Automatic Repository Disk Replacement (ARR) in an IBM PowerHA cluster

Pooja Bableshwar
Dishant J. Doriwala
Prasad Dasari

March 02, 2016

This article explains the Automatic Repository Disk Replacement (ARR) feature of IBM® PowerHA® SystemMirror®. This article can help customers to configure backup repositories for automatic repository update. The PowerHA SystemMirror ARR feature helps to prevent a cluster from moving to the restricted mode in case a cluster repository disk is inaccessible or failed, ensuring that the cluster always remains in the stable state.

Introduction

IBM PowerHA SystemMirror version 7.2 introduces a new feature called Automatic Repository Disk Replacement (ARR).

This feature, when configured, can prevent a cluster from moving into the restricted mode, when an active cluster repository disk fails or is inaccessible. When a repository disk failure is detected, Cluster Aware AIX (CAA), a subsystem of PowerHA updates the cluster repository to any other available backup repositories in backup list and the cluster will be in the working mode.

Definition of repository disk and restricted mode

Repository disk is a disk that is shared across all nodes of the cluster and acts as a central repository for configuration and cluster management operations.

In earlier versions of PowerHA, if a repository disk fails, the cluster moves into the restricted mode. When the cluster is in the restricted mode, only critical cluster configuration operations, such as moving a resource group from an active node to a standby node, are allowed. Most of the topology-related operations such as adding a node to the cluster or synchronizing a cluster are not allowed. To get the cluster out of the restricted mode, manual intervention is required by an administrator, in which an alternate repository disk has to be provisioned manually.
How can ARR be used to prevent the cluster from going into the restricted mode?

With PowerHA version 7.2 or later releases, administrator can configure up to a maximum of six backup disks that can be used as repository disks. When a repository disk fails, the CAA subsystem of PowerHA will automatically replace and rebuild the failed repository with an alternative predefined backup repository. This swap will log a notification and information in the syslog.caa file.

If you have a linked cluster, which has two sites with their respective CAA clusters and associated repository disks, you can configure six backup repository disks per site. Because standard and stretched clusters have a single CAA cluster, you can configure six backup repository disks.

Software version supported and prerequisites

The ARR feature requires the following versions of PowerHA and AIX software:

- PowerHA SystemMirror Version 7.2, or later.
- One of the following versions of the IBM AIX® operating system:
  - AIX version 7.1.4 or later
  - AIX version 7.2.0 or later
- Corresponding Reliable Scalable Cluster Technology (RSCT) version along with the AIX version.

PowerHA cluster configuration for ARR

**Figure 1: Stretched cluster with cluster repository and backup disk from same storage**

In this article we will consider a stretched cluster with two sites, where each site consists of one node. SiteA is the production site and SiteB is the secondary site. Node1 is associated with SiteA and Node2 is associated with SiteB.
Following is the cluster configuration with one network and one resource group.

```bash
(0) root @ Node1: /
# cltopinfo
Cluster Name:    Node1_cluster
Cluster Type:    Stretched
Heartbeat Type:  Multicast
Repository Disk: hdisk1 (00f601736b563ee7)
Cluster IP Address: 228.40.1.43
Cluster Nodes:
  Site 1 (SiteA):
    Node1
  Site 2 (SiteB):
    Node2

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

NODE Node1:
  Network net_ether_01
    Node1  10.40.1.43

NODE Node2:
  Network net_ether_01
    Node2  10.40.1.44

Resource Group RG1
  Startup Policy   Online On Home Node Only
  Failover Policy  Failover To Next Priority Node In The List
  Fallback Policy  Fallback To Higher Priority Node In The List
  Participating Nodes      Node1 Node2
```

In this cluster configuration, `hdisk1 (00f601736b563ee7)` has caavg which is the cluster repository for Node1 and Node2, and all cluster related operations are performed from Node1. After the synchronization is complete from Node1, all changes to the cluster will be propagated throughout the cluster nodes.

```bash
(0) root @ Node1: /
# lspv
hdisk0 00f6017354cc8d9 rootvg active
hdisk1 00f601736b563ee7 caavg_private active

(0) root @ Node2: /
# lspv
hdisk0 00f6017354cc8d9 rootvg active
hdisk1 00f601736b563ee7 caavg_private active
```

This can also be viewed with the help of the PowerHA `clmgr` utility command.

```bash
(0) root @ Node1: /usr/es/sbin/cluster/utilities
# clmgr query repository
hdisk1 (00f601736b563ee7)

(0) root @ Node2: /usr/es/sbin/cluster/utilities
# clmgr query repository
hdisk1 (00f601736b563ee7)
```

For PowerHA version 7.2 release and later versions, running the query command for the cluster specifies whether ARR is made available.
Figure 2 shows the output of the **HACMPsircol** Object Data Manager (ODM), object on one of the cluster nodes before configuring ARR. As you can observe, the `backup_repository` field is empty. **HACMPsircol** is one of the PowerHA-related objects in ODM. This object stores the cluster information along with the repository disk information.

**Figure 2: HACMPsircol output before added backup repository disk**

You can add a backup repository for ARR using the System Management Interface Tool (SMIT) menu or the PowerHA `clmgr` utility command-line interface. In this article, we will see how SMIT is used to add backup repository disks. Refer to the following command to add backup repository disks using the `clmgr` utility.

```bash
(0) root @ Node1: / # clmgr add repository -?
cmgr add repository <disk>[,<backup_disk#2>,...] \
    [ SITE=<site_label> ] \
    [ NODE=<reference_node> ] \
    [ DISABLE_VALIDATION={false|true} ]
add => create, make, mk
repository => rp
```

Open the SMIT interface using the `smit hacmp` command at the command prompt and select the following options to add a backup repository disk (as shown in Figure 3).

**Smit hacmp → Cluster Nodes and Networks → Manage Repository Disks → Add a Repository Disk.**
This operation is performed on Node1. After this, you need to verify and sync for the changes to take place across the cluster.

**Figure 3: Adding a backup repository disk**

![Manage Repository Disks](image)

Next, you need to select the backup repository from the list of disks available. In this scenario, four backup disks are selected, as shown in Figure 4. ARR allows up to six disks.

**Figure 4: Selecting a backup desk from the available shared disks**

![Repository Disk](image)

After the disks are added, a command status message (as shown in Figure 5) is displayed in case the disk is added successfully to cluster.

**Figure 5: Command status after adding a backup disk**

![Command status](image)

The next step is to synchronize the cluster, so that the configuration is reflected across all the nodes of the cluster. This can be done using the `verify` and `synchronization` options provided by PowerHA.

After the cluster is synchronized, the configuration changes are reflected across all the nodes of the cluster. Use the `clmgr view report repository` command on each of the nodes to verify this...
information. This enables you to identify the disks that are added as backup disks and also the disk on which the current cluster is active.

**Node1 on Site A**

```
(0) root @ Node1: / 
# clmgr view report repository
Node1_cluster :
   00f601736b563ee7 hdisk1(Node2)   active
   00f601736b563dad hdisk2(Node2)   backup
   00f601736b563cba hdisk3(Node2)   backup
   00f601736b563b84 hdisk4(Node2)   backup
   00f601736b563aa4 hdisk5(Node2)   backup
```

**Node2 on Site B**

```
(0) root @ Node2: / 
# clmgr view report repository
Node1_cluster :
   00f601736b563ee7 hdisk1(Node2)   active
   00f601736b563dad hdisk2(Node2)   backup
   00f601736b563cba hdisk3(Node2)   backup
   00f601736b563b84 hdisk4(Node2)   backup
   00f601736b563aa4 hdisk5(Node2)   backup
```

Here, hdisk2, hdisk3, hdisk4, and hdisk5 are the backup repository disks.

The HACMPsircol ODM object on each of the cluster nodes now contains the physical volume ID (PVID) of the repository disks. Figure 6 shows the output along with repository disk and backup disk list.

**Figure 6: HACMPsircol output after adding backup disks**

```
(0) root @ Node1: /usr/es/sbin/cluster/utilities
# odmget HACMPsircol
HACMPsircol: 
  name = "Node1_cluster_sircol"
  id = 0
  uuid = "00f601736b5633ee7"
  ip_address = "228.40.1.43"
  repository = "00f601736b563ee7"
  backup_repository = "00f601736b563dad 00f601736b563cba 00f601736b563b84 00f601736b563aa4"
```

After setting up the backup disk, start the cluster services and wait for cluster stability.

**Demonstration of ARR by failing disk**

As shown in Figure 6, the disk with the PVID "00f601736b563ee7" (hdisk1) is a disk where the cluster repository is active. For demonstration purpose, we will fail I/O on hdisk1, and this can be done by removing the disk from Virtual I/O Server (VIOS) if that disk is a virtual Small Computer System Interface (VSCSI) or else it can be unmapped from storage if the corresponding active cluster repository disk is N_Port ID Virtualization (NPIV). If the disk is from back-end storage, failure of the disk can be done using the `portdisable` command in Fibre Channel switch.
Once the cluster service are started, state of cluster will be active.

```bash
(0) root @ Node1: /home/f/Tools
# clcmd lssrc -ls clstrmgrES | grep state
Current state: ST_STABLE
Current state: ST_STABLE
```

For the demonstration of ARR in this article, failing I/O operations on hdisk1 is done using kernel extension application.

```bash
(0) root @ Node1: /home/Tools
# lke fail_io_kext
a0256000
(0) root @ Node1: /home/Tools/
# fail_io -e /dev/hdisk1 ----------------------- failed I/O enabled
I/O fail ON
(0) root @ Node1: /home/Tools/
# dd if=/dev/hdisk1 of=/dev/null count=10
dd: 0511-051 The read failed.
: There is an input or output error. ---------------------- disk is inaccessible
0+0 records in.
0+0 records out.
```

Disk failed event is logged in syslog.caa which is from CAA. Most of the work of ARR is done by CAA, but SystemMirror needs to be connected to CAA so that configuration changes are propagated either from SystemMirror to CAA, or from CAA to SystemMirror. Figure 7 and Figure 8 show the logs of syslog.caa. Figure 7 shows that when hdisk1 fails, ARR is enabled, whereas, Figure 8 shows an event when an active cluster repository disk is replaced with the next disk in backup list (which is hdisk2).

**Figure 7: syslog.caa log with ARR enabled**
Figure 8: syslog.caa log when active disk is replaced with a backup disk

![syslog.caa log](image)

After failing active cluster repository disk, active disk is replaced by a disk from back up repository list. This can even be verified using the `lspv` command. Figure 9 shows that after hdisk1 is failed caavg is replaced and will be active on hdisk2, this is automatically updated by CAA, It can also be verified by checking the HACMPsircol ODM values. Comparing with Figure 6, the hdisk1 (00f601736b563ee7) repository was active, whereas, after automatic repository update, hdisk2 (00f601736b563dad) is changed to active repository and hdisk1 (00f601736b563ee7) is added to the backup list, as shown in figure 10.

Figure 9: lspv output after repository disk replacement

![lspv output](image)

Figure 10: HACMPsircol output after automatic repository replacement

![HACMPsircol output](image)

This can also be verified using the CAA command, `lscluster -d` and `clmgr view report repository` command. The `lscluster -d` command shows a list of disks, where hdisk2 is the active repository disk and the others are backup disks.

```bash
(0) root @ Node1: / 
# lscluster -d 
Storage Interface Query 
Cluster Name: Node1_cluster 
```
Cluster UUID: 8d6a2434-ccdd-11e5-8077-9a9da6c0850c
Number of nodes reporting = 2
Number of nodes expected = 2

Node Node1.ausprv.stglabs.ibm.com
Node UUID = 8d5cb556-ccdd-11e5-8077-9a9da6c0850c
Number of disks discovered = 5

    hdisk2:
        State : UP
        uDid : 200B75TL7711A0207210790003IBMfcp
        uUid : ba63c805-b68d-6157-91bb-b065d22c8c0b
        Site uUid : 51735173-5173-5173-5173-517351735173
        Type : REPDISK

    hdisk3:
        State : UP
        uDid : 200B75TL7711A0307210790003IBMfcp
        uUid : 58633a29-cedf-aa49-0495-56d72a198b55
        Site uUid : 51735173-5173-5173-5173-517351735173
        Type : BACKUP_DISK

    hdisk4:
        State : UP
        uDid : 200B75TL7711A0407210790003IBMfcp
        uUid : 90728701-4766-c3bf-b14c-7406ba8eabe0
        Site uUid : 51735173-5173-5173-5173-517351735173
        Type : BACKUP_DISK

    hdisk5:
        State : UP
        uDid : 200B75TL7711A0507210790003IBMfcp
        uUid : 45405d44-d4da-7e18-181b-c2543f826382
        Site uUid : 51735173-5173-5173-5173-517351735173
        Type : BACKUP_DISK

Node Node2.ausprv.stglabs.ibm.com
Node UUID = 8d4eb15e-ccdd-11e5-8077-9a9da6c0850c
Number of disks discovered = 5

    hdisk2:
        State : UP
        uDid : 200B75TL7711A0207210790003IBMfcp
        uUid : ba63c805-b68d-6157-91bb-b065d22c8c0b
        Site uUid : 51735173-5173-5173-5173-517351735173
        Type : REPDISK

    hdisk3:
        State : UP
        uDid : 200B75TL7711A0307210790003IBMfcp
        uUid : 58633a29-cedf-aa49-0495-56d72a198b55
        Site uUid : 51735173-5173-5173-5173-517351735173
        Type : BACKUP_DISK

    hdisk4:
        State : UP
        uDid : 200B75TL7711A0407210790003IBMfcp
        uUid : 90728701-4766-c3bf-b14c-7406ba8eabe0
        Site uUid : 51735173-5173-5173-5173-517351735173
        Type : BACKUP_DISK

    hdisk5:
        State : UP
        uDid : 200B75TL7711A0507210790003IBMfcp
        uUid : 45405d44-d4da-7e18-181b-c2543f826382
        Site uUid : 51735173-5173-5173-5173-517351735173
        Type : BACKUP_DISK

    hdisk1:
        State : UP
Automatic Repository Disk Replacement (ARR) in an IBM PowerHA cluster

If the backup repository is inaccessible, CAA will replace and rebuild any of the backup disks available in the list, as the active repository.

Conclusion

You can use the ARR feature of IBM PowerHA SystemMirror to prevent a cluster from moving into the *restricted mode* when an active cluster repository disk fails or is inaccessible, ensuring that the cluster always remains in the *stable* state.

References

- Automatic Repository Replacement in PowerHA v7.2
- Cluster repository disks

© Copyright IBM Corporation 2016
Trademarks
(www.ibm.com/developerworks/ibm/trademarks/)