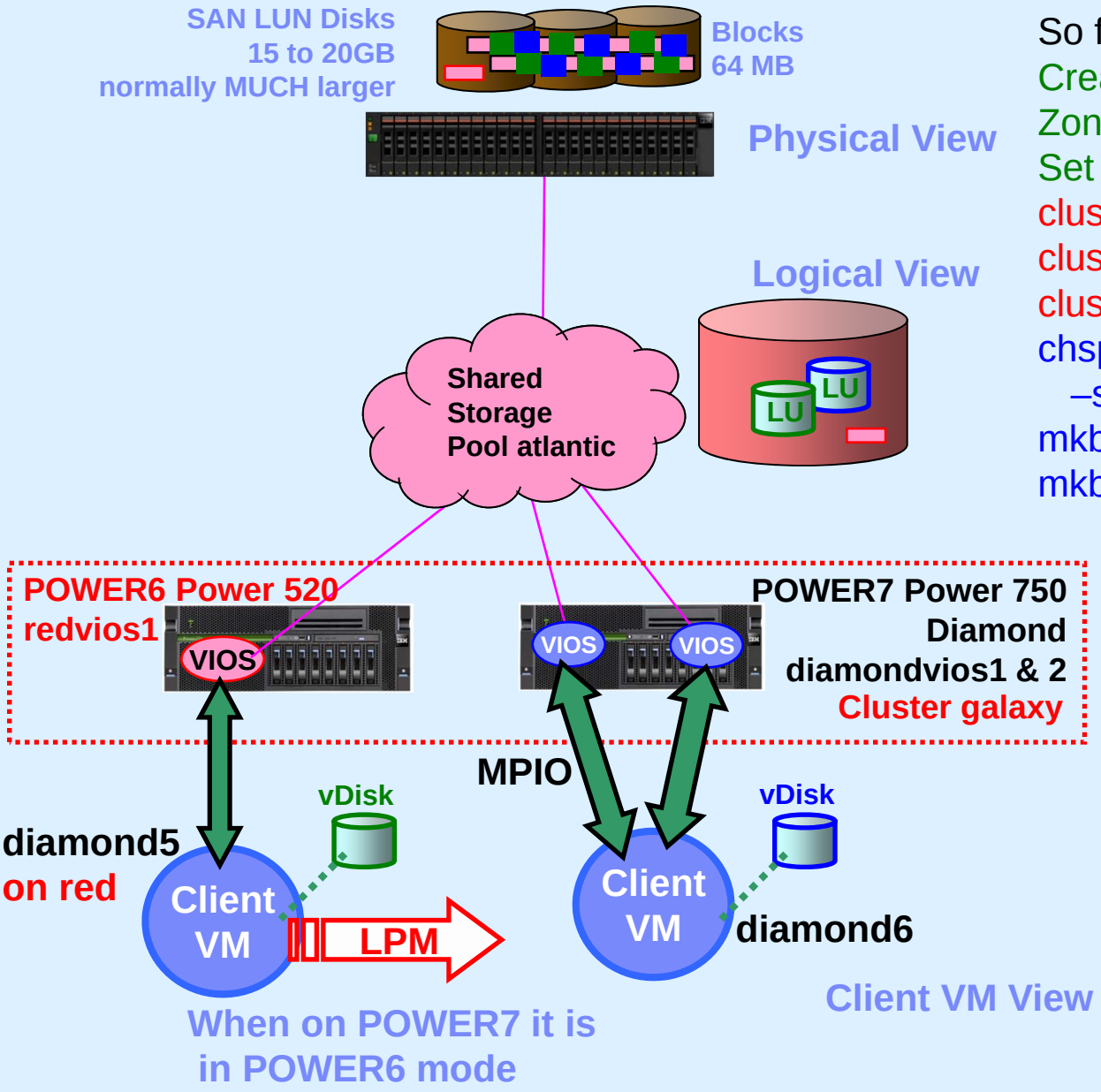
7. `alert -list -clustername galaxy -spname atlantic`

8. `snapshot -create diamond5s.snap -clustername galaxy -spname atlantic -lu vdisk_diamond5a`
   `snapshot -list -clustername galaxy -spname atlantic`
   `lssp -clustername galaxy -sp atlantic -bd`

9. `lsmap -clustername galaxy –all`

10. **HMC user interface & perhaps LPM (no GUI change here)!**

Demonstration - Shared Storage Pool phase 2



So far:

Created LUNs

Zoned to all VIOS

Set no_reserve

cluster -create … on diamondvios1

cluster -addnode … diamondvios2

cluster -addnode … redvois1

chsp –add –clustername galaxy \
–sp atlantic hdisk8

mkbdsp … for diamond5

mkbdsp … diamond6 + vSCSI link