Achieving KSYS high availability through IBM PowerHA

Monitor and manage KSYS daemon using PowerHA sample scripts

Overview

Challenge
How to achieve high availability for the KSYS daemon in a VM Recovery Manager HA/DR environment when ksysnode is not accessible and cannot monitor the data center VMs?

Solution
This high availability can be provided by managing the KSYS daemon with PowerHA SystemMirror software.

KSYS is a major component in the virtual machine (VM) Recovery Manager HA/DR solution, which monitors and manages the complete environment health. Hence providing high availability (HA) for KSYS is helpful to handle any scenario where a KSYS daemon hanged or the KSYS node itself went down. High availability can be provided by managing the KSYS daemon with IBM® PowerHA® SystemMirror® software.

To achieve this, you need to configure PowerHA SystemMirror to monitor and manage the KSYS daemon using custom scripts. This paper explains how to provide high availability to KSYS using PowerHA SystemMirror thereby removing the single point of failure.

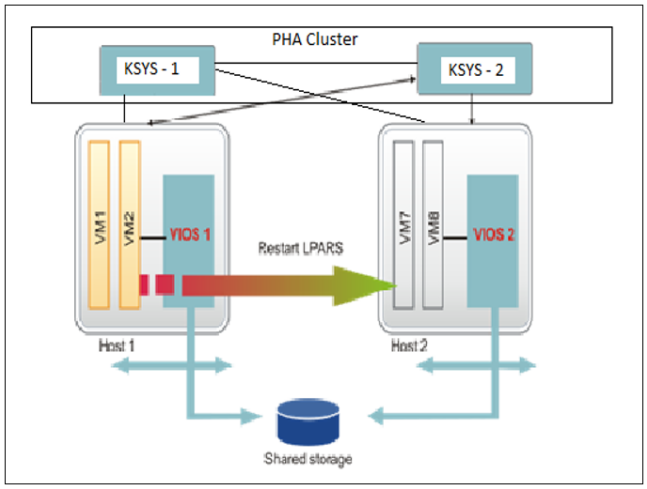

Figure 1. VM Recovery Manager HA

Architecture

Hardware

- IBM POWER7 processor-based servers or later
- For VM Recovery Manager and PowerHA specific hardware

Software

- IBM VM Recovery Manager DR-HA version 1.3.0.2 or later
- IBM PowerHA System Mirror version 7.1.3 or later

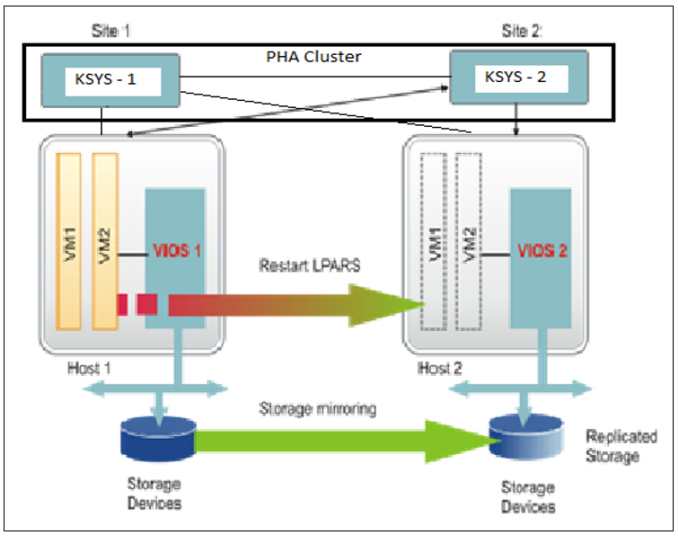

Figure 2. VM Recovery Manager DR

When a PowerHA cluster is created between two or more nodes, a Reliable Scalable Cluster Technology (RSCT) cluster is created underneath with all resource monitoring framework and resources available (including KSYS RM resources). Hence KSYS can use this RM cluster and framework for its own purpose instead of creating a new one. The advantage of this is, the configuration and data saved or modified in one KSYS node reflects in other KSYS nodes. A KSYS daemon can be monitored by custom scripts which will be part of the PowerHA resource group. A resource group will be online on one node at a time. If the KSYS node is down, RG will move to another node. This will make sure that another KSYS node (where RG is moved) starts monitoring and managing the environment. This ensures high availability between the KSYS nodes. Both VM Recovery Manager HA and disaster recovery (DR) configurations can be supported in this method.

PowerHA installation

To create KSYS HA using IBM PowerHA, first install the PowerHA files sets on both the KSYS nodes.

Run the following installation commands or upgradation of PowerHA software on the KSYS nodes:

```
# installp -acgXYd . ALL
```

KSYS installation

After the successful installation of PowerHA, install KSYS file sets on all the KSYS nodes. This installs the KSYS daemon (IBM.VMR daemon) along with the other file sets.

Then, install VIOS and VM file sets on the managed VIOS and VMs.

KSYS file set installation

Command to install KSYS packages:

```
# installp -ac -F -Y  -d . ksys.ha.license ksys.hautils.rte
ksys.license ksys.main.cmds ksys.main.msg.en_US.cmds
ksys.main.rte
```

VIOS file set installation:

Command to install VIOS file installation:

```
# installp -ac -FXYd. ALL
# sync
# reboot -q
```

Command to install host monitor file sets:

```
# installp -ac -FXYd. ksys.hsmon.rte
```

VM file set installation:

Command to install virtual machine monitor file sets:

```
# installp -ac -FXYd. ksys.vmmon.rte
```

Prerequisites for KSYS installation

Make sure that the following prerequisites are fulfilled for KSYS installation:

- The KSYS logical partition must be running IBM® AIX® 7.2 with Technology Level 1 Service Pack 1 (7200-01-01) or later.
- The virtual machines, which must be recovered during a disaster situation, must be running on an IBM POWER7® processor-based server, or later.
- The VM Recovery Manager HA solution requires Hardware Management Console (HMC) Version 9 Release 9.1.0, or later.
- The VM Recovery Manager HA solution requires VIOS Version 3.1.0.1, or later, with all the subsequent patches.
- Each logical partition (LPAR) in the host must have one of the following operating systems:
  - AIX version 6.1 or later
  - Red Hat Enterprise Linux (little endian, big endian) version 7.4 or later
  - SUSE Linux Enterprise Server version 12 or later
  - Ubuntu Linux distribution Version 16.04
  - IBM i Version 7.2 or later

PowerHA configuration

During PowerHA setup, application controller and custom monitor scripts are added. These are used by the resource group to monitor the KSYS application and handle failover scenarios.

The following two application monitors are added during setup:

- ksysmonstartup – It is the first monitor that runs when the resource group comes online on a node.
- ksysmonlongrun – It monitors the continuous functioning of KSYS.

Both the monitors use the following custom scripts to perform their functions:

- **startupmonksys.sh**
  This custom script checks the existence of a temporary file. Initially there is no temporary file in any node. This script would pass if the temporary script is present, or else it would fail and exit after creating a temporary file.

- **startksys.sh**
  As there is no temporary file initially, startupmon.sh will always fail and exit by creating the temporary file. The failure of startupmonksys.sh will call the start.sh script.
  This script first checks if the current node is the group leader node or not.
  If yes, it will run the remaining script or else it would start the KSYS daemon on this node and exit, indicating resource group failover to the next available node.
  In startksys.sh, the status of the KSYS daemon (IBM.VMR) is checked. If it is in the *inoperative* state, it is started manually. Else, if it is in the *running* state it will be stopped and restarted. This will sync the configuration changes made in the cluster.

- **longrunmonksys.sh**
  After the startksys.sh script is successfully passed, the **startupmonksys.sh** script is again called and this time the temporary file is present on the node so the startupmon.sh passes successfully. This event calls the longrunmonksys.sh script.
  The longrunmonksys.sh script deletes the temporary file. It monitors the state of IBM.VMR on the group leader node. If the node is not the group leader node or the state of IBM.VMR is not *active*, it will fail, and the resource group will fall over to the next available KSYS node.

Resource group

Resource group is a feature in PowerHA that enables users to add or remove controllers and monitors used to monitor the cluster nodes and the applications. During PowerHA cluster creation, the phaksyssetup.sh script checks for an existing resource group. If not present, it creates a resource group mapped to the application controller created before the resource group.

The STARTUP and FALLOVER attributes for the resource group are also set during this setup.

File collection

Using the file collection feature, all the scripts are synced with all the cluster nodes. This feature helps to maintain consistency in the scripts.
If any change occurs in any script on a node, then it will be made available in the other nodes of the cluster.

PowerHA cluster creation

As the setup is ready with PowerHA and KSYS installation, you can move on to the creation of the PowerHA cluster for the KSYS nodes.
Creating a cluster of multiple KSYS nodes leads to provide HA for it.
To create a PowerHA cluster, you can develop a script that automates the process and makes it easy to implement.

File path: /opt/IBM/ksys/samples/pha

The setup script provides a choice to the user to either have a standard (HA) cluster or a linked cluster (DR) for the KSYS nodes. This is provided by using the menu options in the script which makes it more user interactive. For a PowerHA standard cluster for VM Recovery Manager HA, the script requires the following details:

- KSYS node names
- Shared disk for repository

For a PowerHA linked cluster for VM Recovery Manager DR, the script requires the following details:

- KSYS node names for both the sites
- Repository disk for site1
- Repository disk for site2

After providing the required details, the sample script creates a standard or a linked cluster based on the chosen option.

The following PowerHA resources are created by the script:

- Resource group – ksysRG
- Startup monitor – ksysmonstartup
- Startup monitor script - /opt/IBM/ksys/samples/pha/startupmonksys
- Long running monitor – ksysmonlongrun
- Long running monitor script – /opt/IBM/ksys/samples/pha/longrunmonksys
- Start script – /opt/IBM/ksys/samples/pha/startksys
- Stop script – /opt/IBM/ksys/samples/pha/stopksys
- File collection – ksysfiles

Functionalities of the phaksyssetup.sh script:

- It takes input such as site details and the disk which is used to store the cluster data based on the chosen option for HA or DR cluster from the menu.
- It validates the PowerHA and KSYS file set installed on the node by checking if the mentioned file sets exist on the node or not.
- Checks if there is an existing PowerHA cluster or not. If not, then it creates the cluster.
- It creates the application controllers using the defined start script and stop script.
- It creates two application monitors that run the *startupmonksys* script and the *longrunmonksys* script.
- Checks the existence of a resource group. If resource group is not existing, it then adds the resource group to the *startupmonksys* and *longrunmonksys* scripts.
- KSYS would fetch the cluster details such as *Name*, *State*, and *Type* from the XML file stored at the location: /var/ksys/config/ksysmgr.xml.
  This script creates the file (ksysmgr.xml) with the correct cluster name and modifies the cluster type as *HA* when the user selects the **Standard Cluster** option from the first menu.
- At the end it makes the cluster online on the KSYS node.

In case of a standard (HA) cluster:

If requirement is there for VM Recovery Manager HA ksysnode high availability, then standard HA cluster need to be selected. Shared disk and list of ksysnode is required to be passed as input to prompt as shown in Figure 3.

```
(0) root @ ksys810: /opt/IBM/ksys/samples/pha
# ./phaksyssetup
        PHA integration with KSYS


        Which PHA cluster do you need?
        ------------------------------------------
        1) Standard (HA)
        2) Linked (DR)

        q) Quit

        Enter your selection: 1

You selected cluster type as HA
Please provide the HA values.

Enter the ksys nodes separated by ',' : ksys810,ksys811
Enter the shared disk: hdisk5
User provided KSYS nodes to be part of PHA: ksys810,ksys811

All filesets are present.

0513-044 The clcomd Subsystem was requested to stop.
0513-059 The clcomd Subsystem has been started. Subsystem PID is 6226308.

ERROR: no cluster is defined.

No PHA cluster is present. Creating PHA cluster with name ksys810_cluster.
Initializing..
```

Figure 3. Standard HA cluster creation

In case of a linked (DR) cluster:

If requirement is there for VM Recovery Manager DR ksysnode High availability. Linked DR cluster option is to be selected. For creating Linked DR cluster shared disk of each site is to be given as input along with list of ksysnodes as shown in Figure 4.

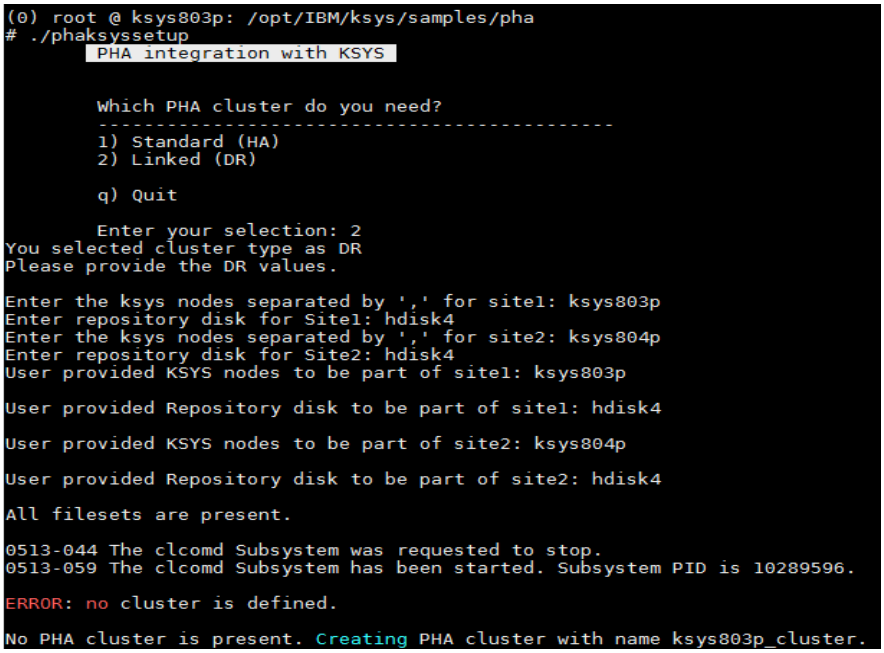

Figure 4: Linked DR cluster creation

Output of the phaksyssetup script for a standard (HA) cluster

```
(0) root @ ksys810: /opt/IBM/ksys/samples/pha
# ./phaksyssetup
        PHA integration with KSYS

        Which PHA cluster do you need?
        ------------------------------------------
-
        1) Standard (HA)
        2) Linked (DR)

        q) Quit

        Enter your selection: 1

You selected cluster type as HA
```

```
Please provide the HA values.


Enter the ksys nodes separated by ',' :
ksys810,ksys811
Enter the shared disk: hdisk5
User provided KSYS nodes to be part of PHA:
ksys810,ksys811


All filesets are present.


0513-044 The clcomd Subsystem was requested to stop.
0513-059 The clcomd Subsystem has been started.
Subsystem PID is 6226308.


ERROR: no cluster is defined.


No PHA cluster is present. Creating PHA cluster with
name ksys810_cluster.
Initializing..
Gathering cluster information, which may take a few
minutes...
Processing...
Storing the following information in file
/usr/es/sbin/cluster/etc/config/clvg_config


ksys810:


Hdisk:          hdisk0
PVID:           00f8309c3a81e72c
VGname:         rootvg
VGmajor:        10
Conc-capable:   No
VGactive:       Yes
Quorum-required:Yes
Hdisk:          hdisk1
PVID:           00f8309c0695258b
VGname:         None
VGmajor:        0
```

```
Conc-capable:   No
VGactive:       No
Quorum-required:No
Hdisk:          hdisk2
PVID:           00f8309c0695cc72
VGname:         None
VGmajor:        0
Conc-capable:   No
VGactive:       No
Quorum-required:No
Hdisk:          hdisk3
PVID:           00f8309c0695ccb3
VGname:         None
VGmajor:        0
Conc-capable:   No
VGactive:       No
Quorum-required:No
Hdisk:          hdisk4
PVID:           00f8309c0695ccf0
VGname:         None
VGmajor:        0
Conc-capable:   No
VGactive:       No
Quorum-required:No
Hdisk:          hdisk5
PVID:           00f8309c0695cd39
VGname:         None
VGmajor:        0
Conc-capable:   No
VGactive:       No
Quorum-required:No
FREEMAJORS:     39...

ksys811:

Hdisk:          hdisk0
```

```
PVID:           00f8309c3a5128e6
VGname:         rootvg
VGmajor:        10
Conc-capable:   No
VGactive:       Yes
Quorum-required:Yes
Hdisk:          hdisk1
PVID:           00f8309c0695258b
VGname:         None
VGmajor:        0
Conc-capable:   No
VGactive:       No
Quorum-required:No
Hdisk:          hdisk2
PVID:           00f8309c0695cc72
VGname:         None
VGmajor:        0
Conc-capable:   No
VGactive:       No
Quorum-required:No
Hdisk:          hdisk3
PVID:           00f8309c0695ccb3
VGname:         None
VGmajor:        0
Conc-capable:   No
VGactive:       No
Quorum-required:No
Hdisk:          hdisk4
PVID:           00f8309c0695ccf0
VGname:         None
VGmajor:        0
Conc-capable:   No
VGactive:       No
Quorum-required:No
Hdisk:          hdisk5
PVID:           00f8309c0695cd39
```

```
VGname:         None
VGmajor:        0
Conc-capable:   No
VGactive:       No
Quorum-required:No
FREEMAJORS:     39...


Warning: since no heartbeating type was specified,
and unicast is not available
         on all nodes, a default heartbeating style
of multicast will be used.
Default Multicast IP address will be assigned during
synchronization
Successfully added a primary repository disk.
To view the complete configuration of repository
disks use:
"clmgr query repository" or "clmgr view report
repository"
Cluster Name:    ksys810_cluster
Cluster Type:    Standard
Heartbeat Type:  Multicast
Repository Disk: hdisk5 (00f8309c0695cd39)
Cluster IP Address: None


There are 2 node(s) and 1 network(s) defined


NODE ksys810:
        Network net_ether_01
                ksys810 10.40.0.57


NODE ksys811:
        Network net_ether_01
                ksys811 10.40.0.58


No resource groups defined


*** The initial cluster configuration information
has been saved. You can now
```

```
define repository disks, along with other
configuration information. When the

cluster configuration is fully defined, verify and
synchronize the cluster to

deploy the configuration to all defined nodes.



Communication path ksys810 discovered a new node.
Hostname is ksys810.ausprv.stglabs.ibm.com. Adding
it to the configuration with Nodename ksys810.

Communication path ksys811 discovered a new node.
Hostname is ksys811.ausprv.stglabs.ibm.com. Adding
it to the configuration with Nodename ksys811.

Discovering IP Network Connectivity


Retrieving data from available cluster nodes.  This
could take a few minutes.


        Start data collection on node ksys810

        Start data collection on node ksys811

        Collector on node ksys811 completed

        Collector on node ksys810 completed

        Data collection complete

        Completed 10 percent of the verification
checks

        Completed 20 percent of the verification
checks

        Completed 30 percent of the verification
checks

        Completed 40 percent of the verification
checks

        Completed 50 percent of the verification
checks

        Completed 60 percent of the verification
checks

Discovered [2] interfaces

        Completed 70 percent of the verification
checks

        Completed 80 percent of the verification
checks

        Completed 90 percent of the verification
checks
```

```
        Completed 100 percent of the verification
checks
IP Network Discovery completed normally



Discovering Volume Group Configuration
Adding KSYS app controller ksysctrler.
Adding KSYS startup app monitor ksysmonstartup.


claddappmon warning:  The parameter
"HUNG_MONITOR_SIGNAL" was not specified.  Will use
9.
claddappmon warning:  The parameter
"RESTART_INTERVAL" was not specified.  Will use 198.
Adding KSYS long running app monitor .


claddappmon warning:  The parameter
"HUNG_MONITOR_SIGNAL" was not specified.  Will use
9.
claddappmon warning:  The parameter
"RESTART_INTERVAL" was not specified.  Will use 198.
The following file collections will be processed:


        ksysfiles


Starting file propagation to remote node ksys811.
clfileprop[1649]: Fetching file modification time
for name=/opt/IBM/ksys/samples/pha/longrunmonksys on
node ksys811
clfileprop[1669]: Updating file modification time
for /opt/IBM/ksys/samples/pha/longrunmonksys on node
ksys811 to 1556175617
Successfully propagated file
/opt/IBM/ksys/samples/pha/longrunmonksys to node
ksys811.
clfileprop[1649]: Fetching file modification time
for name=/opt/IBM/ksys/samples/pha/startupmonksys on
node ksys811
clfileprop[1669]: Updating file modification time
for /opt/IBM/ksys/samples/pha/startupmonksys on node
ksys811 to 1556175617
```

```
Successfully propagated file
/opt/IBM/ksys/samples/pha/startupmonksys to node
ksys811.

clfileprop[1649]: Fetching file modification time
for name=/opt/IBM/ksys/samples/pha/startksys on node
ksys811

clfileprop[1669]: Updating file modification time
for /opt/IBM/ksys/samples/pha/startksys on node
ksys811 to 1556175617

Successfully propagated file
/opt/IBM/ksys/samples/pha/startksys to node ksys811.

clfileprop[1649]: Fetching file modification time
for name=/opt/IBM/ksys/samples/pha/stopksys on node
ksys811

clfileprop[1669]: Updating file modification time
for /opt/IBM/ksys/samples/pha/stopksys on node
ksys811 to 1556175617

Successfully propagated file
/opt/IBM/ksys/samples/pha/stopksys to node ksys811.

clfileprop[1649]: Fetching file modification time
for name=/var/ksys/config/ksysmgr.xml on node
ksys811

clfileprop[1669]: Updating file modification time
for /var/ksys/config/ksysmgr.xml on node ksys811 to
1556175617

Successfully propagated file
/var/ksys/config/ksysmgr.xml to node ksys811.

Total number of files propagated to node ksys811: 5


Saving existing /var/hacmp/clverify/ver_mping.log to

/var/hacmp/clverify/ver_mping.log.bak

Verifying clcomd communication, please be patient.

ERROR: Skipping multicast communication as enough
nodes are not available.

Multicast communication verification between nodes
passed.


Committing any changes, as required, to all
available nodes...

Adding any necessary PowerHA SystemMirror entries to
/etc/inittab and

/etc/rc.net for IPAT on node ksys810.
```

```
cldare: Configuring a 2 node cluster in AIX may take
up to 2 minutes. Please
        wait.


1 tunable updated on cluster ksys810_cluster.
Adding any necessary PowerHA SystemMirror entries to
/etc/inittab and
/etc/rc.net for IPAT on node ksys811.


Verification has completed normally.


Verification to be performed on the following:
        Cluster Topology
        Cluster Resources


Retrieving data from available cluster nodes.  This
could take a few minutes.


        Start data collection on node ksys810
        Start data collection on node ksys811
        Waiting on node ksys810 data collection, 15
seconds elapsed
        Waiting on node ksys811 data collection, 15
seconds elapsed
        Collector on node ksys811 completed
        Collector on node ksys810 completed
        Data collection complete



For nodes with a single Network Interface Card per
logical
network configured, it is recommended to include the
file
'/usr/es/sbin/cluster/netmon.cf' with a "pingable"
IP address as described in the 'PowerHA SystemMirror
Planning Guide'.
WARNING: File 'netmon.cf' is missing or empty on the
following nodes:
ksys810
ksys811
```

```
        Completed 10 percent of the verification
checks

        Completed 20 percent of the verification
checks

WARNING: Duplicate host found on node ksys810 in
/etc/hosts: p9zze-dsail.aus.stglabs.ibm.com

WARNING: Duplicate host found on node ksys810 in
/etc/hosts: goatp04.upt.austin.ibm.com

WARNING: Duplicate host found on node ksys810 in
/etc/hosts: storage249.aus.stglabs.ibm.com

WARNING: Duplicate host found on node ksys811 in
/etc/hosts: p9zze-dsail.aus.stglabs.ibm.com

WARNING: Duplicate host found on node ksys811 in
/etc/hosts: goatp04.upt.austin.ibm.com

WARNING: Duplicate host found on node ksys811 in
/etc/hosts: storage249.aus.stglabs.ibm.com

WARNING: There are IP labels known to PowerHA
SystemMirror and not listed

in file /usr/es/sbin/cluster/etc/clhosts.client on
node: ksys810.

Verification can automatically populate this file to
be used on a client

node, if executed in auto-corrective mode.

WARNING: There are IP labels known to PowerHA
SystemMirror and not listed

in file /usr/es/sbin/cluster/etc/clhosts.client on
node: ksys811.

Verification can automatically populate this file to
be used on a client

node, if executed in auto-corrective mode.

        Completed 30 percent of the verification
checks

        Completed 40 percent of the verification
checks

        Completed 50 percent of the verification
checks

        Completed 60 percent of the verification
checks

        Completed 70 percent of the verification
checks


Verifying XD Solutions...
```

```
        Completed 80 percent of the verification
checks
        Completed 90 percent of the verification
checks
        Completed 100 percent of the verification
checks


Verification has completed normally.
lscluster: Cluster services are not active.
WARNING: refreshing clxd daemon failed.
WARNING: refreshing clxd daemon failed.
Please wait for clxd to stabilize...
Please wait for clxd to stabilize...


Warning: "WHEN" must be specified. Since it was not,
a default of "now" will be
         used.



Warning: "MANAGE" must be specified. Since it was
not, a default of "auto" will
         be used.



Verifying Cluster Configuration Prior to Starting
Cluster Services.



WARNING: No backup repository disk is UP and not
already part of a VG for nodes :
WARNING: File 'netmon.cf' is missing or empty on the
following nodes:
ksys810
ksys811
WARNING: Duplicate host found on node ksys810 in
/etc/hosts: p9zze-dsail.aus.stglabs.ibm.com
WARNING: Duplicate host found on node ksys810 in
/etc/hosts: goatp04.upt.austin.ibm.com
WARNING: Duplicate host found on node ksys810 in
/etc/hosts: storage249.aus.stglabs.ibm.com
```

```
WARNING: Duplicate host found on node ksys811 in
/etc/hosts: p9zze-dsail.aus.stglabs.ibm.com

WARNING: Duplicate host found on node ksys811 in
/etc/hosts: goatp04.upt.austin.ibm.com

WARNING: Duplicate host found on node ksys811 in
/etc/hosts: storage249.aus.stglabs.ibm.com

WARNING: There are IP labels known to PowerHA
SystemMirror and not listed

in file /usr/es/sbin/cluster/etc/clhosts.client on
node: ksys810.

Verification can automatically populate this file to
be used on a client

node, if executed in auto-corrective mode.

WARNING: There are IP labels known to PowerHA
SystemMirror and not listed

in file /usr/es/sbin/cluster/etc/clhosts.client on
node: ksys811.

Verification can automatically populate this file to
be used on a client

node, if executed in auto-corrective mode.

ksys810: start_cluster: Starting PowerHA
SystemMirror

ksys810:   3735962      -  0:09 syslogd

ksys810: Setting routerevalidate to 1

ksys811: start_cluster: Starting PowerHA
SystemMirror

ksys811:   4325810      -  0:13 syslogd

ksys811: Setting routerevalidate to 1


The cluster is now online.



Starting Cluster Services on node: ksys810

This may take a few minutes.  Please wait...

ksys810: Apr 25 2019 02:02:44Starting execution of
/usr/es/sbin/cluster/etc/rc.cluster

ksys810: with parameters: -boot -N -b -P
cl_rc_cluster -A

ksys810:
```

```
ksys810: Apr 25 2019 02:02:44usage: cl_echo
messageid (default) messageApr 25 2019
02:02:44usage: cl_echo messageid (default)
messageApr 25 2019 02:02:45

ksys810: /usr/es/sbin/cluster/utilities/clstart:
called with flags -m -G -b -P cl_rc_cluster -B -A

ksys810:

ksys810:        Apr 25 2019 02:02:46

ksys810: Completed execution of
/usr/es/sbin/cluster/etc/rc.cluster

ksys810: with parameters: -boot -N -b -P
cl_rc_cluster -A.

ksys810: Exit status = 0

ksys810:



Starting Cluster Services on node: ksys811

This may take a few minutes.  Please wait...

ksys811: Apr 25 2019 02:02:46Starting execution of
/usr/es/sbin/cluster/etc/rc.cluster

ksys811: with parameters: -boot -N -b -P
cl_rc_cluster -A

ksys811:

ksys811: Apr 25 2019 02:02:46usage: cl_echo
messageid (default) messageApr 25 2019
02:02:46usage: cl_echo messageid (default)
messageApr 25 2019 02:02:48

ksys811: /usr/es/sbin/cluster/utilities/clstart:
called with flags -m -G -b -P cl_rc_cluster -B -A

ksys811:

ksys811:        Apr 25 2019 02:02:48

ksys811: Completed execution of
/usr/es/sbin/cluster/etc/rc.cluster

ksys811: with parameters: -boot -N -b -P
cl_rc_cluster -A.

ksys811: Exit status = 0
```

KSYS configuration for a HA cluster

After the PowerHA cluster is stable, user needs to create the KSYS cluster. Before starting to configure ksyscluster, verification checks are required to ensure that the PowerHA cluster is stable.

Before configuring KSYS, verify the following details:

- The output of the `ksysmgr q cl` command should display the respective cluster name and type as **HA** in both nodes (if not, /var/ksys/config/ksysmgr.xml may be incorrect).

- On all KSYS nodes, IBM.VMR should be in the active state. To check the state, run the following command:
  `clcmd lssrc -s IBM.VMR`

- PowerHA resource group is online only on the group leader node. To check the group leader node, run the following command:
  `lssrc -ls IBM.ConfigRM | grep -w GroupLeader`

Perform the following tasks on group leader of the KSYS nodes to create the KSYS cluster:

1. Add HMC.
   ```
   ksysmgr add hmc <HMC_NAME> login=<username>
   password=<password>
   hostname=<HMC_NAME>.aus.stglabs.ibm.com
   ```

2. Add the hosts.
   ```
   ksysmgr add host <HOST_NAME>
   ```

3. Query for free shared VIOS disks.
   ```
   ksysmgr query viodisk
   vios=<viosname1>,<viosname2>,<viosname3>,<viosname4
   >
   ```

4. Manage Virtual I/O Server (VIOS) instances and the virtual machines (VMs).
   ```
   ksysmgr unmanage vios
   <viosname1>,<viosname2>,<viosname3>,<viosname4>
   ksysmgr unmanage vm
   <vmname1>,<vmname2>,<vmname3>,<vmname4>
   ```

5. Create a host group.
   ```
   ksysmgr add host_group <HOST_GROUP_NAME>
   hosts=<HOST_NAME1>,<HOST_NAME2>
   ha_disk=<disk_uuid1> repo_disk=<disk_uuid2>
   ```

6. Modify ha_monitor for the managed VMs.
   ```
   ksysmgr modify vm <vm_name> ha_monitor=enable
   ```

7. Modify ha_monitor for the system.
   ```
   ksysmgr modify system ha_monitor=enable
   ```

8. Run the `discovery` command to successfully manage the VIOS instances and VMs.
   ```
   ksysmgr -t discovery host_group <HOST_GROUP_NAME>
   ```

KSYS configuration for a DR cluster

After the PowerHA cluster and the configuration is stable, user needs to create the KSYS cluster.

Before configuring KSYS, verify the following information:

1. `ksysmgr q cl` should be provided appropriate cluster names and **type** as DR in both nodes (if not /var/ksys/config/ksysmgr.xml may be incorrect).
2. On all KSYS nodes, IBM.VMR should be in the *active* state. To check the state, run the following command:
   `clcmd lssrc -s IBM.VMR`
3. PowerHA resource group is online only on the group leader node to check the group leader node, run the follwoing command:
   `lssrc -ls IBM.ConfigRM | grep -w GroupLeader`

Perform the following steps on the group leader of the KSYS nodes to create the KSYS DR cluster:

1. Create a ksyscluster with cluster type as DR.
2. Add a home site and a backup site.
3. Add the source HMC and the target HMC.
4. Add a new host on source site and backup site.
5. Pair the active host with the backup host.
6. Add a Source site storage agent and a target site storage agent.
7. Unmanage the required source site VM, if any.
8. Add the host to the host group.
9. Run discovery and verify.

Validation of PowerHA cluster stability

PowerHA cluster stability can be verified to ensure HA for ksysnode.

1. Validate the state of the PowerHA cluster using the cluster manager command.

```
(0) root @ ksys7001: /
# clcmd lssrc -ls clstrmgrES | egrep "NODE|state"
NODE ksys7002.ausprv.stglabs.ibm.com
Current state: ST_STABLE
NODE ksys7001.ausprv.stglabs.ibm.com
Current state: ST_STABLE
```

2. Validate if the group leader node is correct and is same as that of online resource group.

```
(0) root @ ksys7002: /
```

```
# lssrc -ls IBM.ConfigRM | grep -w GroupLeader |
head -1 | cut -f1 -d',' | cut -f2 -d':'
ksys7001.ausprv.stglabs.ibm.com
```

3. Validate the resource group to know which ksysnode is online to perform the KSYS operation.

```
(0) root @ ksys7001: /
# clRGinfo
-----------------------------------------------------
Group Name         State           Node
-----------------------------------------------------
RG1               ONLINE           ksys7001
                  OFFLINE          ksys7002
```

4. Validate if VMR KSYS daemon is active and running.

```
(0) root @ ksys7001: /
# clcmd lssrc -s IBM.VMR

-------------------------------
NODE ksys7002.ausprv.stglabs.ibm.com
-------------------------------
Subsystem         Group            PID       Status
 IBM.VMR          rsct_rm          12845502  active

-------------------------------
NODE ksys7001.ausprv.stglabs.ibm.com
-------------------------------
Subsystem         Group            PID       Status
IBM.VMR          rsct_rm          15991140   active
```

VM failure scenario in a standard (HA) cluster

HA of ksysnode during VM failure

During this process if the current KSYS node crashes, the resource group moves to the next available KSYS node. The next available KSYS node resumes the migration task.

For the next KSYS node to resume the task, the IBM.VMR daemon is stopped and restarted.

When a managed VM crashes, it is migrated to the next available host in the host group. The VM is then restarted on the next host with the same configuration as earlier. This migration is handled by KSYS. When the VM crashes, the failure detection engine detects the crash and initiates the *restart and cleanup* phase for that VM.

During this process, if the current KSYS node crashes, the resource group moves to the next available KSYS node. The next available KSYS node resumes the migration task. For the next KSYS node to resume the task, the IBM.VMR daemon is stopped and restarted.

This is achieved by the start and stop script that runs when the resource group moves to the new node.

The start script stops the daemon and restarts it. This synchronizes the data on the new current KSYS node, and it resumes the migration task which the previous group leader node left. Hence the VM migration process is successfully completed using KSYS HA through PowerHA.

Refer to the detailed explanation that need to be performed in each step.

Step 1:

Create a PowerHA cluster and a KSYS cluster. On successfully completing discovery, check for daemon status on all the nodes of the PowerHA cluster. Once the cluster state is stable, start with adding the ksyscluster resources from the group leader node.

KSYS node	KSYS daemon	Group leader
KSYS 7001	Active	Yes
KSYS 7002	Active	No

Table 1. Daemon status and group leader identification

Figure 5 shows the current status of the KSYS daemon and the group leader node.

```
(0) root @ ksys7001: /
# clcmd lssrc -s IBM.VMR

------------------------------
NODE ksys7002.ausprv.stglabs.ibm.com
------------------------------
Subsystem         Group            PID          Status
 IBM.VMR          rsct_rm          15466956     active

------------------------------
NODE ksys7001.ausprv.stglabs.ibm.com
------------------------------
Subsystem         Group            PID          Status
 IBM.VMR          rsct_rm          13238600     active

(0) root @ ksys7001: /
# clcmd lssrc -ls clstrmgrES | egrep "NODE|state"
NODE ksys7002.ausprv.stglabs.ibm.com
Current state: ST_STABLE
Last event run was JOIN_NODE_CO  on node 2
NODE ksys7001.ausprv.stglabs.ibm.com
Current state: ST_STABLE
Last event run was JOIN_NODE_CO  on node 2

(0) root @ ksys7001: /
# lssrc -ls IBM.ConfigRM | grep -w GroupLeader | head -1 | cut -f1 -d',' | cut -f2 -d':'
 ksys7001.ausprv.stglabs.ibm.com

(0) root @ ksys7001: /
# clRGinfo
-----------------------------------------------------------------------------
Group Name                   State            Node
-----------------------------------------------------------------------------
ksysRG                       ONLINE           ksys7001
                             OFFLINE          ksys7002
```

Figure 5. Daemon status and group leader identification

Step 2:

Fail one of the VMs from HMC using KDB mode=ON. Console output is displayed in Figure 6.

```
AIX Version 7
Copyright IBM Corporation, 1982, 2019.
Console login: Debugger entered via keyboard.
h_cede_end_point+000000             ori    r0,r0,0            <0000000000000000> r0=0
KDB(0)>
 Connection has closed
```

Figure 6. Crashing virtual machine using KDB

Restart KSYS 7001 from HMC after the recovery phase is started for the VM.

Run the following command on the HMC CLI to restart the KSYS node:

```
hscroot@vmhmc5:~> chsysstate -r lpar -m ksys7_8246-L2C-10018EA -o shutdown --immed --restart -n ksys7001
```

Step 3:

As soon as the current KSYS node crashes, the group leader node changes to the next available KSYS node and resource group comes online on the new group leader node.

KSYS node	KSYS daemon	Group leader
KSYS 7001	Inoperative	No
KSYS 7002	Active	Yes

Table 2. New status of the KSYS daemon and the group leader node

Figure 7 shows new group leader and daemon nodes after crashing the VM. It also shows resource group state ONLINE acquired by the other node.

```
(1) root @ ksys7002: /
# clRGinfo
-----------------------------------------------------------------------------
Group Name                  State           Node
-----------------------------------------------------------------------------
ksysRG                      OFFLINE         ksys7001
                            ACQUIRING       ksys7002


(0) root @ ksys7002: /
# lssrc -ls IBM.ConfigRM | grep -w GroupLeader | head -1 | cut -f1 -d',' | cut -f2 -d':'
 ksys7002.ausprv.stglabs.ibm.com

(0) root @ ksys7002: /
# clRGinfo
-----------------------------------------------------------------------------
Group Name                  State           Node
-----------------------------------------------------------------------------
ksysRG                      OFFLINE         ksys7001
                            ONLINE          ksys7002


(0) root @ ksys7002: /
# clcmd lssrc -s IBM.VMR

------------------------------
NODE ksys7002.ausprv.stglabs.ibm.com
------------------------------
Subsystem         Group            PID          Status
 IBM.VMR          rsct_rm          20840732     active

------------------------------
NODE ksys7001.ausprv.stglabs.ibm.com
------------------------------
Subsystem         Group            PID          Status
 IBM.VMR          rsct_rm          7799180      active
```

Figure 7. Console output of resource group and daemon status

Step 4:

When the resource group is online on the new node, the startupmonksys script is called. This script fails because there is no temporary file present in the ksysnode. Further, the same script will create the required temporary file to restart the daemon.

Due to prior failure, the startupmonksys script is called again to checks the existence of the temporary file and successfully passes this time as the temporary file is present.

After the *startupmonksys* script is successfully executed, run the *longrunmonksys* script to monitor the KSYS daemon periodically.

The following output shows the stop and restart of the KSYS daemon on the new node.

```
(0) root @ ksys7002: /
# lssrc -s IBM.VMR
Subsystem         Group            PID        Status
IBM.VMR           rsct_rm          12845502   stopping

(0) root @ ksys7002: /
# lssrc -s IBM.VMR
Subsystem         Group            PID        Status
IBM.VMR           rsct_rm          12845510   active
```

KSYS node	**KSYS daemon**	**Group leader**
KSYS 7001	Inoperative	No
KSYS 7002	Inoperative	Yes

KSYS node	**KSYS daemon**	**Group leader**
KSYS 7001	Inoperative	No
KSYS 7002	Active	Yes

Table 3. The start script toggles the KSYS daemon on the next group leader

From Table 3, you can notice that the KSYS daemon toggles between the *Inoperative* and *Active* states, and this is required for the new group leader KSYS node to resume the VM migration process.

Deleting a KSYS and PowerHA cluster and configuration

Perform the following steps to delete a KSYS and Power HA cluster and its corresponding configuration:

1. Make the PowerHA cluster offline (using the `clmgr offline cl` command).
2. Check the status of the peer domain using the `lsrpdomain` command. If the status is offline, turn it online using the `startrpdomain` command.
3. Delete all host groups from the KSYS configuration by using the `ksysmgr delete hg` command.
4. Stop peer domain (using the `stoprpdomain` command).
5. Delete the PowerHA cluster (using the `clmgr del cl` command).

These steps will delete both KSYS and the PowerHA configuration from the environment.

Limitations

- During KSYS configuration for HA, it is possible that the group leader node may not accept configuration changes, such as adding hosts to the KSYS cluster. This can be resolved by performing the following steps:Set the value of the environment variable.
  `“CT_MANAGEMENT_SCOPE” = 2`
- Run the following command to include HA as cluster type. By default, it is DR.
  `chrsrc -c IBM.VMR_SITE ClusterType='HA'`

Get more information

- IBM VM Recovery Manager HA for Power Systems
  https://www.ibm.com/support/knowledgecenter/en/SSHQN6_1.3/navigation/welcome.html
- IBM VM Recovery Manager DR for Power Systems
  https://www.ibm.com/support/knowledgecenter/en/SSHQV4_1.3/base/vmrm_introduction.html
- Planning PowerHA System mirror
  https://www.ibm.com/support/knowledgecenter/en/SSPHQG_7.2/plan/ha_plan.htm

Summary

From VM Recovery Manager configuration for either a HA or a DR cluster, KSYS is the single point of failure. If KYS goes during an operation, the operation either fails or halts, which can lead to a messed up environment. This paper explained how to provide high availability to KSYS using PowerHA SystemMirror thereby removing the single point of failure. With help from PowerHA SystemMirror, KSYS can be made available on two nodes. If one KSYS node goes down, the other KSYS node takes over and continues the operation.

About the authors

Dishant Doriwala is a developer in the VM Recovery Manager product team. He has more than 5 years' experience working with the IBM Systems Power platform including PowerHA SystemMirror and VM Recovery Manager. You can reach Dishant at dishantdoriwala@in.ibm.com

Alok Chandra Mallick is a developer in the VM Recovery Manager product team, working for Altran Technologies. He has more than 9 years' experience working with the IBM Systems Power platform including PowerHA SystemMirror and VM Recovery Manager. You can reach Alok at alok.mallick@altran.com

Harsh Ailani is a developer in the VM Recovery Manager product team, working for Altran Technologies. He has almost 2 years' experience working with the IBM Systems Power platform including VM Recovery Manager. You can reach Harsh at harsh.ailani@altran.com

IBM Systems
3039 Cornwallis Road
RTP, NC 27709

Produced in the United States of America

Please recycle