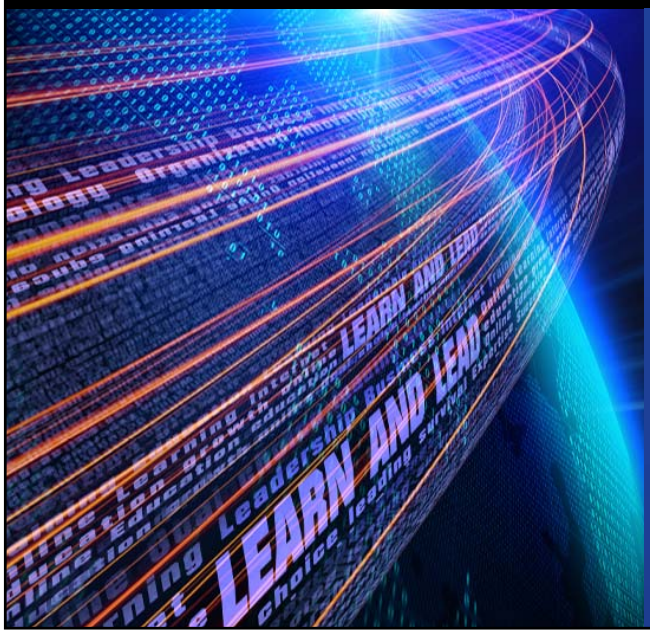
TechU



Care and Feeding of VIO Servers

Part 2 – Backups, Storage and Monitoring

Jaqui Lynch
Flagship Solutions Group
jlynch@flagshipsg.net

2020 IBM Systems Technical University
October 26-29, 2020| Virtual Conference

IBM

1

Agenda

- ***Presentation***
  - Backup and recovery
  - Storage
  - Network
  - Monitoring
- ***Documentation***
  - Useful Commands
  - Useful Links
  - Backup Material
-

2

2

Backup and recovery

3

3

IBM Supported Backup and Restore Methods for VIO Servers

- https://www.ibm.com/support/knowledgecenter/9009-22G/p9hb1/p9hb1_vios_backup_backup.htm
- Note that IBM does not support (even at v3) backup and restore with **USB sticks**

Table 1. Backup and restoration methods for the VIOS

Backup method	Media	Restoration method
To tape	Tape	From tape
To DVD	DVD-RAM	From DVD
To remote file system	nim_resources.tar image	From an HMC using the Network Installation Management (NIM) on Linux facility and the **installios** command
To remote file system	mksysb image	From an AIX 5L™ NIM server and a standard mksysb system installation
Tivoli Storage Manager	mksysb image	Tivoli Storage Manager

You can backup and restore AIX from USB but VIO is not supported – see below for AIX information
https://www.ibm.com/support/pages/using-and-taking-advantage-usb-devices-and-aix

4

4

Backing up VIOS

- The VIO server version of viosupgrade will take an extra viosbr backup for you

- You may also want to use snap to grab other critical data
  - #snap -gtknc
- Mount NFS filesystem to backup to (in my case /backups)
- mkdir /backups/viosa
- umount /var/vio/VMLibrary if you are using the media library or your mksysb will be huge

- Then as padmin run backupios which automatically calls savevgstruct:
- **backupios** -file {*File name*} [-mksysb] [-nopak] [-nosvg] [-nomedialib]
- backupios –file /backups/viosa
- The above creates a nim_resources.tar package in that directory and it can be used to clone or restore VIO servers using installios (NIMOL) from the HMC

- You can also back it up as a mksysb file that can be used to restore from your NIM server
- backupios -file /backups/viosa.mksysb –mksysb
- When the **-mksysb** flag is used, the NIMOL resources are not saved in the image.

- If the media library is large and is on rootvg, then you can add the –nomedialib flag, but still unmount it prior to the backup

5

5

Backing up VIOS from root

As root (login as padmin then oem_setup_env) run viosave.sh (coming up)

#su - padmin -c "ioscli viosbr –backup –file /home/padmin/viosabr.backup”

Mount the NFS repository for the backups (/backups)
#mount /backups
#su – padmin –c “ioscli backupios –file /backups/vio2-sep0919.mksysb -mksysb -nomedialib”

This backs it up to a bootable mksysb file

If using NIM to clone VIO servers don’t forget:
#mkdir /backups/nimbkups
#su – padmin –c “ioscli backupios -file /backups/nimbkups -nomedialib”

This creates a nim_resources.tar file that can be used for restores described at:
http://public.dhe.ibm.com/software/server/vios/docs/backupios_mod.pdf

Create a daily backup once a day and keep up to 7 in /home/padmin/cfgbackups
#su - padmin -c "ioscli viosbr –backup -file viobkup –frequency daily numfiles 7”

If you use alt_disk_copy to clone your rootvg disk you have a very fast failback – consider using this as a preupdate backup

6

6

Backup

If you have set the system up to automatically do a daily viosbr backup then you don't need to do it here

Create a daily backup once a day and keep up to 7 in /home/padmin/cfgbackups

#su - padmin -c "ioscli viosbr –backup -file viobkup –frequency daily numfiles 7"

I usually do 2 backups

A regular mksysb

A backup that is for NIM

Both are done to an NFS mount

If you are using the file backed optical (media repository) there seems to be a bug where specifying nomedialib does not stop it backing up the media library. Since mine is over 100GB that is a problem so here is what I do

Prior to the backup I unmount /var/vio/VMLibrary and I remount it after the backup

My 3.1.1.25 backups are around 19GB but during the backup they can need as much as 40GB

So for 2 x VIO servers I need 80GB to back them up as I do both the NIM and the HMC type of backups for each VIO

But I need closer to double that until the compress is run at the end of the backups

Once you are done you can remount /var/vio/VMLibrary

7

7

Backup Script to put in crontab

```
#! /bin/sh
#
machine=`uname -n`
mount /usr/local/backups
mkdir /usr/local/backups/$machine
umount /var/vio/VMLibrary
su - padmin -c "ioscli backupios -file /usr/local/backups/$machine -nomedialib"
su - padmin -c "ioscli backupios -file /usr/local/backups/vio-mksysbs/$machine.mksysb -nomedialib -mksysb"
mount /var/vio/VMLibrary
#
exit 0
```

NOTES

The above can be put in root's crontab to run regularly

Don't forget to set up an NFS mount to the VIO from your NIM or NFS server

Do not allow ANY NFS mount to mount automatically at boot in case the NIM or NFS server is down at the time of boot

Also, regularly grab an HMCScanner report

https://www.ibm.com/support/pages/hmc-scanner-power-server-config-and-performance-stats

https://www.ibm.com/support/pages/sites/default/files/inline-files/$FILE/hmcScanner-0.11.42.zip

8

8

Check the virtual backup

The following adds a cron entry and will backup your VIO virtual definitions every day and keep the last 7 copies in /home/padmin/cfgbackups
You only need to run it once

viosbr –backup -file viobkup –frequency daily numfiles 7

You can view the backups taken using viosbr –view (next slide)

You can list what is in a backup using:
viosbr -view -file viosname.01.tar.gz

#crontab –l
0 3 1 * * /usr/local/bin/viobackup.sh >/usr/local/logs/viobackup.txt >2&1
0 3 15 * * /usr/local/bin/viobackup0.sh >/usr/local/logs/viobackup0.txt >2&1
0 0 * * * (/usr/ios/cli/ioscli viosbr -backup -file viosname -frequency daily -numfiles 7)

The above runs my VIO backup on the 1st and 15th and it runs the virtual definitions backup daily

9

9

Check the virtual backup

```
$ viosbr -view -list
autoviosbr_vio2.tar.gz
vio2_16056368.tar.gz
viosname.01.tar.gz
viosname.02.tar.gz
viosname.03.tar.gz
viosname.04.tar.gz
viosname.05.tar.gz
viosname.06.tar.gz
viosname.07.tar.gz
$ ls -al /home/padmin/cfgbackups
total 160
drwxr-xr-x    2 root     staff          4096 Sep 12 00:00 .
drwxr-x---    7 padmin   system         4096 Aug 06 09:08 ..
-rw-r--r--    1 padmin   staff          7463 Jun 30 13:00 autoviosbr_vio2.tar.gz
-rw-r--r--    1 padmin   staff          6900 Apr 13 2019  vio2_16056368.tar.gz
-rw-r--r--    1 padmin   staff          7558 Sep 06 00:00 viosname.01.tar.gz
-rw-r--r--    1 padmin   staff          7560 Sep 07 00:00 viosname.02.tar.gz
-rw-r--r--    1 padmin   staff          7560 Sep 08 00:00 viosname.03.tar.gz
-rw-r--r--    1 padmin   staff          7559 Sep 09 00:00 viosname.04.tar.gz
-rw-r--r--    1 padmin   staff          7558 Sep 10 00:00 viosname.05.tar.gz
-rw-r--r--    1 padmin   staff          7559 Sep 11 00:00 viosname.06.tar.gz
-rw-r--r--    1 padmin   staff          7497 Sep 12 00:00 viosname.07.tar.gz
```

10

10

Check the backups

1. MKSYSB for use with NIM

```
#ls –al /usr/local/backups/vio-mksysbs/vio2*
-rw-r--r--    1 root     system   19430502400 Sep 01 03:29 vio2.mksysb
# du -sg vio2*
18.10   vio2.mksysb
```

2. nim_resources.tar for use with installios from the HMC

```
#ls –al /usr/local/backups/vio2
total 38870368
drwxr-xr-x    2 root     system          256 Sep 01 03:20 .
drwxr-xr-x   10 root     system         4096 Sep 02 11:13 ..
-rw-r--r--    1 root     staff    19898746880 Sep 01 03:20 nim_resources.tar
# du -sg *
18.53   nim_resources.tar
```

11

11

Document VIO Information – save-viostuff.sh

```
#! /bin/sh
#
day="`/bin/date +'%d'`"
month="`/bin/date +'%m'`"
year="`/bin/date +'%y'`"
set -- Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
shift $month
lmonth="$1"
machine=`uname -n`
directory="`/bin/date +'%m%d%Y_%H%M'`"
machine_directory=`printf "%s_%s" $machine $directory`
mkdir /home/padmin/saveit
cd /home/padmin/saveit
logit="/home/padmin/saveit/$machine"
logit1="/home/padmin/saveit/$machine"
su - padmin -c "ioscli ioslevel" >>$logit1.ioslevel.txt
su - padmin -c "ioscli lsdev -type disk" >>$logit1.viodisk.txt
su - padmin -c "ioscli lsdev -type adapter" >>$logit1.vioadapter.txt
su - padmin -c "ioscli lsdev -vpd" >>$logit1.viovpd.txt
su - padmin -c "ioscli lsdev -slots" >>$logit1.vioslots.txt
su - padmin -c "ioscli lsmap -all" >>$logit1.violsmapall.txt
su - padmin -c "ioscli lsmap -all -npiv" >>$logit1.violsmapall.npiv.txt
su - padmin -c "ioscli lsdev -virtual" >>$logit1.violsdevv.txt
su - padmin -c "ioscli cfgnamesrv -ls " >cfgname.txt
su - padmin -c "ioscli entstat -all ent9 "  >entstat.txt
su - padmin -c "ioscli hostmap -ls" >hostmap.txt
su - padmin -c "ioscli lsuser" >user.txt
su - padmin -c "ioscli netstat -routinfo" >routinfo.txt
su - padmin -c "ioscli optimizenet -list" >optimize.txt
su - padmin -c "ioscli viosecure -firewall view" >firewall.txt
su - padmin -c "ioscli viosecure -view -nonint" >view.txt
oslevel -s >$logit1.oslevel.txt
getlvodm -C > $logit1.disktmp.txt
while read label line
do
  echo "\n" >>$logit1.viodisks.txt
  echo "Hdisk is $label" >>$logit1.viodisks.txt
  echo " " >>$logit1.viodisks.txt
  su - padmin -c "ioscli lsdev -dev $label -attr" >>$logit1.viodisks.txt
done <"$logit1.disktmp.txt"
#
exit 0
```

12

12

Output created by save-viostuff.sh

Back it up:
./save-viostuff.sh
mkdir: 0653-358 Cannot create /home/padmin/saveit.
/home/padmin/saveit: Do not specify an existing file.

ls -l /home/padmin/saveit
total 392
-rw-r--r-- 1 root staff 23 Sep 12 10:29 cfgname.txt
-rw-r--r-- 1 root staff 38085 Sep 12 10:29 entstat.txt
-rw-r--r-- 1 root staff 240 Sep 12 10:29 firewall.txt
-rw-r--r-- 1 root staff 403 Sep 12 10:29 hostmap.txt
-rw-r--r-- 1 root staff 5970 Sep 12 10:29 optimize.txt
-rw-r--r-- 1 root staff 713 Sep 12 10:29 routinfo.txt
-rw-r--r-- 1 root staff 258 Sep 12 10:29 user.txt
-rw-r--r-- 1 root staff 46 Sep 12 10:29 view.txt
-rw-r--r-- 1 root staff 28 Sep 12 10:29 vio2.disktmp.txt
-rw-r--r-- 1 root staff 9 Sep 12 10:29 vio2.ioslevel.txt
-rw-r--r-- 1 root staff 16 Sep 12 10:29 vio2.oslevel.txt
-rw-r--r-- 1 root staff 3989 Sep 12 10:29 vio2.vioadapter.txt
-rw-r--r-- 1 root staff 245 Sep 12 10:29 vio2.viodisk.txt
-rw-r--r-- 1 root staff 13336 Sep 12 10:29 vio2.viodisks.txt
-rw-r--r-- 1 root staff 4192 Sep 12 10:29 vio2.violsdevv.txt
-rw-r--r-- 1 root staff 13315 Sep 12 10:29 vio2.violsmapall.npiv.txt
-rw-r--r-- 1 root staff 6171 Sep 12 10:29 vio2.violsmapall.txt
-rw-r--r-- 1 root staff 2836 Sep 12 10:29 vio2.vioslots.txt
-rw-r--r-- 1 root staff 49470 Sep 12 10:29 vio2.viovpd.txt

13

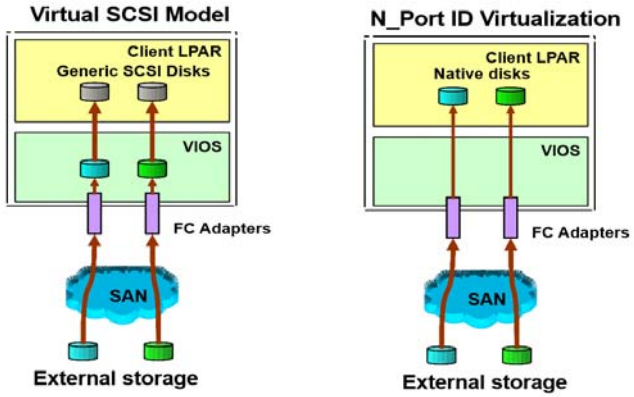
13

Storage

14

14

The context: PowerVM disk access



This is relevant for LPM as for:

- *Zoning*
- *LUN-masking*

15

15

Zoning and Mapping

- ***Zoning***
  - This is when the switch is configured to allow the switch port to talk to the storage and the WWPN for the LPAR or server
- ***Mapping (masking)***
  - This is when the storage is updated to allow the host (LPAR or server) WWPNs access to the specific LUNs provisioned
- LUNs must be provisioned at the storage, then mapped and zoned before they can be used in an LPAR
- For direct attach we zone and map the WWNs for the real adapters, for NPIV we use the WWPNs on the virtual adapters
- WWNs tend to start with 10 or 20
- WWPNs (NPIV) start with C0
- These can be found in an HMCScanner report or by logging onto the LPAR or VIO or from the HMC
- Check the VIO connection to the switch is NPIV enabled:

```
$ lsnports
name             physloc                        fabric tports aports swwpns  awwpns
fcs0             U78C9.001.WZS0234-P1-C12-T1         1     64     63   2048    2046
fcs1             U78C9.001.WZS0234-P1-C12-T2         1     64     63   2048    2046
fcs2             U78C9.001.WZS0234-P1-C6-T1          1     64     56   3088    3062
fcs3             U78C9.001.WZS0234-P1-C6-T2          1     64     56   3088    3062
```

16

16

LPM Zoning and mapping - NPIV

- **Do not confuse zoning with mapping (masking)**
- Regular and LPM WWPNs must be zoned at the switch and mapped at the storage
- Each virtual fibre adapter for an LPAR has 2 x WWPNs
  - The first is the default one that is used
  - The second is used by LPM – it normally does not login unless LPM has been used
  - Both WWPNs must be zoned and mapped
- If they are not mapped at the storage and you do an LPM you will damage your boot image
  - You can avoid this problem after 2.2.4 by setting 2 parameters on viosIpm0 on all VIO LPARs
- You should also do your zoning by zoning all WWPNs for the LPAR to both switches. Keep zoning simple and have a zone that is LPARname and all the WWPNs. This will avoid problems during LPM when you allocate fiber ports to each VIO for dual VIO systems.

17

17

vSCSI and NPIV

- vSCSI
  - Use WWN of the VIO server when zoning then map at VIO server to the client LPAR
  - lscfg –vpl fcs0 | grep Network will show something like: 10000090fa530975
  - Those WWNs belong to the VIO not the client LPAR
  - MPIO drivers are installed in the VIO
  - Mirrored in client LPAR
  - Disks are seen at and assigned from the VIO
    - View using “lsmap –all”
  - Shows as vSCSI in client LPAR
  - FCS tunables set in VIO
- NPIV
  - Use WWPNs that are created when client LPAR is created
  - You can find them in the client’s profile for the virtual adapters and they look like: c0507607dbd80028
  - Those WWPNs (also called VFCs) migrate with the client LPAR
  - Disks are not seen at the VIO so MPIO drivers are installed in the client LPAR
  - Fibre adapters get mapped from the VIO
    - View using “lsmap –all –npiv”
  - Shows as fibre adapters (FCS?) in the client LPAR
  - FCS tunables set in VIO and client LPAR – client LPAR settings must be <= to settings in the VIO servers
  - On V5000 and some other storage arrays you may also have to zone the real WWNs for the adapters, not just the client WWPNs

18

18

Simple Zoning for LPM

SIMPLE ZONE

SWITCH1
zone: NPIV_AIX1
c0:50:76:03:ca:b6:00:28
c0:50:76:03:ca:b6:00:29
c0:50:76:03:ca:b6:00:2a
c0:50:76:03:ca:b6:00:2b
50:05:07:68:02:16:2f:c3
50:05:07:68:02:16:2f:c4
10:00:00:90:fa:19:04:40

SWITCH2
zone: NPIV_AIX1
c0:50:76:03:ca:b6:00:28
c0:50:76:03:ca:b6:00:29
c0:50:76:03:ca:b6:00:2a
c0:50:76:03:ca:b6:00:2b
50:05:07:68:02:26:2f:c3
50:05:07:68:02:26:2f:c4
10:00:00:90:fa:19:15:a9
20:02:00:0e:11:13:06:67

I zone all the WWPNs for the client on both switches – reduces problems with LPM
If you don't do this, you have to be certain to assign the right adapter to the right VIO when using LPM
The only difference is the storage subsystem zones (50:05) and tape drive units they can see

19

19

NPIV

https://www.ibm.com/support/knowledgecenter/TI0002C/p8edm/chnportlogin.html

chnportlogin -o login -m ServerName --id 3
Above logs in all ports for LPAR 3

Can also use -d (0-5 default it 1) to get more detail and -v for verbose mode
-o logout logs them out
-n profilename logs in only those in the specified profile otherwise it uses the current running profile

When performing a login operation, all inactive WWPNs will be activated, including the second WWPN in the pair assigned to each virtual Fibre Channel client adapter. When performing a logout operation, all WWPNs not in use will be deactivated.

https://www.ibm.com/support/knowledgecenter/TI0002C/p8edm/lsnportlogin.html
lsnportlogin -m Server-8286-41A-SN123452X --filter "lpar_names=jaqui" -F lpar_name:wwpn:wwpn_status

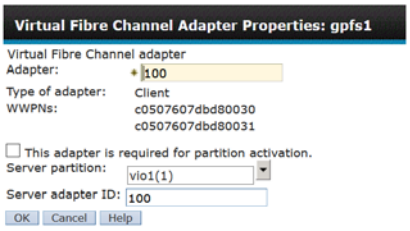
wwpn_status
The WWPN status. Possible values are:
0 - WWPN is not activated
1 - WWPN is activated
2 - WWPN status is unknown

20

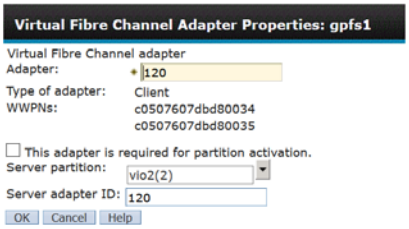
20

NPIV Finding WWPNs

On the HMC – VIO1



On the HMC – VIO2

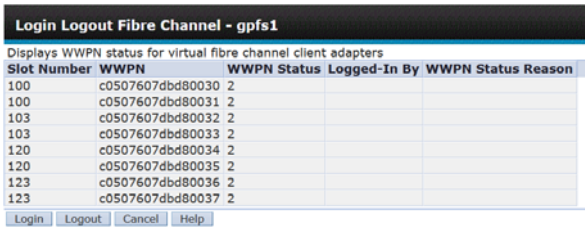


21

21

NPIV – check WWPN status on client

http://www-01.ibm.com/support/docview.wss?uid=isg3T1024487
On the HMC go into the client lpar profile
Select one of the virtual fibre definitions then actions, advanced, login logout fibre channel
You can log them in, log them out or just look at them



wwpn_status
The WWPN status. Possible values are:
0 - WWPN is not activated
1 - WWPN is activated
2 - WWPN status is unknown

22

22

NPIV Mapping at VIO as padmin

NPIV

$ vfcmap -vadapter vfchost20 -fcp fcs0

$ lsmap -vadapter vfchost20 -npiv
Name Physloc ClntID ClntName ClntOS
------------- ---------- ------ -------------- -------
vfchost20 U8286.41A.215D3AV-V1-C108 17 aixtest1 AIX

Status:LOGGED_IN
FC name:fcs0 FC loc code:U78C9.001.WZS0234-P1-C7-T1
Ports logged in:3
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs0 VFC client DRC:U8286.41A.215D3AV-V17-C108

23

23

Zoning and Mapping – adapters on VIO2

```
# lsdev -C | grep fcs
fcs0           Available 00-00     8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs1           Available 00-01     8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs2           Available 01-00     PCIe3 2-Port 16Gb FC Adapter (df1000e21410f103)
fcs3           Available 01-01     PCIe3 2-Port 16Gb FC Adapter (df1000e21410f103)
fcs4           Available 02-04     PCIe3 10Gb 4-Port FCoE Adapter (df1060e214101004)
fcs5           Available 02-05     PCIe3 10Gb 4-Port FCoE Adapter (df1060e214101004)
# lscfg -vpl fcs*| grep Network
        Network Address.............10000000C98CE5B8
        Network Address.............10000000C98CE5B9
        Network Address.............10000090FA530975
        Network Address.............10000090FA530976
        Network Address.............10000090FA740155
        Network Address.............10000090FA740156
# lscfg -vpl fcs*| grep fcs
  fcs0             U78C9.001.WZS0234-P1-C12-T1  8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
  fcs1             U78C9.001.WZS0234-P1-C12-T2  8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
  fcs2             U78C9.001.WZS0234-P1-C6-T1   PCIe3 2-Port 16Gb FC Adapter (df1000e21410f103)
  fcs3             U78C9.001.WZS0234-P1-C6-T2   PCIe3 2-Port 16Gb FC Adapter (df1000e21410f103)
  fcs4             U78C9.001.WZS0234-P1-C8-T1   PCIe3 10Gb 4-Port FCoE Adapter (df1060e214101004)
  fcs5             U78C9.001.WZS0234-P1-C8-T2   PCIe3 10Gb 4-Port FCoE Adapter (df1060e214101004)
```

24

24

Zoning and Mapping – adapters on VIO2

```
$ lsmap -all -npiv | grep gpfs1
vfchost12     U8286.41A.215D3AV-V2-C120              13 gpfs1          AIX
vfchost15     U8286.41A.215D3AV-V2-C123              13 gpfs1          AIX
$ lsmap -vadapter vfchost12 -npiv
Name          Physloc                            ClntID ClntName       ClntOS
------------- ---------------------------------- ------ -------------- -------
vfchost12     U8286.41A.215D3AV-V2-C120              13 gpfs1          AIX

Status:LOGGED_IN
FC name:fcs2                    FC loc code:U78C9.001.WZS0234-P1-C6-T1
Ports logged in:3
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs2            VFC client DRC:U8286.41A.215D3AV-V13-C120

$ lsmap -vadapter vfchost15 -npiv
Name          Physloc                            ClntID ClntName       ClntOS
------------- ---------------------------------- ------ -------------- -------
vfchost15     U8286.41A.215D3AV-V2-C123              13 gpfs1          AIX

Status:LOGGED_IN
FC name:fcs3                    FC loc code:U78C9.001.WZS0234-P1-C6-T2
Ports logged in:3
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs3            VFC client DRC:U8286.41A.215D3AV-V13-C123
```

25

25

Zoning and Mapping – adapters on NPIV client

```
gpfs1:> lsdev -C | grep fcs
fcs0        Available 00-T1       Virtual Fibre Channel Client Adapter
fcs1        Available 03-T1       Virtual Fibre Channel Client Adapter
fcs2        Available 20-T1       Virtual Fibre Channel Client Adapter
fcs3        Available 23-T1       Virtual Fibre Channel Client Adapter
gpfs1:> lscfg -vpl fcs* | grep Network
        Network Address.............C0507607DBD80030
        Network Address.............C0507607DBD80032
        Network Address.............C0507607DBD80034
        Network Address.............C0507607DBD80036
gpfs1:> lscfg -vpl fcs* | grep fcs
  fcs0             U8286.41A.215D3AV-V13-C100-T1  Virtual Fibre Channel Client Adapter
  fcs1             U8286.41A.215D3AV-V13-C103-T1  Virtual Fibre Channel Client Adapter
  fcs2             U8286.41A.215D3AV-V13-C120-T1  Virtual Fibre Channel Client Adapter
  fcs3             U8286.41A.215D3AV-V13-C123-T1  Virtual Fibre Channel Client Adapter
```

26

26

Zoning and Mapping – vSCSI Mapping on VIO

```
$ lsmap -vadapter vhost13
SVSA            Physloc                                      Client Partition ID
--------------- -------------------------------------------- ------------------
vhost13         U8286.41A.215D3AV-V1-C133                    0x00000010

VTD                   vtopt13
Status                Available
LUN                   0x8300000000000000
Backing device
Physloc
Mirrored              N/A

VTD                   vtscsi4
Status                Available
LUN                   0x8100000000000000
Backing device        hdisk9
Physloc               U78C9.001.WZS0234-P1-C15-T1-LC05EC8F200-L0
Mirrored              false

VTD                   vtscsi5
Status                Available
LUN                   0x8200000000000000
Backing device        hdisk10
Physloc               U78C9.001.WZS0234-P1-C15-T1-LA05EC8F200-L0
Mirrored              false
```

27

27

LPM's use of the two WWPNs

- Each virtual fibre adapter for an LPAR has 2 x WWPNs
  - The first is the default one that is used
  - The second is used by LPM – it normally does not login unless LPM has been used
- Prior to an LPM the default WWPN is used
- After the LPM the second WWPN is used
- After the next LPM it goes back to the default WWPN
- i.e. it flip flops between them
- EXCEPT
  - If you perform an inactive LPM then it stays with whatever the WWPNs were that it used last

28

28

Get rid of annoying FCS errors on 10/1Gb cards

- The 4 port 10Gb/1Gb cards can act as network cards or can be used for San. Most people use them as network cards and then see lots of FCS errors.
- You can stop this from happening as follows:

#lsdev -C | grep fcs

Look for the 10Gb cards - in my case they showed as fcs4 and fcs5

```
# lsdev -C | grep fcs
fcs0           Available 00-00       8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs1           Available 00-01       8Gb PCI Express Dual Port FC Adapter (df1000f114108a03)
fcs2           Available 01-00       PCIe3 2-Port 16Gb FC Adapter (df1000e21410f103)
fcs3           Available 01-01       PCIe3 2-Port 16Gb FC Adapter (df1000e21410f103)
fcs4           Available 02-04       PCIe3 10Gb 4-Port FCoE Adapter (df1060e214101004)
fcs5           Available 02-05       PCIe3 10Gb 4-Port FCoE Adapter (df1060e214101004)
```

Check for the converged cards

```
# lsdev -C | grep Converged
ent0           Available 02-00       PCIe3 10GbE SFP+ SR 4-port Converged Network Adapter (df1020e214100f04)
ent1           Available 02-01       PCIe3 10GbE SFP+ SR 4-port Converged Network Adapter (df1020e214100f04)
ent2           Available 02-02       PCIe3 100/1000 Base-TX 4-port Converged Network Adapter (df1020e214103c04)
ent3           Available 02-03       PCIe3 100/1000 Base-TX 4-port Converged Network Adapter (df1020e214103c04)
```

29

29

Get rid of annoying FCS errors on 10/1Gb cards

As padmin:
$rmdev -dev fcs4 -recursive -ucfg
$rmdev -dev fcs5 -recursive -ucfg
$chdev -dev fscsi4 -attr autoconfig=defined
$chdev -dev fscsi5 -attr autoconfig=defined
After you should see:

```
# lsdev -C | grep fscsi
fscsi0         Available 00-00-01    FC SCSI I/O Controller Protocol Device
fscsi1         Available 00-01-01    FC SCSI I/O Controller Protocol Device
fscsi2         Available 01-00-01    FC SCSI I/O Controller Protocol Device
fscsi3         Available 01-01-01    FC SCSI I/O Controller Protocol Device
fscsi4         Defined   02-04-01    FC SCSI I/O Controller Protocol Device
fscsi5         Defined   02-05-01    FC SCSI I/O Controller Protocol Device
```

As root:
Note I have a log filesystem called /usr/local/logs – change this to wherever you want to save these files
#errpt >/usr/local/logs/errpt-sep122020.txt
#errpt -a >/usr/local/logs/errpta-sep122020.txt
#errclear 00
#cfgmgr
#errpt

30

30

HBA Settings

31

31

HBA Tuning

- Make the same tuning changes you would make on AIX, but VIO must be set at least as high as clients
- Set num_cmd_elems and max_xfer_size on the fiber adapters on VIO
  chdev -l fcs0 -a max_xfer_size=0x200000 -a num_cmd_elems=1024 -P
  chdev -l fcs1 -a max_xfer_size=0x200000 -a num_cmd_elems=1024 -P

Check these numbers are supported by your disk vendor

- If NPIV also set on clients
- **Client setting cannot be higher than the VIOs**
- **VIO must be rebooted to at least the client value prior to client change.**
- Pay attention to adapter layout and priorities

- NOTE – as of AIX v7.1 tl2 (or 6.1 tl8) num_cmd_elems is limited to 256 on the VFCs so set num_cmd_elems to the high number on the VIO but to no more than 256 on the NPIV clients
- See: http://www-01.ibm.com/support/docview.wss?uid=isg1IV63282
- **Increased again to 2048 in July 2016**
- http://www-01.ibm.com/support/docview.wss?uid=isg1IV76270
- This upper limit is set in the client LPAR not the VIO server
- BUT the client setting MUST NOT be larger than what is set in the VIO server

32

32

HBA max_xfer_size

The default is
0x100000 /* Default io_dma of 16MB */

After that, 0x200000,0x400000,0x80000 gets you 128MB

After that 0x1000000 checks for bus type, and you may get 256MB, or 128MB

There are also some adapters that support very large max_xfer sizes which can possibly allocate 512MB

VFC adapters inherit this from the physical adapter (generally)

Unless you are driving really large IO's, then max_xfer_size on the HBA is rarely changed beyond 0x200000 which provides a 128MB DMA

33

33

Adapter Tuning 1/2

fcs0

bus_intr_lvl	115	Bus interrupt level	False	
bus_io_addr	0xdfc00	Bus I/O address	False	
bus_mem_addr	0xe8040000	Bus memory address	False	
init_link	al	INIT Link flags	True	
intr_priority	3	Interrupt priority	False	
lg_term_dma	0x800000	Long term DMA	True	
max_xfer_size	0x100000	Maximum Transfer Size	True	(16MB DMA)
num_cmd_elems	200	Maximum number of COMMANDS to queue to the adapter	True	
pref_alpa	0x1	Preferred AL_PA	True	
sw_fc_class	2	FC Class for Fabric	True	

Changes I often make (test first)

max_xfer_size	0x200000	Maximum Transfer Size	True	*128MB DMA area for data I/O*
num_cmd_elems	1024	Maximum number of COMMANDS to queue to the adapter	True	

Often I raise this to 2048 – check with your disk vendor first
lg_term_dma is the DMA area for control I/O

34

34

Adapter Tuning 2/2

Check these are ok with your disk vendor!!! And also for the adapter.

chdev -l fcs0 -a max_xfer_size=0x200000 -a num_cmd_elems=1024 -P
chdev -l fcs1 -a max_xfer_size=0x200000 -a num_cmd_elems=1024 -P

After AIX 6.1 TL2, VFCs will always use a 128MB DMA memory area even with default max_xfer_size

DMA area (max_xfer_size) controls the max IO size the adapter can send to the disk subsystem (default is 16MB). To use the full bandwidth of the adapter this needs to be 128MB.

Remember to make changes to both VIO servers and client LPARs if using NPIV.
VIO server setting must be at least as large as the highest client setting and rebooted prior.

Remember VFCs on the client may be limited to num_cmd_elems=256 after AIX 6.1 tl8 or 7.1 tl2

35

35

My VIO Server and NPIV Client Adapter Settings

VIO SERVER connected to V7000

#lsattr -El fcs0

lg_term_dma	0x800000	Long term DMA	True
max_xfer_size	0x200000	Maximum Transfer Size	True
num_cmd_elems	1024	Max number of COMMANDS to queue to the adapter	True

NPIV Client

#lsattr -El fcs0

lg_term_dma	0x800000	Long term DMA	True
max_xfer_size	0x200000	Maximum Transfer Size	True
num_cmd_elems	256	Maximum Number of COMMAND Elements	True

NOTE NPIV client must be <= to settings on VIO

36

36

MPIO

- MPIO
  - IBM is now recommending using the AIXPCM rather than SDDPCM. No new fixes will be provided for SDDPCM after June 30, 2020. See:
    - https://www.ibm.com/support/pages/node/1106937?myns=s035&mync=E&cm_sp=s035-_-NULL-_-E
  - They have a new MPIO best practices document at:
    - https://developer.ibm.com/articles/au-aix-mpio/
  - There is a good description of AIXPCM here:
    - https://www.ibm.com/support/knowledgecenter/en/ssw_aix_72/com.ibm.aix.osdevice/devmpio.htm
  - Migration notes
    - http://www-01.ibm.com/support/docview.wss?uid=ssg1S1010646
  - Article on MPIO resiliency and problem determination
    - https://developer.ibm.com/articles/au-aix-multipath-io-mpio/
  - If you need to change defaults (i.e. from single_path to no_reserve)
    - chdef -t mpioosdisk -c disk -s fcp -a reserve_policy=no_reserve
    - Use chdef –H to check for all attributes that have been changed from defaults

37

37

Ensure you install correct drivers and MPIO

Install correct drivers for SAN and other disks
Ensure all disks are set to no_reserve and round_robin (or shortest_queue)
lsattr -El hdisk0
Check MPIO drivers
lslpp -l | grep mpio

manage_disk_drivers -l
Switch to new AIXPCM driver
manage_disk_drivers -d IBMSVC -o AIX_AAPCM
manage_disk_drivers -l

bosboot -a -d hdisk0
bootlist -m normal hdisk0
Shutdown and reactivate

```
# manage_disk_drivers -l
Device              Present Driver        Driver Options
2810XIV             AIX_AAPCM             AIX_AAPCM,AIX_non_MPIO
DS4100              AIX_APPCM             AIX_APPCM
DS4200              AIX_APPCM             AIX_APPCM
DS4300              AIX_APPCM             AIX_APPCM
DS4500              AIX_APPCM             AIX_APPCM
DS4700              AIX_APPCM             AIX_APPCM
DS4800              AIX_APPCM             AIX_APPCM
DS3950              AIX_APPCM             AIX_APPCM
DS5020              AIX_APPCM             AIX_APPCM
DCS3700             AIX_APPCM             AIX_APPCM
DCS3860             AIX_APPCM             AIX_APPCM
DS5100/DS5300       AIX_APPCM             AIX_APPCM
DS3500              AIX_APPCM             AIX_APPCM
XIVCTRL             MPIO_XIVCTRL          MPIO_XIVCTRL,nonMPIO_XIVCTRL
2107DS8K            NO_OVERRIDE           NO_OVERRIDE,AIX_AAPCM,AIX_non_MPIO
IBMFlash            NO_OVERRIDE           NO_OVERRIDE,AIX_AAPCM,AIX_non_MPIO
IBMSVC              NO_OVERRIDE           NO_OVERRIDE,AIX_AAPCM,AIX_non_MPIO
```

Now correct any single paths if still needed:
chdev -l hdisk0 -a algorithm=round_robin -a reserve_policy=no_reserve -P
chdev -l hdisk1 -a algorithm=round_robin -a reserve_policy=no_reserve -P
etc
bosboot -a -d hdisk0
bootlist -m normal hdisk0
shutdown -r now

38

38

lsmpio

```
# lspv | grep hdisk1
hdisk1          00f95d3a425513d5                    rootvg          active

# lsdev -C | grep hdisk1
hdisk1          Available 90-T1-01    MPIO IBM 2076 FC Disk

# lsmpio -ql hdisk1
Device:  hdisk1
        Vendor Id:  IBM
       Product Id:  2145
         Revision:  0000
         Capacity:  50.00GiB                                                    Shows size
     Machine Type:  2078
     Model Number:  124
      Host Group:  P8NIM                                                        Disk subsystem Group name
     Volume Name:  NIM_ROOTVG                                                   Disk subsystem volume name
   Volume Serial:  60050763808100F70000000000000009  (Page 83 NAA)              Sows serial number for LUN
```

39

39

lsmpio

Shows if paths are optimized or not
lsmpio -l hdisk1

```
# lsmpio -l hdisk1
name    path_id  status   path_status  parent  connection

hdisk1  0        Enabled  Non          fscsi0  500507680d048ef6,1000000000000
hdisk1  1        Enabled  Sel,Opt      fscsi0  500507680d048ef7,1000000000000
hdisk1  2        Enabled  Non          fscsi1  500507680d088ef6,1000000000000
hdisk1  3        Enabled  Sel,Opt      fscsi1  500507680d088ef7,1000000000000
hdisk1  8        Enabled  Non          fscsi2  500507680d108ef6,1000000000000
hdisk1  9        Enabled  Sel,Opt      fscsi2  500507680d108ef7,1000000000000
hdisk1  10       Enabled  Non          fscsi3  500507680d0c8ef6,1000000000000
hdisk1  11       Enabled  Sel,Opt      fscsi3  500507680d0c8ef7,1000000000000
```

Use chpath if you need to change priorities for paths

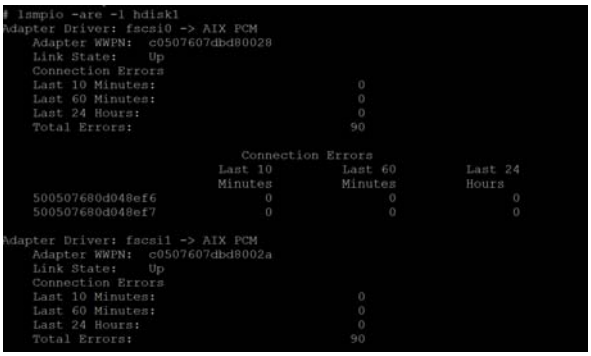
You can also find the parent for an hdisk and get statistics on each of the paths using:
lsmpio -are -l hdisk1
This will show a long report that includes all the paths. An example of part of one of those reports is on the next slide:

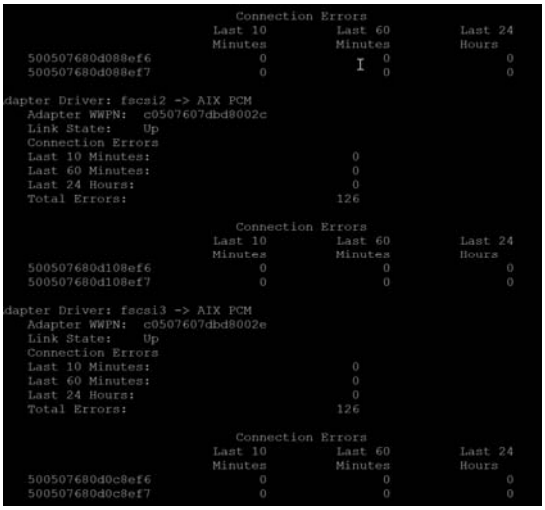
Other options:
lsmpio -l hdisk1 –Sd Detailed path statistics for the hdisk

40

40

lsmpio –are –l hdisk1





41

41

lspath Output

lspath was updated with new –t and –i flags. The –t flag ensures the pathid is listed at the end, and the –i flag allows you to specify the pathid and only get information for devices on that specific path. The mkpath and rmpath commands were also updated to include the option to process specific pathids.

lspath –t –l hdisk2	Include pathid in report
lspath -l hdisk2 –i 4	Only show path 4
lspath –i 4	Report on all disks on path 4

```
# lspath -l hdisk2
Enabled hdisk2 fscsi2
Enabled hdisk2 fscsi2
Enabled hdisk2 fscsi3
Enabled hdisk2 fscsi3
# lspath -t -l hdisk2
Enabled hdisk2 fscsi2 4
Enabled hdisk2 fscsi2 5
Enabled hdisk2 fscsi3 6
Enabled hdisk2 fscsi3 7
# lspath -l hdisk2 -i 4
Enabled hdisk2 fscsi2
# lspath -i 4
Enabled hdisk0 fscsi2
Enabled hdisk1 fscsi2
Enabled hdisk2 fscsi2
Enabled hdisk3 fscsi2
```

42

42

Remove Missing or Failed Paths

The script below creates a script that you can check and then run to clean up Missing or Failed paths

```
#!/bin/ksh
# badpaths - create cleanup scripts
>removepaths.sh
disks=$(lspv | awk '{print $1}')
for loop in $disks
do
lspath -l $loop -H -F "name:parent:connection:status" |grep Missing| awk -F: '{print "rmpath -dl",$1,"-p", $2, "-w",
$3}'>>removepaths.sh
lspath -l $loop -H -F "name:parent:connection:status" |grep Failed| awk -F: '{print "rmpath -dl",$1,"-p", $2, "-w",
$3}'>>removepaths.sh
done
exit 0
```

Based on article by David Tansley
https://ibmsystemsmag.com/Power-Systems/12/2014/removing-failed-missing-paths

43

43

lsmpio –are –l hdisk2

```
# lsmpio -are -l hdisk2
Adapter Driver: fscsi3 -> AIX PCM
    Adapter WWPN:  10000090fa530976
    Link State:    Up
    Connection Errors
    Last 10 Minutes:                    0
    Last 60 Minutes:                    0
    Last 24 Hours:                      0
    Total Errors:                       0

                          Connection Errors
                      Last 10         Last 60         Last 24
                      Minutes         Minutes         Hours
    500507680d0c8ef6        0               0               0
    500507680d0c8ef7        0               0               0

Adapter Driver: fscsi2 -> AIX PCM
    Adapter WWPN:  10000090fa530975
    Link State:    Up
    Connection Errors
    Last 10 Minutes:                    0
    Last 60 Minutes:                    0
    Last 24 Hours:                      0
    Total Errors:                       0

                          Connection Errors
                      Last 10         Last 60         Last 24
                      Minutes         Minutes         Hours
    500507680d108ef6        0               0               0
    500507680d108ef7        0               0               0
```

44

44

VIOS and NVMe

One or more NVMe devices can be assigned to your VIOS partition
S922 and S924 support up to 14 (four U.2 NVMe plus up to ten PCIe add in cards) NVMe adapters
S924 supports the base plus up to seven PCIe add in cards (max of 11)
Depending on backplane can have 0, 2 or 4 NVMe drives on the backplane –
NVMe is a high-speed flash storage which comes in various levels of write endurance so f you plan to use them get 4 and mirror

1. Device can be used as a VIOS boot device
2. Device can be configured as a local read cache in the SSP (shared storage pool)
3. NVMe disks cannot be used in the actual SSP pool
4. LVs (logical volumes) can be carved out and assigned to clients as LV backed vSCSI devices
5. NVMe Physical volumes CANNOT be assigned to a client as a PV backed vSCSI device
6. Client LPARs backed by vSCSI NVMe devices cannot be used in LPM operations
7. NVMe devices cannot be used for AMS (active memory sharing) devices

NVMe devices used as VIOS boot devices should be mirrored
In some Power 9 systems expansion cards are used to hold M.2 form factor NVMe devices. It is advisable to mirror across expansion cards to protect from expansion card failure

IBM S922, S914 and S924 Technical Overview and Introduction Featuring PCIe Gen4 Technology
http://www.redbooks.ibm.com/redpapers/pdfs/redp5595.pdf

Blog on NVMe Support for VIOS
https://community.ibm.com/community/user/power/blogs/ninad-palsule1/2020/07/25/nvme-device-support-in-virtual-io-server-vios

45

45

NVMe and VIOS

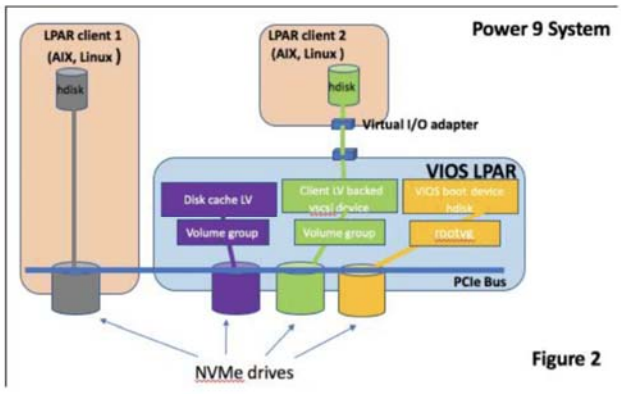


Figure 2

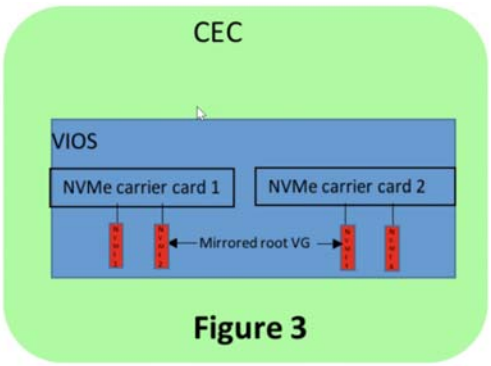


Figure 3

Note: Each NVMe drive is a separate PCIe endpoint and can be assigned individually to a unique AIX, VIOS, or Linux logical partition (LPAR).

Diagrams from Blog on NVMe Support for VIOS
https://community.ibm.com/community/user/power/blogs/ninad-palsule1/2020/07/25/nvme-device-support-in-virtual-io-server-vios

46

46

Network

47

47

Virtual Ethernet

Link aggregation

- Put vio1 aggregate on a different switch to vio2 aggregate
- Provides redundancy without having to use NIB
- Allows full bandwidth and less network traffic (NIB is pingy)
- Basically SEA failover with full redundancy and bandwidth

Pay attention to entitlement

- VE performance scales by entitlement not VPs (in VIO and client)

If VIOS is only handling the network, then disable network threading on the virtual Ethernet

- chdev –dev ent? thread=0
- Non threaded improves LAN performance
- Threaded (default) is best for mixed vSCSI and LAN

http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/perf.html

48

48

Large Receive and Send

Turn on large send on VE adapters
chdev –dev ent? –attr large_send=yes
Turn on large send on the SEA
chdev –dev entx –attr largesend=1

NOTE do not do this if you are supporting Linux or IBM i LPARs with the VE/SEA without checking first

See http://tinyurl.com/gpe5zgd for information for Linux and Large send/receive
This provides information on correctly using Large send and Large receive with Linux

Also http://tinyurl.com/lm6x5er for info for large send in general and also IBM i

LRO (large receive offload) is enabled by default on virtual ethernet resources
The issue with IBM i and LRO was resolved in the base code for IBM I 7.2 so you are safe to enable LRO at that level
Below IBM i 7.2 there are patches you need to install

49

49

VIO 2.2.3 and above SEA Changes

Traditional SEA setup
ent0-3 are the physical adapters
ent4 is the virtual adapter defined at the HMC with external access
(SEA goes here)
VIO1 is priority 1 and VIO2 is priority 2
ent5 is the virtual adapter on Vlan 1 with no external
(IP will go here)
ent6 is the control channel on vlan 255 or you can leave this out and let it default to 4095 on mkvdev
OLD
Add a virtual network to the profile to be used for the control channel (used vlan 255 in this case)
mkvdev –sea ent0 –vadapter ent4 –default ent4 –defaultid 1 –attr ha_mode=auto ctl_chan=ent6
Creates ent7 as the SEA and uses ent6 for the control channel
NEW
mkvdev –sea ent0 –vadapter ent4 –default ent4 –defaultid 1 –attr ha_mode=auto
Above creates ent7 as SEA and defaults to vlan 4095 for control channel

Do not mess up priorities or ctl_chan or you will cause a spanning tree loop

Update with 2.2.3
See chapter 4 of SG248198- Redbook on 2.2.3 Enhancements

SEA setup has been simplified
Requirement removed for dedicated control channel and VLAN ID for each SEA failover configuration
Multiple SEA pairs can now share VLAN 4095 within the same virtual switch and no ctl_chan is needed
HMC (>= 7.8) reserves 4095 for internal management traffic
Requires VIOS 2.2.3, HMC 7.7.8 and firmware 780 or higher
Not available on 770/780 B models

50

50

Check SEA and Control Channel

```
# lsdev -C | grep ent
ent0           Available 02-00       PCIe3 10GbE SFP+ SR 4-port Converged Network Adapter (df1020e214100f04)
ent1           Available 02-01       PCIe3 10GbE SFP+ SR 4-port Converged Network Adapter (df1020e214100f04)
ent2           Available 02-02       PCIe3 100/1000 Base-TX 4-port Converged Network Adapter (df1020e214103c04)
ent3           Available 02-03       PCIe3 100/1000 Base-TX 4-port Converged Network Adapter (df1020e214103c04)
ent4           Available             Virtual I/O Ethernet Adapter (l-lan)
ent5           Available             Virtual I/O Ethernet Adapter (l-lan)
ent6           Available             Virtual I/O Ethernet Adapter (l-lan)
ent7           Available             Virtual I/O Ethernet Adapter (l-lan)
ent8           Available             EtherChannel / IEEE 802.3ad Link Aggregation
ent9           Available             Shared Ethernet Adapter
```

```
# entstat -d ent9 | grep "Control Channel"


    Control Channel PVID: 4095
Control Channel Adapter: ent4
# entstat -d ent9 | grep "VLAN ID"


Enabled VLAN IDs: None
Enabled VLAN IDs: None
Invalid VLAN ID Packets: 0
Port VLAN ID:     1
Invalid VLAN ID Packets: 0
Port VLAN ID:     1
```

```
$ entstat -all ent9 | grep "VLAN Tag ID"
 VLAN Tag IDs:     2     5    20    21    22    23    24    25    26
 VLAN Tag IDs:     2     5    20    21    22    23    24    25    26
```

51

51

Correct setup of switch ports for SEA Failover

- Incorrect setup can cause SEA failover delays which may disconnect users
- https://www.ibm.com/support/pages/sea-failover-delayed-and-supported-methods-test-failover
- If Spanning Tree is turned on, then Portfast also should be turned on.
  - Portfast is a Cisco term
  - The IBM switch equivalent is Spanning Tree Edge
- To reduce failback time you may want to turn Spanning Tree off

52

52

SEA Failover Testing

After you set up a VIO pair you should test that failover is working
The quick way is to ssh to the primary and reboot the secondary and see if you lose your ssh connection
Then ssh to the secondary and reboot the primary and see if you lose the ssh connection

If the switch ports are set up correctly and the SEA is defined correctly there should be no issues

Also check errpt to make sure that the primary and secondary (backup) correctly become primary and backup during the process

When performing maintenance on the primary network VIO LPAR, updating network adapter firmware on the primary or rebooting it, I normally force a failover manually by doing the following:
If the SEA is set up correctly then this is not necessary, but it ensures the connectivity of the client partitions that are using the SEA in PRIMARY state do not lose connectivity temporarily. It also lets me quickly know if there is an issue.

Prior to maintenance on the primary VIOS (reboots and updating I/O firmware)
1. Set ha_mode to standby on primary VIOS with chdev command:
$ chdev -dev entX -attr ha_mode=standby

After reboot or maintenance is complete:
2. Reset it back to auto and the SEA should fail back to the primary VIOS:
$ chdev -dev entX -attr ha_mode=auto

53

53

Starter set of tunables - Network

Typically we set the following – they are not set by default:

NETWORK
no -p -o rfc1323=1
no -p -o tcp_sendspace=262144
no -p -o tcp_recvspace=262144
no -p -o udp_sendspace=65536
no -p -o udp_recvspace=655360

Also check the actual NIC interfaces and make sure they are set to at least these values
You can't set udp_sendspace > 65536 as IP has an upper limit of 65536 bytes per packet

54

54

Network Performance and Throughput

- Depends on:
  - Available CPU power
    - **Scales by entitlement not by VPs**
  - MTU size
  - Distance between receiver and sender
  - Offloading features
  - Coalescing and aggregation features
  - TCP configuration
  - Firmware on adapters and server
  - Ensuring all known efixes are on for 10GbE issues

- Network Performance Presentation at:
- http://youtu.be/8pth2ujGWK0
- http://www.circle4.com/movies/networkperf/networkperf.pdf

55

55

More complex Networking

Resources, All Systems, Server, Virtual Networks

The hypervisor provides an IEEE 802.1Q virtual LAN style virtual Ethernet switch. When you add a virtual network, you can add a virtual switch. The default is ethernet0.
You can add Virtual networks, switches, bridges and VLANs

You can also create your etherchannels here
Networks can be internal or bridged
They can have no tagging or 802.1q tagging
They can be set for single VIO, dual VIO, loadsharing (load groups) or just regular
You can create VLANs and spread them across networks

https://www.ibm.com/support/knowledgecenter/TI0002C/p8efd/p8efd_virt_switch_concept.htm
https://www.ibm.com/support/knowledgecenter/en/9119-MHE/p8efd/p8efd_add_new_virt_net_wizard_task.htm

There are also options such as vnic and SR-IOV – Alexander Paul has great presentations dedicated to these
NOTE – map out all your networks and vlans before you even consider going down this path – it gets complicated very fast

56

56

SEA Loadsharing

Allows you to have some VLANs primary on VIO1 and backed up on VIO2, with other VLANs primary on VIO2 and backed up on VIO1.
This lets you take advantage of the full bandwidth of the adapters
https://www.ibm.com/support/pages/how-setup-sea-failover-load-sharing-configuration

2 options
1. Use ha_mode=sharing as per the above – let you have a single SEA with multiple VLANs
   ha_mode=sharing must be set on the primary SEA before the backup
   SEA evenly divides traffic up by adapter (not VLAN) between the two VIO LPARs

2. Define two SEAs (with their own adapters)
   SEA1 is primary on VIO1 (priority=1)
   SEA2 is primary on VIO2 (priority=1)
   Assign SEAs to clients depending on which vio you want to be primary for that client

57

57

Monitoring

58

58

CPU and Memory

- Remember VIO scales by entitlement not VPs
- Ensure sufficient entitlement
- Watch for VCSWs – this is a sign of entitlement shortage
- If running close to entitlement on average increase entitlement
  - Check your %idle first as you can go over entitlement if you have too many VPs, yet not be using all the threads on the cores
- If running close to VPs on average increase entitlement and VPs
- Consider running dedicated

- NEVER EVER let your VIO server page
- Clean up the VIO server page spaces
- Plan for cores and memory for VIO servers when sizing systems
  - At least 2-3 cores for a pair and 8Gb minimum each

59

59

nmon Monitoring

- nmon -ft –AOPV^dML -s 15 -c 120
  - Grabs a 30 minute nmon snapshot
  - A is async IO
  - M is mempages
  - t is top processes
  - L is large pages
  - **O is SEA on the VIO**
  - P is paging space
  - V is disk volume group
  - d is disk service times
  - ^ is fibre adapter stats
  - W is workload manager statistics if you have WLM enabled you can add this

If you want a 24 hour nmon use:

nmon -ft –AOPV^dML -s 150 -c 576

May need to enable accounting on the SEA first – this is done on the VIO
chdev –dev ent* -attr accounting=enabled

Can use entstat/seastat or topas/nmon to monitor – this is done on the vios
topas –E
nmon -O

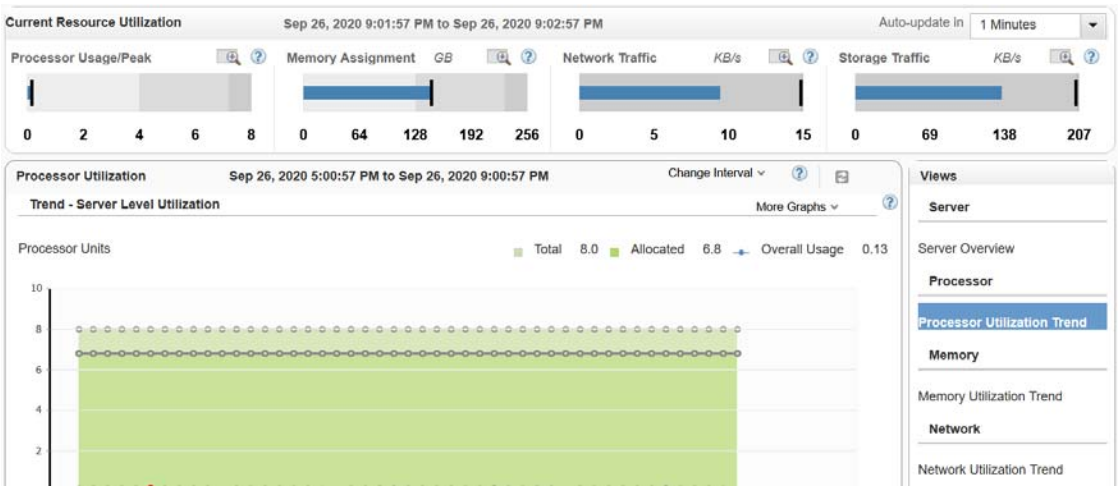
VIOS performance advisor also reports on the SEAs

60

60

HMC PCM Monitoring

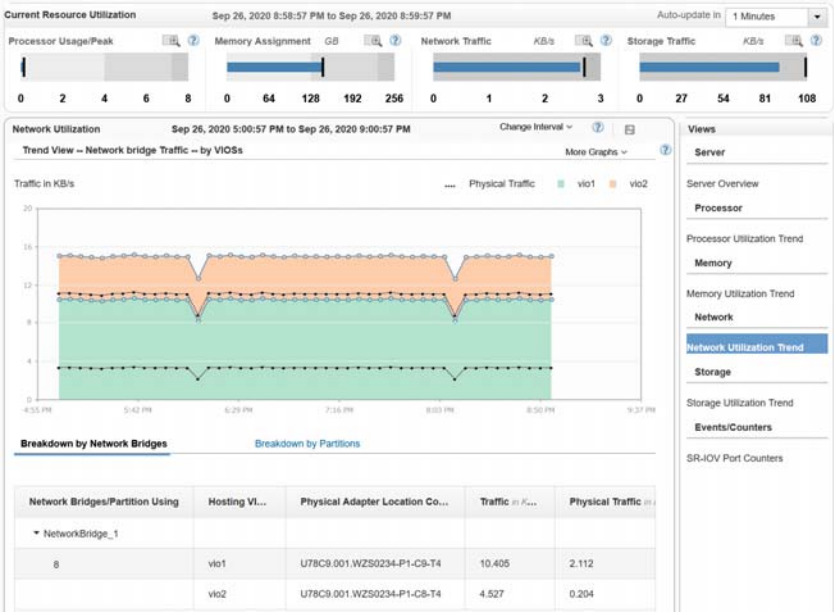
You can use the HMC to get an overall view of the LPARs and VIO servers
https://www.ibm.com/support/knowledgecenter/POWER8/p8efe/p8efe_kickoff.htm



61

61

HMC PCM Monitoring



62

62

Shared Processor Pool Monitoring

Turn on "Allow performance information collection" on the LPAR properties
- This is a dynamic change
- Without this being set on every LPAR the cross LPAR statistics won't be correct
- This includes APP and other statistics reported by nmon and lparstat, etc

topas –C

Uses the xmquery system so it needs to be uncommented
- grep xm /etc/inetd.conf
- xmquery dgram udp6 wait root /usr/bin/xmtopas xmtopas -p9

Limited to lpars on same server

Most important value is app – available pool processors
- This represents the current number of free physical cores in the pool

nmon option p for pool monitoring
- To the right of PoolCPUs there is an unused column which is the number of free pool cores

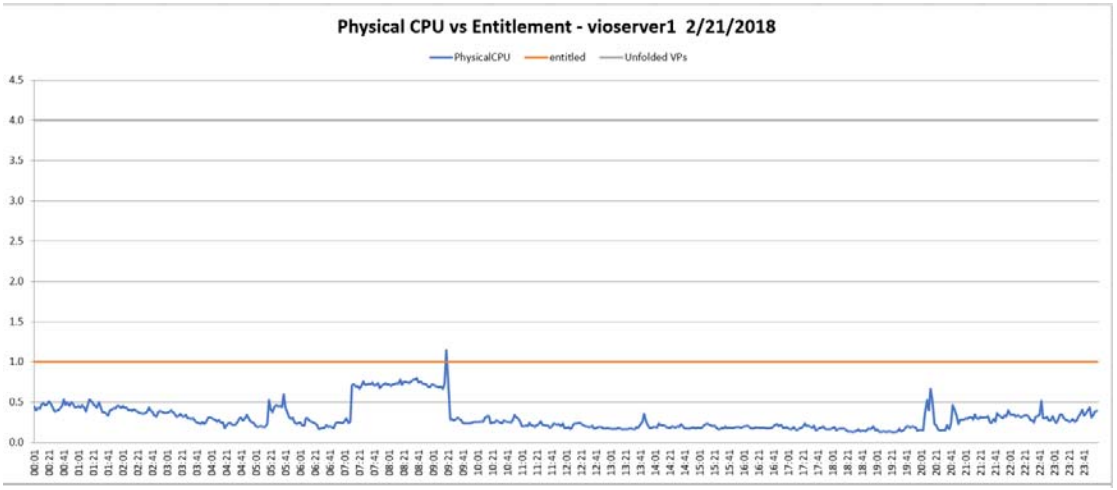
nmon analyser LPAR Tab

lparstat
- Shows the app column and poolsize

63

63

nmon Analyser LPAR Tab



64

64

NPIV Statistics

- Normally need to use nmon to get information at each client LPAR
- Could also use –O when recording

- BUT as of v2.2.3
- VIOS Performance advisor supports NPIV aggregation information
- http://www-01.ibm.com/support/knowledgecenter/POWER7/p7hcg/fcstat.htm?cp=POWER7%2F1-8-3-8-2-60
- fcstat –n wwpn device_name
- i.e. fcstat –n C05012345678000 fcs0
- Provides statistics at the WWPN for the virtual adapter
- You can also try fcstat –client as padmin

Shows all clients for the vio, the WWPNs, statistics and error counts

- Also check out NPIVGRAPH for visualizing NPIV mappings:
- http://npivgraph.sourceforge.net/

- Review options on fcstat – fcstat –d and fcstat –e provide additional statistics on adapter usage
- https://www.ibm.com/support/knowledgecenter/en/ssw_aix_61/com.ibm.aix.cmds2/fcstat.htm

65

65

fcstat -client on VIO1

```
$ fcstat -client
             hostname   dev              wwpn     inreqs     outreqs ctrlreqs         inbytes        outbytes  DMA_errs Elem_errs Comm_errs
                 vio1  fcs0  0x10000090FA530BE2  249514977      382899 12358894   2914664220462      2433855304         0         0         0
              aix1nim  fcs0  0xC0507607DBD80028     526500      682335   291385     15700556832     30795659264         0         0         0

                 vio1  fcs1  0x10000090FA530BE3  232751337      519335 12359140   2791124369628     13255021624         0         0         0
              aix1nim  fcs1  0xC0507607DBD8002A     266405      579636   233128      8799626368     23545063936         0         0         0
```

fcstat -client on VIO2

```
                 vio2  fcs2  0x10000090FA530975  207302289      837522 14032486   2217157199880      4353958928         0         0         0
              aix1nim  fcs2  0xC0507607DBD8002C     266109      580393   233208      8817861760     23620933120         0         0         0

                 vio2  fcs3  0x10000090FA530976  224909138     1038592 14032077   2311405392854     13237805720         0         0         0
              aix1nim  fcs3  0xC0507607DBD8002E     266156      580341   233216      8776368768     23605806080         0         0         0
```

66

66

SEA Statistics

- Output in part command
- SEA Adapter Statistics
  - Provides details on the meaning of statistics for the SEA
  - https://www.ibm.com/support/knowledgecenter/POWER9/p9hb1/p9hb1_statssea.htm
- SEA Adapter Failover Statistics
- Provides details on meaning of output from entstat -all specific to SEA Failover Statistcs
  - https://www.ibm.com/support/knowledgecenter/POWER9/p9hb1/p9hb1_statsseafailover.htm
- VIO Server Network Attributes Information
  - https://www.ibm.com/support/knowledgecenter/POWER9/p9hb1/p9hb1_vios_managing_sea_attr.htm

67

67

netstat –v (vio or client Virtual Ethernet)

SEA

Transmit Statistics:	Receive Statistics:
Packets: 83329901816	Packets: 83491933633
Bytes: 87482716994025	Bytes: 87620268594031
Interrupts: 0	Interrupts: 18848013287
Transmit Errors: 0	Receive Errors: 0
Packets Dropped: 0	**Packets Dropped: 67836309**
	Bad Packets: 0

Max Packets on S/W Transmit Queue: 374
S/W Transmit Queue Overflow: 0
Current S/W+H/W Transmit Queue Length: 0

Elapsed Time: 0 days 0 hours 0 minutes 0 seconds

Broadcast Packets: 1077222	Broadcast Packets: 1075746
Multicast Packets: 3194318	Multicast Packets: 3194313
No Carrier Sense: 0	CRC Errors: 0
DMA Underrun: 0	DMA Overrun: 0
Lost CTS Errors: 0	Alignment Errors: 0
Max Collision Errors: 0	**No Resource Errors: 67836309**

check those tiny, etc Buffers
Check on client LPARs too

Virtual I/O Ethernet Adapter (l-lan) Specific Statistics:

Hypervisor Send Failures: 4043136
Receiver Failures: 4043136
Send Errors: 0
Hypervisor Receive Failures: 67836309

"No Resource Errors" can occur when the appropriate amount of memory can not be added quickly to vent buffer space for a workload situation.
You can also see this on LPARs that use virtual Ethernet without an SEA

68

68

Buffers – check at vio and client

Virtual Trunk Statistics
Receive Information
Receive Buffers

Buffer Type	Tiny	Small	Medium	Large	Huge
Min Buffers	512	512	128	24	24
Max Buffers	2048	**2048**	256	64	64
Allocated	513	2042	128	24	24
Registered	511	506	128	24	24
History					
Max Allocated	532	**2048**	128	24	24
Lowest Registered	502	354	128	24	24

"Max Allocated" represents the maximum number of buffers ever allocated
"Min Buffers" is number of pre-allocated buffers
"Max Buffers" is an absolute threshhold for how many buffers can be allocated

chdev –l <veth> -a max_buf_small=4096 –P
chdev –l <veth> -a min_buf_small=2048 –P
Above increases min and max small buffers for the virtual ethernet adapter configured for the SEA above
Needs a reboot

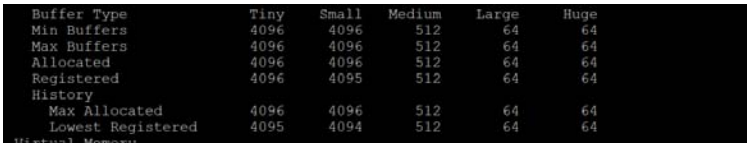
Max buffers is an absolute threshold for how many buffers can be allocated
Use entstat –d (-all on vio) or netstat –v to get this information
entstat –d ent7 (where ent7 is the SEA) gets you the information for ent7 only

69

69

Tune Virtual Buffers

The new default is buf_mode=max_min
This sets the min and max settings to be equal – the issue is you cannot tell if you have a problem as max allocated is always the same as allocated
I create a script called /etc/tunables/rc-tunebufs.sh
It is customized for the virtual adapters I want to control
#! /bin/ksh
#
chdev -l ent5 -a buf_mode=min -P
chdev -l ent4 -a buf_mode=min -P
chdev -l ent5 -a max_buf_tiny=4096 -P
chdev -l ent5 -a max_buf_small=4096 -P
chdev -l ent5 -a max_buf_medium=512 -P
chdev -l ent4 -a max_buf_tiny=4096 -P
chdev -l ent4 -a max_buf_small=4096 -P
chdev -l ent4 -a max_buf_medium=512 -P



Buffer Type	Tiny	Small	Medium	Large	Huge
Min Buffers	4096	4096	512	64	64
Max Buffers	4096	4096	512	64	64
Allocated	4096	4096	512	64	64
Registered	4096	4095	512	64	64
History					
Max Allocated	4096	4096	512	64	64
Lowest Registered	4095	4094	512	64	64

70

70

part command

https://www.ibm.com/support/knowledgecenter/en/POWER8/p8hcg/p8hcg_part.htm
Part is used to gather VIO specific statistics and to create a .xml file that shows the information gathered

On vios as padmin run:

part {-i *interval* | -i *filename*} [-t *level*] [**-help** | -?]
Minimum of 10 minuted, maximum of 30 minutes

part –i 30
part: Reports are successfully generated in vio1_181007_15_22_16.tar
Runs a 30 minute nmon and gathers other data during that 30 minutes
Creates the .nmon file and a .xml file then tars them up into the current directory

You can run it against a current .nmon file as follows:
part –f filename.nmon

Output fields are described at:
https://www.ibm.com/support/knowledgecenter/en/POWER8/p8hb1/p8hb1_vios_perf_adv_reports.htm

71

71

part command

Advisory Report

System - Configuration

Name	Value
Processor Family	Architecture PowerPC Implementation POWER7_COMPAT_mode 64 bit
Server Model	IBM 8286-41A
Server Frequency	3724.0 MHz
Server - Online CPUs	3.0 cores
Server - Maximum Supported CPUs	3.0 cores
VIOS Level	2.2.6.21
VIOS Advisor Release	0.1

VIOS - I/O Activity

Name	Value
Disk I/O Activity	Insufficient Data from recording
Network I/O Activity	[Average Send: 0 @ 0.0 MBps , Average Receieve: 0 @ 0.0MBps] [Peak Send: 0 @ 0.0 MBps , Peak Receive: 0 @ 0.0MBps]

72

72

part command

VIOS - Disk Adapters

Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
	FC Adapter Count	2		10/07/2018 03:22 PM			
	FC I/O Operations per second	569 @ 38 KB		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	FC Adapter Utilization	optimal					
	NPIV Client Utilization - fcs1	High: 0.00 % Average: 0.00 %		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	NPIV Client Utilization - fcs0	High: 0.00 % Average: 0.00 %		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	FC I/O Operations Blocked	optimal		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	FC Port Speeds	running at full speed		10/07/2018 03:22 PM	10/07/2018 03:32 PM		

VIOS - Disk Drives

Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
	Physical Drive Count	19		10/07/2018 03:22 PM			
	I/O Operations Blocked	pass		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	Long I/O Latency	pass		10/07/2018 03:22 PM	10/07/2018 03:32 PM		

73

73

part command

VIOS - Disk Adapters

Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
	FC Adapter Count	2		10/07/2018 03:22 PM			
	FC I/O Operations per second	569 @ 38 KB		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	fcs1	Average : 282 @ 19 KB		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	fcs0	Average : 287 @ 18 KB		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	FC Adapter Utilization	optimal					
	FC Adapter Utilization (fcs1)	high:12.9% (799.0 @ 73.9K)		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	FC Adapter Utilization (fcs0)	high:13.3% (822.0 @ 67.2K)		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	NPIV Client Utilization - fcs1	High: 0.00 % Average: 0.00 %		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	aix1nim	Average 0 iops @ 0 KB Peak: 0 iops @ 0 KB		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	[redacted]	Average 0 iops @ 0 KB Peak: 0 iops @ 0 KB		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	NPIV Client Utilization - fcs0	High: 0.00 % Average: 0.00 %		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	aix1nim	Average 0 iops @ 0 KB Peak: 0 iops @ 0 KB		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	[redacted]	Average 0 iops @ 0 KB Peak: 0 iops @ 0 KB		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	FC I/O Operations Blocked	optimal		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
	FC Port Speeds	running at full speed		10/07/2018 03:22 PM	10/07/2018 03:32 PM		

74

74

part command

VIOS - Processor

Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
✓	CPU Capacity	1.5 ent		10/07/2018 03:22 PM			
i	CPU consumption	Average:1.4% (cores:0.1) High:2.7% (cores:0.1)		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
i	Processing Mode	Shared CPU, (UnCapped)		10/07/2018 03:22 PM			
✓	Variable Capacity Weight	255		10/07/2018 03:22 PM			
✓	Virtual Processors	3		10/07/2018 03:22 PM			
✓	SMT Mode	SMT4		10/07/2018 03:22 PM			

System - Shared Processing Pool

Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
✓	Shared Pool Monitoring	enabled		10/07/2018 03:22 PM			
i	Shared Processor Pool Capacity	8.0 ent.		10/07/2018 03:22 PM			
✓	Free CPU Capacity	average_free:7.7 ent. lowest_free:7.3 ent.					

75

75

part command

VIOS - Memory

Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
✓	Real Memory	8.000 GB		10/07/2018 03:22 PM			
i	Available Memory	1.867 GB		10/07/2018 03:22 PM			
✓	Paging Rate	0.0 MBps Paging Rate		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
✓	Paging Space Size	4.000 GB		10/07/2018 03:22 PM	10/07/2018 03:32 PM		
i	Free Paging Space	3.880 GB free		10/07/2018 03:22 PM			
⚠	Pinned Memory	5.488 GB pinned	less than 4.000 GB pinned	10/07/2018 03:22 PM	10/07/2018 03:32 PM	1	4

VIOS - Shared Ethernet Adapters

Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
i	SEA Adapter Count	1		10/07/2018 03:22 PM			
⚠	⊞ SEA (ent7)	Mapping: Physical :(ent2,ent3),Virtual : (ent4,ent4)		10/07/2018 03:22 PM			

76

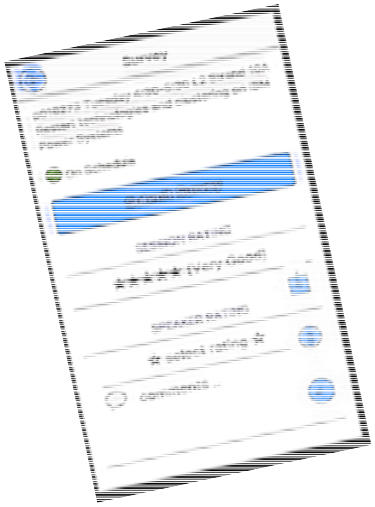
76

Thank you!

Jaqui Lynch

jlynch@flagshipsg.net

Please complete the Session Evaluation!



77

77

Thank you for your time

If you have questions please email me at:
jaqui@circle4.com or jlynch@flagshipsg.net

Also check out:
http://www.circle4.com/movies/

Copy of presentation at:
http://www.circle4.com/ptechu/ vioscare-part2-oct042020.pdf

And the Virtual User Group
https://www.ibm.com/support/pages/node/1120377

78

78

78

Useful Commands, Links and Documentation

79

79

USEFUL COMMANDS

80

80

Useful Commands

Command History
$ fc -l
725 lsrep
726 backupios -file /usr/local/backups/b750viobkp
727 exit
728 lsmap -vadapter vhost0
729 fc –l

Global command log
$ lsgcl | grep "Aug 9 2013"
Aug 9 2013, 08:25:35 root ioslevel
Aug 9 2013, 08:59:22 padmin license
Aug 9 2013, 09:00:29 padmin lsmap -vadapter vhost0
Aug 9 2013, 09:01:29 padmin lsgcl

Redirecting output when running as padmin
lsmap –all –npiv | tee npivdata.txt

81

81

Useful Commands

vSCSI Commands

mkvdev -vdev hdisk2 -vadapter vhost0

mkvdev –fbo –vadapter vhost0

NPIV

Setup NPIV mappings

vfcmap –vadapter vfchost0 –fcp fcs0

lsmap –npiv –all

lsmap –vadapter vfchost0 –npiv

lsdev –virtual

lsnports

lsdev –slots

lscfg –vpl vfchost0

82

82

Useful Commands

$ lsdev -virtual

```
name            status      description
ent5            Available   Virtual I/O Ethernet Adapter (l-lan)
ent6            Available   Virtual I/O Ethernet Adapter (l-lan)
ent7            Available   Virtual I/O Ethernet Adapter (l-lan)
vasi0           Available   Virtual Asynchronous Services Interface (VASI)
vbsd0           Available   Virtual Block Storage Device (VBSD)
vfchost0        Available   Virtual FC Server Adapter
vfchost1        Available   Virtual FC Server Adapter
vhost0          Available   Virtual SCSI Server Adapter
vhost1          Available   Virtual SCSI Server Adapter
vsa0            Available   LPAR Virtual Serial Adapter
b740ios1_rv1    Available   Virtual Target Device - Logical Volume
b740l1_rv1      Available   Virtual Target Device - Logical Volume
vtopt0          Available   Virtual Target Device - File-backed Optical
vtopt1          Available   Virtual Target Device - File-backed Optical
vtscsi0         Available   Virtual Target Device - Disk
vtscsi1         Available   Virtual Target Device - Disk
vtscsi2         Available   Virtual Target Device - Disk
vtscsi3         Available   Virtual Target Device - Disk
ent8            Available   Shared Ethernet Adapter
```

83

83

Useful Commands

$ lsmap -vadapter vhost0

```
SVSA            Physloc                                     Client Partition ID
--------------- ------------------------------------------- ------------------
vhost0          U8205.E6B.1093XXX-V1-C21                    0x00000003

VTD             b740l1_rv1
Status          Available
LUN             0x8300000000000000
Backing device  lv_b740l1
Physloc
Mirrored        N/A

VTD             vtopt0
Status          Available
LUN             0x8200000000000000
Backing device
Physloc
Mirrored        N/A

VTD             vtopt1
Status          Available
LUN             0x8100000000000000
Backing device
Physloc
Mirrored        N/A
```

84

84

Useful Commands

$ lsmap -vadapter vfchost0 -npiv

```
Name        Physloc                          ClntID ClntName       ClntOS
----------- -------------------------------- ------ -------------- -------
vfchost0    U8205.E6B.1093XXX-V1-C31              3

Status:NOT_LOGGED_IN
FC name:fcs0                FC loc code:U78AA.001.WZSG8XX-P1-C5-T1
Ports logged in:0
Flags:4<NOT_LOGGED>
VFC client name:            VFC client DRC:
```

$ lsmap -vadapter vfchost4 -npiv

```
Name        Physloc                          ClntID ClntName       ClntOS
----------- -------------------------------- ------ -------------- -------
vfchost4    U8205.E6B.1093XXX-V1-C36              8 b740nl1       AIX

Status:LOGGED_IN
FC name:fcs0                FC loc code:U78AA.001.WZSG8XX-P1-C5-T1
Ports logged in:3
Flags:a<LOGGED_IN,STRIP_MERGE>
VFC client name:fcs0        VFC client DRC:U8205.E6B.1093XXX-V8-C36
```

85

85

Useful Commands

$ lsnports

```
name            physloc                        fabric tports aports swwpns awwpns
fcs0            U78AA.001.WZSG8XX-P1-C5-T1          1     64     63    2048    2041
```

$ lsdev -slots

```
# Slot                  Description       Device(s)
HEA 1                   Logical I/O Slot  lhea0 ent0
U8205.E6B.1093XXX-V1-C0     Virtual I/O Slot  vsa0
U8205.E6B.1093XXX-V1-C11    Virtual I/O Slot  ent5
U8205.E6B.1093XXX-V1-C12    Virtual I/O Slot  ent6
U8205.E6B.1093XXX-V1-C13    Virtual I/O Slot  ent7
U8205.E6B.1093XXX-V1-C21    Virtual I/O Slot  vhost0
U8205.E6B.1093XXX-V1-C22    Virtual I/O Slot  vhost1
U8205.E6B.1093XXX-V1-C23    Virtual I/O Slot  vhost2
U8205.E6B.1093XXX-V1-C31    Virtual I/O Slot  vfchost0
U8205.E6B.1093XXX-V1-C32    Virtual I/O Slot  vfchost1
U8205.E6B.1093XXX-V1-C33    Virtual I/O Slot  vfchost2
U8205.E6B.1093XXX-V1-C32769 Virtual I/O Slot  vasi0
U8205.E6B.1093XXX-V1-C32773 Virtual I/O Slot  vasi1
U8205.E6B.1093XXX-V1-C32774 Virtual I/O Slot  vasi2
U8205.E6B.1093XXX-V1-C32775 Virtual I/O Slot  vasi3
U8205.E6B.1093XXX-V1-C32776 Virtual I/O Slot  vasi4
```

86

86

Documentation on VIOS 3.1 upgrades

- What's new in Virtual I/O Server commands
- https://www.ibm.com/support/knowledgecenter/en/9040-MR9/p9hcg/p9hcg_whatsnew.htm

- Virtual I/O Server release notes – include USB Memory/Flash key install
- https://www.ibm.com/support/knowledgecenter/en/9040-MR9/p9eeo/p9eeo_ipeeo_main.htm
  - USB Memory/Flash key install
  - Minimum size for a VIOS

- **VIOS viosupgrade** command in VIOS 2.2.6.30
- https://www.ibm.com/support/knowledgecenter/en/9009-42A/p9hcg/p9hcg_viosupgrade.htm
  - Hint – upgrade to at least 2.2.6.32 prior to trying to upgrade to v3

- **NIM viosupgrade** command on the NIM AIX 7.2 TL3 + sp
- https://www.ibm.com/support/knowledgecenter/en/ssw_aix_72/com.ibm.aix.cmds6/viosupgrade.htm
  - This one is buried in the AIX commands reference for AIX Commands of AIX 7.2

87

87

Useful Links

- Jaqui Lynch Articles
  - http://www.circle4.com/jaqui/eserver.html
  - https://ibmsystemsmag.com/Authors/jaqui-lynch

- Nigel Griffiths AIXpert Blog
  - https://www.ibm.com/support/pages/aixpert-blog-nigel-griffiths-mrnmon
- Nigel Griffiths Twitter – mr_nmon
  - https://twitter.com/mr_nmon
- Nigel Griffiths YouTube
  - https://www.youtube.com/nigelargriffiths
- Gareth Coates – Tricks of the POWER Masters
  - https://www.ibm.com/support/pages/node/1116939
- Gareth Coates Twitter – power_gaz
  - https://twitter.com/power_gaz
- Jaqui's Movie Replays
  - http://www.circle4.com/movies
- IBM US Virtual User Group
  - https://www.ibm.com/support/pages/node/1120377
- Power Systems UK User Group
  - https://www.ibm.com/support/pages/node/1110195

88

88

Useful Links

- ESS Website to download base software
  - https://www.ibm.com/servers/eserver/ess/index.wss?lnk=msdDO-enss-usen
- HMC Scanner
  - https://www.ibm.com/support/pages/node/1117515
  - https://www.ibm.com/support/pages/sites/default/files/inline-files/$FILE/hmcScanner-0.11.42.zip
- AIX 7.2 Performance Guide
  - https://www.ibm.com/support/knowledgecenter/ssw_aix_72/performance/performance_pdf.pdf
  - https://www.ibm.com/support/knowledgecenter/en/ssw_aix_72/navigation/performance.html
- VIOS Advisor
  - https://www.ibm.com/support/knowledgecenter/TI0002C/p8hcg/p8hcg_part.htm
  - https://www.ibm.com/support/knowledgecenter/TI0003N/p8hb1/p8hb1_vios_perf_adv.htm
  - https://www.ibm.com/support/knowledgecenter/TI0003M/p8hb1/p8hb1_vios_perf_adv_reports.htm
- SG24-8171 – Power Systems Performance Optimization including POWER8
  - http://www.redbooks.ibm.com/redbooks/pdfs/sg248171.pdf
- SG24-8453 - AIX Modernization and Enhancements
  - http://www.redbooks.ibm.com/redbooks/pdfs/sg248453.pdf

89

89

Useful Articles

- Conduct an end of year AIX Health Check (Dec 2019)
  - https://ibmsystemsmag.com/Power-Systems/12/2019/Conduct-AIX-Systems-Health-Check
- Using NIM with VIO Servers
  - https://ibmsystemsmag.com/Power-Systems/09/2019/Using-NIM-with-VIO-Servers
- PowerVM v3 Installation and Upgrade Experience
  - https://ibmsystemsmag.com/Power-Systems/05/2019/powervm-experience
- Systems Management Tips
  - https://ibmsystemsmag.com/Power-Systems/08/2019/2019-AIX-Systems-Management-Tips
- 2019 AIX System Management Tips
  - https://ibmsystemsmag.com/Power-Systems/08/2019/2019-AIX-Systems-Management-Tips
- Secure your VIO Server
  - http://archive.ibmsystemsmag.com/aix/administrator/security/secure-your-vio-server/
- Upgrading your VIO server – July 2018
  - https://ibmsystemsmag.com/Power-Systems/12/2018/powervm-3-1-update
  - https://ibmsystemsmag.com/Power-Systems/05/2019/powervm-experience
- Maintaining the HMC
  - http://ibmsystemsmag.com/aix/administrator/systemsmanagement/hmc-maintenance/
- LPM
  - https://ibmsystemsmag.com/Power-Systems/10/2018/guide-live-partition-mobility
- HMC Enhanced GUI Links
  - https://www.ibm.com/support/pages/enhanced-gui-links-documentation

90

90

VIOS Specific References

- VIO Server Support
  - https://www14.software.ibm.com/support/customercare/sas/f/vios/home.html
- SDD and SDDPCM Specific procedures for VIOS
  - http://www-01.ibm.com/support/docview.wss?uid=ssg1S7002686&aid=1
- SG24-7940 - PowerVM Virtualization - Introduction and Configuration
  - http://www.redbooks.ibm.com/redbooks/pdfs/sg247940.pdf
- SG24-7590 – PowerVM Virtualization – Managing and Monitoring
  - http://www.redbooks.ibm.com/redbooks/pdfs/sg247590.pdf
- SG24-8080 – Power Systems Performance Guide – Implementing and Optimizing
  - http://www.redbooks.ibm.com/redbooks/pdfs/sg248080.pdf
- SG24-8062 – PowerVM Best Practices
  - http://www.redbooks.ibm.com/redbooks/pdfs/sg248062.pdf
- SEA Load Sharing
  - https://www.ibm.com/support/pages/how-setup-sea-failover-load-sharing-configuration
  - https://www.ibm.com/support/pages/shared-ethernet-adapter-sea-fail-over-load-balancing
- POWERVM Enhancements – what is new in 2013
  - http://www.redbooks.ibm.com/redbooks/pdfs/sg248198.pdf
- Capturing Debug output for padmin
  - http://www-01.ibm.com/support/docview.wss?uid=isg3T1012362

91

91

VIOS Specific References - Network

- SEA Failover Statistics
  - https://www.ibm.com/support/knowledgecenter/POWER9/p9hb1/p9hb1_statsseafailover.htm
- SEA Statistics
  - https://www.ibm.com/support/knowledgecenter/POWER9/p9hb1/p9hb1_statssea.htm
- Enhanced GUI Links
  - https://www.ibm.com/support/pages/enhanced-gui-links-documentation
  - Includes many Developerworks documents related to the HMC enhanced GUI
  - Includes how to dynamically add and remove virtual ethernets and VLANs
- Configure VIO Server using VLAN Tagging
  - https://www.ibm.com/support/knowledgecenter/POWER8/p8hb1/p8hb1_vios_scenarios_network_two.htm
- VLAN Tagging – Load sharing with 10Gb adapters (PPT)
  - https://www.ibm.com/support/knowledgecenter/POWER8/p8hb1/p8hb1_vios_scenarios_network_two.htm

92

92

Backup Slides

93

93

Changes to Fix Central

- IBM has moved from anonymous FTP to Secure FTP
- http://www-01.ibm.com/support/docview.wss?uid=isg3T1024541

- On AIX this means you will be provided with a userid and password to login when you request the fixes
- ftp –s –i delivery04-mul.dhe.ibm.com
- When prompted for userid and password use the ones provided
- passive (to set passive mode)
- binary (to download as binary)
- mget * (to download fixes)
- Quit

You can also use sftp – i.e. once they give you a userid and password:
sftp user@delivery04-mul.dhe.ibm.com
Put in password when prompted then type in “mget *” then quit when done

94

94

MPIO Algorithms

I/O algorithms

One attribute that can be changed is the **algorithm** attribute. It determines the way I/O is distributed across the paths that are available for a device. There are three different algorithms for I/O distribution:

fail_over	In this mode, I/O is sent over one enabled path, and another path is used only if this path fails. The initial path that is selected is the one with the highest priority (the lowest path priority value). If this path fails, the next highest priority path is selected for I/O operations. This is the default algorithm for iSCSI and FC devices.
round_robin	In this mode, I/O is distributed across multiple enabled paths. For any device that has preferred and non-preferred paths or active and passive paths, only a subset of paths is used. Paths that have a higher path priority receive a larger share of I/O operations in this mode.
shortest_queue	In this mode, an I/O path is distributed across multiple enabled paths. For any device that has preferred and non-preferred paths or active and passive paths, only a subset of paths is used. In this mode, path priority is ignored, and paths are selected based on the number of pending I/O operations only. The path with the lowest number of pending I/O operations is selected for I/O.

From: SG24-8453 - AIX Modernization and Enhancements - http://www.redbooks.ibm.com/redbooks/pdfs/sg248453.pdf

95

95

MPIO Health Check Mode

Health check mode and interval

Another key device attribute is the **hcheck_mode** attribute. This attribute determines which paths are probed when the health check capability is used. Health checking is only performed on devices that have a state of open. A device that is not in use does not have its paths health checked. Health checking is also not performed on any disabled or missing paths. There are three health check modes:

Enabled	In this mode, the **healthcheck** command is sent to all paths that are enabled for the device, which includes paths that failed.
Failed	In this mode, the **healthcheck** command is sent to all paths that are in a failed state for the device.
Nonactive	In this mode, the **healthcheck** command is sent to all paths that do not have any active I/O, which includes paths that are in enabled and failed states. This is the default health check mode that is configured on AIX.

Along with **hcheck_mode**, you can also configure how often the health check is performed by configuring the **hcheck_interval** value. This attribute can be set to any value 0 - 3600, and it represents the time in seconds between polling. If a value of 0 is specified, it indicates that health checking should be disabled on the device. The default value for **hcheck_interval** is set to perform health checking every 60 seconds.

From: SG24-8453 - AIX Modernization and Enhancements - http://www.redbooks.ibm.com/redbooks/pdfs/sg248453.pdf

96

96

MPIO Reservation Policies

Reservation policies

If your disks require concurrent access from multiple initiators, another attribute you might need to modify is the device attribute **reserve_policy**. This device attribute is required for all MPIO devices regardless of the PCM in use. This value describes the type of reservation policy that is set on a device. For MPIO devices, the following reservation policies exist:

no_reserve	This policy does not apply any reservation on the target device allowing initiators (paths) on the same system, and on other systems, access to the target device. This is the recommended policy for devices where disks are shared between hosts and devices that have the **shortest_queue** or **round_robin** algorithms configured.
single_path	This is the default policy when using AIXPCM. This policy places an SCSI2 reserve on a target device so that the device can be accessed only on the path it was reserved on. This policy prevents other paths on the same system from accessing the storage without first sending a bus device reset to release the reserve on the device.
PR_exclusive	This policy applies an SCSI3 persistent-reserve with exclusive-host methodology on the device when the device is opened to exclusively lock it to a single host. A **PR_key_value** attribute must also be set on the device when using this mode to uniquely identify the host.
PR_shared	This policy applies an SCSI3 persistent-reserve with shared-host methodology when the device is opened. Initiators from other host systems must register before they can access the device. A **PR_key_value** attribute must also be set on the device when using this mode to uniquely identify the host.

From: SG24-8453 - AIX Modernization and Enhancements - http://www.redbooks.ibm.com/redbooks/pdfs/sg248453.pdf

97

97

SEA with link Aggregate and dual VIO Servers

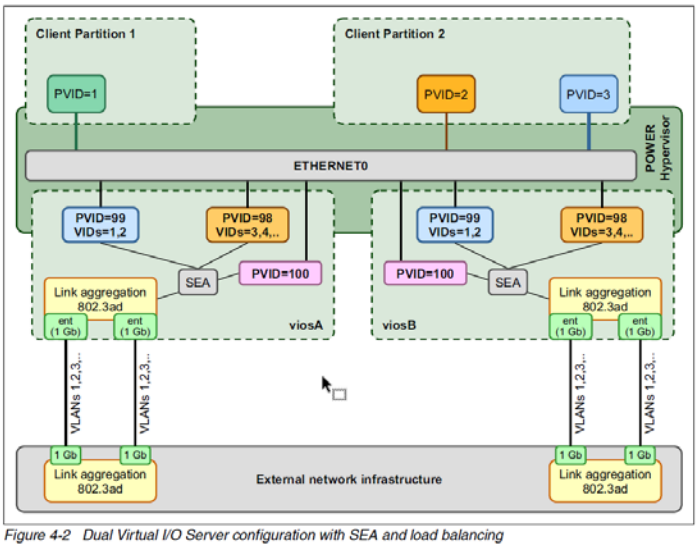


Figure 4-2 Dual Virtual I/O Server configuration with SEA and load balancing

The SEA is connected to two virtual Ethernet adapters; each adapter has a different set of VLAN IDs. Virtual Ethernet adapters on different Virtual I/O Servers must have the same set of VLANs and different trunk priorities.

In addition, the SEA is connected to another virtual Ethernet adapter with a PVID=100 that is used as a SEA control channel.

The control channel is used for SEA heartbeating and exchanging information between the two SEA adapters on the set of VLAN IDs that each SEA bridges.

From PowerVM Best Practices Red Book SG24-8062

NOTE – as of VIO 2.2.3 we let the control channel default to VLAN 4095 – we no longer define it.

98

98

SEA Loadsharing

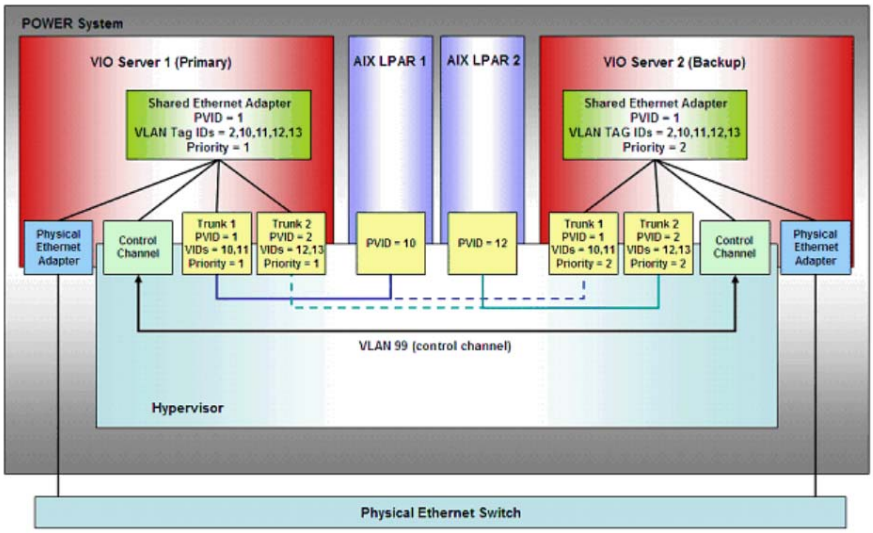


Figure 4: SEA with Load Sharing

99

99