VIOS 2.2.4
Shared Storage Pools
Phase 5 with Tiers
and loads more

Marketing: VIOS Shared Storage Pools
1. Enormous reduction in storage man-power
2. Independence from underlying SAN technology & team!
3. Sub-second disk space allocate & connect
   – lu command: create, map, unmap, remove
   – snapshot: create/delete/rollback
4. Autonomic disk mirrors & resilver with zero VM effort
5. Live Partition Mobility ready by default
6. Simple Pool management:
   pv & failgrp, lssp, alert, VIOS logs
7. DR capability to rebuild a VM quickly on a Remote Pool Copy
8. HMC GUI for fast SSP disk setup across dual VIOS
   – No more: VIOS slot numbers, Cnn or vhosts
YouTube Videos

1. Shared Storage Pool (SSP) Intro 17 mins
2. Shared Storage Pools (SSP2) Getting Started 10 mins
3. Shared Storage Pools (SSP2) Thin Provisioning Alerts 17 mins
4. Shared Storage Pools (SSP3) New Features 24 mins
5. Looking Around a Shared Storage Pool SSP3 15 mins
6. Live Partition Mobility (LPM) with Shared Storage Pool SSP3 7 mins
7. SSP3 Recover a Crashed Machine’s LPAR to Another Machine 25 mins
8. Migrating to Shared Storage Pool (SSP3) & then LPM 18 mins
10. Shared Storage Pools 4 (SSP4) Hands On 19 mins
11. PowerVC 1.2.1 with Shared Storage Pools 20 mins
12. Shared Storage Pool in 3 Commands in 3 Minutes 8 mins
13. Shared Storage Pools Repository is bullet proof 13 mins
14. Shared Storage Pool Remote Pool Copy Activation for Disaster 22 mins
15. PowerVM VUG 33 VIOS Shared Storage Pools Phase4 100 mins

11,500 views in total so far (Q4 2015)

- YouTube search: Shared Storage Pools Nigel Griffiths

Recent new information: AIXpert Blog


- Look for entries with these titles:

**SSP4 Best Practice & FAQ**
- 35 recommendations + 20 questions

**SSP Hands-on Fun with LU by Example**
- Rename a LU (offline)
- Backup & restore LU + Backup a snapshot
- Slim down a Thin LU (offline)
- Move a LU between SSPs
- Check if a LU is mapped across whole SSP cluster
- No Testing in Production (TIP) please!!!
New information: AIXpert Blog


- Look for entries with these titles:

  SSP4 a better lu -list command
  - Script to make the LUs in order + better format

  SSP4 Pool Expansion
  - Grow the LUNs and the pool grows

  SSP4 Cheat Sheet
  - Learn the commands by simple examples

How many SSP in the world?
- SSP = no charge option of VIOS so we don’t know
- I guess: SSP=1000’s, 10’s of TB & loads in production

Reminder
### Feature
- Number of VIOS Nodes in Cluster: 1–16
- Number of LUNs in Pool: 1–1024
- Number of Virtual Disks (LUs) Pool: 1–8192
- Number of Client LPARs per VIOS node: 1–200
- Each LUN in Pool size: 10GB–16TB
- Total Pool size: 10GB–512TB
- Virtual Disk (LU) size from the Pool: 1GB–4TB
- Number of Repository Disks: 1–1
- Capacity of Repository Disk: 512MB–1016GB

### Nigel’s Recommendation
- 8 LUN minimum
- LUNs of 128 GB in the pool up to 8 TB
- Larger LUN for large pools
- Repository LUN size 1 GB & spare repository LUN of 2 GB

### SSP - Architecture

**VIOS Cluster**
- Concurrent access to the pool LUNs

**Virtualdisks thin or thick provisioned over vSCSI**

**Shared Storage Pool**
- Create + zone LUNs then add to the pool once

**Pool space allocation & mapping to VM performed on the VIOS & takes ~1 second**

- **Pool & “chunk” level**: 1 MB
- **Pool LUN level**: 128 GB
- **SAN Hardware level**: V7000
SSP is simple – allocate some LUNs

SSP is simple – Zone them to all VIOS’
SSP is simple – Initialise the SSP on 1 VIOS

```
cluster -create -clusternname stellar -spname stellar
-repopvs hdisk15 -sppvs hdisk7 hdisk8
-hostname orangevios1.domain.com
```

SSP is simple – 1st VIOS SSP, next add another

Two LUN sets
1 for each failgrp

Two V7000 units
SSP is simple – add more VIOS’s

cluster -addnode -clusternname stellar
-hostname silvervios1.domain.com

SSP is simple - allocate + assign space to VM's

lu -create -lu blue1 -size 64G -vadapter vhost3 [optional: -thick]
**SSP is simple – add SSP mirror for all LU’s**

```
failgrp -create -fg V7000b: hdisk11 hdisk12
```

Mirrors need zero VM changes

VIOS SSP does re-silvering after problems

---

**SSP4 – is Simple and flexible**

```
“pv” command to add additional LUNs to the pool to grow its size as space is used up or to remove LUNs later
pv -add -fg a: hdisk10 b: hdisk14
```

---

Two LUN sets
1 for each failgrp
Two V7000 units
SSP4 – is Simple and fast to operate

SSP4 Advanced Functions

- Live Partition Mobility (LPM)
  - The default with no more work

- Migrate fixed / internal disks to SSP
  - For high performance AND LPM

- Simple manual server crash recovery
  - “Get out of Jail FREE” card

  - See net 3 charts . . .
4 Advanced Functions: Live Partition Mobility

- SSP VIOS’s already have the LUNs online
  → no SAN zoning issues

- Provided you have Virtual Ethernet & no physical adapters
  → you are LPM ready
  → default is LPM ready
  → no additional work

- Assuming your machines have PowerVM Enterprise
  = LPM

5 Advanced Functions: Migrating to SSP

Got old local disks, VIOS LV or hdisk via vSCSI, NPIV but want to use SSP !!!

Actions:
- Add VIOS to SSP
- Add LU disks
- With AIX
  - migratepv live to SSP
  - bosboot
  - bootlist
  - Remove old disks/adapters
  - and you are LPM ready
Advanced Functions: Box Crash Recovery

- Total box lost!
  - SSP will disks survive

Actions:
- Make a new LPAR
- Map in the SSP LU
- Connect to right network
- Set the bootlist
- Reboot
- and you are running again in, say, 2 minutes!

New Stuff

To install Shared Storage Pool phase 5
- Just upgrade/install VIOS 2.2.4 (or later)
- Available 5th December 2015
VIOS 2.2.4

1. SSP LU Resize (grow = saves admin time)
2. Command: lu -list in alphabetical order
3. Removing pointless -clustename -spname in commands
4. SSP Tiers (multiple pools only better)
   - 10 tiers (think grouping not levels)
   - Fast, medium, slow or IBM, HDS, EMC or prod, in-house, test
5. SSP mirrored now at tier level (was whole SSP)
6. Move a LU between tiers
7. HMC GUI extended for tiers
8. SSP Tier advanced features

1 SSP LU Resize

Why resize, when you can add a new LU virtual Disk?
   - Answer: new LU = manual spread data across disks at OS

   - New option to the lu command: -resize
   - Example: lu -resize -lu myLU -size 35G [G = GB or M = MB]

   - LU shrink not possible – give you a polite error
   - No “add a bit” option (+4GB) → you state the new total size

   - Online = Live with the VM using the LU
     - Actually, the VM will not notice more blocks at the end
1 SSP LU Resize on the VIOS

$ lu -list -attr lu_name=vm96boot

POOL_NAME: spiral
TIER_NAME: prod
LU_NAME SIZE(MB) UNUSED(MB) UDI D
vm96boot 38912 36353
4634dca8b41654ddb39893177d61060e

$ lu -resize -lu vm96boot -size 40G

Logical unit vm96boot with udid '4634dca8b41654ddb39893177d61060e' has been successfully changed.

$ lu -list -attr lu_name=vm96boot

POOL_NAME: spiral
TIER_NAME: prod
LU_NAME SIZE(MB) UNUSED(MB) UDI D
vm96boot 40960 38401
4634dca8b41654ddb39893177d61060e

1 SSP LU Resize on the VM

– OS level details
  – AIX: chvg -g rootvg → AIX then finds the larger disk space
  – AIX client has a VG PP size = minimum you need to grow
    – 32 GB rootvg has default PP size of 64 MB

  – IBM i: Does not support LUN/LU resize → actually “dangerous!”
    – For IBM i just give it a new LU – it knows what to do.

  – Linux: depends on the volume manager & filesystem in use
    – Good luck!
1 SSP LU Resize on the AIX VM

# lsvg rootvg | grep "PP SIZE"
VG STATE: active        PP SIZE: 64 megabyte(s)
# lsvg rootvg | grep "TOTAL PP"
VG PERMISSION: read/write TOTAL PPs: 607 (38848 megabytes)

Resize on the VIOS here
# chvg -g rootvg

# lsvg rootvg | grep "TOTAL PP"
VG PERMISSION: read/write TOTAL PPs: 639 (40896 megabytes)

Have to grow at least this size & preferably a multiple Or grow by GB’s
Pre-resize

2 lu -list in alphabetical order

– Wow, about time, right!
– I pointed out the random order in the beta testing 😊
– Developers said “No way. Oh cumbs! We will get that fixed.”
– And they have
Spot the THREE user hostile features? IMHO

$ lu -list
POOL_NAME: spiral
TIER_NAME: test
LU_NAME                SIZE(MB)    UNUSED(MB)  UDID
"testa"                 32768        32770        4091ba1cd22bb7a401040f307689636dd21
SNAPSHOTS
2015-09-30T11:26:32
testb                 40960        40962        5c2abfbf20e7a4edac8d166cb6bec33c1
SNAPSHOTS
before-upgrade
SNAPSHOTS
after-upgrade
v234567890123456789012~ 8192        8192        ebf fb84803e5ced5401ebf1ed7d6c2fc
vm97boot                38912       36352       a1c cdea9dac4ed18b4ea546de9a69bocc
vm97data                8256        8233        bb033473f8ec78752550ba0fbe940f27

2 lu -list

$ lu -list
POOL_NAME: spiral
TIER_NAME: test
LU_NAME                SIZE(MB)    UNUSED(MB)  UDID
"testa"                 32768        32770        4091ba1cd22bb7a401040f307689636dd21
SNAPSHOTS
2015-09-30T11:26:32
testb                 40960        40962        5c2abfbf20e7a4edac8d166cb6bec33c1
SNAPSHOTS
before-upgrade
SNAPSHOTS
after-upgrade
v234567890123456789012~ 8192        8192        ebf fb84803e5ced5401ebf1ed7d6c2fc
vm97boot                38912       36352       a1c cdea9dac4ed18b4ea546de9a69bocc
vm97data                8256        8233        bb033473f8ec78752550ba0fbe940f27

(1) On what planet is that useful in the default output? 32 hexadecimal digits

(2) Mixed in snapshot names = confusing

(3) Truncated LU name at 22 letters
2 lu -list

- Trouble is, everyone has their own “perfect layout”
- Fortunately, the SSP designers are very clever
- The lu options allow you any format

```bash
$ lu -list -field LU_SIZE LU_NAME -fmt :
32768:testa
40960:testb
38912:vm97boot
8256:vm97data
8192:testc
8192:v23456789012345678901234567890
38912:vm96boot
8256:vm96data
```

- OK not pretty but nothing awk can't sort out
- In Tier then alphabetical order = actually makes sense

```bash
© 2015 IBM
```

- For a full list of field names use: lu -list -verbose

```bash
field names use:
```

```bash
© 2015 IBM
```

```
```bash
2 lu -list

echo "SizeMB UsedMB Used% Type Tier Name"
/usr/ios/cli/ioscli lu -list -fmt :
-field LU_SIZE LU_USED_SPACE LU_USED_PERCENT \ 
LU_PROVISION_TYPE TIER_NAME LU_NAME \ 
| awk -F: '{ printf "%6d %6d %4d% %5s %7s %s\n",$1,$2,$3,$4,$5,$6}'

<table>
<thead>
<tr>
<th>SizeMB</th>
<th>UsedMB</th>
<th>Used%</th>
<th>Type</th>
<th>Tier</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>32768</td>
<td>0</td>
<td>0%</td>
<td>THIN</td>
<td>SYSTEM</td>
<td>testa</td>
</tr>
<tr>
<td>40960</td>
<td>0</td>
<td>0%</td>
<td>THIN</td>
<td>SYSTEM</td>
<td>testb</td>
</tr>
<tr>
<td>38912</td>
<td>2562</td>
<td>6%</td>
<td>THIN</td>
<td>SYSTEM</td>
<td>vm97boot</td>
</tr>
<tr>
<td>8256</td>
<td>23</td>
<td>0%</td>
<td>THIN</td>
<td>SYSTEM</td>
<td>vm97data</td>
</tr>
<tr>
<td>8192</td>
<td>8192</td>
<td>100%</td>
<td>THICK</td>
<td>prod</td>
<td>testc</td>
</tr>
<tr>
<td>8192</td>
<td>0</td>
<td>0%</td>
<td>THIN</td>
<td>prod</td>
<td>v23456789012345678901234567890</td>
</tr>
<tr>
<td>38912</td>
<td>2561</td>
<td>6%</td>
<td>THIN</td>
<td>prod</td>
<td>vm96boot</td>
</tr>
<tr>
<td>8256</td>
<td>26</td>
<td>0%</td>
<td>THIN</td>
<td>prod</td>
<td>vm96data</td>
</tr>
</tbody>
</table>
```

© Copyright IBM Corporation 2015
2 nlu – download AIXpert blog

```bash
help()
{
    echo "$0 "Bipal’s lu command with improved layout and column ordering"
    echo "$0 "(-sizemb | -usedmb | -used | -type | -tier | -name (default))"
    exit 0
}

if [[ $(whoami) == "padmin" ]]
then
    command=${whoami} $0
    if [[ $DEBUG != 1 ]] then
debug $0 as padmin so restart $command again as the root user
echo "$command" $1 | oem_setup_env
    fi
else
    debug $0 now I am root
    # lowercase the parameter with tr to avoid input case errors
    case "echo $1 | tr "a-zA-Z" "a-z" " in
        1 | -sizemb) COLUMN="-k 1" ;
        2 | -usedmb) COLUMN="-k 2" ;
        3 | -used) COLUMN="-k 3" ;
        4 | -type) COLUMN="-k 4" ;
        5 | -tier) COLUMN="-k 5" ;
        6 | -name) COLUMN="-k 6" ;
        *) help ;;
    esac
    echo "$0 " SizeMB UsedMB Used% Type Tier Name"
    /usr/ios/cli/iocll \
    lu -list -field LU_SIZE LU_USED_SPACE LU_USED_PERCENT LU_PROVISION_TYPE TIER_NAME LU_NAME -fmt : \
    -name)              COLUMN="-k 6" ;
      5 | -tier)              COLUMN="-k 5" ;
      6 | -sizemb)            COLUMN="-nk 1" ;
      *) help ;;
    *) COLUMN="-k 4" ;
    echo "$0 " Nigel’s lu command with improved layout and column ordering"
    /usr/ios/cli/iocll \
    lu -list -field LU_SIZE LU_USED_SPACE LU_USED_PERCENT LU_PROVISION_TYPE TIER_NAME LU_NAME -fmt : \
    -sizemb | -usedmb | -used | -type | -tier | -name (default)
    sort $COLUMN
fi
exit 0
```

© Copyright IBM Corporation 2015

---

2 nlu with sort by field name

```
$ nlu -usedmb
<table>
<thead>
<tr>
<th>SizeMB</th>
<th>UsedMB</th>
<th>Used%</th>
<th>Type</th>
<th>Tier</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>8192</td>
<td>0</td>
<td>0%</td>
<td>THIN</td>
<td>prod</td>
<td>testa</td>
</tr>
<tr>
<td>32768</td>
<td>0</td>
<td>0%</td>
<td>THIN</td>
<td>test</td>
<td>testa</td>
</tr>
<tr>
<td>40960</td>
<td>0</td>
<td>0%</td>
<td>THIN</td>
<td>test</td>
<td>testb</td>
</tr>
<tr>
<td>8256</td>
<td>23</td>
<td>0%</td>
<td>THIN</td>
<td>test</td>
<td>vm97data</td>
</tr>
<tr>
<td>8256</td>
<td>26</td>
<td>0%</td>
<td>THIN</td>
<td>test</td>
<td>vm96data</td>
</tr>
<tr>
<td>38912</td>
<td>2573</td>
<td>6%</td>
<td>THIN</td>
<td>test</td>
<td>vm97boot</td>
</tr>
<tr>
<td>40960</td>
<td>2579</td>
<td>6%</td>
<td>THIN</td>
<td>test</td>
<td>vm96boot</td>
</tr>
<tr>
<td>39936</td>
<td>39936</td>
<td>100%</td>
<td>THICK</td>
<td>prod</td>
<td>testc</td>
</tr>
</tbody>
</table>
```

© Copyright IBM Corporation 2015
3 Remove -clusternam -spname

- Small item – more of a clean up change
- Only one cluster per VIOS AND one pool per VIOS
  - The VIOS knows the only one possible name
  - QED: don’t make the user type it all day

- Removed (optional) from most commands including:
  lu, failgrp, tier, pv, alert, snapshot

- Nigel’s recommendation
  KISS: make the cluster name & pool name the Same

---

3 Remove -clusternam -spname

Still needed for

- cluster -create -cluster XXX -sp YYY
  - to set the names later seen in command output

- NOTE: it was -spname and now is -sp

- Older “list storage pool” command lssp has it
  lssp -clusternam XXX -sp YYY -bd

- Same command used for local pool disks (rootvg LV’s)
- Most people use the lu command instead via a script

- Note: was always -sp
4 Tiers

SSP5 – Multiple Pools verses Tiers

- Regular question; I need **multiple pools**, WHEN?
  - With the current version just one pool per VIOS and “hose it all about” policy – I can’t separate workloads

- Different reasons:
  1. Two pools for high speed IBM SAN disks & older slower EMC ones 😊
  2. Different pools for FlashSystem 9000 (prod) & V7000 (test/dev)
  3. Policy to separate production data & “other” workload disks
  4. Have local speedy & remote FC “dark fibre” storage
     - need to set which VM gets which type
     - By the way tiers can do all this
SSP5 – Multiple Pools verses Tiers

- Regular question; I need **multiple pools**, WHEN?
  - With the current version just one pool per VIOS and “hose it all about” policy – I can’t separate workloads

- But multiple pools has inherit issues:
  - Live moving a LU between pools is impossible to do safely
  - What if you fill a pool and other is 95% empty
  - Implement in the slow pool but urgently now need fast disks
  - What if VIO Servers have different mixture of pools
    - LPM could be messy & Zone complexity
    - SSP is meant to reduce complexity and save people time

SSP5 – Multiple Pools verses Tiers

Conclusion:
- What you wanted multiple pool for, can be done by tiers
- Tiers can do more than multiple pools
- Tiers keep everything simple & fast to operate

- Regular expression: 1,$s/multiple pools/Tiers/g
  - Over a beer, we could argue they are the same thing!

- Tiers Win! – Hurray!
SSP5 – Tiers

- Ten tiers inside the single shared storage pool
- Tier made up of a set of LUN – giving you data separation
- LUN can be from different disk units
- Use Tiers any way you like:
  1. Speed: FAST, Medium, slow
  2. Vendor: IBM, HDS, EMC
  3. Importance: Critical, Prod, in-house, test, dev
  4. Isolation: customerA and customerB
  5. Isolation: marketing, sales, support and proper techies
  6. Location: Local, computer-room-C, across-campus
  7. Functionality: RDBMS, web-service, archive, video-collection
- Can live / dynamically move an LU between tiers
  - Due to one common set of meta data

SSP5 – Tiers

- Misconception 1 Tiers do not imply layers = no top or bottom

- Misconception 2 Move from a tier to any other tier

- Misconception 3 There is no special tier*
  Well there is a tier which has the meta data (SYSTEM) & the Default tier (changeable)
SSP – Tiers creation

- A Tier is just a collection of LUNs
- An SSP with one tier works exactly like SSP4
  - The only is called SYSTEM includes the meta data and is the default

- SSP create single first tier
  \[\text{cluster -create -clustername FRED -repopvs hdisk20 -sp FRED -appvs hdisk31 hdisk32 ...}\]

- SSP create with two tier “starter pack”
  \[\text{cluster -create -clustername SALLY -repopvs hdisk20 -sp SALLY -systier RED: hdisk30 hdisk31 ... -usrtier GREEN: hdisk40 hdisk41 ...}\]

- then you can add more tiers

SSP – Later Tier creation

- Add one Tier at a time
  - \text{tier -create -tier BLUE: hdisk50 hdisk51}

- Other Options follow SSP conventions
  - \text{tier -remove -tier NAME}
  - \text{tier -modify -tier NAME -attr ATTR=VALUE}
  - \text{tier -list [-fmt : | -field | verbose] ...}
SSP5 – Tiers LU allocation

- One Tier is marked as “Default”
- If you make a LU as before it goes in the Default Tier
  
  $ lu -create -lu vm22boot -size 64G

- Add options to allocate from your choice of Tier
  
  $ lu -create -lu vm23boot \
  -size 64G -tier RED

- A LU is only in one Tier

- IMHO: Makes sense for the LU’s of one VM to be in the same Tier

5 SSP5 – Tiers Mirrors

- Mirrors (a second failgrp) are now at the Tier level
- Same failgrp command but with a -tier parameter
  
  failgrp -create -tier SALLY -fg SALLY2: hdisk60 hdisk61 ...

- As before you need a 2nd set of disks – on a different disk unit
- So you can have mixed mirrored tiers & unmirrored tiers
  
  - IMHO that would be unusual
- Have to mirror each tier, of course
$ tier -list
POOL_NAME: testsp
<table>
<thead>
<tr>
<th>TIER_NAME</th>
<th>SIZE(MB)</th>
<th>FREE_SPACE(MB)</th>
<th>FAILOVER_SIZE(MB)</th>
<th>FAILOVER_FREE_SPACE(MB)</th>
<th>MIRRORS</th>
<th>STATE</th>
<th>ERASURE_CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>mytier</td>
<td>10112</td>
<td>8000</td>
<td>NOT_MIRRORED</td>
<td></td>
<td>1</td>
<td>SYNCHD</td>
<td>MIRROR2</td>
</tr>
</tbody>
</table>

$ tier -list -vbose
POOL_NAME: testsp
<table>
<thead>
<tr>
<th>TIER_NAME</th>
<th>SIZE(MB)</th>
<th>FREE_SPACE(MB)</th>
<th>OVERCOMMIT_SIZE(MB)</th>
<th>TOTAL_LU_SIZE</th>
<th>PG_COUNT</th>
<th>MIRROR_STATE</th>
<th>ERASURE_CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>mytier</td>
<td>10110</td>
<td>8000</td>
<td>0</td>
<td>2112</td>
<td>1</td>
<td>NOT_MIRRORED</td>
<td>MIRROR2</td>
</tr>
</tbody>
</table>

6 SSP5 – Tiers: LU Live move
- Live storage migrate a LU's blocks to a different Tier
- With the LU client VM running
  - $ lu -move -l1u vm42boot -dsttier bluetier
- dst = destination = IMHO ghastly
SSP5 – Tiers: LU move

- Effectively Live moving the LU data to a different disk set
- For example:
  - faster SAN disks like Flash
  - more reliable SAN disks
  - similar disks but different location

Some ghastly slow old non-IBM SAN disk array with small disk count

If the workload is disk bound then this is a non-disruptive way to tune the disks or even experiment

Shiny fast V7000 SAN disk array with many more disks & caching

Effectively Live moving the LU data to a different disk set

SSP5 – Tiers: LU attributes

```
$ lu -list -verbose
POOL_NAME: sp1
TIER_NAME: SYSTEM
TIER_RELATION: PRIMARY
ADDITIONAL_TIERS: N/A
LU_NAME: test
LU_UDID: 4b9ab8ac36f99fc6d81720528a5dd64b
LU_SIZE(MB): 10
LU_USED_PERCENT: 0
LU_UNUSED_SPACE(MB): 10
LU_PROVISION_TYPE: THIN
LU_UDID_DERIVED_FROM: N/A
LU_MOVE_STATUS: N/A
LU_SNAPSHOTS: N/A
```

There is no lu -modify
i.e. you can’t directly change anything
"tier" SSP virtual disk command / state map

- tier -create
- tier -list
- tier -list - verbose
- failgrp -create
- failgrp -remove
- tier -modify -attr
  name=x
  type=system | comingled
  default=x
- lu -move -dsttier

SSP5
Worked Example
Shared Storage Pool phase 4

VIOS Client VM LU virtual disks vSCSI thin or thick provisioned

VIOS Cluster concurrent access to the pool LUNs

Pool space allocation & mapping to VM performed on the VIOS & takes ~1 second

Shared Storage Pool create LUNs+ zone then add to the pool once

Pool & “chunk” level
Pool LUN level
Hardware level

Shared Storage Pool phase 4

VIOS Client VM LU virtual disks vSCSI thin or thick provisioned

VIOS Cluster concurrent access to the pool LUNs

Pool space allocation & mapping to VM performed on the VIOS & takes ~1 second

Shared Storage Pool

Multiple Tiers create LUNs+ zone then add to the tier once

Pool & “chunk” level
Pool LUN level
Hardware level
SSP4 to SSP5 upgrade?

- SSP4 = VIOS 2.2.3.x
- SSP5 = VIOS 2.2.4.x

- Online VIOS upgrade – business as usual
  - Check dual path VIOS
  - clstartstop -stop …
  - updateios … plus a VIOS reboot
  - clstartstop -start …
  - Upgrade last VIOS → Tier functions work

- Online:
  - tier -create …
  - lu -move …

7 SSP5
HMC Classic & Enhanced+

Need to recapture with HMC 840
as it has much more SSP support including Tiers
our current HMC 840 has no SSP attached
SSP5 – HMC Support Classic view

Virtual Storage Management

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.

SSP5 – HMC Support Classic view

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 can also create a new virtual disk.

Create Virtual Disk

To create a virtual disk, enter a name and a size for the new disk, and select a storage pool from which to create the new disk. You can also assign the new disk to a logical partition. This task can take several minutes to complete if you are creating a virtual disk in a file-based storage pool.

Virtual disk name:

Storage pool name:

Virtual disk size:

Assigned partition:

Disk type:

Map to VIOS(s):

Select Virtual IO Server:

- rubyvios3
- rubyvios4
SSP5 – HMC Support Enhanced+ view

 SSP5 – HMC Support Enhanced+ view

Manage Shared Storage Pool Cluster
The table lists all the shared storage pool clusters that are assigned to the managed system. You can manage the clusters that are listed or add additional clusters to your managed system. Right-click a cluster in the table to view the manage task, and to remove the cluster from the list. Learn more →

Right Click
SSP5 – HMC Support Enhanced+ view

Shared Storage Pool Name: spiral

Repository Disk (1 required)

Number Of Nodes (max 16)

Physical Volume

SSP Volumes

Add / Remove Nodes Cluster

Cluster Name: spiral

The table lists only the Virtual I/O Servers (VIOS) that can be configured to the shared storage pool of volume number (0,0). Select the VIOS from the table to configure them on the cluster. Clear the check-box for any VIOS that you want to remove from the cluster.

Add Tier

Tiers

Mirrored?

Alert Thresholds

Repository Disk

VIOS Nodes

HMC 840 – with Tier support

Click these for different information
8 SSP5 – Advanced

- Tier modify options
- Separate SYSTEM tier
- lu -move -nonrecursive
- Alert is tier based – warnings on SSP near full or grossly over-committed

SSP5 – Tier – modify name & Default tier attribute

Tier Name:
Change tier name (like lu & failgrp modify)
$/tier -modify -tier FRED -attr tier_name=BERT

Note: 1st tier is called “Default” but might not be the Default tier!

Default tier:
Use tier -list -verbose
- And look for TIER_DEFAULT:YES
- lu -create without -tier option go here
Only one tier can be the Default tier - set using
$/tier -modify -tier GREEN -attr default=yes
  – Note: can use either default=yes or tier_default=yes
SSP5 – separate SYSTEM Tier

If your Shared Storage Pool is
  – Large (10’s of TB)
  – High rates of new or changing LUs
  – High rates of Thin LU extending – dynamic space allocation
  – High numbers of LU tier moves
Then SSP meta-data update I/O rate can be large = effecting performance

The SSP developers really like the idea of a separate SYSTEM tier disks
1. Separate the meta-data I/O to a different disk set
   – To reduce latency i.e. meta-data I/O not queued behind data I/O
2. Turbo charge SYSTEM tier by using FC Flash storage

SSP5 – separate SYSTEM Tier

SYSTEM TIER “Rules of thumb”
1. When separate SYSTEM meta-data is recommended
   ➔ If you have some fast LUNs and much slower larger User tier LUNs
   ➔ If you have access to limited Flash LUNs

2. The size of the SYSTEM tier needed
   ➔ 0.3% of the User Tiers
SSP5 – SYSTEM Tier

Tier types:
1. SYSTEM  → SSP meta-data only
2. COMINGLED  → SSP meta-data and LUs
3. USER  → All other Tiers for LU data

1st tier has to be type SYSTEM or COMINGLED (only one)

Set Tier type “system” or “system comingled”
- `$ tier -modify -tier RED -attr type=system`
- `$ tier -modify -tier RED -attr type=comingled`

If Type SYSTEM
- Can `lu -move a LU from it` used to remove LUs from this SYSTEM Tier until no LUs left in it= only meta-data
- Can't `lu -move a LU to it`

SSP5 – SYSTEM Tier

Number of Tiers

1. Comingled  (SYSTEM + User LUs)
   Default

2. Comingled  (SYSTEM + User)

2. SYSTEM
   No user LUs

many
   SYSTEM
   No user LUs

cluster -create single tier

cluster -create two tier then add more tiers

User LUs Default

User LUs

User LUs

User LUs

SSP metadata is much smaller than the LU data so a SYSTEM only tier could be smaller
SSP4 to SSP5 – adding a tier (new or reused LUNs)

With SSP4 this is all you have
1 tier, comingled and default

1 Add more LUNs for new tier -create

2 lu -move VM to new tier or create more LUs

OR

1 pv -remove LUNs from SYSTEM tier
2 tier -create with these released disks

Reuse existing LUNs for new tier
Assuming:
SSP has spare capacity

Upgrading SYSTEM tier to Faster/Flash LUNs

- You can’t directly move the SYSTEM tier
- For example: to faster LUNs to gain performance

- You have to rotate your disks (LUNs)
- Scary the first time!!
- Next is two ways to get this done . . .
Upgrading SYSTEM tier to Faster/Flash LUNs

- You can move the SYSTEM tier


  2. Then remove slow: `pv -remove -pv hdisk42 hdisk52`

  3. Then add to User tier: `pv -add -tier user -pv hdisk42 hdisk52`

- One step Swap out:

  1. Swap out: `pv -replace -oldpv hdisk42 -newpv hdisk89`
     `pv -replace -oldpv hdisk52 -newpv hdisk99`

  2. Then add to User tier: `pv -add -tier user -pv hdisk42 hdisk52`
**Fast/Flash System Tier in practice**

Don’t worry – be happy!

- Good to monitor SYSTEM meta-data disk I/O
- ATS has a 128GB min LUN $\rightarrow 128 / 0.3\% = 39$TB
  - Result just one LUN (pair) for the tier = not good for I/O
  - Flash is expensive $\rightarrow$ smaller LUN size so lots of LUNs
- VIOS Advisor spots hot disks, small queue depth …
- SYSTEM I/O rate only an issue on
  - Very large SSP and very busily changing SSP config
  - High numbers of VIOS (future!)

**SSP5 – Tiers: LU Advanced move with Clones**

- `lu -move -lu vm42 -dsttier blue -nonrecursive`

Background

- SSP system admin can’t normally create LU clones
- The commands are not documented $\rightarrow$ ignoring “under the hood” fiddlers!
- But System Director and PowerVC use clones & clones of clones
- Then the clones can have snapshots or further clones

![Diagram of LU Advanced move with Clones](image-url)
SSP5 – Tiers: LU Recursive move (default)
SSP5 – Tiers: LU **Recursive** move (default)

Could be a massive move of data
LU=100GB and clones ~50% difference
Could move 600GB & same space use

Could be a large-ish move of data
Could move 200GB
50 to 100 GB more space overall

Move 100GB as expected
Source gets more space + 50GB more target space

SSP5 – Tiers: LU **Non-Recursive** move (-nonrecursive)

Move 100GB as expected
Source gets more space + 50GB more target space
SSP5 – Tiers: LU Non-Recursive move (-nonrecursive)

Source Tier                                            Target Tier

Note: 1 master images becomes 3!
Free space goes down

Source Tier                                            Target Tier

Source master clone gone space used goes up ~50GB
Will move 100GB + 100GB more target space

Source intermediate clone gone space used goes up ~100GB
Will move 100GB + 100GB more target space

Move as expected, source 50GB more space
Will move 100GB + 100GB more target space
SSP5 – Summary lu-move with clones

- No clones = no complexity
- With clone hierarchy it can get complex
  - Probably a problem that you which I had never told you about 😊
- It can result in
  - higher than expected large data moves
  - extra space on source and/or target
- But it can be used to break a PowerVC clone away
  - Example: clone AIX LU & then install Linux on it = no point being a clone
  - lu -move -lu vm42 -dsttier blue -nonrecursive
  - lu -move -lu vm42 -dsttier red ← Optional move back to original tier

alert -tier mandatory → Messaged sent to the HMC

- By example - my tier is called “prod”
  - alert -set -tier prod -type threshold -value 10
  - alert -set -tier prod -type overcommit -value 50
  - Threshold is the Pool “free space” getting low
    i.e. 10 means alert when Pool free space crosses to below the 10%
  - Overcommit warns when you go “to far” & risk problems later on.
    50% might be accept on not busy VM's but 500% is sure to bite you!
    $ alert -list
      PoolName:spiral
      PoolId:000000009893ED900000000560AA6E3
      TierName:prod
      ThresholdPercent:10
      OverCommitPercent:50
      PoolName:spiral
      PoolId:000000009893ED900000000560AA6E3
      TierName:test
      ThresholdPercent:35
      OverCommitPercent:N/A

The defaults

Nigel’s Recommendation: Set alert values but monitor in coming Alerts via HMC (which every customer should be watching) !!
Summary VIOS 2.2.4 with SSP Tiers

1. SSP LU Resize (grow = saves admin time)
2. DIY lu -list
3. SSP Tiers (multiple pools only better)
   - 10 tiers (think grouping not levels)
   - Fast, medium, slow or IBM, HDS, EMC or prod, in-house, test
   - lu -create -lu fred -size 32G -tier prod lu -move -lu fred -dsttier test
4. SSP mirrored now at tier level (was whole pool)
5. SSP LU move between tiers
6. Possible separate SYSTEM tier LUNs
7. HMC GUI Support