

IBM

2015

IBM Systems Technical University

IBM z Systems • IBM Power Systems • IBM Storage

October 5-9 | Hilton Orlando, Florida

Lab Exercise:

Configuring a PowerHA SystemMirror 7.1 cluster

Session ID: pBA3020

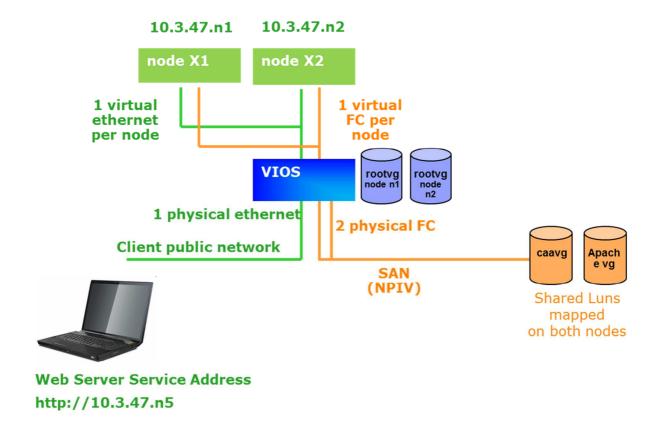
Speaker Name: Michel Passet

Introduction:

This exercise covers the configuration of a two-node typical PowerHA cluster, to protect a web service. You'll see a failover of the web service onto the backup node.PowerHA uses this failover to recover automatically the service if an issue occurs on the primary server.

1. Lab environment for a team

Infrastructure



Two AIX nodes with PowerHA

These fully virtualized nodes are provided by IBM Montpellier, France cloud apachevg and caavg_private are shared Luns. rootvg disks are local to the nodes.

The HTML page is located on the apachevg disk PowerHA is using a disk (caavg_private) as a topology repository disk

When the PowerHA cluster is configured and started, the primary AIX server (node n1) will run an Apache web server, managed and secured by PowerHA. The web server is using a service IP address, ending in 5. Note that the service address is different from native nodes addresses, ending in 1 or 2. Service address an IP alias managed by PowerHA, hosted on node n1 or n2.

Upon failure of the primary server (node n1), PowerHA will automatically failover on the standby server (node n2). The service IP address as well as the apachevg disk are moved to the standby server to restart the apache web server. From the browser running on your laptop, you'll see only a very brief outage. The service address does not change. The HTML page will display which AIX server (node n1 or n2) the apache server is running on.

During this lab, you will setup and configure a PowerHA cluster to manage a resilient Apache Web server. Then, you'll failover to node n2 and see how fast the web server is restored (a few seconds)!

Lab access

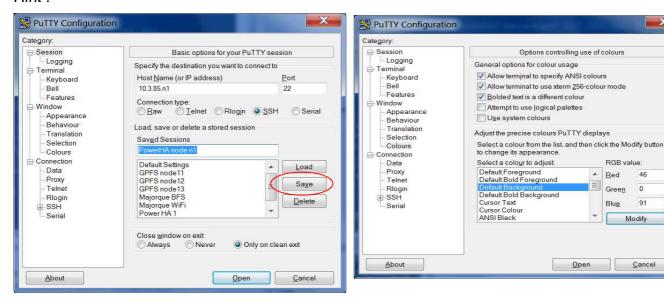
Use putty terminal on your laptop to open shells on AIX nodes, as root user.

The IP addresses for you and the passwords are provided on a separate sheet.

The AIX nodes addresses, ending in 1 and 2 are to be used with putty.

The service address, ending in 5 is to be used with a browser. This address does not work until the PowerHA cluster is up and running.

Hint:



Save your AIX nodes IP addresses, to avoid keying IP addresses multiple times

You can also change the background color of your putty session to avoid mistakes, one color per node.

91

You'll avoid issuing a command on a server while you think to be on another!

Use this configuration panel: Windows / Colors / Default Background / Modify

Don't forget to save afterward

2. Preliminary infrastructure requirements and checks

Open a putty session on both nodes. It is a good habit to have both windows opened when working on a cluster, as some commands must be executed on both servers.

<u>Important:</u> The node names and IP addresses listed in all this document does not match your own hostnames and addresses. In your AIX commands, use the names and addresses provided to you on a separate sheet.

__ 1. Check /etc/hosts file (both nodes)

Verify that your own IP addresses for your lab are listed and in sync in /etc/hosts files on both nodes. *n* is the number corresponding to your laptop letter (A up to V)

The **service address** is used by PowerHA to secure the Apache web server. This address is an IP alias created on nodeX1 or nodeX2 when PowerHA cluster starts. Use this service address from your browser to display the HTML page, later in this Lab.

___ 2. Check connectivity between the two nodes (both nodes)

Hostnames and IP addresses listed in this document are just examples, they are different from the ones you use. The apache-service address does not ping until the PowerHA cluster is configured and started.

```
{D-pwrha-nodeX1:root}/ # ping D-pwrha-nodeX2
PING D-pwrha-nodeX2 (10.3.47.2): 56 data bytes
64 bytes from 10.3.47.2: icmp_seq=0 ttl=255 time=0 ms
```

___ 3. Make sure that there is no /etc/resolv.conf file (both nodes)

If you find one, move it to another name (/etc/resolv.conf.bak for instance). PowerHA manages the service address alone. No need for DNS change to mask a hosting server change to the clients.

```
{D-pwrha-nodeX1:root}/ # ls /etc/resolv.conf
/etc/resolv.conf not found
```

4. Edit /etc/cluster/rhosts (both nodes)

This file enables the cluster secure communication between the nodes. Edit with your own hostnames. This file name, rhosts, is confusing. It is absolutely not related to rsh. This configuration file should have been named more wisely like clcomd.conf, for instance!

```
{D-pwrha-nodeX1:root}/ # vi /etc/cluster/rhosts
D-pwrha-nodeX1
D-pwrha-nodeX2

{D-pwrha-nodeX2:root}/ # vi /etc/cluster/rhosts
D-pwrha-nodeX1
D-pwrha-nodeX2
```

__ 5. Start the clcomd daemon (both nodes)

PowerHA does not use rsh neither ssh to communicate between the nodes, but its own mechanism, that needs clcomd daemon to be started on both nodes. This service permits the PowerHA communication only between the nodes listed in /etc/cluster/rhosts.

To start the clcomd daemon (both nodes):

```
{D-pwrha-nodeX1:root}/ # startsrc -s clcomd
0513-059 The clcomd Subsystem has been started. Subsystem PID is 7209064.
```

To see if the clcomd is running:

```
{D-pwrha-nodeX1:root}/ # lssrc -s clcomd
Subsystem Group PID Status
clcomd caa 7209064 active
```

To load the newly updated configuration file into a running clcomd daemon:

```
{D-pwrha-nodeX1:root}/ # refresh -s clcomd
0513-095 The request for subsystem refresh was completed successfully.
```

If you want to stop the clcomd daemon (normally not):

```
{D-pwrha-nodeX1:root}/  # stopsrc -s clcomd 0513-044 The clcomd Subsystem was requested to stop.
```

___ 6. Synchronize the time between nodeX1 and nodeX2 (nodeX2)

There is no time server in this case, just synchronize the two nodes time.

```
{D-pwrha-nodeX2:root}/ # setclock D-pwrha-nodeX1
Wed Oct 01 15:40:56 2014
```

__ 7. Check that you have the same shared LUNs by verifying the PVIDs on the two server nodes (both nodes)

{D-pwrha-no	deX1:root}/ # lspv		
hdisk0	00f86c371cd39579	rootvg	active
hdisk1	00f86c375b34c6a7	None	
hdisk2	00f86c375b34c5b2	None	
{D-pwrha-no	deX2:root}/ # lspv		
hdisk0	00f86c3755739aaf	rootvg	active
hdisk1	00f86c375b34c6a7	None	
hdisk2	00f86c375b34c5b2	None	

The same LUN must be seen on the two nodes, for hdisk1 and hdisk2

hdisk0 (rootvg) is local to each node

hdisk1 will be used for Cluster Aware AIX (CAA) topology repository disk (caavg_private) hdisk2 is targeted for the HTML page, used by Apache web server (apachevg) It is possible that on some setups, the hdisk numbers are different. For example, rootvg can be on hdisk1. It is not an issue.

3. PowerHA and Cluster Aware AIX (CAA) base configuration

In this section, you will create a PowerHA SystemMirror 7.1.3 cluster

This lab is a simple configuration with a single network interface (en0). In such case, edit /usr/es/sbin/cluster/netmon.cf to provide PowerHA with an address to ping. It helps to determine if a lack of connectivity comes from the interface (en0) or from the network itself. Virtualized environments also require netmon.cf

___ 8. Edit /usr/es/sbin/cluster/netmon.cf (both nodes)

Edit with your own hostnames, but do not modify the required syntax. The address provided is the default gateway. Check your default gateway address, and ping it to be sure. Edit the default gateway in netmon.cf file if needed.

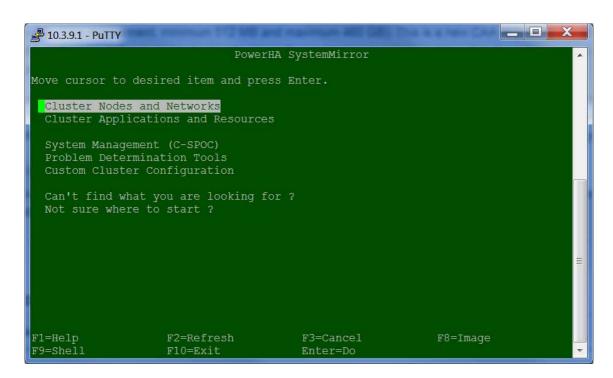
```
{D-pwrha-nodeX1:root}/ # netstat -rn
Routing tables
Destination
                 Gateway
                                  Flags
                                         Refs
                                                  Use If
                                                           Exp Groups
Route tree for Protocol Family 2 (Internet):
                 10.3.47.254
                                                 1634 en0
default
                                  IJĠ
                                          0
10.3.47.0
                                                  0 en0
                 10.3.47.11
                                  UHSb
10.3.47/24
                 10.3.47.11
                                  IJ
                                           0
                                                    0 en0
                                                  221 100
10.3.47.11
                                  UGHS
                                           1
                 127.0.0.1
10.3.47.255
                                           2.
                 10.3.47.11
                                  UHSb
                                                   72 en0
127/8
                127.0.0.1
                                          14
                                                8302 100
{D-pwrha-nodeX1:root}/ # ping 10.3.47.254
PING 10.3.47.254 (10.3.47.254): 56 data bytes
64 bytes from 10.3.47.254: icmp_seq=0 ttl=255 time=0 ms
{D-pwrha-nodeX1:root}/ # vi /usr/es/sbin/cluster/netmon.cf
!REQD D-pwrha-nodex1 10.3.47.254
!REQD D-pwrha-nodex2 10.3.47.254
{D-pwrha-nodeX2:root}/ # vi /usr/es/sbin/cluster/netmon.cf
!REQD D-pwrha-nodeX1 10.3.47.254
!REQD D-pwrha-nodex2 10.3.47.254
```

__ 9. Start the cluster configuration

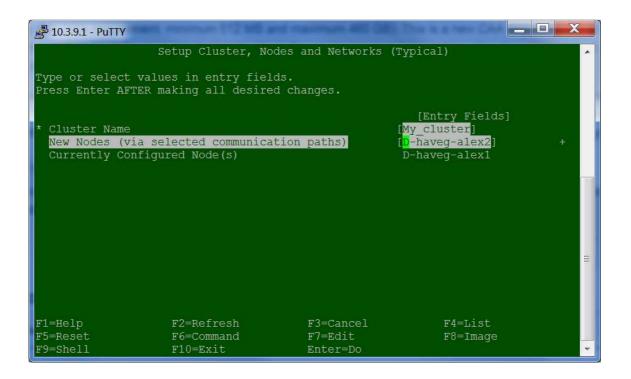
Create the following configuration only on nodeX1. After the change, the Cluster Verification and Synchronization process will propagate the configuration clusterwide.

```
{D-pwrha-nodeX1:root}/ # smit sysmirror / Cluster Nodes and Networks / Standard
Cluster Deployment / setup a Cluster, Nodes and Networks
```

You should navigate through the screens as follows:



The suggested cluster name is *My_cluster*, and you only have to select your second node to add in the cluster, in New Nodes field. Use F4 picklist.



Upon completion, scan the cltopinfo output for errors.

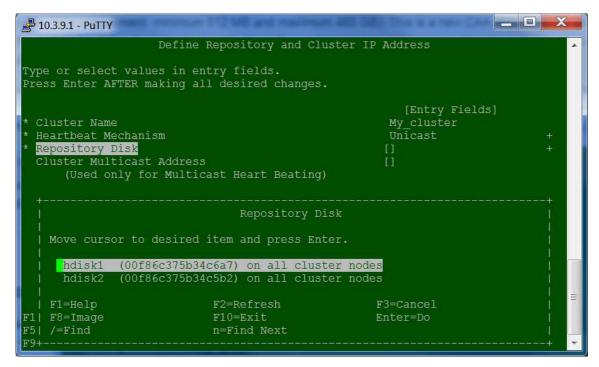
You can verify the configuration done so far on nodeX1. The other node will be configured only after the synchronization.

No resource groups defined

10. Define the Cluster Aware AIX (CAA) repository disk

Now the CAA disk must be defined. It is a disk shared between the cluster nodes. hdisk1 is free for CAA repository. (The hdisk number may be different on your server)

{D-pwrha-nodeX1:root}/ # smit sysmirror / Cluster Nodes and Networks / Standard
Cluster Deployment / Define Repository Disk and Cluster IP address



hdisk1 is the target disk for CAA repository

The first step of the cluster configuration is now complete. The PowerHA SystemMirror cluster is defined on node 1. Node 2 is unaware of the cluster configuration.

<u>Hint:</u> If it fails, it is probably due to the fact that the repository disk has been used before. To fix it, ensure that the right disk is being used, then, with care, use chpv –C hdiskn to clear. Normally hdisk0 is for rootvg, hdisk1 for CAA repository, hdisk2 for apachevg.

The nodeX2 is still unaware of the cluster definition, cluster services are not running, but the cluster communication process (clcomd) is ready to communicate with nodeX1.

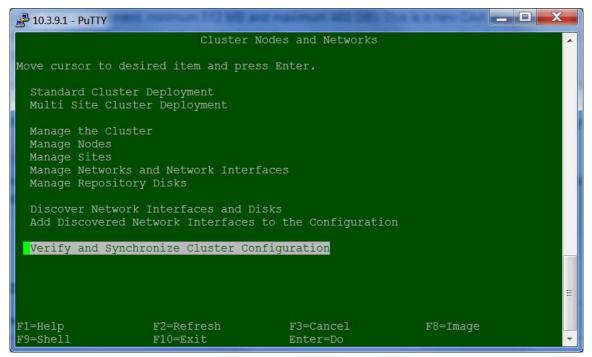
```
{D-pwrha-nodeX2:root}/ # cltopinfo
Warning: There is no cluster found.

{D-pwrha-nodeX2:root}/ # lscluster -c
lscluster: Cluster services are not active.

{D-pwrha-nodeX2:root}/ # lssrc -g caa
Subsystem Group PID Status
clcomd caa 6750432 active
clconfd caa inoperative
```

__ 11. Synchronize the cluster configuration

D-pwrha-nodeX1:root}/ # smit sysmirror / Cluster Nodes and Networks / Verify
and Synchronize Cluster Configuration



Be patient, this can take a few minutes.

___ 12. Note that now, the CAA cluster services are started on both nodes.

```
{D-pwrha-nodeX2:root}/ # cltopinfo
```

Cluster Name: My_cluster Cluster Type: Standard Heartbeat Type: Unicast

Repository Disk: hdisk1 (00f86c375b34c6a7)

```
There are 2 node(s) and 1 network(s) defined
NODE D-pwrha-nodeX1:
       Network net_ether_01
                D-pwrha-nodeX1
                                 10.3.47.1
NODE D-pwrha-nodeX2:
       Network net_ether_01
                D-pwrha-nodeX2
                                 10.3.47.2
No resource groups defined
{D-pwrha-nodeX2:root}/ # lscluster -c
Cluster Name: My_cluster
Cluster UUID: 1e461064-38ea-11e4-b5b6-2a7851ceeb0d
Number of nodes in cluster = 2
        Cluster ID for node D-pwrha-nodeX1: 1
        Primary IP address for node D-pwrha-nodeX1: 10.3.47.1
        Cluster ID for node D-pwrha-nodeX2: 2
        Primary IP address for node D-pwrha-nodeX2: 10.3.47.2
Number of disks in cluster = 1
        Disk = hdisk1
{D-pwrha-nodeX2:root}/ # lssrc -g caa
Subsystem
                  Group
                                                Status
 clconfd
                  caa
                                   15990868
                                                active
 clcomd
                                   9175040
                                                active
                  caa
```

___ 13. Verify that the CAA repository disk has been created and is active

```
{D-pwrha-nodeX1:root}/ # lsvg
caavg_private
rootvg
{D-pwrha-nodeX1:root}/ # lspv
                00f86c371cd39579
hdisk0
                                                     rootvq
                                                                      active
hdisk1
                00f86c375b34c6a7
                                                     caavg_private
                                                                      active
hdisk2
                00f86c375b34c5b2
                                                     apachevg
{D-pwrha-nodeX2:root}/ # lspv
hdisk0
                00f86c3755739aaf
                                                     rootvg
                                                                      active
hdisk1
                00f86c375b34c6a7
                                                     caavg_private
                                                                      active
hdisk2
                00f86c375b34c5b2
                                                     apachevg
```

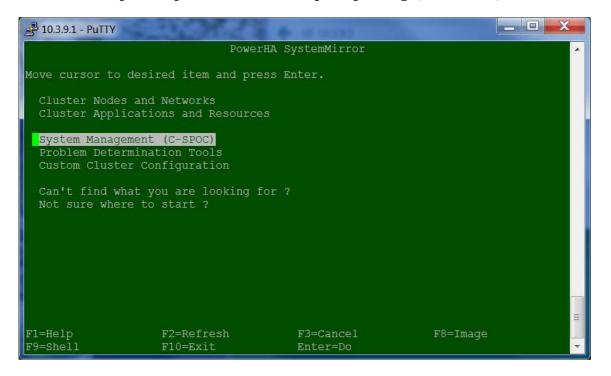
4. Import the apache volume group into PowerHA

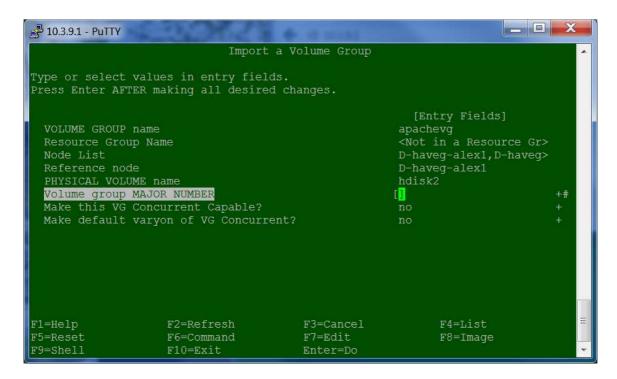
A volume group, called apachevg, available on the first node, contains the /apache filesystem. It holds the HTML source file, necessary for the Apache server. In case of failover, PowerHA moves this volume group to the backup node to recover the Apache service. The volume group is online on a single node at a time, managed by Power HA.

___ 14. Use C-SPOC to import apachevg into PowerHA configuration

Cluster Single Point Of Control (C-SPOC) is an interesting feature. You can control your entire cluster from any node in a single action. You do not have to repeat the command on each cluster node.

{D-pwrha-nodeX1:root}/ # smit sysmirror / System Management (C-SPOC) / Storage / Volume Groups / Import a Volume Group / apachevg (on hdisk2)





Do not specify major number.

The apachevg volume group is now known from PowerHA configuration. This means that after a synchronization, nodeX2 will be aware about this volume group for Apache server. No additional commands are needed to be executed on nodeX2.

5. PowerHA SystemMirror Application and Resource configuration

In PowerHA terminology, a Resource Group is a set of information and resources needed to run the application you want to protect with the cluster.

A resource group is what Power HA manages and makes Highly Available.

Resource Group =

- Application controller =
 - Start script used by PowerHA to start your service (apache server in this lab)
 - Stop script
- Service IP address. The one used by your client browser. It is not the base IP addresses of your nodes. This service IP address can be served by either host, as an alias on top of its base IP address.
- Volume group to provide the necessary file systems for the application. The data are not duplicated. The same LUN is moving from one node to the other.
- Startup / fallover / fallback policies, for PowerHA to know what you want to do in each case.

Information:

You should not start the application yourself. This will be managed by PowerHA directly. When starting the cluster, the resource group will be brought online on the node you choose, and start script will be executed. The service IP address will show up, and the volume group and its filesystems mounted.

Same for stopping, use C-SPOC to control your application

___ 15. Check start/stop scripts

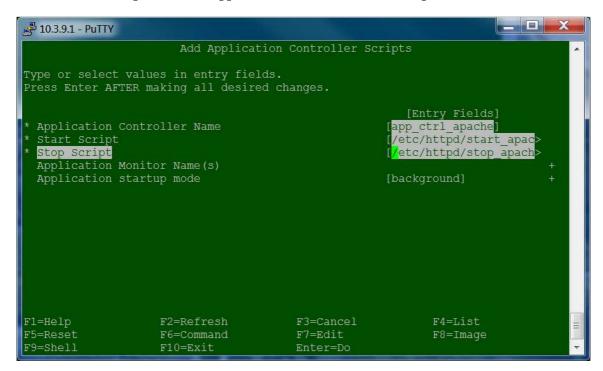
These scripts are provided in /etc/httpd directory on each node. They should be on rootvg, and not on the shared disk (apachevg). They exist on each node, identical and executable.

```
{D-pwrha-nodeX1:root}/etc/httpd # cat start_apache.ksh
print "$(hostname)" > /apache/httpd/htdocs/hostname 2>
/var/httpd/logs/start_apache.log
/opt/httpd/sbin/apachectl start >> /var/httpd/logs/start_apache.log 2>&1

{D-pwrha-nodeX1:root}/etc/httpd # cat stop_apache.ksh
/opt/httpd/sbin/apachectl stop >> /var/httpd/logs/start_apache.log 2>&1
```

__ 16. Define the Application Controller

{D-pwrha-nodeX1:root}/ # smit sysmirror / Cluster Application and Resources /
Resources / Configure User Applications (Scripts and Monitors) / Application
Controller Scripts / Add Application Controller Scripts

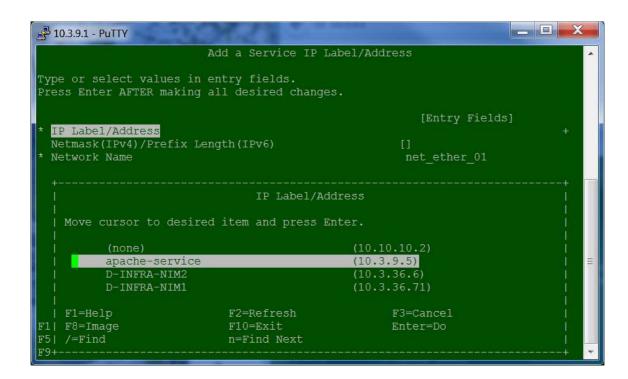


__ 17. Define the Service IP address

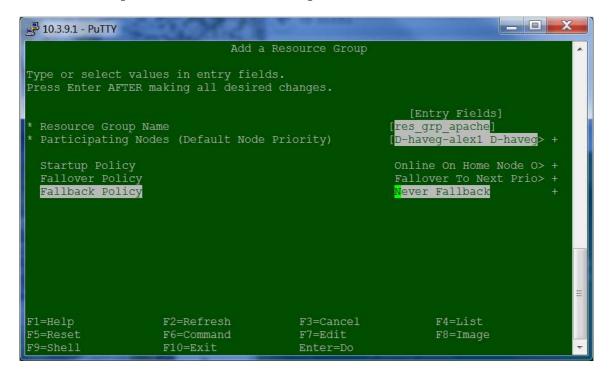
{D-pwrha-nodeX1:root}/ # smit sysmirror / Cluster Application and Resources /
Resources / Configure Service IP Labels/Addresses / Add a Service IP
Label/Address

Network name is net_ether_01 (10.3.47.0/24). The service address is an alias on top of this network

IP Label/Address is apache-service n5



__ 18. Define the Resource Group



The participating nodes list defines the home node for the application: the first in the list. It can be nodeX2! You can edit this list if you want nodeX1 to be the first.

Startup Policy: When the cluster starts, determines on which node the Apache server

will start. Home node is the first node in the list provided.

Fallover Policy: In case of failure on the first node, where to restore the service. By

default the next server on the list

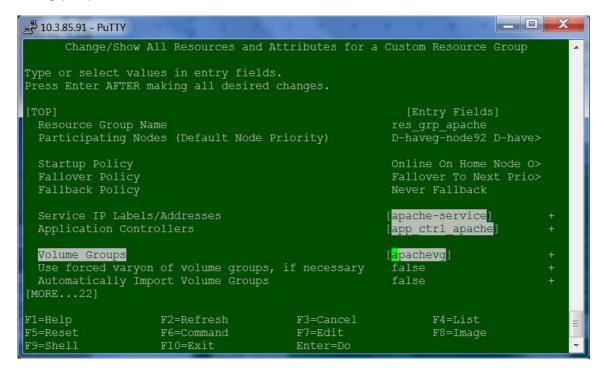
Fallback Policy: When the first node is up again, what to do. By default, Never Fallback,

to avoid an unnecessary outage. Fallback is manual in this case.

___ 19. Add service IP address and apachevg volume group to the Resource Group

{D-pwrha-nodeX1:root}/ # smit sysmirror / Cluster Application and Resources /
Resource Groups / Change/Show Resources and Attributes for a Resource Group

Choose the resource group you have created at the step before, for instance res_grp_apache



Congratulations, you are now almost done with the configuration.

20. Check your final PowerHA cluster configuration

{D-pwrha-nodeX1:root}/ # cltopinfo

Cluster Name: My_cluster Cluster Type: Standard Heartbeat Type: Unicast

Repository Disk: hdisk1 (00f86c375b34c6a7)

```
There are 2 node(s) and 1 network(s) defined
NODE D-pwrha-nodeX1:
       Network net_ether_01
               apache-service 10.3.47.5
                               10.3.47.1
               D-pwrha-nodeX1
NODE D-pwrha-nodeX2:
       Network net_ether_01
               apache-service 10.3.47.5
               D-pwrha-nodeX2
                               10.3.47.2
Resource Group res_grp_apache
       Startup Policy Online On Home Node Only
       Fallover Policy Fallover To Next Priority Node In The List
        Fallback Policy Never Fallback
                                D-pwrha-nodeX1 D-pwrha-nodeX2
       Participating Nodes
       Service IP Label
                                apache-service
{D-pwrha-nodeX1:root}/ # cllsres
APPLICATIONS= "app_ctrl_apache'
SERVICE_LABEL="apache-service"
VOLUME_GROUP= "apachevg"
```

Question:

If you execute these commands on your second node, you'll not have the same result. Why?

Answer: because you didn't synchronize

21. Synchronize the PowerHA configuration

 ${D-pwrha-nodeX1:root}/$ # smit sysmirror / Cluster Applications and Resources / Verify and Synchronize Cluster Configuration

This is a quite long process. Take some time to review the output, search for ERROR, and fix before trying again!

When the synchronization is OK, it means that PowerHA checked and validate all your configuration, and that both nodes are in sync.

6. Start PowerHA SystemMirror Services

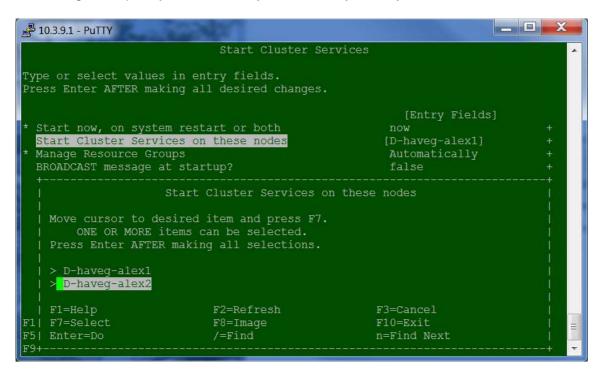
Currently, the cluster services are not started. This means that the Apache server is not started, the volume group (containing the /apache filesystem) and the service IP address (finishing by .5) are both offline.

Upon starting the cluster will start the Apache web server :

__ 22. Start the PowerHA SystemMirror using C-SPOC

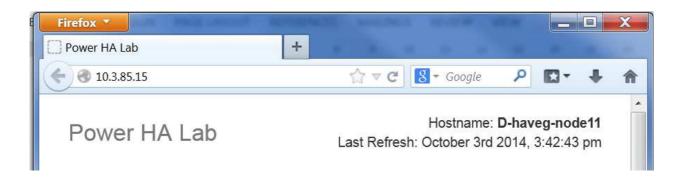
{D-pwrha-nodeX1:root}/ # smit sysmirror / System Management (C-SPOC) / PowerHA
SystemMirror Services / Start Cluster Services

Don't forget to specify both nodes, you have only one by default.



- 23. How to check that the application has successfully started?
 - __ a. Use the application, a web server in this Lab: http://10.3.47.n5

On your laptop, open a browser, and type the *service address* of your Apache server (finishing by .5). You'll immediately see if it works! The goal of PowerHA System Mirror is to protect this Apache server against failures. Once PowerHA detects that there is a concern, it fails over to the backup server to restore the service. After a little outage, the web server is up again, automatically!



The hostname on right top corner shows which host is running the web server. The date is updated while the apache server is up and running.

b. PowerHA command clRGinfo

This command shows where PowerHA is running the resource, which is the Apache server in our case.

```
{D-pwrha-nodeX1:root}/ # clRGinfo

Group Name Group State Node

res_grp_apache ONLINE D-pwrha-nodeX1

OFFLINE D-pwrha-nodeX2

__ c. qha command

{D-pwrha-nodeX1:root}/ # gha
```

This is a script provided by Alex Abderrazag (who organize the IBM Power Server track of this conference). See Redbook SG24-8167-00, "IBM PowerHA SystemMirror 7.1.3 for AIX Reference Guide".

```
Cluster: My_cluster (7130) echo 15:41:43 030ct14

D-haveg-node11 iState: ST_STABLE
res_grp_apache ONLINE

D-haveg-node12 iState: ST_STABLE
```

__ d. AIX command if config -a

This command shows where the service address is hosted. This address is an IP alias, and can be served also on the other node of the cluster, in case of failure of the primary node. The other node does not display the service address.

```
{D-pwrha-nodeX1:root}/ # ifconfig -a
en0:
flags=le084863,10480<UP,BROADCAST,NOTRAILERS,RUNNING,SIMPLEX,MULTICAST,GROUPRT,
64BIT,CHECKSUM_OFFLOAD(ACTIVE),CHAIN>
        inet 10.3.47.5 netmask 0xfffffff00 broadcast 10.3.47.255
        inet 10.3.47.1 netmask 0xffffff00 broadcast 10.3.47.255
        tcp_sendspace 262144 tcp_recvspace 262144 rfc1323 1
```

__ e. AIX command Isvg -o

Power HA sets online the volume group apachevg on the node hosting the service. This volume group contains the web pages code, in the /apache filesystem. This volume group is offline on the other node.

```
{D-pwrha-nodeX1:root}/ # lsvg -o
apachevg
caavg_private
rootvg
```

f. Check if the cluster is started

The command to run is:

```
{D-pwrha-nodeX1:root}/ # clshowsrv -v | grep "Current state"
Current state: ST_STABLE
```

If you see Current state: NOT_CONFIGURED, this means that the cluster is not started on this node.

7. Move the web server on the second node

Now, the Apache server is up and running on the first node, the primary one. To highlight the PowerHA capabilities, it's worth moving the resource (the Apache server) to the backup node.

Moving the resource means:

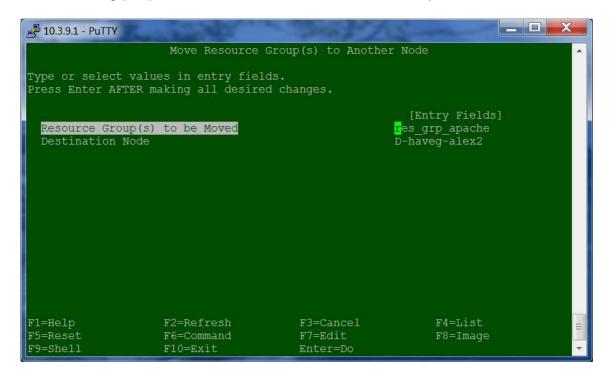
- A stop of the application using the application controller script provided to PowerHA
- Move of service IP address and volume group to the other node
- A restart of the application using the application controller script

In our case, all of these steps are completed in a few seconds. So watch closely the top right hand corner of your browser to monitor the web server outage. It is so short that you can easily miss it!

__ 24. Use PowerHA C-SPOC to move the resource C-SPOC means Cluster Single Point Of Control.

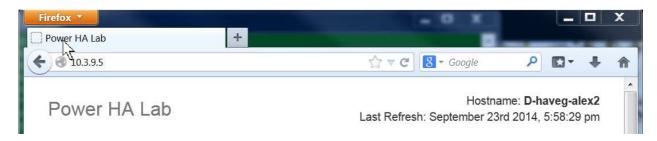
{D-pwrha-nodeX1:root}/ # smit sysmirror / System Management (C-SPOC) / Resource Groups and Applications / Move Resource Groups to Another Node

Choose res_grp_apache as the resource to be moved and your second node



_ a. Browser on your laptop

While PowerHA is moving the web server to the other node, monitor your browser. Look the top right corner. When the time is not updated, it means that the web server is not responding. During this little outage, the hostname displays an error. But very fast, the web server is restarted on the second node. The time remains current (updated each two seconds), and the hostname is the one of the backup server. The move is completed.



b. PowerHA command clRGinfo

This command shows where PowerHA is running the resource, which is the Apache server in our case.

__ c. AIX command if config -a on the second node

This command shows where the service address is hosted. This address is an IP alias, and can be served also on the other node of the cluster, in case of failure of the primary node.

End of Lab