Power Systems Virtualisation from IBM - Technical Webinar User Group

Today

27th → Shared Storage Pools - Phase 3

Starting at 10:00 am UK time by Nigel Griffiths

Smart Meeting → Put questions into the Chat box or AT&T Toll Free phone for better audio

- 0800-368-0638 = UK Toll Free
- 0203-059-6451 = UK but you pay for the call
- Then 6403785# Participant Code
- Other countries see chat box for the website
- Please Mute with *6

Previous Sessions:

Electric Server Agent
RDX Removable disks
Dynamic Platform Optimiser
PowerSC
POWER Advisors
POWER7 Affinity and Perf.
Updating Power, I/O & HMC
VPM for IBM i
ISD VMControl
- Capture & Deploy

Future Sessions → http://tinyurl.com/newUK-PowerVM-VUG

- **July 31st: Whole POWER Machine Monitoring**

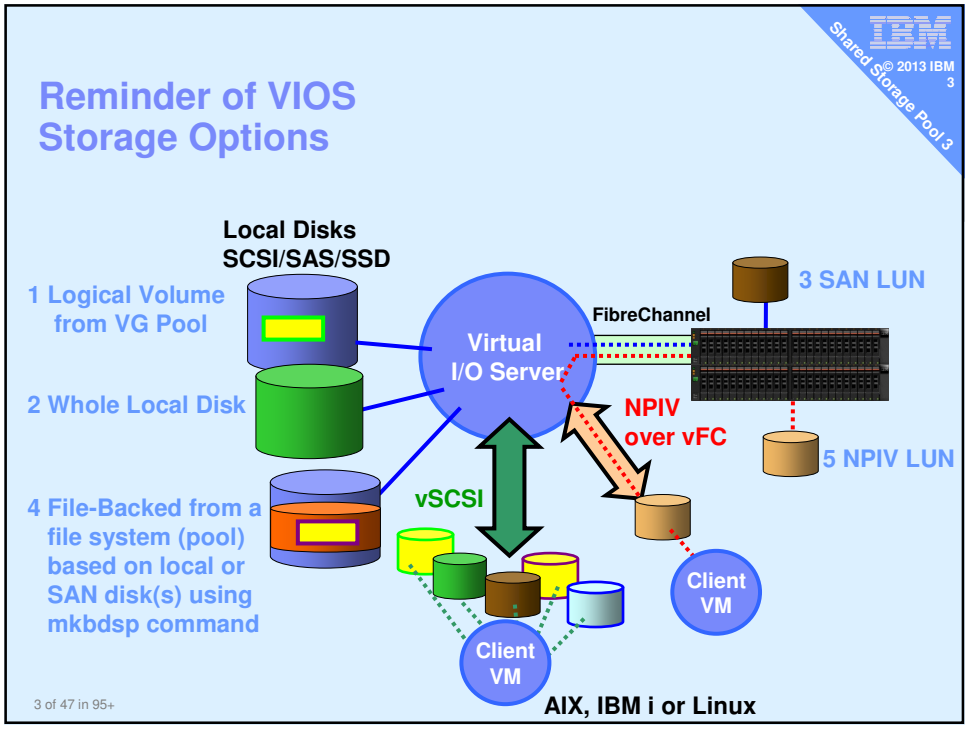
Twitter:
Nigel Griffiths @mr_nmon
Jyoti Dodhia @JyotiDodhia
Website http://tinyurl.com/newUK-PowerVM-VUG

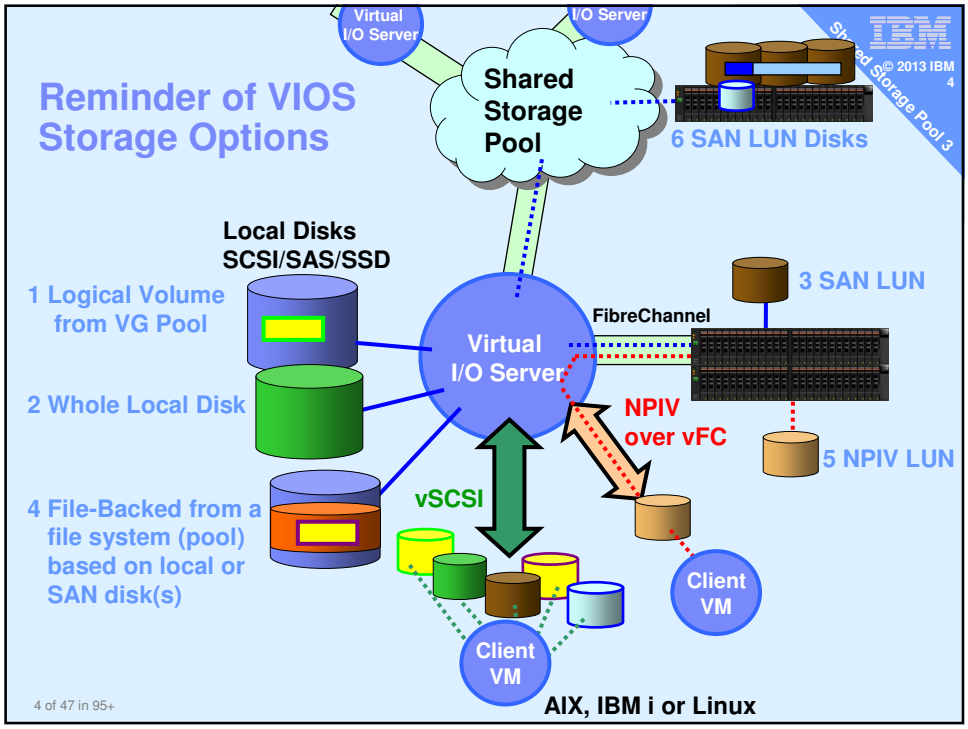
VIOS Shared Storage Pools

Phase 3 (also called SSP3)

→ Q4 2012

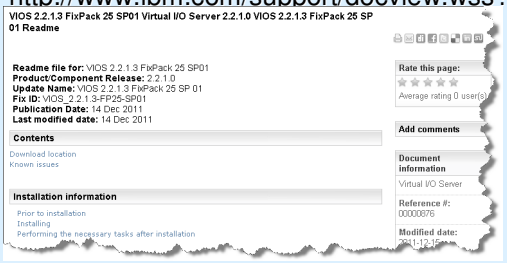
Nigel Griffiths
IBM Power Systems
Advanced Technology Support, Europe





VIO Shared Storage Pool phase 3

- New function delivered with VIOS 2.2.2.1
  now 2.2.2.2 fix
- Please read the Readme notes:
  - http://www.ibm.com/support/docview.wss?uid=hpc1vios117f5701
  - http://www.ibm.com/support/docview.wss?uid=hpc1vios610c6192



- Then read VIOS 2.2.2 User Guide
  - http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/topic/p7hb1/p7hb1.pdf

Shared Storage Pool phase 3 Requirements 1 of 3

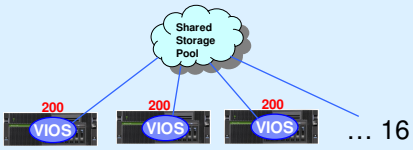
Read the Readme Notes

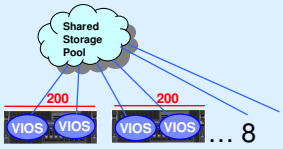
- Platforms: **POWER6** & **POWER7** only (includes **Blades**)
- VIOS Storage Pool (minimums):
  - Direct fibre-channel attached LUNs:
  - **1 for repository ~10 GB (NEW INFO: 1GB is enough) &**
  - **1 or more for data, 10 GB → in practice lots more [like 1TB+]**
- Pool Storage Redundancy: Repository & pool storage must be **RAIDed**
- VIOS **name resolution** to resolve hostnames
- Virtual I/O Server(s):
  - **Minimum CPU: Entitlement=1+, VP=1+** (shared, uncapped is OK)
  - **Minimum Memory: 4+ GB** (no skinny VIOS)

Shared Storage Pool phase 3 Limits

2 of 3

- Max VIOS nodes: **16** (was 4)





	Phase 3	Phase 2
Max physical disks (LUNs) in a pool:	1024	(was 256)
Max virtual disks (LUs) in a cluster:	8192	(was 1024)
Number of Client LPARs per VIOS (or pair)	200	(was 40)
Capacity of Physical Disks in Pool (each)	16TB	(was 4TB)
Storage Capacity of Storage Pool (total)	512TB	(was 128TB)
Capacity of each Virtual Disk (LU) in Pool	1GB to 4TB	(same)
Number of Repository Disks	1 (CAA limit) → new recovery options	(same)

7 of 47 in 95+

Shared Storage Pool phase 3 Limits

2 of 3

- Max VIOS nodes:

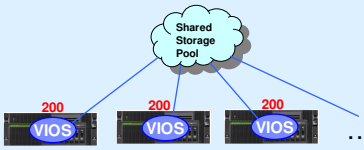
16

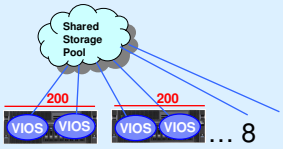
8 of 47 in 95+

Shared Storage Pool phase 3 Limits

2 of 3

- Max VIOS nodes: **16** (was 4)





	Phase 3	Phase 2
Max physical disks (LUNs) in a pool:	1024	(was 256)
Max virtual disks (LUs) in a cluster:	8192	(was 1024)
Number of Client LPARs per VIOS (or pair)	200	(was 40)
Capacity of Physical Disks in Pool (each)	16TB	(was 4TB)
Storage Capacity of Storage Pool (total)	512TB	(was 128TB)
Capacity of each Virtual Disk (LU) in Pool	1GB to 4TB	(same)
Number of Repository Disks	1 (CAA limit) → new recovery options	(same)

Shared Storage Pool phase 3 Restrictions

3 of 3

Network

- Reliable & not congested
- DNS should use local /etc/hosts first
- Forward & reverse lookup must work
- Recommended to synchronise clocks
- SEA must use default threaded mode

Storage

- 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

Lots of restrictions were dropped for this release

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 ?

Reminder

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.

Reminder

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

Reminder

Skipping SSP phase 2 …

- Cluster command for details of the VIOSs
- Setting up Dual VIOS multiple pathing to SSP
- Add a new LUN to grow the pool
- Replace a LUN also allows migration

- Thin Provisioning of disk space at 64 MB chunk
  - Optional regular provisioning with -thick ☺
- Alerting on Pool Space getting LOW !!
  - There is a new Alert for extreme Over-Commit level
- Snapshots
  - Learnt a rollback to earlier snapshot removes later ones

Reminder

Monitoring Disk use with lssp – SSP3

15

```
$ lssp -clustername stars -sp atlantic -bd
Lu Name      Size(mb) ProvisionType %Used Unused(mb) Lu Udid
orange7a     32768    THIN             9% 29615      7d58538152 …
orange7b     32768    THIN             0% 32770      76136907aa …

$ lssp -clustername stars
POOL_NAME:        atlantic
POOL_SIZE:        130944
FREE_SPACE:       125514
TOTAL_LU_SIZE:    65536
OVERCOMMIT_SIZE: 0
TOTAL_LUS:        2
POOL_TYPE:        CLPOOL
POOL_ID:          0000000009893E510000000050740962
```

See who is using most disk & who might run out

Monitoring Disk use with lssp – SSP3

16

```
$ lssp -clustername stars -sp atlantic -bd
Lu Name      Size(mb) ProvisionType %Used Unused(mb) Lu Udid
orange7a     32768    THIN             9% 29615      7d58538152 …
orange7b     32768    THIN             0% 32770      76136907aa …

$ lssp -clustername stars
POOL_NAME:        atlantic
POOL_SIZE:        130944
FREE_SPACE:       125514
TOTAL_LU_SIZE:    65536
OVERCOMMIT_SIZE: 0
TOTAL_LUS:        2
POOL_TYPE:        CLPOOL
POOL_ID:          0000000009893E510000000050740962
```

See who is using most disk & who might run out

Over-commit
Good to know the “worst case”

Four User Interface

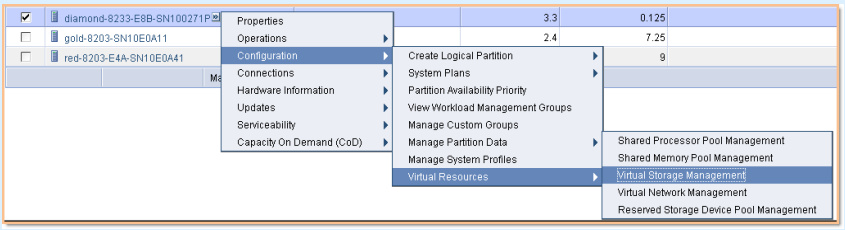
- VIOS Command Line (CLI)
  - Already shown in this presentation by example
  - Some feature make sense via CLI → like: cluster –create/-addnode
- VIOS cfgassist menu
  - This is the VIOS version of smitty
  - Menu driven interface for CLI
  - Fully covers SSP functions
- HMC Graphical User Interface (HMC 7.7.4+)
  - SSP virtual disk list, create and connect to your VMs
    - Note: don't have any VIOS vSCSI slots in "Any node" = bad practice anyway
  - See screen shots …
- System Director – **New in SSP3**
  - 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 Linked-Cloning

Reminder

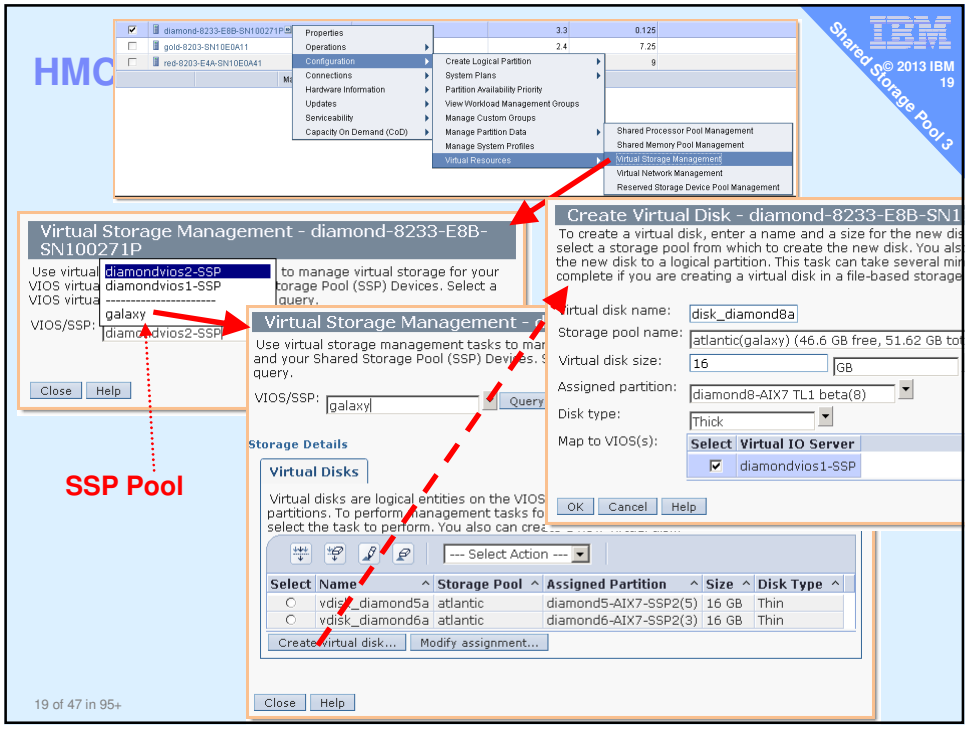
New

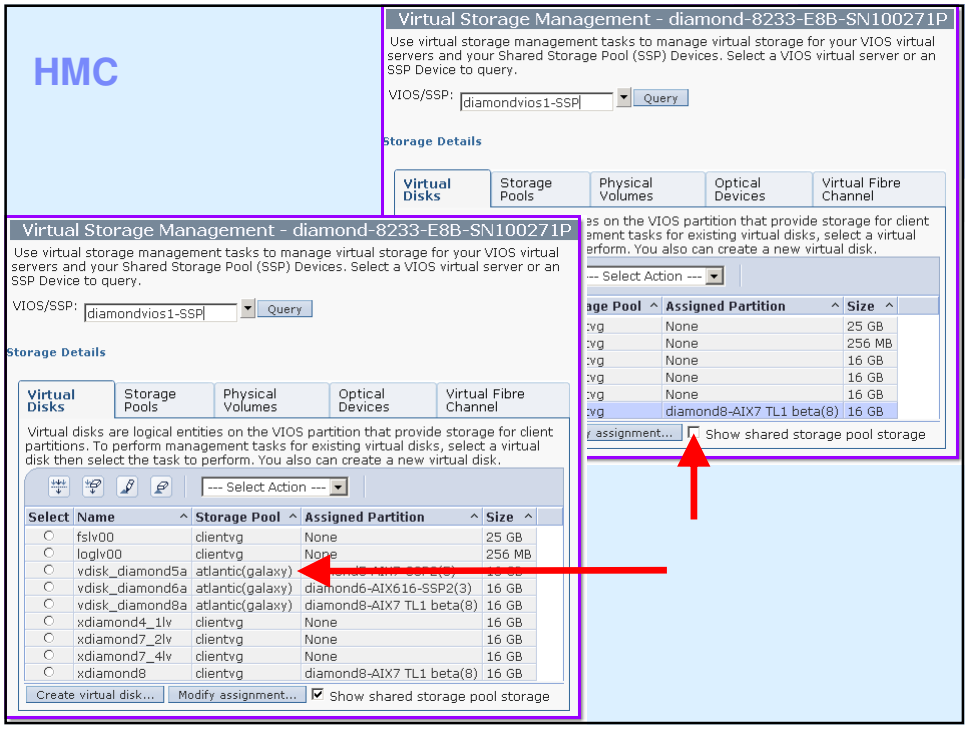
Hardware Management Console

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



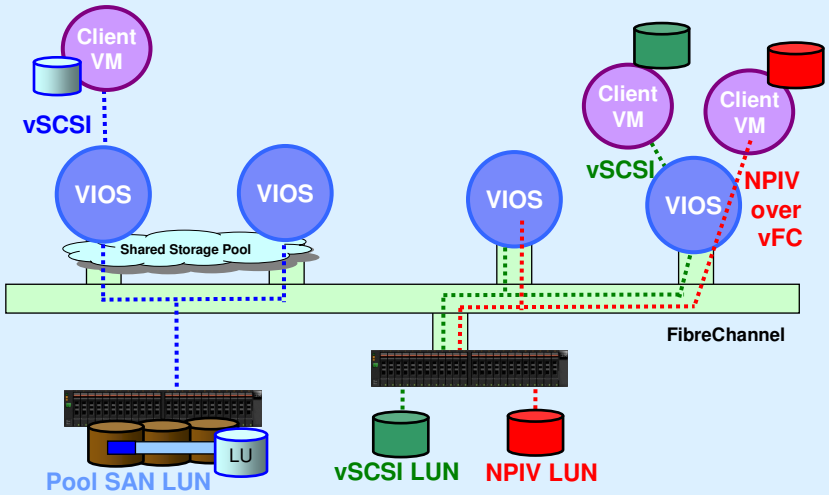
- Shipped with HMC upgrade V7 R7.4 SP0+
  - Not part of the VIOS package



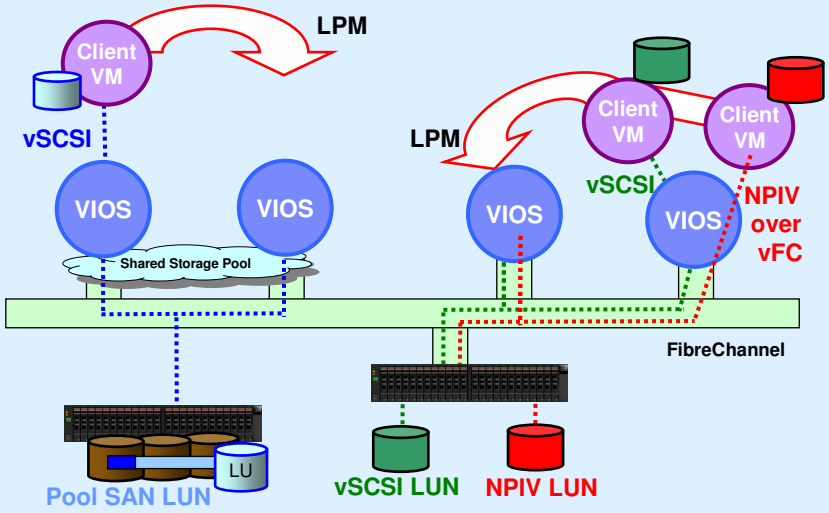


SSP3 in practice

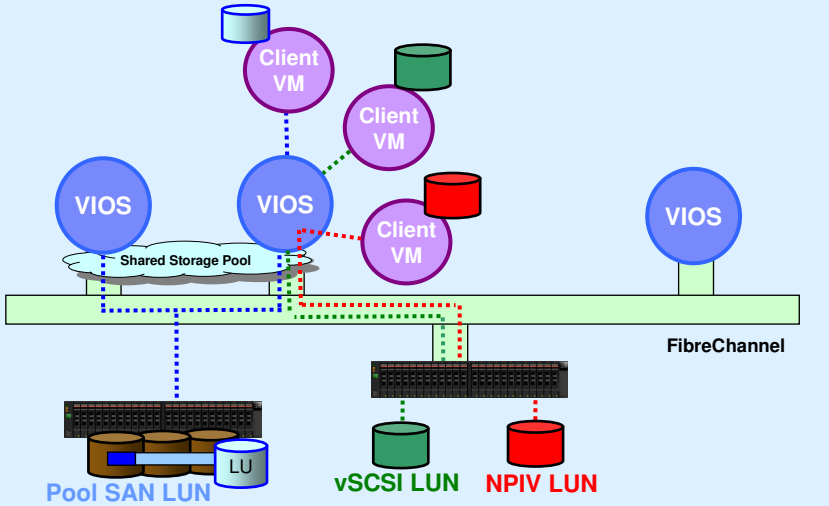
Mixed Environment SSP + vSCSI + NPIV



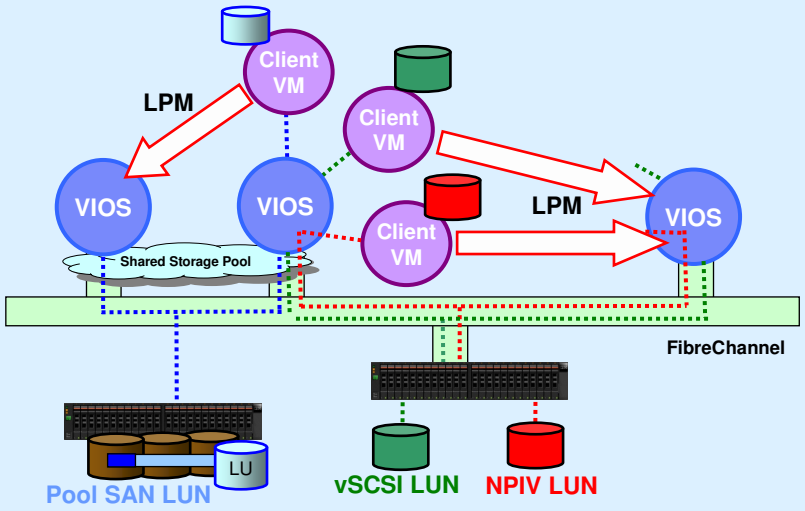
LPM SSP→SSP or vSCSI→vSCSI or NPIV→NPIV



Can mix on the same VIOS



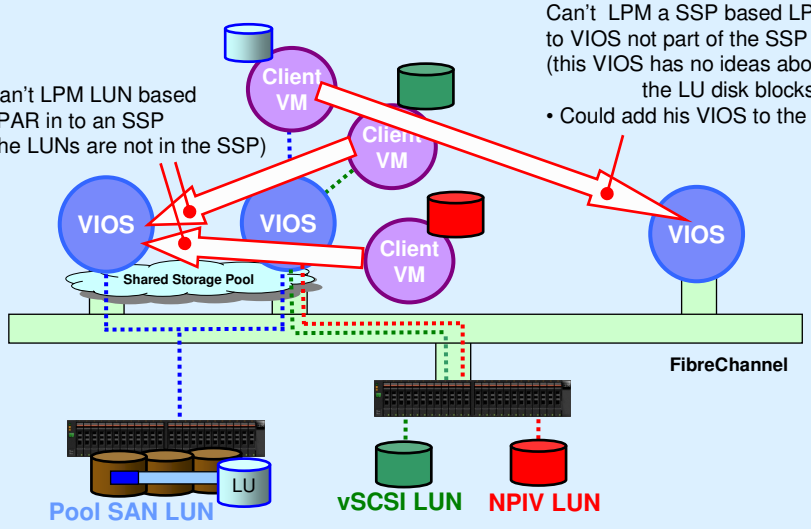
Can LPM



Can't LPM → if it means a different disk layout

Can't LPM LUN based LPAR in to an SSP (the LUNs are not in the SSP)

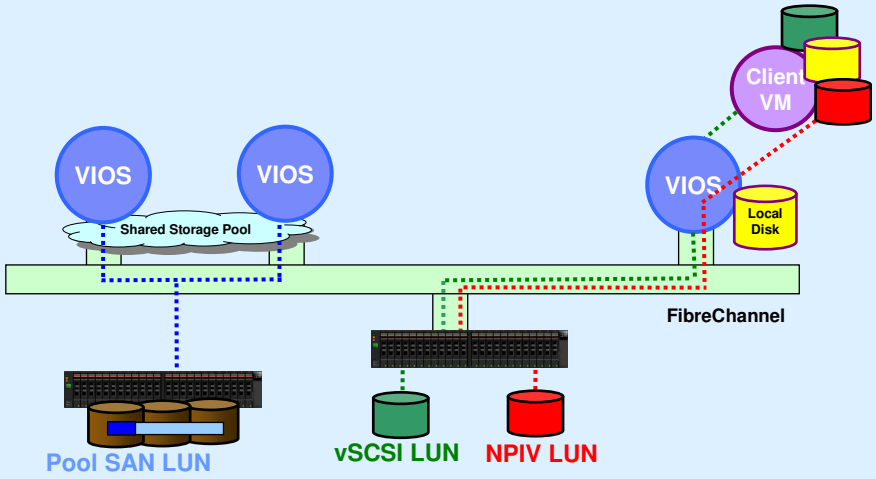
Can't LPM a SSP based LPAR to VIOS not part of the SSP pool (this VIOS has no ideas about the LU disk blocks)
- Could add his VIOS to the SSP 1st



Adopt an AIX LPAR

Shared Storage Pool 3

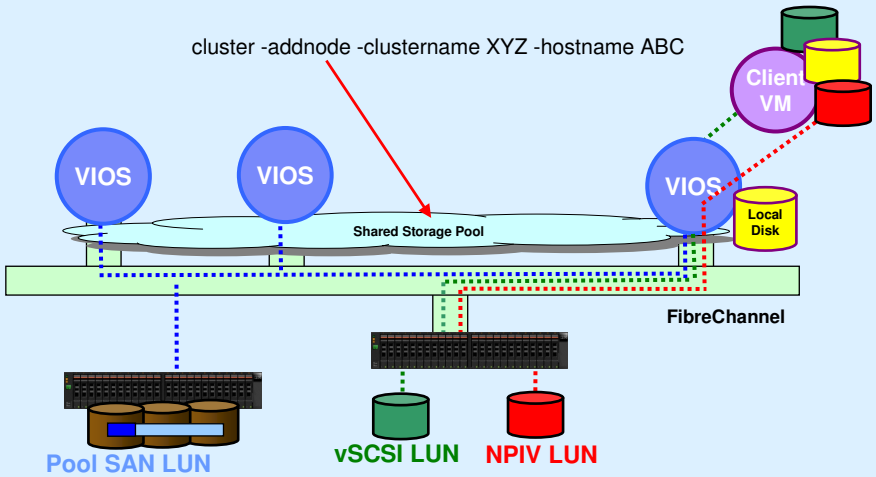
Want to move your client VM to use the SSP!



Adopt an AIX LPAR

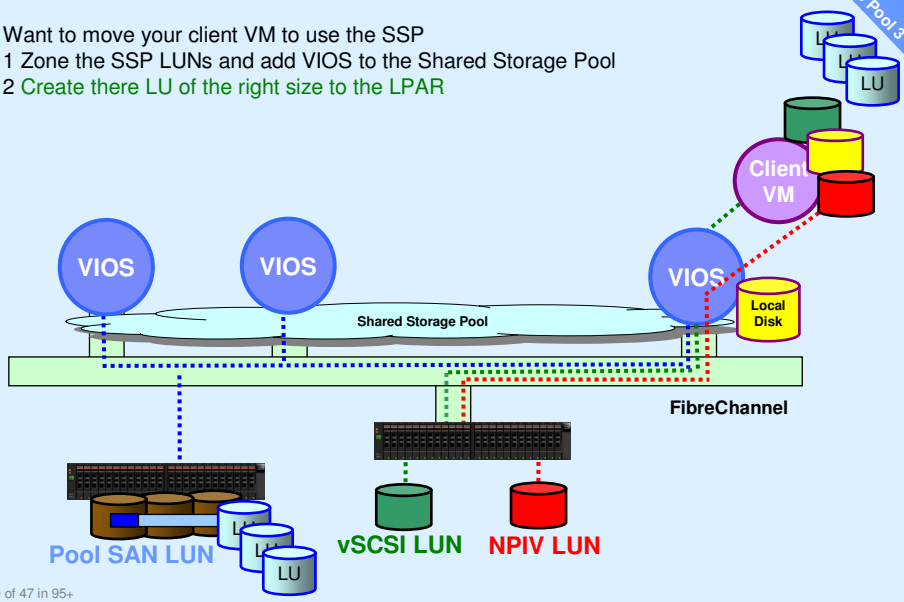
Shared Storage Pool 3

Want to move your client VM to use the SSP
1 Zone the SSP LUNs and add VIOS to the Shared Storage Pool



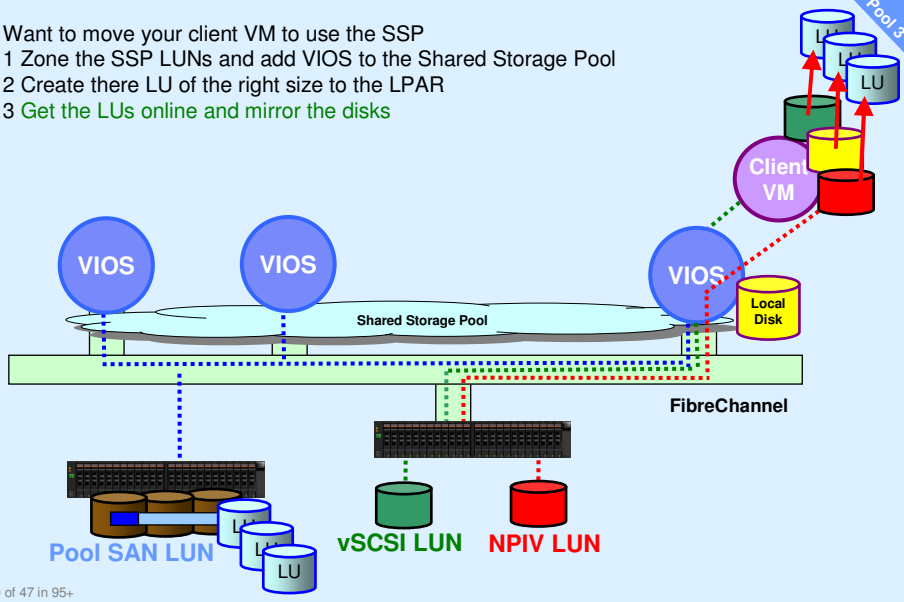
Adopt an AIX LPAR

Want to move your client VM to use the SSP
1 Zone the SSP LUNs and add VIOS to the Shared Storage Pool
2 Create there LU of the right size to the LPAR



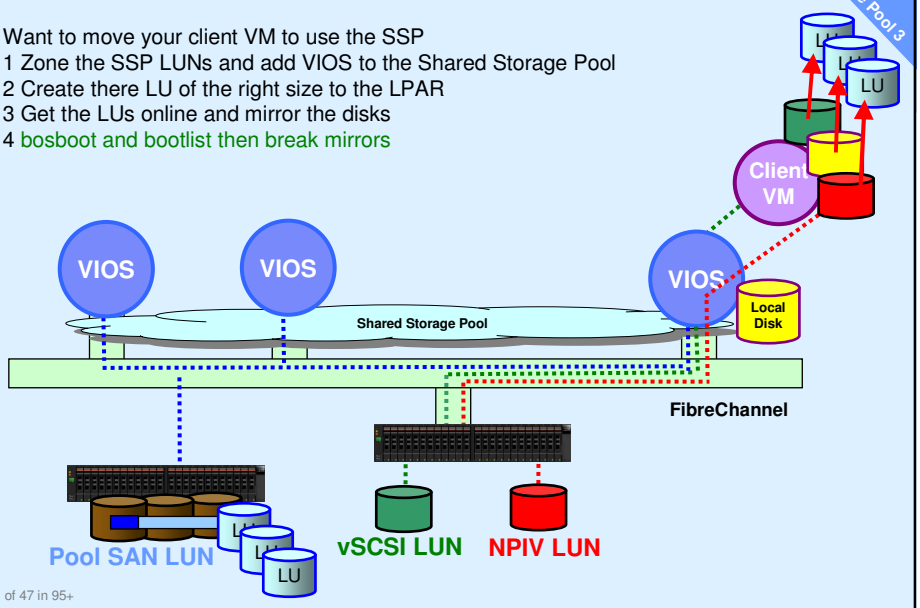
Adopt an AIX LPAR

Want to move your client VM to use the SSP
1 Zone the SSP LUNs and add VIOS to the Shared Storage Pool
2 Create there LU of the right size to the LPAR
3 Get the LUs online and mirror the disks



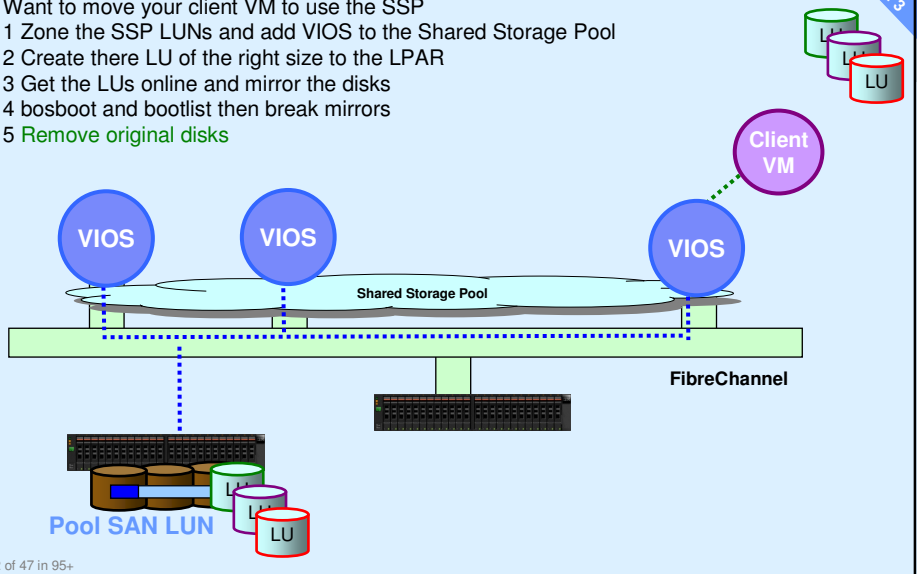
Adopt an AIX LPAR

Want to move your client VM to use the SSP
1 Zone the SSP LUNs and add VIOS to the Shared Storage Pool
2 Create there LU of the right size to the LPAR
3 Get the LUs online and mirror the disks
4 bosboot and bootlist then break mirrors



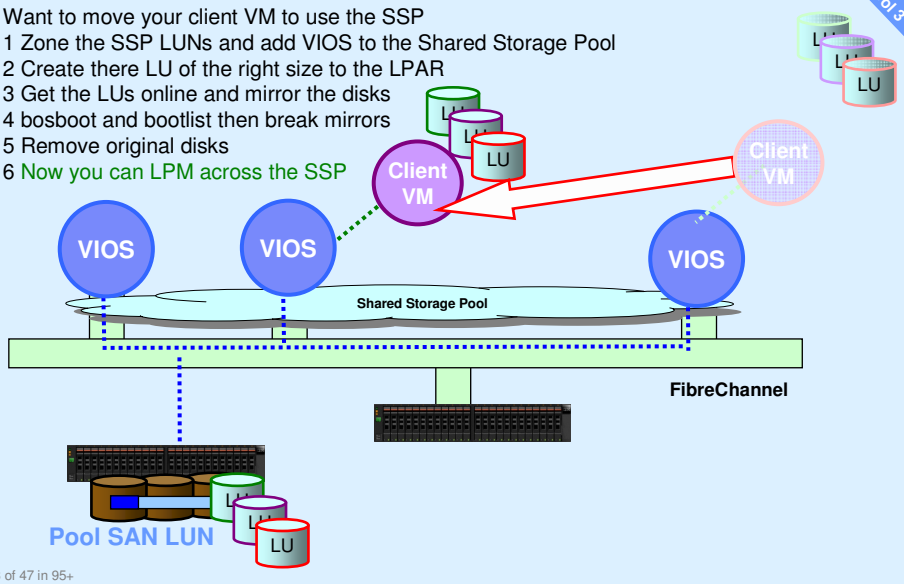
Adopt an AIX LPAR

Want to move your client VM to use the SSP
1 Zone the SSP LUNs and add VIOS to the Shared Storage Pool
2 Create there LU of the right size to the LPAR
3 Get the LUs online and mirror the disks
4 bosboot and bootlist then break mirrors
5 Remove original disks



Adopt an AIX LPAR

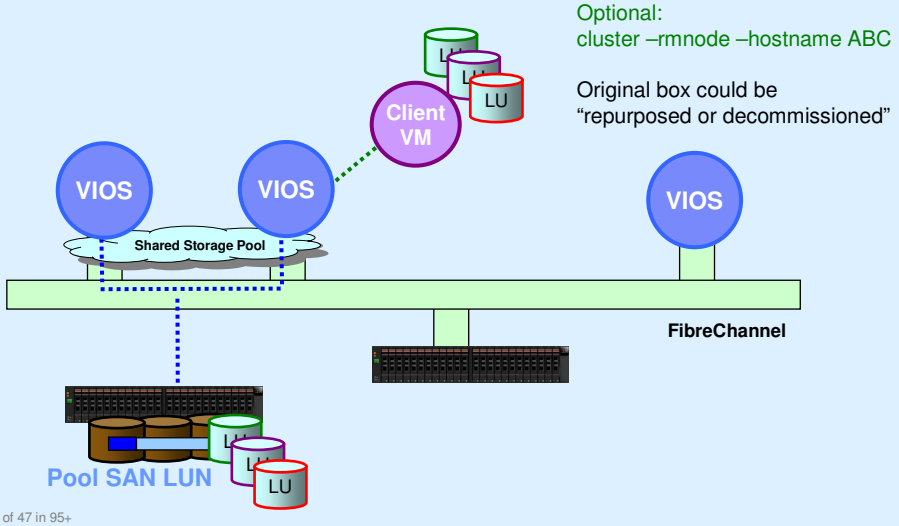
Want to move your client VM to use the SSP
1 Zone the SSP LUNs and add VIOS to the Shared Storage Pool
2 Create there LU of the right size to the LPAR
3 Get the LUs online and mirror the disks
4 bosboot and bootlist then break mirrors
5 Remove original disks
6 Now you can LPM across the SSP



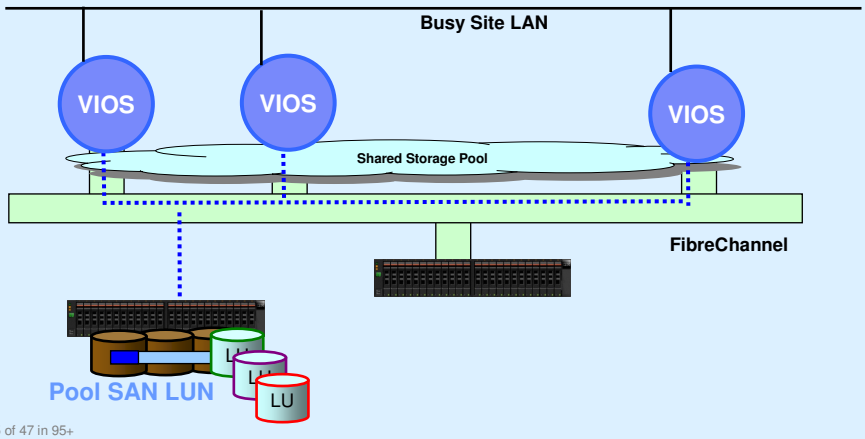
Adopt an AIX LPAR = "You will be assimilated!" The Borg.

Optional:
cluster –rmnode –hostname ABC

Original box could be "repurposed or decommissioned"

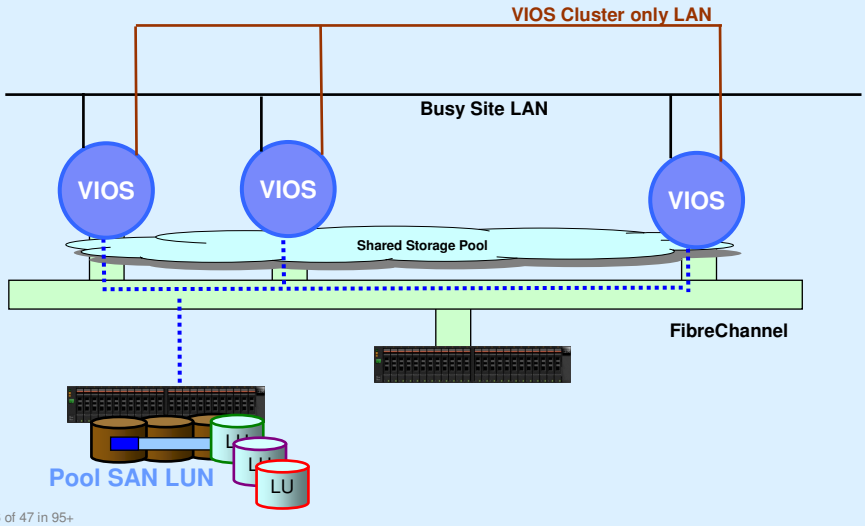


Dedicated VIOS Cluster network



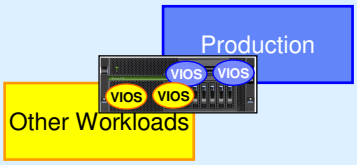
Dedicated VIOS Cluster network

Setup network route to the other VIOS to use this new network interface
Not really a VIOS SSP feature but now tested & supported



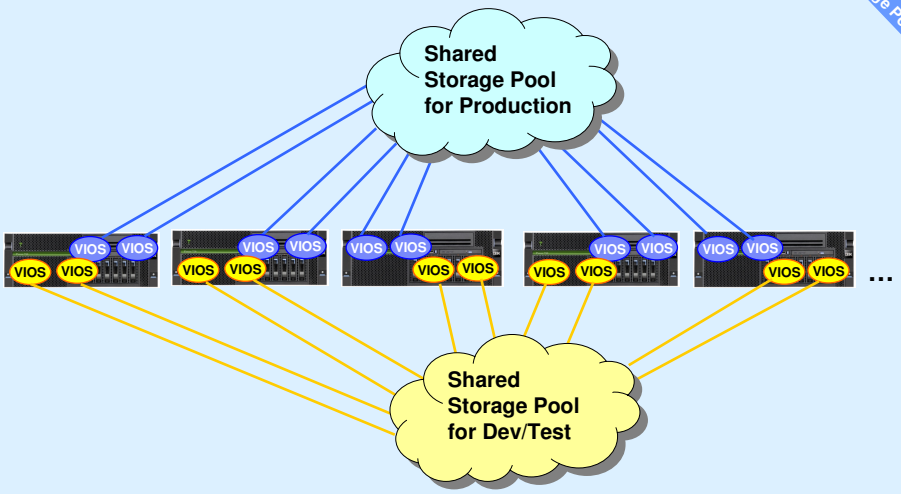
Production Separation

Popular configuration is:
dual VIOS for Production
dual VIOS for Other workloads

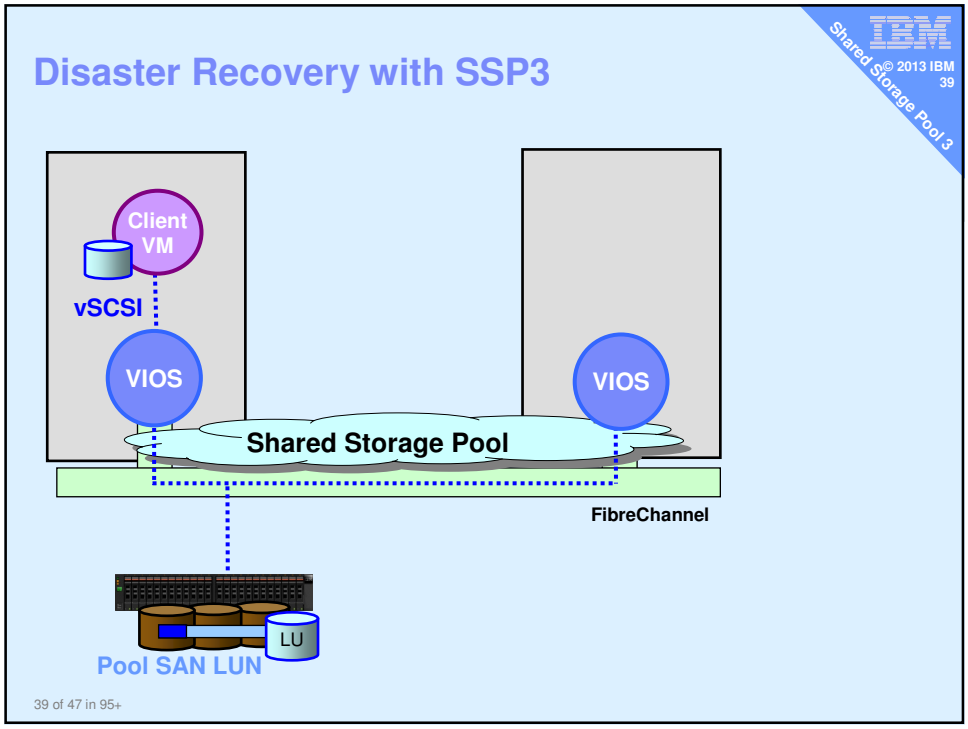


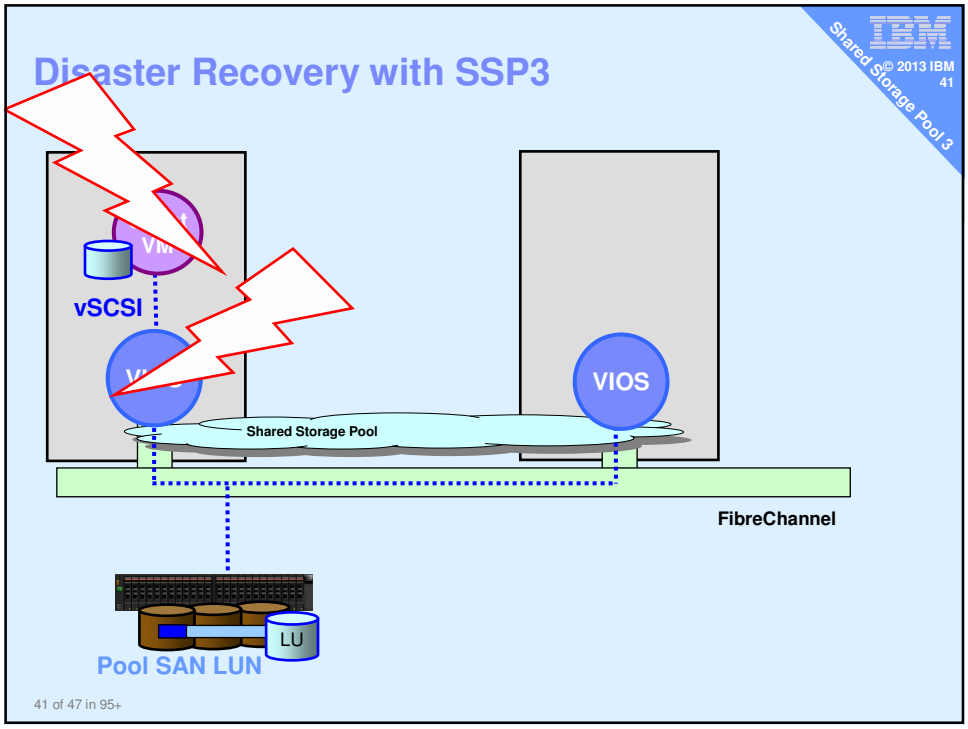
How can we add Shared Storage Pool and keep them separate ?

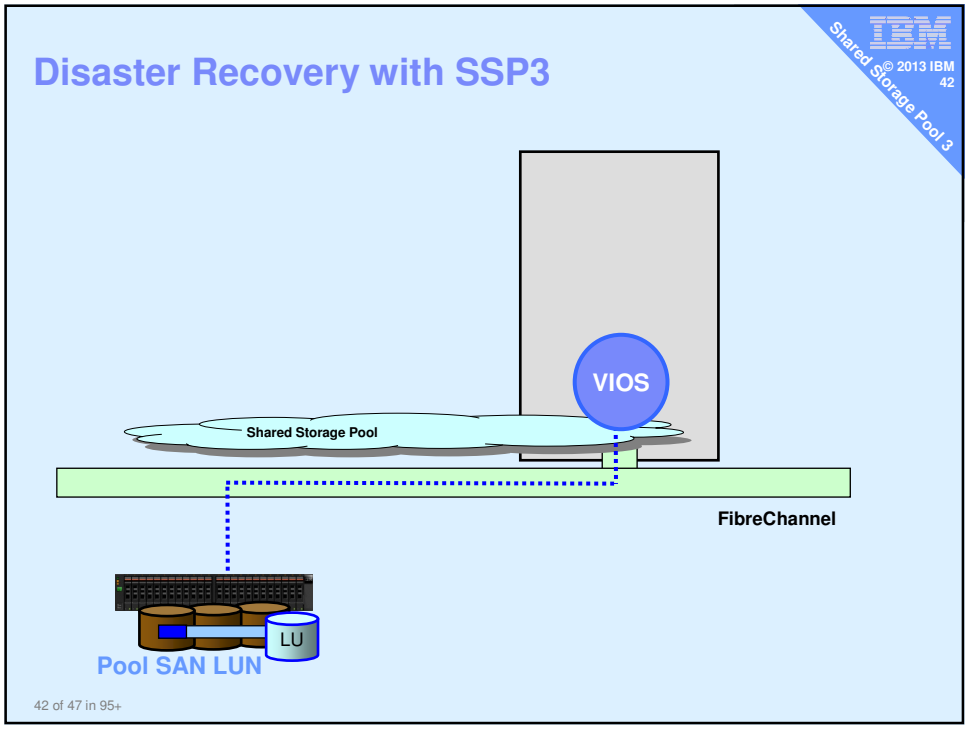
One machine on multiple pools

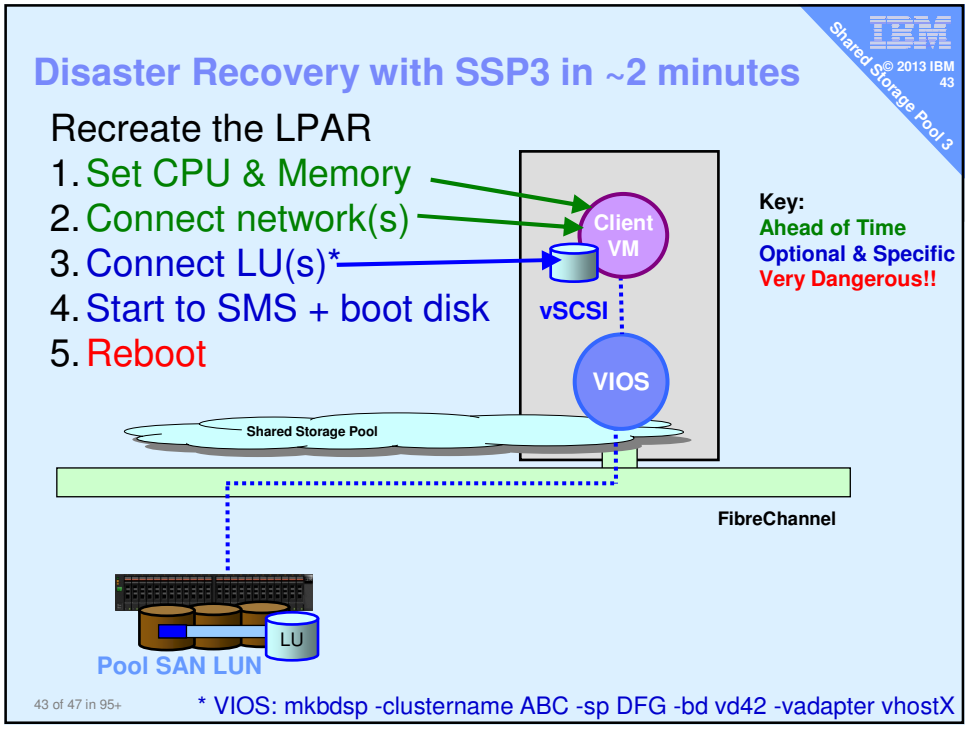


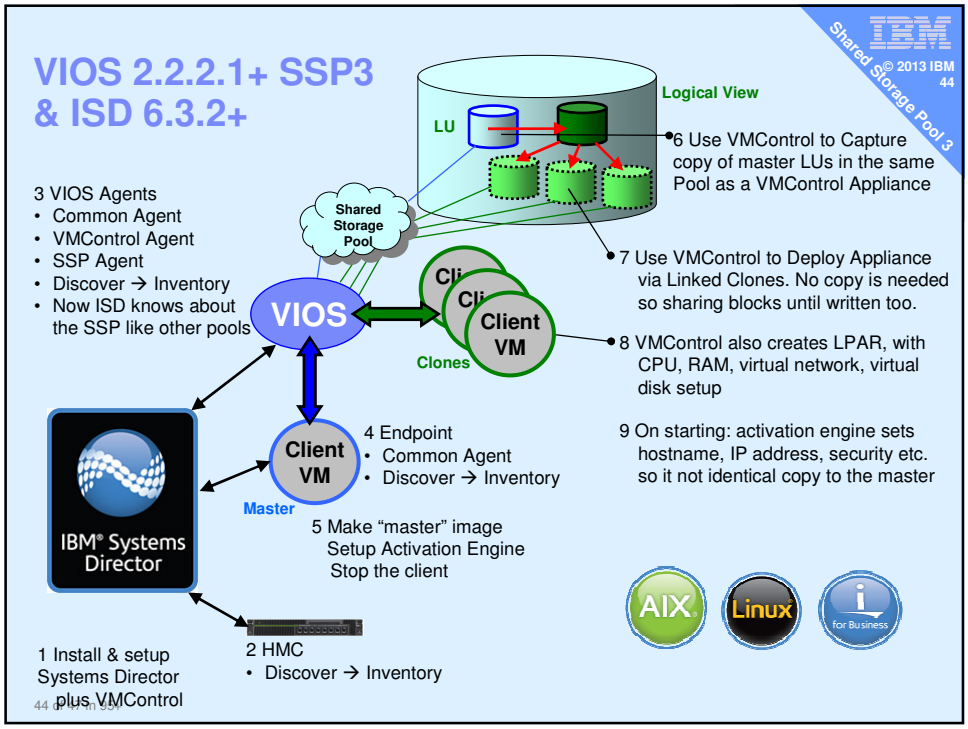
Note: Any one VIOS can only be part of only one SSP cluster
but different VIOS's can be on difference SSP clusters











Shared Storage Pool phase 3 – 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. Decide your VIOS cluster architecture?
3. Get to VIOS 2.2.2.2 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

Want more on Shared Storage Pools?

Five videos to Get You Started Today

- AIXpert blog covering these videos
  - http://www.ibm.com/developerworks/community/blogs/aixpert/entry/shared_storage_pools_five_videos_to_get_you_started_today

1. Overview Presentation on what's new.
   - Shared Storage Pool phase 3 (SSP3) New Features
2. Get a LPAR on local disk to SSP in 5 minutes and then LPM to a new machine in 1 minute
   - Migrating to Shared Storage Pool (SSP3) & then LPM
3. Not everything needs HACMP but it would be nice to get those LPAR running again quickly
   - SSP3 Recover a Crashed Machine's LPAR to Another Machine
4. Load balance across the computer room, evacuate a machine for maintenance & use a new machine on day 1
   - Live Partition Mobility (LPM) with Shared Storage Pool SSP3 -
5. See how easy it is to understand and operate SSP

 Looking around a Shared Storage pool via commands and HMC -

SSP2 command cheat sheet Reference Only


```
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 commands
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 -type threashold -value 10
16. alert -list -clustername galaxy -spname atlantic
17. errlog -ls
18. snapshot -create name   -clustername galaxy -spname atlantic -lu LU42
19. snapshot -delete name   -clustername galaxy -spname atlantic -lu LU42
20. snapshot -rollback name -clustername galaxy -spname atlantic -lu LU42
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  …
```


Questions

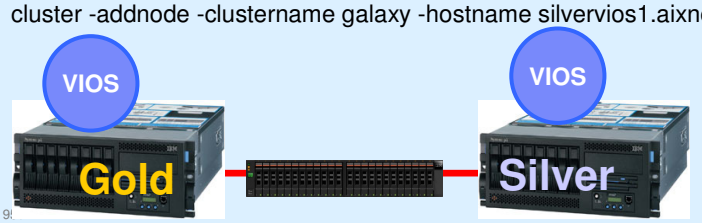
The Shared Storage Pools future

- Phase 1 Q4 2010
- Phase 2 Q4 2011
- Phase 3 Q4 2012
- Phase 4 Q? 201?

- 3 new features listen carefully …

Demo part 1 - Commands

1. Look around my cluster: cluster, lssp, lscluster
   1. `cluster -list`
   2. `cluster -status -clustername galaxy`
   3. `lssp -clustername galaxy`
   4. `lssp -clustername galaxy -sp atlantic -bd`
   5. `lscluster -c (nodes) -i (network) -d (disk)`
   6. `lscluster -m (more!) -s (local stats)`
2. Grow the cluster
   1. Add a node to the galaxy cluster called: silvervios1
   2. Show LUNs PVID's on silvervios1 with: lspv –size
   3. cluster -addnode -clustername galaxy -hostname silvervios1.aixncc.uk.ibm.com



Demo part 2 - Add or Setup LU disks FAST

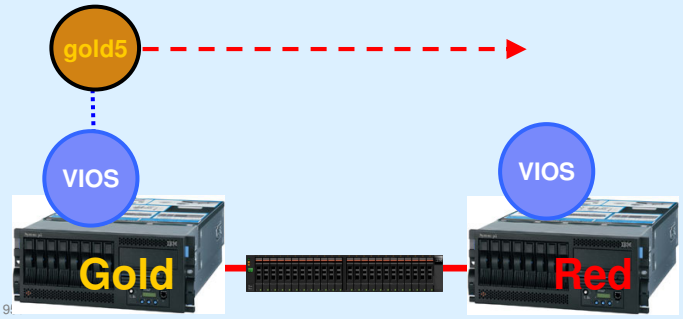
3. Add new SSP disk to AIX, cfgmgr, find new disk
   - gold6 add vdisk_gold6c, AIX cfgmgr x2
   - Check lspv on gold6
   - HMC Gold box → Config → Virtual → Storage→ Create disk
   - CLI add disk

     mkbdsp -clustername galaxy -sp atlantic 16G
     -bd vdisk_gold6d -vadapter vhost5

4. Newly created LPAR
   - HMC add SSP + virtual Optical gold4
   - Gold4 has only CPU, RAM and vSCSI = no disks
     - goldvios: lsmap -all and check C14
   - Add SSP disk(s) and attach virtual optical
   - Start LPAR and console, install AIX

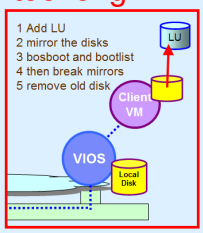
Demo part 3 - LPM

5. Live Partition Mobility (LPM)
   - Run worms
   - Start LPM (gold5 ←→ red)
   - Via Gold5 → Ops → Mobility →
     Validate check then Migrate

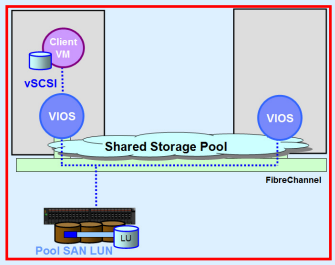


Demo part 4 – Move to SSP and DR

5. You will be assimilated - mirroring takes too long
   1. Internal disks,
   2. add SSP disks bronze2
   3. mirrorvg rootvg
   4. Sync the mirrors,
   5. bosboot/bootlist,
   6. break mirror,
   7. remove disks then LPM!



6. LPAR Disaster Recovery (SSP3 resurrection)
   - gold2 to bronze machine GOLD2reborn
   - Script to write to file
   - gold2: halt -q
   - Get to bronzevio1 (new host)
   - Find gold2 SSP disks
   - Attach them to new LPAR HMC GUI or CLI
     mkbdsp -clustername galaxy -sp atlantic
     -bd vdisk_gold2a -vadapter vhost5
   - Start LPAR in SMS to select a boot list

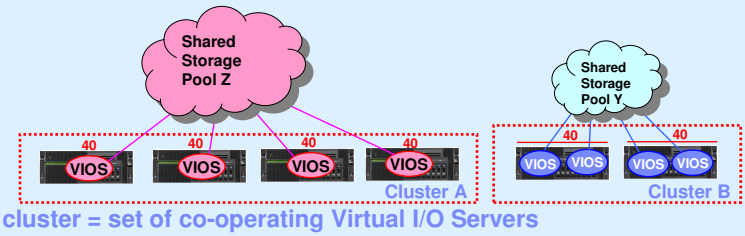


BACKUP – 50 more slides

Topics
- Clusters
- Snapshot
- Working with Dual VIOS
- Thin Provisioning
- Alerts when the pool is nearly empty
- Storage mobility and repository recovery
- SSP commands reminder sheet

- Demo details – now 4 major movies available

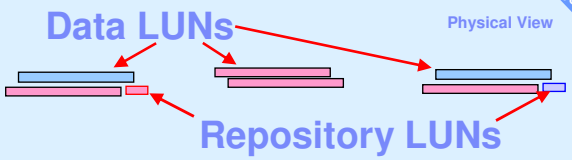
Terms Shared Storage Pool 3



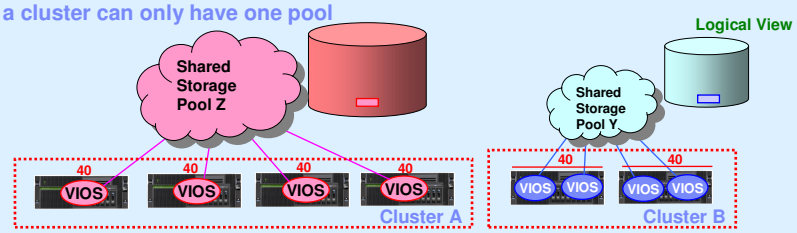
1 SSP2 cluster = set of co-operating Virtual I/O Servers
The VIOS can only be in one cluster

Here we show two clusters

Terms Shared Storage Pool 3



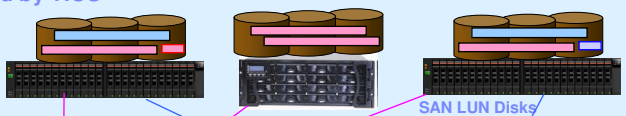
2 SSP2 pool = set of LUNs
Data LUNs + a special Repository LUN = cluster meta-data
Currently a cluster can only have one pool



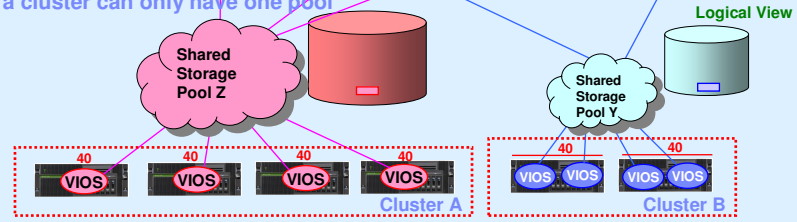
1 SSP2 cluster = set of co-operating Virtual I/O Servers
Currently a VIOS can only be in one cluster

Terms Shared Storage Pool 3

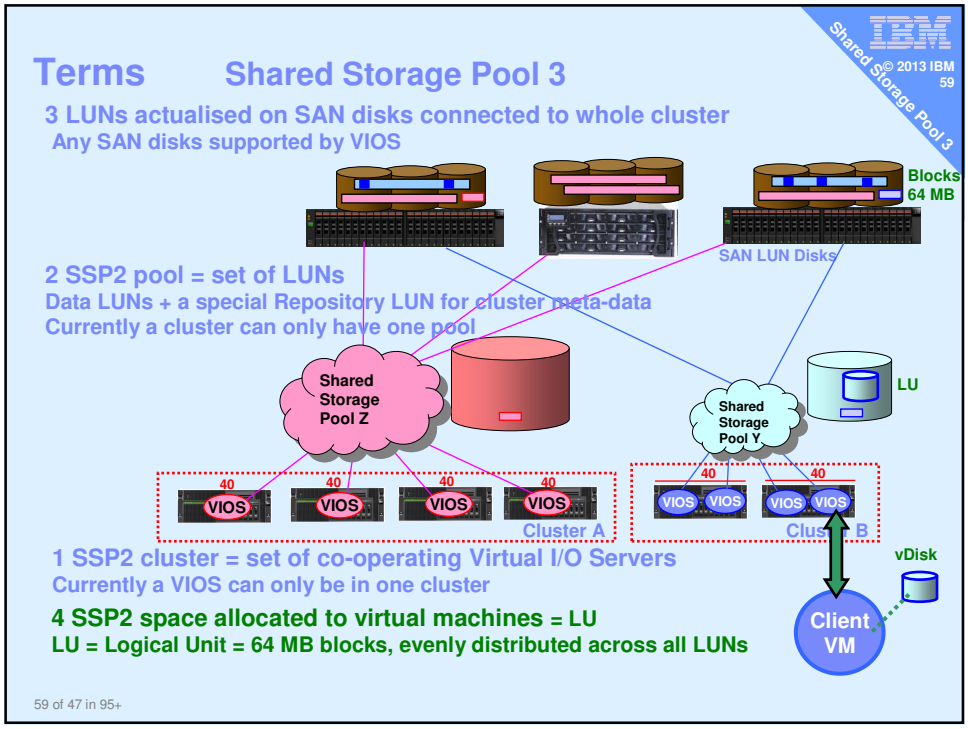
3 LUNs actualised on SAN disks connected to whole cluster
Any SAN disks supported by VIOS

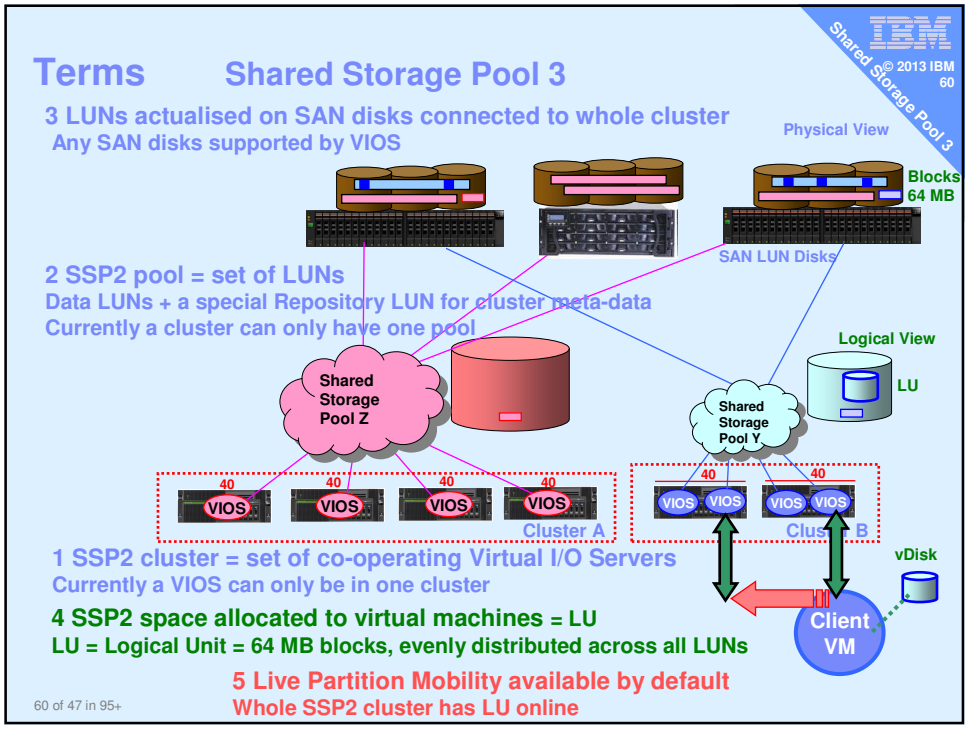


2 SSP2 pool = set of LUNs
Data LUNs + a special Repository LUN for cluster meta-data
Currently a cluster can only have one pool



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





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 to make the
  low level disk attribute change

List cluster & cluster nodes


```
$ cluster -list
CLUSTER_NAME:     galaxy
CLUSTER_ID:       64517962b01c11e1ac6aba367e934e03
$
$ 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 client VMs
- Remove 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
```


Pool Disk Space Management

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

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 2 steps (create LU then assign it)

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 "LU42" 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

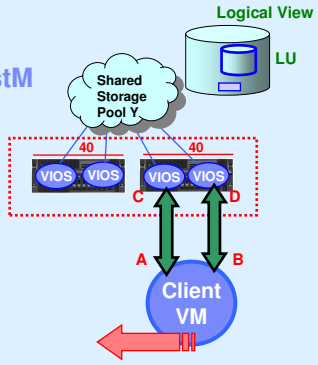
3 on VIOS D: use lsmap -all to map slot to vhostM
- mkbdsp -clustername galaxy -sp atlantic
  -bd vdisk_red6a -vadapter vhostM

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

By name

```
$ rmbdsp -clustername galaxy -sp atlantic
  -bd vdisk_diamond6a
```

By LU hexadecimal id

```
$ rmbdsp -clustername galaxy -sp atlantic
  -luudid 858152297879adfe0d75b05f586d36ee
```

House keeping

Add more physical LUNs to the Pool

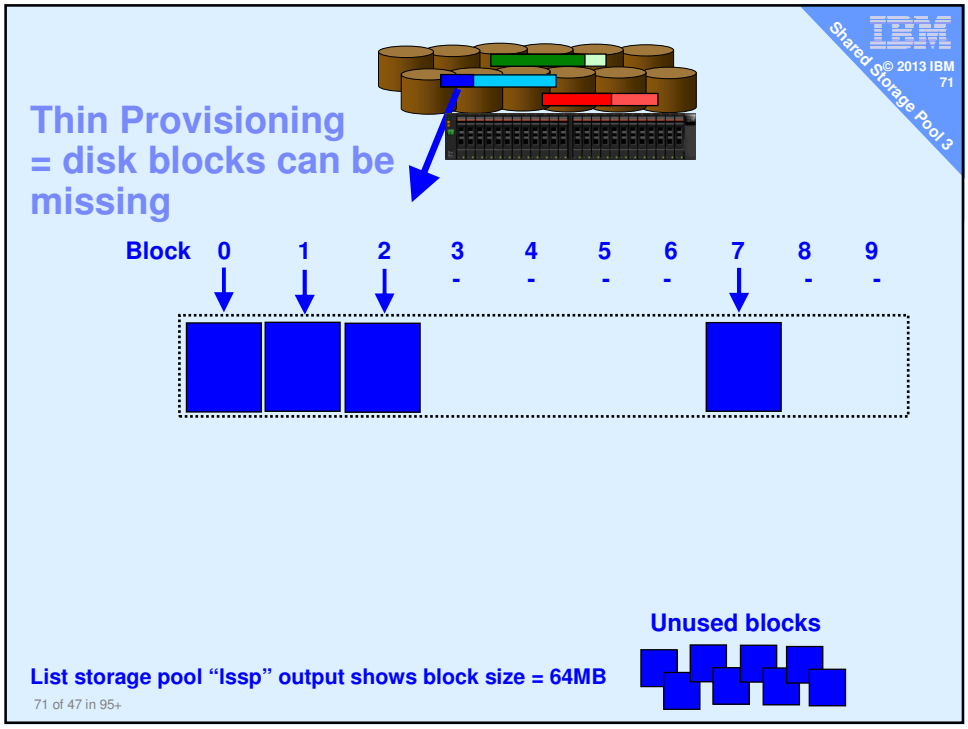
```
$ chsp -add -clustername galaxy -sp atlantic
  hdisk8
```

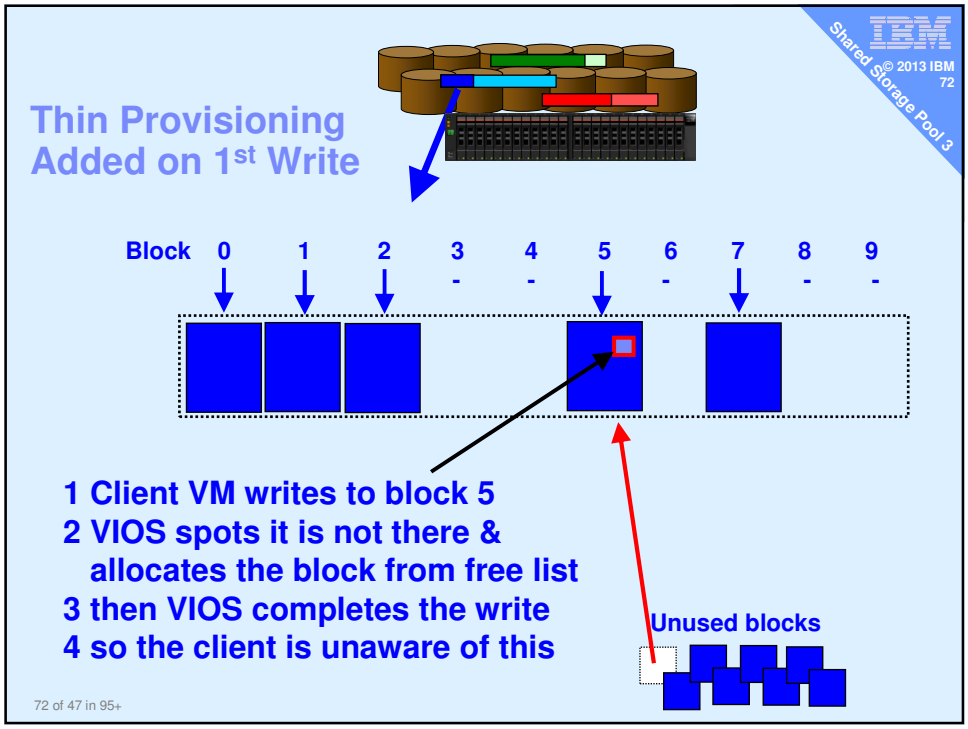
Remove a physical LUN from the Pool
- You can not …. with this release
- We can replace a disk but not remove one
  - Replacement disk - equal or larger size

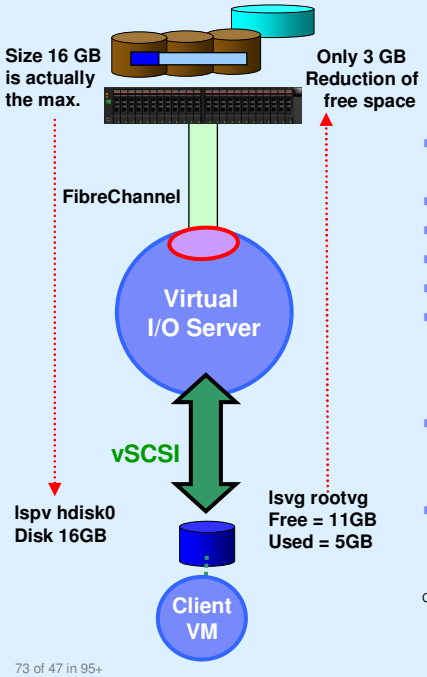
69 of 47 in 95+

Experiments in Thin provisioning

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







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”

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

- No problem, 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"

Shared Storage Pool 3

```
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 (was in the lssp output)

Monitoring Disk use with lssp – SSP2

Shared Storage Pool 3

```
$ 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

Monitoring Disk use with lssp – SSP3


```
$ lssp -clustername stars -sp atlantic -bd
Lu Name      Size(mb) ProvisionType %Used Unused(mb) Lu Udid
orange7a     32768    THIN             9% 29615      7d58538152 …
orange7b     32768    THIN             0% 32770      76136907aa …

$ lssp -clustername stars
POOL_NAME:        atlantic
POOL_SIZE:        130944
FREE_SPACE:       125514
TOTAL_LU_SIZE:    65536
OVERCOMMIT_SIZE:  0
TOTAL_LUS:        2
POOL_TYPE:        CLPOOL
POOL_ID:          0000000009893E510000000050740962
```

See who is using most disk & who might run out

Over-commit
Good to know the "worst case"

Thin provisioning risks running out of space → 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

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

```
$ alert -list -clustername galaxy -spname atlantic
PoolName:          atlantic
PoolID:            0000000009893EDD000000004F174D22
ThresholdPercent: 35       ←these are the defaults
OverCommitPercent: n/a
```

- Set alerts to warn on free pool space getting too low %
  - **alert -set -clustername** *galaxy* **-spname** *atlantic* **-type threshold -value 10**
- Set alerts to warn on overcommit getting too large %
  - **alert -set -clustername** *galaxy* **-spname** *atlantic* **-type overcommit -value 30**
- To remove the alert:
  - **alert -unset -clustername** *galaxy* **-spname** *atlantic* **-type threshold**
  - Threshold is set to 0 (zero) – it will not happen!!

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
...
```

  - $ errlog –ls | more

```
See example on the next page
```

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

$ 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
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**. (4) (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

81 of 47 in 95+

81

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" ops@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
  - Look and check the larger VM realy need the space.

82 of 47 in 95+

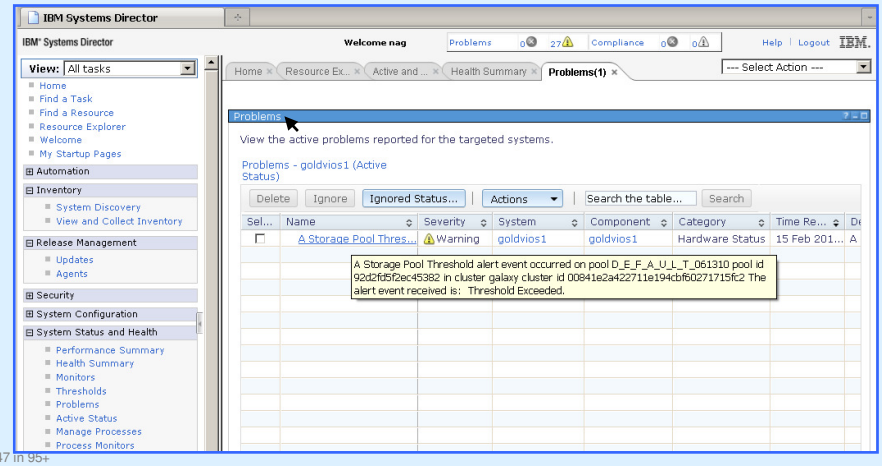
82

House keeping - Thin Provisioning Alerts

Shared Storage Pool 3

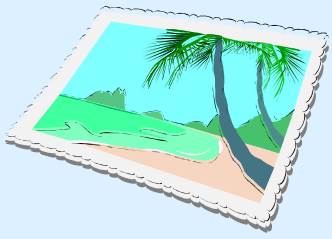
IBM Systems Director captures these events

- Discover, access and Inventory all VIOS.



Snapshot

Shared Storage Pool 3



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 . . .
```


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

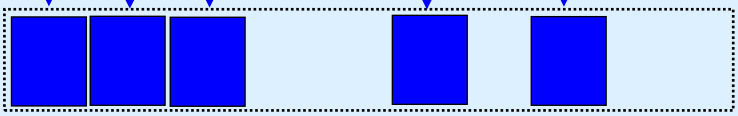
```
$ snapshot -clustername galaxy -rollback diamond5t.snap
  -spname atlantic -lu vdisk_diamond5a
```

Warning:
You loose any updates you made since that snapshot
Any snapshots since that snapshot are removed

Snapshot Model

Original Set

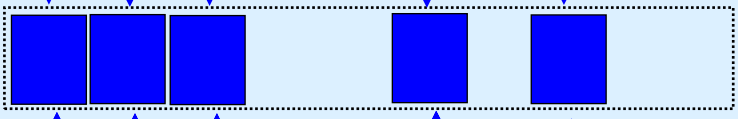
Block	0	1	2	3	4	5	6	7	8	9
				-	-		-		-	-



Snapshot Model

Original Set becomes the Snapshot

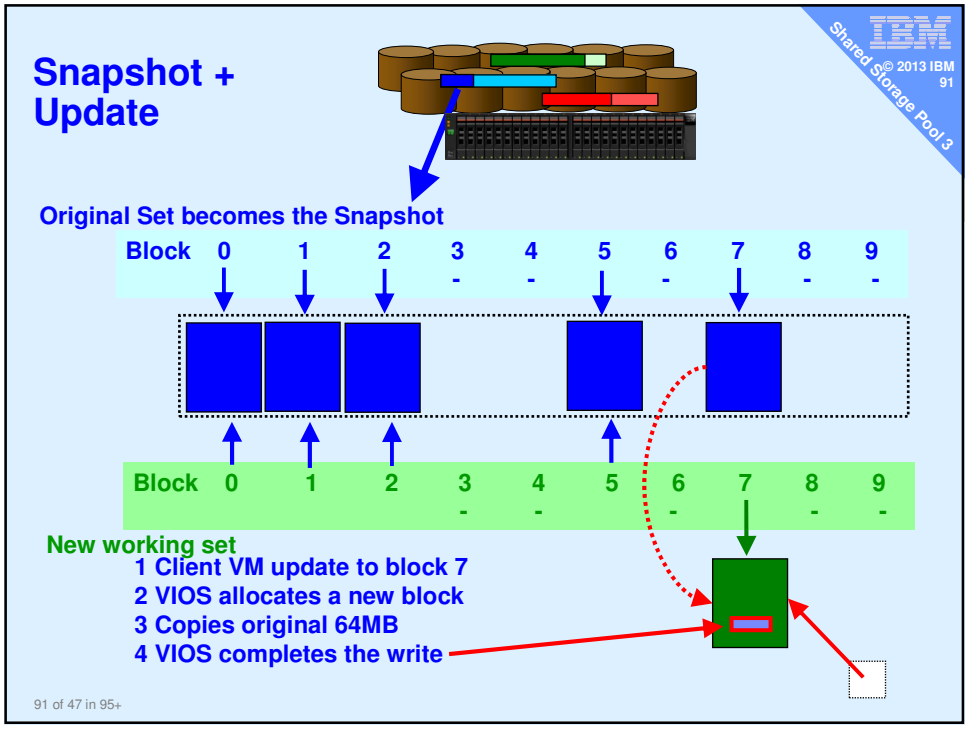
Block	0	1	2	3	4	5	6	7	8	9
				-	-		-		-	-

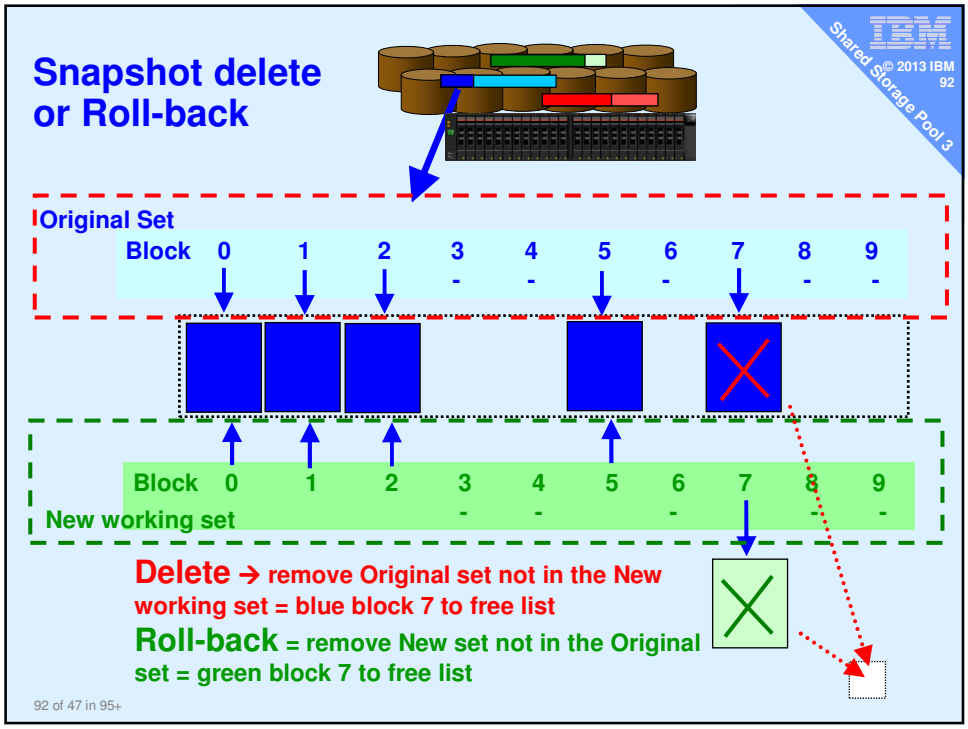


Block	0	1	2	3	4	5	6	7	8	9
				-	-		-		-	-

New working set

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





Storage Management

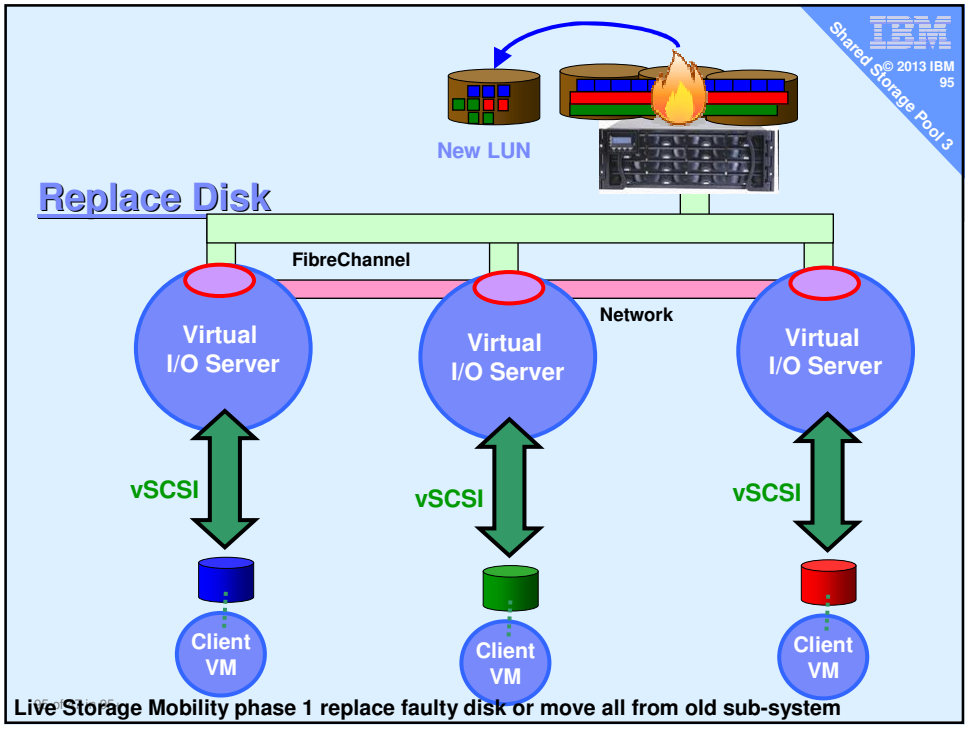
- Reminder currently,
  - One pool of large LUNs
  - Pool can be on a mix of brands or generations of disk sub-systems
  - 64 MB chunks are spread as evenly as possible across LUNs

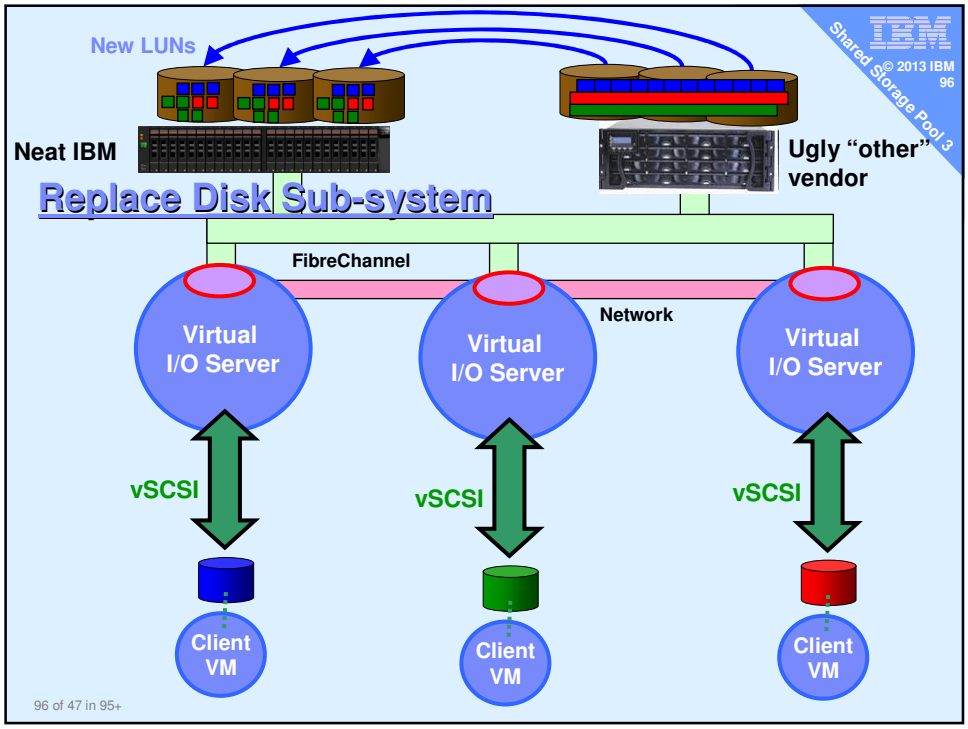
Live Storage Mobility

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





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