CATE CitiSystems Server and Operating Environment

PowerHA SystemMirror (HACMP) for AIX Product Version 7.1 patchset 2

Release for Production

Document Version 7.4

1 Table of Contents

....1

1   TABLE OF CONTENTS...2

   1.1   VERSION CONTROL 4

2   APPLICABILITY 4

3   CERTIFICATION MATRIX...4

4   PACKAGES AND CHECKSUMS...5

5   RFP – POWERHA SYSTEMMIRROR (HACMP) FOR AIX...6

   5.1   INTRODUCTION 6
   5.2   INTENDED AUDIENCE 6
   5.3   OVERVIEW 6
   5.4   BENEFIT TO THE FIRM 6
   5.5   VTM IDS ADDRESSED 6
   5.6   DEPENDENCIES 6
      5.6.1   Networks 6
      5.6.2   Storage 7
      5.6.3   Software 7
      5.6.4   NIM 7
   5.7   OUT OF SCOPE 7
   5.8   KNOWN ISSUES 7
   5.9   SUMMARY OF NEW/CHANGED FEATURES 7
      5.9.1   Version 7.1 Patchset 2...7
      5.9.2   Version 7.1 Patchset 1...7
      5.9.3   Version 6.1 Patchset 3...8
      5.9.4   Version 6.1 Patchset 2...8
      5.9.5   Version 6.1 Patchset 1...8
      5.9.6   Version 5.5 Patchset 2...8
      5.9.7   Support for HACMP using the VIO Server 8
      5.9.8   Version 5.5 Patchset 1...8
   5.10   LIMITATIONS 8

6   POWERHA SYSTEMMIRROR DESCRIPTION 9

7   NEW FEATURES IN POWERHA 7.1 10

   7.1   CLUSTER AWARE AIX (CAA)..10
   7.2   CENTRAL REPOSITORY DISK 10
   7.3   HEARTBEAT NETWORKS 10
   7.4   COMMAND LINE INTERFACE 11
   7.5   ROOTVG SYSTEM EVENT 11
   7.6   RESOURCE MANAGEMENT ENHANCEMENTS 11

8   PRODUCT SPECIFICATION...11

   8.1   POWERHA IMPLEMENTATION MODEL 11
   8.2   NIM SERVER FILESET 11
   8.3   NIM CLIENT PACKAGES INSTALLED ON THE POWERHA NODES 11

9   NIM-MASTER SETUP..11

   9.1   PRE-INSTALLATION TASKS 11
   9.2   INSTALLATION 12
   9.3   DEINSTALLATION 12

10   INSTALLATION OF THE POWERHA SOFTWARE ON THE CLUSTER NODES 12

1.1 Version Control

Version		Date	Author	Comments
Product	**Document**			
5.5.0.0	1.0	January 2010	Chris Howarth	Release Notes
5.5.0.0	1.1	March 2010	Chris Howarth	Updated to add VTM fix 11980
5.5.0.0	2.0	May 2010	Chris Howarth	Added Support for HACMP with VIO Server using NPIV
5.5.0.0	3.0	August 2010	Chris Howarth	Patchset 2 Release Notes
5.5.0.0	3.1	September 2010	Chris Howarth	HDS storage with NPIV tested. Included new SAN stack configuration
6.1.0.0	4.0	December 2010	Chris Howarth	Release for Production
6.1.0.1	5.0	May 2011	Chris Howarth	Release for Production – Added support for vscsi devices, AIX 5.3
6.1.0.0	6.0	August 2011	Chris Howarth	Patchset 3 issued RFP. Updated IHS.hacmp to remove ODM entry
6.1.0.0	6.1	October 2012	Chris Howarth	Added support for AIX 7.1
7.1.0.0	7.0	January 2013	Chris Howarth	Release for Testing of PowerHA 7.1
7.1.0.0	7.1	May 2013	Chris Howarth	Release for Production of PowerHA 7.1
7.1.0.0	7.2	June 2013	Chris Howarth	Documentation update for multicast traffic
7.1.3.0	7.3	Sep 2014	Ravi Kumar Korukonda	Release for Testing of PowerHA 7.1.3
7.1.3.0	7.4	Nov 2014	Ravi Kumar Korukonda	Release for Production

2 Applicability

Business/Region affected	Global
Requestor/VR number	
Planview Project ID	CAT130049116
Platform	pSeries AIX
Operating system version	AIX 6.1 patchset 13 or later, AIX 7.1 patchset 6 or later
Security model tier	-
CAP	-
VTM ID	-
CVSS2 Risk Rating	-
Software Location/Package Name	https://gdsunix.nj.ssmb.com/cgi-bin/down.cgi?lookup=IHS.nim.lpp.hacmp71 nfs://stealth.nj.ssmb.com//export1/home1.localhost/sw/AIX/PKG/IHS.nim.lpp.hacmp 71.p2.1.0.0.0.bff

3 Certification Matrix

Hardware	pSeries Power5, Power6. Power7
Software	PowerHA SystemsMirror 7.1
SCxacl (Tier 3) Certification	-
SOE Patchlevel	AIX 6.1 patchset 13 or later, AIX 7.1 patchset 6 or later

4   Packages and Checksums

The following checksums were generated and may be validated using the indicated tools available in AIX :

CRC32:   `/usr/bin/cksum <filename>`
SHA1:   `/usr/bin/csum –h SHA1 <filename>`
MD5SUM: `/usr/bin/csum –h MD5 <filename>`

Package name	Size	CRC32	SHA1 Sum	MD5 SUM
IHS.nim.lpp.hacmp71.p2.1.0.0.0.bff	92032000	3119824047	8a38d5797765e4216f3826b0f607cc58ce6bf08e	e5e32e59662849a7a9bbcbe59bbdef71

Document Classification: Internal

5 RFP – PowerHA SystemMirror (HACMP) for AIX

CitiServer Platform Engineering – Ravi Kumar Korukonda – 14th Nov 2014

5.1 Introduction

CATE CitiSystems Server and Operating Environment Engineering is publishing the Release for Production of PowerHA SystemMirror 7.1 patchset 2.

5.2 Intended Audience

This document is intended to the Global AIX SOE Task Force, Global Security communities and anyone with a vested interest in the solution. For further questions, please email Unix Engineering at *CATE Global Engineering Support - HACMP.

5.3 Overview

PowerHA SystemMirror is an IBM product for AIX clustering which is widely used within Citi. PowerHA SystemMirror 7.1 is the latest incarnation of this software and includes a number of key architectural differences. This document provides a high level description of the new features of PowerHA 7.1. It details product installation and upgrade procedures from previous versions of HACMP.

PowerHA SystemMirror 7.1 patchset2 is based on version 7.1.3 Service Pack 1, this release added support:

- Support for unicast communication
- Hyphen and leading digit support in node labels
- Capability to dynamically modify the host name of a clustered node, etc

Note: Prior to configuring PowerHA 7.1 it is necessary to:
1. Ensure that multicast traffic is enabled on the network connecting the two LPARs **if the plan is to use multicast** communication (see section 5.6.1). Note that **unicast** communication is supported with this patchset 2. Recommendation is to use Unicast communication.
2. A cluster repository disk has been provisioned (see section 5.6.2)

5.4 Benefit to the Firm

PowerHA SystemsMirror provides a clustered environment on AIX improving application resilience.

5.5 VTM IDs Addressed

5.6 Dependencies
- AIX 6.1 patchset 13 or later, AIX 7.1 patchset 6 or later

5.6.1 Networks

- It is important to ensure multicast traffic is supported between the clustered LPARs if the **plan is to use multicast communication**. Please verify with Network Operations prior to configuring HACMP. Details are as follows:

  It is necessary for certain base multicast configuration to be enabled on the routers which provide L3 connectivity for HACMP cluster VLANs. The configuration should be added by Network Engineering. It may be necessary to raise a request against the appropriate regional team to have the configuration added, if it is not already present. The configuration syntax required on the routers varies according to the make and model of router in use, but essentially boils down to a minimum requirement of enabling IP multicast routing at a global level, and enabling IP PIM Sparse-Mode on the HACMP VLAN interfaces. There is no need to enable any PIM RP functionality for the HACMP multicast groups as these do not need to be

routed. More detail on the configuration syntax can be found in the GNAST-approved multicast standards documents at the URL below.

https://catecollaboration.citigroup.net/domains/networks/transportdom/transportstdshmpgs/multicast.aspx

The reason for requiring PIM to be enabled on the HACMP VLANs is due to the behaviour of IGMP snooping on the modern Cisco platforms used for server access in Citi. With these platforms, the IGMP snooping process does not fall back to flooding behaviour in the absence of an mrouter, and therefore an mrouter is necessary for consistent delivery of multicast traffic, even within a single VLAN.

5.6.2 Storage

- A shared LUN for use as a cluster repository disk which should be a minimum of 1 GB in size

5.6.3 Software

- AIX 7.1 SOE Patchset 6 or later, AIX 6.1 SOE Patchset 13 or later
- PowerHA – version 7.1.3, Service Pack 1

5.6.4 NIM

- NIM Server with AIX patchsets available at levels listed in Software section.
  Disk space as follows:
  /export/lpp_source   150 MB

5.7 Out of Scope

The following items are not in scope as part of this product certification:

- PowerHA System Mirror Enterprise Edition. This is IBM's long range clustering product.
- PowerHA SystemMirror on WPARs is not a certified or recommended configuration.

5.8 Known Issues

- The cluster repository disk cannot be mirrored. However at Citi this disk is a SAN LUN and so has some redundancy. PowerHA 7.1.1 permits the recreation of a new repository disk should this disk fail for some reason without downtime to the cluster.
- It is necessary to install APAR fix if rolling migration method is chosen to migrate PowerHA from 6.1 to 7.1.3. Instructions on how to install the ifix are described in the installation instructions.
- Verification and clstart may show libodm error message, this is harmless and may be ignored.
- Cluster.es.server.rte fileset failed to install if syslogd service is not working on AIX systems, refer this Alert https://catecollaboration.citigroup.net/domains/platstor/osunix/stdsrelateddocs/Alert-Syslogd-Bug-Fix-AIX-and-VIOS-v1.0-d1.0.pdf and fix syslogd issue prior to the installation of PowerHA filesets.

5.9 Summary of New/Changed Features

5.9.1 Version 7.1 Patchset 2

- Support for unicast communication
- Hyphen and leading digit support in node labels
- Capability to dynamically modify the host name of a clustered node, etc

5.9.2 Version 7.1 Patchset 1

- As this is a major version update the new features are detailed in section 7 below

5.9.3 Version 6.1 Patchset 3

- Several bug fixes. Of particular note is IZ93613 whereby if an application fails over to a secondary node, the filesystems of the primary node are not unmounted and the enhanced concurrent volume group is not varied off. Consequently the filesystem remains mounted on both nodes. This could potentially cause data corruption if the primary node were to write to the filesystem.
- Prior releases of PowerHA SystemMirror would not failover if all connectivity to SAN storage from the primary node were lost. This was because PowerHA assumed that data replication was handled by LVM mirroring. There was no intelligence built in to handle SAN data replication. As a workaround an ODM entry was added by the package IHS.hacmp to cater for scenarios where all paths to the SAN storage is lost for any disk. This update removes this ODM entry as HACMP 6.1 now recognizes errors reported by Powerpath and will also distinguish handle individual disk access issues resulting in failure at the resource group level rather than failing an entire node.

5.9.4 Version 6.1 Patchset 2

- Several bug fixes including two items which have been encountered on Citi HACMP clusters. This release also includes the installation of an APAR fix to disable the emaixos subsystem as this is no longer required by HACMP and can result in a core dump of this subsystem or the errpt system to flood with messages relating to failures starting this subsystem.

5.9.5 Version 6.1 Patchset 1

- Added support for IPv6. Improvements to the User Interface. Enhancements to the DLPAR functionality. Inclusion of Smart Assists for Oracle, DB2 and Websphere.

5.9.6 Version 5.5 Patchset 2

- Bug fixes and a vulnerability fix. This update is based on Service Pack 6
- Bug fix to IHS.hacmp

5.9.7 Support for HACMP using the VIO Server

- Documentation only update to include support for clustering of fully virtualized LPARs using the VIO servers and NPIV.

5.9.8 Version 5.5 Patchset 1

- Support for the use of host IP addresses for network heartbeat
- Asynchronous replication support
- Multiple bug fixes

5.10 Limitations

- PowerHA does not support 'Shared Concurrent Cluster Filesystem' for simultaneous read/write operations from multiple nodes. This functionality is provided by GPFS which is a separate licensed product from IBM.

6 PowerHA SystemMirror Description

In today's complex environments, providing continuous service for applications is a key component of a successful IT implementation. High availability is one of the components that contributes to providing continuous service for the application and clients, by making or eliminating both planned and unplanned systems and application downtime.

This is achieved through the elimination of hardware and software single points of failure (SPOF). A high availability solution will ensure that the failure of any component of the solution be it hardware, software, or system management, will help in reducing the outage time an of an application and its data, thereby increasing the availability to the user.

Figure 1shows provides a high level architecture of 2-node PowerHA cluster including the core components.
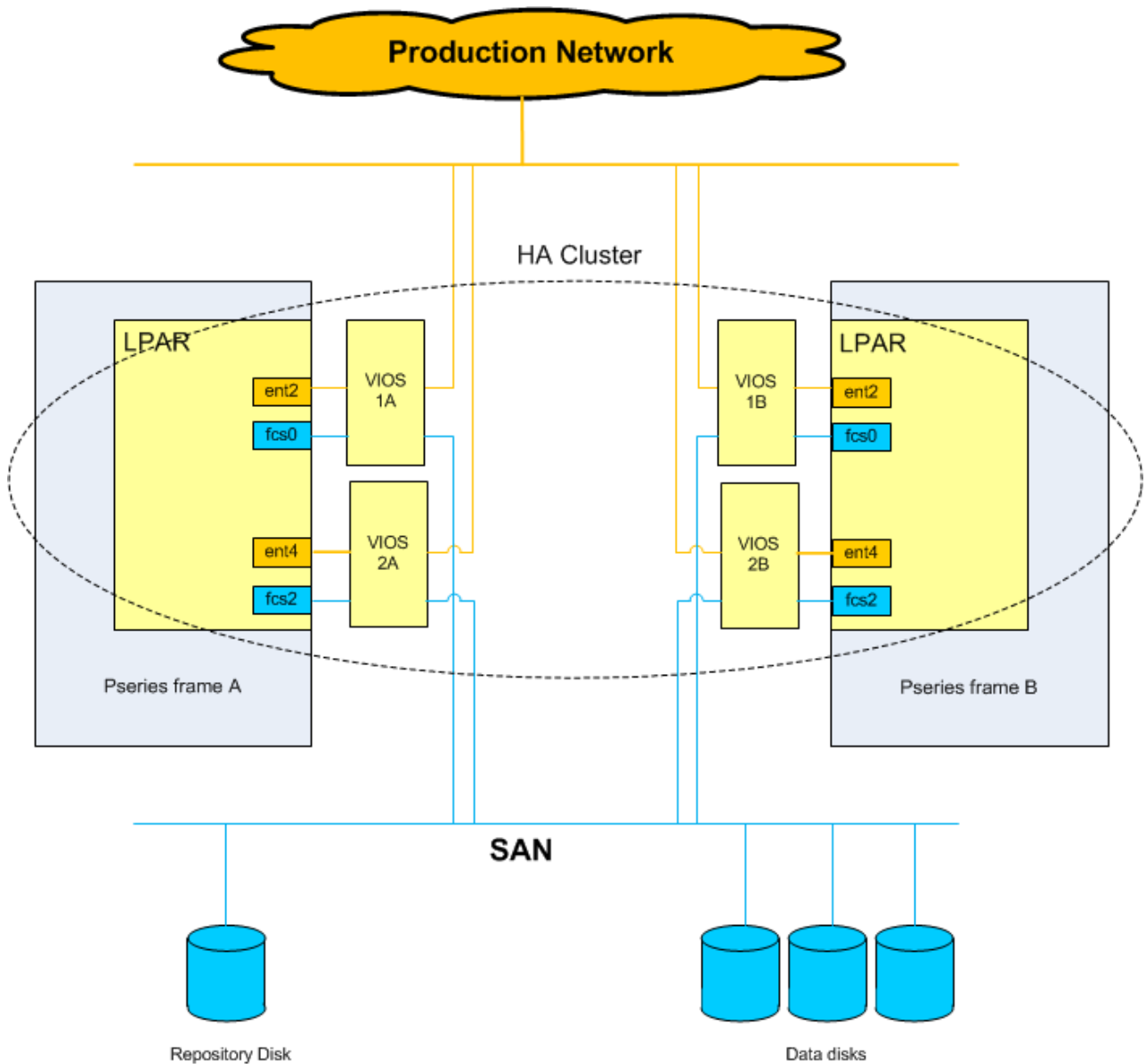
Figure 1: A Typical 2-Node PowerHA Cluster

The above configuration shows fully virtualized LPARs, but direct attached network and storage designs are also supported. For virtualized storage, the use of both NPIV and vscsi disks are supported.

The cluster nodes communicate with each other using heartbeats over the network. They also use a disk heartbeat using the cluster repository disk. Heartbeating across HBA adapters is also possible, but this is an option we are not intending to use at Citi as described in section 7.3 below. These heartbeat mechanism enable one to distinguish between a network failure and a cluster node failure.

A full description of PowerHA and its functionality is described in:
http://www.redbooks.ibm.com/redpieces/abstracts/sg248030.html

7   New Features in PowerHA 7.1

PowerHA 7.1 is a major update to previous releases and has a significant number of new features. The key items are summarized below:

7.1 Cluster Aware AIX (CAA)

Cluster Aware AIX (CAA) is a core component of the latest releases of AIX. It introduces a number of clustering capabilities into the base operating system. In particular the cluster communications now falls within the remit of CAA. The cluster topology services used for heartbeating, configuration information and live notification events all now use CAA. In previous releases this role was performed by RSCT.

7.2 Central Repository Disk

A central repository disk which is shared between the cluster nodes is now required as part of PowerHA7.1. This disk is used by CAA for configuration management, messaging and heartbeating. This is a dedicated disk and cannot be shared with application data. It is not the same as the heartbeat disk used in PowerHA 6.1.

7.3 Heartbeat Networks

There are a number of changes to the heartbeat networks as follows:

- **SAN-based communication interface for heartbeats**.
  This is a heartbeat between physical HBA devices belonging to the cluster LPARs. We are opting not to use this heartbeat mechanism for the following reasons:
    - This requires the HBA adapters which heartbeat to reside within the same SAN zone. This is contrary to the current standard SAN configuration standards used at Citi
    - Use of the SAN heartbeat with virtualized LPARs requires the creation of additional virtual network adapters in cluster LPARs which leads to non-standard LPAR configurations
    - There is no obvious benefit from a redundancy perspective unless the SAN heartbeat uses different switches/HBAs to those used for the repository disk connection.

- **Central cluster repository disk communication**.
  This disk is used for heartbeating.

- **Multicast / Unicast Based Heartbeat.**
  The Cluster Aware AIX (CAA) environment does have the option to select IP unicast or IP multicast for heartbeat exchanges.

  Multicast option is selected, PowerHA 7.1 uses multicast traffic for communication. The PowerHA software generates the multicast address for use automatically or one can be manually assigned. Following consultation with the Network Engineering team it has been decided to use the multicast address automatically assigned by PowerHA.

7.4 Command Line Interface

Arguably the most significant enhancement to PowerHA is the introduction of a simple CLI which greatly eases the configuration of HACMP clusters and enables scripted cluster configuration. In previous releases of HACMP the smit menus were the only officially supported configuration method and the underlying commands were liable to vary from one release to the next.

7.5 Rootvg system event

The rootvg system event allows for the monitoring of loss of access to the rootvg volume group. Previous releases were unable to detect a loss of the rootvg. For example if a SAN disk is lost which hosts the rootvg volume group.

7.6 Resource Management Enhancements

Additional resource group dependencies have been added to PowerHA to enable "Start After" and "Stop After" functionality.

8 Product Specification

8.1 PowerHA Implementation Model

Installation of the PowerHA software is a two stage process:

1. The IHS.nim.hacmp71 package is installed on the NIM master
2. Install the PowerHA software on the cluster nodes

8.2 NIM Server Fileset

The PowerHA software packages should first be installed on the nim server:

```
IHS.nim.lpp.hacmp71.p2.7.1.0.0.bff
IHS.nim.lpp.hacmp71.p2.rte HACMP 7.1.3 with Service Pack 1
```

This package contains:

a) IBM provided PowerHA filesets
b) A Citigroup engineering package (IHS.hacmp71) providing installation and de-installation scripts.

8.3 NIM Client Packages Installed on the PowerHA Nodes

In addition to the PowerHA filesets provided by IBM, the following Citigroup engineering package is installed:

```
IHS.hacmp71.1.0.0.1.bff
IHS.hacmp71.rte 1.0.0.1 Citigroup HACMP Package
```

9 NIM-MASTER SETUP

9.1 Pre-Installation Tasks

Setup the NIM master completely before installing this package. See the SOE AIX Installation Infrastructure document for more information. Verify sufficient disk space is available.

9.2 Installation

Install the filesets using standard AIX commands. Using the command line, after changing directory where the packages resides, run :

```
# installp –ad . IHS.nim.lpp.hacmp71.p2.rte
```

This will install the filesets that are included with the package into /export/lpp_source/hacmp71-p2.

Verify that the filesets are installed using lslpp. They should be in the committed state:

```
# lslpp -L IHS.nim.lpp.hacmp71.p2.rte
  Fileset                     Level  State  Type  Description (Uninstaller)
  ----------------------------------------------------------------------
  IHS.nim.lpp.hacmp71.p2.rte  1.0.0.0   C     F      HACMP 7.1.3 Service Pack 1
```

Add the lpp_source entry to the nim master:
```
# nim -o define -t lpp_source -a location=/export/lpp_source/hacmp71-p2 -a comments="HACMP 7.1.3
Service Pack 1" -a server=master lpp_hacmp71-p2
```

Note this will issue a warning lpp_source does not have the "simages" attribute, but this may be safely ignored. Verify the lpp_source has been created:

```
# lsnim -l lpp_hacmp71-p2
 lpp_hacmp71-p2:
   class        = resources
   type         = lpp_source
   comments     = HACMP 7.1.3 Service Pack 1
   arch         = power
   Rstate       = ready for use
   prev_state   = unavailable for use
   location     = /export/lpp_source/hacmp71-p2
   alloc_count  = 0
   server       = master
```

9.3 Deinstallation

To remove the package from the nim master, run the following command:

```
# installp –u IHS.nim.lpp.hacmp71.p2.rte
```

Remove the lpp_source entry

```
# nim -o remove lpp_hacmp71-p2
```

10 Installation of the PowerHA Software on the Cluster Nodes

The following procedure needs to be applied to all nodes in the cluster

10.1 Installation of the Citigroup Package – IHS.hacmp71.rte

First install this package from the nim master:

```
# nimclient –o allocate -a lpp_source=lpp_hacmp71-p2
# mount <nimserver>:/export/lpp_source/hacmp71-p2 /mnt
# installp -ad /mnt IHS.hacmp71
# umount /mnt
# nimclient -o deallocate -a lpp_source=lpp_hacmp71-p2
```

10.2 Installation of the PowerHA filesets

A script is provided to automate the installation of PowerHA. Prior to installation Seos should be disabled on USM servers. This script will:

1. Make some preliminary checks prior to software installation
2. Install some prerequisite O.S. filesets required by the PowerHA software from the nim server. Note you will need to know which O.S. and patchset your server is at (e.g. aix71-p6). For example, type:
   # oslevel –s
   7100-03-03-1415

   which corresponds to patchset 6. See
   https://catecollaboration.citigroup.net/domains/platstor/osunix/wpages/ProductDetail.aspx?Name=AIX&Publish=1 for patchset level details.

3. Install the PowerHA software from the nim server:

 # /opt/IHS.hacmp71/sbin/install all
    Enter patchset level (e.g. aix61-p12): **aix71-p6**
    Performing preliminary checks...
    ...
    ...
    Enter hacmp version (e.g. hacmp71-p1): **hacmp71-p2**
    ...
    ...

4. Install ifix to handle rolling migration known issue (this is required only if rolling migration method is chosen to upgrade PowerHA from 6.1 7.1.3)

   # nimclient –o allocate –a lpp_source=lpp_hacmp71-p2
   # mount <nim-server>:/export/lpp_source/hacmp71-p2 /mnt

   For AIX 6.1

   # emgr -e /mnt/emgr/AIX61.IV61060s3a.140627.epkg

   For AIX 7.1

   # emgr -e /mnt/emgr/AIX71.IV60736s3a.140627.epkg

   # emgr -l
   # umount /mnt
   # nimclient –o deallocate –a lpp_source=lpp_hacmp71-p2

5. Reboot the nodes following software installation

10.3 Deinstallation of the PowerHA filesets

A script is provided to automate the deinstallation of the PowerHA and the O.S. prerequisite filesets. Note that in removing the cluster filesets the entire PowerHA directory structure is removed. Consequently if a snapshot of the cluster has been created and is to be kept, it should be copied into a directory other than /usr/es/sbin/cluster/snapshots.

To deinstall the PowerHA software run:

 # /opt/IHS.hacmp/sbin/deinstall

Finally remove the Citi installation package:

 # installp –u IHS.hacmp71

11 Upgrading from PowerHA 6.1 to PowerHA 7.1.1

There are two methods to upgrade from HACMP 6.1 or later to HACMP 7.1:

- **Snapshot migration** – this requires the cluster to be shutdown prior to upgrade
- **Rolling migration** – each node in the cluster can be migrated one node at a time, moving the applications between nodes to minimize downtime during the migration.

Prior to performing any upgrade, ensure that the cluster is in a stable state with all nodes synchronized. Test application fallovers work. Perform a cluster snapshot and copy into safe location (snapshots are in /usr/es/sbin/cluster/snapshots). Create a mksysb image for each cluster node. Also make a copy of the rootvg on an alternate disk (using alt_disk_copy for example).

NOTE: Prior to upgrading you will need to ensure an additional shared LUN is provisioned to the clustered LPARS. This is the shared repository disk. It should be a minimum of 1 GB in size. Note this cannot be the same disk which is used for PowerHA 6.1 cluster heartbeats.

Ensure that scsi reservations are disabled for the cluster repository disk on all nodes e.g.:

```
chdev –l hdiskpower3 –a reserve_lock=no_reserve
```

11.1 Snapshot Migration

1. Make sure the cluster is running and then create a snapshot on first node

   ```
   smit cm_add_snap.dialog
   ```

   ```
   Cluster Snapshot Name                              [labcluster1_28Aug]          /
     Custom-Defined Snapshot Methods                  []                           +
     Save Cluster Log Files in snapshot               No                           +
   * Cluster Snapshot Description                     [labcluster1 upgrade to 7.1.3]
   ```

   Copy the snapshot to a different location:

   ```
   cp /usr/es/sbin/cluster/snapshots/labcluster1_28Aug.* /var/tmp/snaps
   ```

2. Stop the cluster on all nodes and verify is is shutdown:

   ```
   # lssrc -ls clstrmgrES
   Current state: ST_INIT
   ```

On both nodes perform steps 3,4 and 5
3. Install the O.S. prerequisite filesets for PowerHA 7.1

   ```
   # /opt/IHS.hacmp71/sbin/install prereq
   Enter patchset level (e.g. aix61-p12): aix61-p13
   ...
   ...
   Name                    Level        Part      Event      Result
   -------------------------------------------------------------------------
   bos.cluster.solid       6.1.7.15     USR       APPLY      SUCCESS
   bos.cluster.solid       6.1.7.15     ROOT      APPLY      SUCCESS
   bos.ahafs               6.1.7.15     USR       APPLY      SUCCESS
   bos.ahafs               6.1.7.15     ROOT      APPLY      SUCCESS
   bos.cluster.solid       6.1.8.15     USR       APPLY      SUCCESS
   bos.cluster.rte         6.1.7.15     USR       APPLY      SUCCESS
   bos.cluster.rte         6.1.7.15     ROOT      APPLY      SUCCESS
   bos.ahafs               6.1.9.15     USR       APPLY      SUCCESS
   ```

```
bos.ahafs                    6.1.9.15          ROOT       APPLY      SUCCESS
bos.cluster.rte              6.1.9.15          USR        APPLY      SUCCESS
bos.cluster.rte              6.1.9.15          ROOT       APPLY      SUCCESS
```

4. Edit /etc/cluster/rhosts on both nodes to include FQDN:

    # cat /etc/cluster/rhosts
    ```
    saixr312.apac.nsroot.net
    saixr313.apac.nsroot.net
    ```

5. Reboot the node

On first node perform steps 6, 7, 8 and 9

6. Run clmigcheck to verify cluster snapshot, assign a cluster repository disk and add a multicast address. For the multicast address use the default assigned by PowerHA.

    # cd /usr/es/sbin/cluster/snapshots
    # clmigcheck

    ```
    ------------[ PowerHA System Mirror Migration Check ]-------------


    Please select one of the following options:


    1       = Check ODM configuration.

    2       = Check snapshot configuration.

    3       = Enter reopsitory disk and multicast IP addresses.



    Select one of the above,"x"to exit or "h" for help:  2

    Enter snapshot name (in /usr/es/sbin/cluster/snapshots): labcluster1_28Aug.odm

    clsnapshot: Removing any existing temporary HACMP ODM entries...
    clsnapshot: Creating temporary HACMP ODM object classes...
    clsnapshot: Adding HACMP ODM entries to a temporary directory..

    ------------[ PowerHA System Mirror Migration Check ]-------------

    CONFIG-WARNING: The configuration contains unsupported hardware:  Disk
    Heartbeat network.  The PowerHA network name is net_diskhb_01.  This will be
    removed from the configuration during the migration
     to PowerHA System Mirror 7.1

    ------------[ PowerHA System Mirror Migration Check ]-------------

    The ODM has no unsupported elements.

    ------------[ PowerHA System Mirror Migration Check ]-------------

    Please select one of the following options:
    ```

```
1       = Check ODM configuration.


2       = Check snapshot configuration.


3       = Enter repository disk and IP addresses.


Select one of the above,"x"to exit or "h" for help:  3


------------[ PowerHA System Mirror Migration Check ]------------


Your cluster can use multicast or unicast messaging for heartbeat.
Multicast addresses can be user specified or default (i.e. generated by AIX).
Select the message protocol for cluster communications:


        1       = DEFAULT_MULTICAST
        2       = USER_MULTICAST
        3       = UNICAST


Select one of the above or "h" for help or "x" to exit:  3
Protocol choice is UNICAST
Hit <Enter> to continue
------------[ PowerHA System Mirror Migration Check ]------------


Select the disk to use for the repository
        1       = 00ccc9d6fbf21591(hdiskpower6)
        2       = 00ccc9d6fbf59f8e(hdiskpower8)
        3       = 00ccc9d6fbfcd0c3(hdiskpower7)


-----------[ PowerHA System Mirror Migration Check ]------------


Please select one of the following options:


1       = Check ODM configuration.


2       = Check snapshot configuration.


3       = Enter repository disk and IP addresses.


Select one of the above,"x"to exit or "h" for help:  x


------------[ PowerHA System Mirror Migration Check ]------------


You have requested to exit clmigcheck.
Do you really want to exit? (y) y
Note - If you have not completed the input of repsitory disks and
multicast IP addresses, you will not be able to install
PowerHA System Mirror


Additional details for this session may be found in
/tmp/clmigcheck/clmigcheck.log.
```

7. Uninstall PowerHA 6.1

```
# installp -u cluster
```

8. Install PowerHA 7.1 filesets

```
# /opt/IHS.hacmp71/sbin/install hacmp
Enter hacmp version (e.g. hacmp71-p1): hacmp71-p2
...
...
cluster.adt.es.client.inclu 7.1.3.1       USR         APPLY        SUCCESS
cluster.es.server.diag      7.1.3.1       USR         APPLY        SUCCESS
cluster.es.cspoc.rte        7.1.3.1       USR         APPLY        SUCCESS
cluster.es.client.rte       7.1.3.1       USR         APPLY        SUCCESS
cluster.es.client.rte       7.1.3.1       ROOT        APPLY        SUCCESS
cluster.es.server.events    7.1.3.1       USR         APPLY        SUCCESS
cluster.es.cspoc.cmds       7.1.3.1       USR         APPLY        SUCCESS
cluster.es.client.lib       7.1.3.1       USR         APPLY        SUCCESS
cluster.es.client.lib       7.1.3.1       ROOT        APPLY        SUCCESS
cluster.es.client.clcomd    7.1.3.1       USR         APPLY        SUCCESS
cluster.es.server.rte       7.1.3.1       USR         APPLY        SUCCESS
cluster.es.server.rte       7.1.3.1       ROOT        APPLY        SUCCESS
cluster.es.server.utils     7.1.3.1       USR         APPLY        SUCCESS
cluster.es.server.utils     7.1.3.1       ROOT        APPLY        SUCCESS
```

9. Restart clcomd

```
# stopsrc -s clcomd;startsrc -s clcomd
```

Repeat steps 7,8 and 9 on second node

On first node

10. Convert the snapshot to PowerHA 7.1

```
# /usr/es/sbin/cluster/conversion/clconvert_snapshot -v 6.1 -s labcluster1_28Aug
Extracting ODM's from snapshot file... done.
Converting extracted ODM's... done.
Rebuilding snapshot file... done.
```

11. Restore the snapshot

```
# smitty cm_apply_snap.select
```

```
                        Restore the Cluster Snapshot

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

                                              [Entry Fields]
  Cluster Snapshot Name                       labcluster1_28Aug
  Cluster Snapshot Description                labcluster1 upgrade to 7.1.3
  Un/Configure Cluster Resources?             [Yes]                    +
  Force apply if verify fails?                [No]                     +
```

On all nodes

12. Run the clmigcleanup script on all nodes one at a time to clean up the cluster configuration by removing heartbeat disks and redundant ODM entries

    ```
    # /usr/es/sbin/cluster/utilities/clmigcleanup
    ```

13. Start the cluster on each node (`smitty clstart`) in turn.

14. Following the upgrade the cluster should be sanity checked. In particular:

    - Check cluster network: `lscluster –m`
    - Check cluster topology: `cltopinfo`
    - Verify you can synchronize the cluster
    - Check that applications starr/stop and you can move resource groups
    - Test failure scenarios such as node failure, network failure, less of SAN connectivity.

15. Remove the package IHS.hacmp which was used for PowerHA 6.1 installations:

    ```
    # installp –u IHS.hacmp
    ```

16. Disable Firstalias option by referring section 15.4.2

11.2 Rolling Migration

1. The cluster should be running on all nodes. In this example we have two nodes called juraviocl4 and niveviocl4.The first node to be upgraded is juravioc4. Consequently any resource groups running on this node should first be migrated to another node; in this case resource groups are migrated to niveviocl4

Perform the following steps on the first node juraviocl4

2. Stop the cluster on juraviocl4 and verify it is stopped. Note that a bug in HACMP 6.1 patchset 3 may require you to stop and restart the clstrmgrES service if it remains in the ST_STABLE state.

    ```
    # lssrc -ls clstrmgrES
    Current state: ST_INIT
    ```

3. Install the O.S. prerequisite filesets for PowerHA 7.1 on juraviocl4

    ```
    # /opt/IHS.hacmp71/sbin/install prereq
    Enter patchset level (e.g. aix61-p10): aix61-p10
    ...
    ...
    Name                      Level          Part       Event       Result
    ----------------------------------------------------------------------------
    bos.cluster.solid         6.1.8.15       USR        APPLY       SUCCESS
    bos.cluster.rte           6.1.7.15       USR        APPLY       SUCCESS
    bos.cluster.rte           6.1.7.15       ROOT       APPLY       SUCCESS
    bos.ahafs                 6.1.9.15       USR        APPLY       SUCCESS
    bos.ahafs                 6.1.9.15       ROOT       APPLY       SUCCESS
    bos.cluster.rte           6.1.9.15       USR        APPLY       SUCCESS
    bos.cluster.rte           6.1.9.15       ROOT       APPLY       SUCCESS
    ```

4. Install ifix to handle rolling migration known issue

    ```
    # nimclient –o allocate –a lpp_source=lpp_hacmp71-p2
    ```

```
# mount <nim-server>:/export/lpp_source/hacmp71-p2 /mnt
```

For AIX 6.1

emgr -e /mnt/emgr/AIX61.IV61060s3a.140627.epkg

For AIX 7.1

emgr -e /mnt/emgr/AIX71.IV60736s3a.140627.epkg

```
# emgr -l
# umount /mnt
# nimclient -o deallocate -a lpp_source=lpp_hacmp71-p2
```

5. Edit /etc/cluster/rhosts to include FQDN on juraviocl4

```
# cat /etc/cluster/rhosts
niveviocl4.eulab.nsroot.net
juraviocl4.eulab.nsroot.net
```

6. Reboot juraviocl4 as the AIX image has been updated (this also restarts clcomd as required)

7. Run clmigcheck to verify ODM configuration, assign a cluster repository disk and IP address. You can choose either multicast or unicast messaging heartbeat. For the multicast address use the default assigned by PowerHA.

```
------------[ PowerHA System Mirror Migration Check ]-------------

Please select one of the following options:

1       = Check ODM configuration.

2       = Check snapshot configuration.

3       = Enter repository disk and IP addresses.


Select one of the above,"x"to exit or "h" for help:  1

------------[ PowerHA System Mirror Migration Check ]-------------

CONFIG-WARNING: The configuration contains unsupported hardware:  Disk
Heartbeat network.  The PowerHA network name is net_diskhb_02.  This will be
removed from the configuration during the migration
 to PowerHA System Mirror 7.1.


Hit <Enter> to continue

------------[ PowerHA System Mirror Migration Check ]-------------

The ODM has no unsupported elements.

------------[ PowerHA System Mirror Migration Check ]-------------

Please select one of the following options:

1        = Check ODM configuration.

2        = Check snapshot configuration.
```

```
3       = Enter repository disk and IP addresses.


Select one of the above,"x"to exit or "h" for help:  3


------------[ PowerHA System Mirror Migration Check ]------------

Your cluster can use multicast or unicast messaging for heartbeat.
Multicast addresses can be user specified or default (i.e. generated by AIX).
Select the message protocol for cluster communications:


        1        = DEFAULT_MULTICAST
        2        = USER_MULTICAST
        3        = UNICAST

Select one of the above or "h" for help or "x" to exit:  3
Protocol choice is UNICAST

------------[ PowerHA System Mirror Migration Check ]------------

Select the disk to use for the repository

        1        = 00ccc9d6fbf21591(hdiskpower6)
        2        = 00ccc9d6fbf59f8e(hdiskpower8)
        3        = 00ccc9d6fbfcd0c3(hdiskpower7)
        4        = 00ccc9d6fbfd6b07(hdiskpower10)

------------[ PowerHA System Mirror Migration Check ]------------

Please select one of the following options:

1       = Check ODM configuration.

2       = Check snapshot configuration.

3       = Enter repository disk and IP addresses.

------------[ PowerHA System Mirror Migration Check ]------------

You have requested to exit clmigcheck.

Do you really want to exit? (y) y

Note - If you have not completed the input of repsitory disks and
multicast IP addresses, you will not be able to install
PowerHA System Mirror

Additional details for this session may be found in
/tmp/clmigcheck/clmigcheck.log.
```

8. Install PowerHA 7.1 filesets on the first node juraviocl4

    # /opt/IHS.hacmp71/sbin/install hacmp
    Enter hacmp version (e.g. hacmp71-p1): **hacmp71-p2**
    ...

...

9. Start PowerHA on node juraviocl4 (`smitty clstart`). Note that at this point the heartbeat disk will still be present:

```
# cltopinfo
Cluster Name:    labcluster1
Cluster Type:    Standard
Heartbeat Type:  Unicast
Repository Disk: hdiskpower8 (00ccc9d6fbf59f8e)

There are 2 node(s) and 3 network(s) defined
NODE saixr312:
     Network net_diskhb_01
         saixr312_hdpwr1 /dev/hdiskpower1
     Network net_ether_01
     Network net_ether_010
         appa    169.187.213.31
         saixr312     169.187.213.29
NODE saixr313:
     Network net_diskhb_01
         saixr313_hdpwr3 /dev/hdiskpower3
     Network net_ether_01
     Network net_ether_010
         appa    169.187.213.31
         saixr313     169.187.213.30

Resource Group hatest
     Startup Policy   Online On Home Node Only
     Fallover Policy  Fallover To Next Priority Node In The List
     Fallback Policy  Fallback To Higher Priority Node In The List
     Participating Nodes     saixr312 saixr313
     Service IP Label          appa

Total Heartbeats Missed:      0
Cluster Topology Start Time:   09/05/2014 12:47:33
```

```
# cltopinfo
Cluster Name: cluslab
Cluster Connection Authentication Mode: Standard
Cluster Message Authentication Mode: None
Cluster Message Encryption: None
Use Persistent Labels for Communication: No
Repository Disk: hdiskpower3
Cluster IP Address:
There are 2 node(s) and 2 network(s) defined
NODE juraviocl4:
        Network net_diskhb_02
                juraviocl4_hdiskpower5_01        /dev/hdiskpower5
        Network net_ether_01
                gironde-appa    10.154.222.173
                juraviocl4      10.154.222.6
NODE niveviocl4:
        Network net_diskhb_02
                niveviocl4_hdiskpower5_01        /dev/hdiskpower5
        Network net_ether_01
                gironde-appa    10.154.222.173
                niveviocl4      10.154.222.249
```

```
Resource Group rgA
        Startup Policy   Online On Home Node Only
        Fallover Policy  Fallover To Next Priority Node In The List
        Fallback Policy  Never Fallback
        Participating Nodes      niveviocl4 juraviocl4
        Service IP Label                 gironde-appa

Total Heartbeats Missed:        0
Cluster Topology Start Time:    08/10/2012 13:14:04
```

10. Having completed juraviocl4, migrate the running applications from niveviocl4 onto juraviocl4.

Perform the following steps on the second node niveviocl4

11. Once no applications are present on niveviocl4 stop the cluster services on niveviocl4 and verify it has stopped

    ```
    # lssrc -ls clstrmgrES
    Current state: ST_INIT
    ```

12. Install the O.S. prerequisite filesets for PowerHA 7.1

    ```
    # /opt/IHS.hacmp71/sbin/install prereq
    Enter patchset level (e.g. aix61-p10): aix61-p10
    ...
    ...
    Name                      Level          Part       Event       Result
    -------------------------------------------------------------------------
    clic.rte.kernext          4.7.0.1        USR        APPLY       SUCCESS
    clic.rte.kernext          4.7.0.1        ROOT       APPLY       SUCCESS
    bos.cluster.solid         6.1.7.15       USR        APPLY       SUCCESS
    bos.cluster.solid         6.1.7.15       ROOT       APPLY       SUCCESS
    bos.ahafs                 6.1.7.15       USR        APPLY       SUCCESS
    bos.ahafs                 6.1.7.15       ROOT       APPLY       SUCCESS
    bos.cluster.rte           6.1.7.15       USR        APPLY       SUCCESS
    bos.cluster.rte           6.1.7.15       ROOT       APPLY       SUCCESS
    ```

13. Edit /etc/cluster/rhosts on niveviocl4 to include FQDN

    ```
    # cat /etc/cluster/rhosts
    niveviocl4.eulab.nsroot.net
    juraviocl4.eulab.nsroot.net
    ```

14. Reboot miveviocl4 as AIX image has been updated (which will also restart clcomd as required)

15. Run clmigcheck to verify the CAA cluster

    ```
    # clmigcheck
    Verifying clcomd communication, please be patient.

    Verifying multicast IP communication, please be patient.

    Verifying IPV4 multicast communication with mping.

    clmigcheck: Running /usr/sbin/rsct/install/bin/ct_caa_set_disabled_for_migration on
    each node in the cluster
    ```

```
------------[ PowerHA System Mirror Migration Check ]------------

About to configure a 2 node CAA cluster, this can take up to 2 minutes.
Creating CAA cluster, please be patient.

------------[ PowerHA System Mirror Migration Check ]------------

You can install the new version of PowerHA System Mirror.


Hit <Enter> to continue
```

Check the CAA cluster was successfully created:

 # lscluster -m

16. Install PowerHA 7.1.3 filesets on niveviocl4

    # /opt/IHS.hacmp71/sbin/install hacmp
    Enter hacmp version (e.g. hacmp71-p1): **hacmp71-p2**
    ...
    ...

17. Start the cluster on niveviocl4 (smitty clstart). This will remove the legacy PowerHA6.1 disk heartbeats from the cluster. Note this will take a few minutes disappear after starting the cluster software on niveviocl4.

18. Verify the cluster has completed the migration on both nodes

    # odmget HACMPcluster | grep cluster_version
            cluster_version = 15

    # odmget HACMPnode | grep version | sort -u
            version = 15

19. Following the upgrade the cluster should be sanity checked. In particular:

    - Check cluster network: lscluster –m
    - Check cluster topology: cltopinfo
    - Check you can verify and synchronize the cluster
    - Check that applications starr/stop and you can move resource groups
    - Test failure scenarios such as node failure, network failure, less of SAN connectivity.

20. Disable Firstalias option by referring the section 15.4.2

11.3 Back out of Upgrade to PowerHA 7.1
Boot back from the alternate rootvg disk. This will have preserved the old cluster configuration including the cluster repository disk.

12 Upgrading from PowerHA 7.1.1 to PowerHA 7.1.3

1. On juraviocl4 node, stop cluster services (smitty clstop) with the option Move resource groups.

2. Install PowerHA 7.1.3 filesets on juraviocl4 node.

```
# /opt/IHS.hacmp71/sbin/install hacmp
Enter hacmp version (e.g. hacmp71-p1): hacmp71-p2
...
...
```

3. Start cluster services (smitty clstart) on juraviocl4 node

4. Verify migration status using following command:

```
 # lssrc -ls clstrmgrES
Current state: ST_STABLE
sccsid = "@(#)36 1.135.1.118
src/43haes/usr/sbin/cluster/hacmprd/main.C,hacmp.pe,61haes_r713,1343A_hacmp713 10/21/"
build = "May  6 2014 15:08:06 1406D_hacmp713"
i_local_nodeid 0, i_local_siteid -1, my_handle 1
ml_idx[1]=0    ml_idx[2]=1
There are 0 events on the Ibcast queue
There are 0 events on the RM Ibcast queue
CLversion: 13  ➜ this means migration is still in progress
local node vrmf is 7131
cluster fix level is "1"
```

5. On niveviocl4, stop the cluster services with option Move resource group

6. Install PowerHA 7.1.3 filesets on niveviocl4 node.

```
# /opt/IHS.hacmp71/sbin/install hacmp
Enter hacmp version (e.g. hacmp71-p1): hacmp71-p2
...
...
```

7. Start cluster services on niveviocl4 node.

8. Verify that the cluster has completed the migration on both nodes as shown in

```
# lssrc -ls clstrmgrES
Current state: ST_STABLE
sccsid = "@(#)36 1.135.1.118
src/43haes/usr/sbin/cluster/hacmprd/main.C,hacmp.pe,61haes_r713,1343A_hacmp713 10/21/"
build = "May  6 2014 15:08:06 1406D_hacmp713"
i_local_nodeid 1, i_local_siteid -1, my_handle 2
ml_idx[1]=0    ml_idx[2]=1
There are 0 events on the Ibcast queue
There are 0 events on the RM Ibcast queue
CLversion: 15  ➜ this means migration is completed
local node vrmf is 7131
cluster fix level is "1"
```

```
# odmget HACMPcluster | grep cluster_version
        cluster_version = 15
```

```
# odmget HACMPnode | grep version | sort -u
        version = 15
```

13 Procedure to convert from multicast to Unicast

Below is the procedure to convert from multicast to Unicast communication if the existing cluster is running with multicast communication.

1. Verify that the existing CAA communication mode is set to multicast as shown

   # lscluster -c
   > Cluster Name: clus_lab
   > Cluster UUID: 4aac6db8-194d-11e4-9143-f6071c1c74e7
   > Number of nodes in cluster = 2
   > Cluster ID for node g10-p750-01-viocl1.nam.nsroot.net: 1
   > Primary IP address for node g10-p750-01-viocl1.nam.nsroot.net: 10.40.38.205
   > Cluster ID for node g10-p750-02-viocl1.nam.nsroot.net: 2
   > Primary IP address for node g10-p750-02-viocl1.nam.nsroot.net: 10.40.38.212
   > Number of disks in cluster = 1
   > Disk = hdiskpower0 UUID = 9d564e76-8f7d-34a0-0507-9c4a3db8f3de cluster_major = 0 cluster_minor = 1
   > Multicast for site LOCAL: IPv4 228.40.38.205 IPv6 ff05::e428:26cd
   > **Communication Mode: multicast**
   > Local node maximum capabilities: HNAME_CHG, UNICAST, IPV6, SITE
   > Effective cluster-wide capabilities: HNAME_CHG, UNICAST, IPV6, SITE

2. Change the heartbeat mechanism from multicast to unicast as below

   # smitty cm_define_repos_ip_addr -> Define Repository Disk and Cluster IP Address

* Cluster Name	clus_lab
* Heartbeat Mechanism	Unicast
Repository Disk	00f6071c67f35f3b
Cluster Multicast Address	228.40.38.205

   Note: Once a cluster has been defined to AIX, all that can be modified is the Heartbeat Mechanism

3. Run verify and sync

4. Verify that the new CAA communication mode is now set to unicast

   # lscluster -c
   Cluster Name: clus_lab
   Cluster UUID: 4aac6db8-194d-11e4-9143-f6071c1c74e7
   Number of nodes in cluster = 2
   > Cluster ID for node g10-p750-01-viocl1.nam.nsroot.net: 1
   > Primary IP address for node g10-p750-01-viocl1.nam.nsroot.net: 10.40.38.205
   > Cluster ID for node g10-p750-02-viocl1.nam.nsroot.net: 2
   > Primary IP address for node g10-p750-02-viocl1.nam.nsroot.net: 10.40.38.212
   Number of disks in cluster = 1
   > Disk = hdiskpower0 UUID = 9d564e76-8f7d-34a0-0507-9c4a3db8f3de cluster_major = 0 cluster_minor = 1
   Multicast for site LOCAL: IPv4 228.40.38.205 IPv6 ff05::e428:26cd
   Communication Mode: unicast
   Local node maximum capabilities: HNAME_CHG, UNICAST, IPV6, SITE
   Effective cluster-wide capabilities: HNAME_CHG, UNICAST, IPV6, SITE

14 PowerHA Server Configuration

This section details some recommendations relating to the PowerHA topology

14.1.1   Networking

- To ensure network resilience each node must have at least two interfaces connected to two separate switches which support the same subnet/VLAN. These interfaces can either be physical or virtual; if virtual then these should be configured as per the VIO server release notes and best practices guide (see VIOS standards) Within a cluster, all the nodes must be connected to the same subnet. This is required for seamless failover of the IP address used by applications, referred to as the Service IP address.

- To simplify the cluster configuration it is recommended to combine two network interfaces using Etherchannel with Network Interface Backup (NIB) on each node.

- PowerHA 7.1.3 uses either unicast/ multicast for network based heartbeats. As such there is no requirement to configure a separate private boot IP network. A persistent IP address is also not required.
- It is necessary to inform PowerHA of a ping address to use for the network adapters for use when a network communication error exists. This is configured using the netmon.cf file and is described below (section 15.4.1)

14.1.2   Storage

- The current SAN Stack should be used:

  http://catecollaboration.citigroup.net/domains/platstor/gdse/products/techstdshmpgs/SANStacks.aspx

  Clustered products are supported by CATE only on EMC Symmetrix and HDS Tagmastor

- HA applications must store their data on SAN shared storage devices. This data must be accessible to both nodes in the HA cluster. Each node must have a minimum of two HBAs attached to the SAN via different fabrics (either directly attached or through a pair of VIO servers – see VIOS standards)

- Enhanced Concurrent Volume Groups must be used.

- After the initial cluster installation and configuration, all shared volume management should be performed using the PowerHA System Management (C-SPOC).

15 Configuring PowerHA SystemMirror

This section details how to configure a 2 node cluster. Figure 2 below shows the example cluster to be created. There are two nodes called niveviocl4 and juraviocl4. niveviocl4 has an application appA running on it as its primary node and juraviocl4 has appB as its primary node. Note that all IP addresses and disks listed are examples only and should not be used.
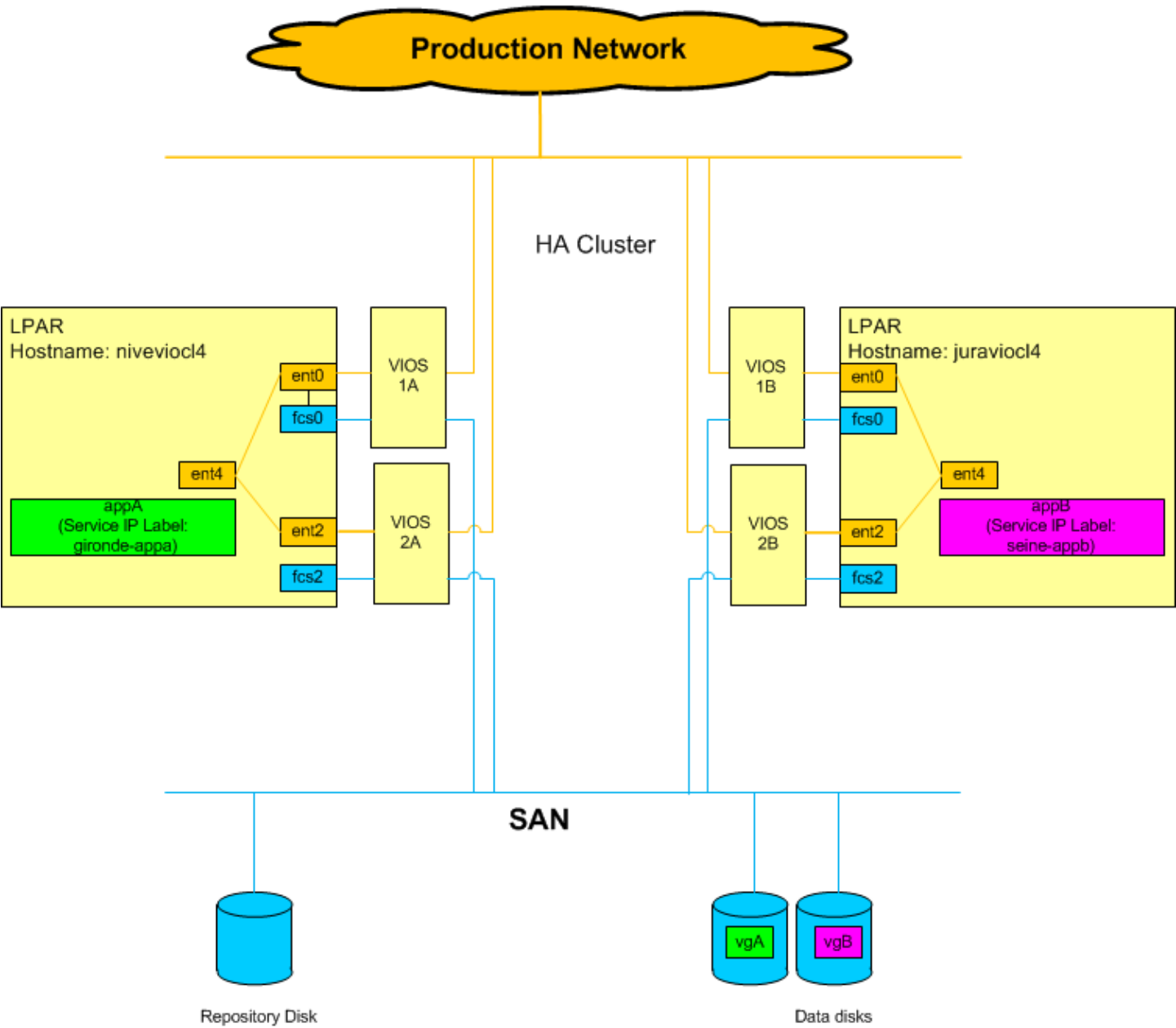

Figure 2: Test Cluster Topology

Table 1 lists the IP addresses used. Note that the hostname and persistent IP address have been configured to be the same:

Cluster: prod1				
Host	niveviocl4		juraviocl4	
Hostname/ IP	10.154.222.249	niveviocl4	10.154.222.6	juraviocl4

Service Address label for appA	10.154.222.173	appa		
Service Address label for appB			10.154.222.174	appb

Table 1: IP Addresses used to Configure Test Cluster

15.1 Configure the Network Interfaces

The following steps should be taken on both nodes to etherchannel the host IP address for resilience

1. Assuming en0 was used for the NIM build, from the console:
 # ifconfig en0 down detach
 # rmdev –dl en0

2. Etherchannel the two interfaces on different adapter cards together adding the gateway as a ping address.

```
# /usr/sbin/mkdev -c adapter -s pseudo -t ibm_ech \
-a adapter_names=ent0 \
-a backup_adapter=ent2 \
-a netaddr={Gateway_IP} \
-a num_retries=3 \
-a retry_time=5

ent4 available
```

3. Add the hostname address back to the node:
 # smitty chinet
```
  Network Interface Name                                en4
  INTERNET ADDRESS (dotted decimal)            [10.154.222.249]
  Network MASK (hexadecimal or dotted decimal) [255.255.248.0]
  Current STATE                                       up                +
```

4. Add the following entries to the /etc/hosts file. Note these files should be identical on both nodes:
```
10.154.222.249  niveviocl4.eulab.nsroot.net niveviocl4
10.154.222.6    juraviocl4.eulab.nsroot.net juraviocl4
10.154.222.173  gironde-appa.eulab.nsroot.net gironde-appa
10.154.222.174  gironde-appb.eulab.nsroot.net gironde-appb
```

15.2 Configure the Cluster Repository Disk

The cluster repository disk must be visible on all cluster nodes and should be at least 1GB in size.

On both nodes:

```
# chdev -l hdiskpower3 -a reserve_policy=no_reserve
Hdiskpower3 changed
```

15.3 Configure the Shared Application Disks

It is necessary to configure the shared SAN disks and ensure they are visible from both cluster nodes:

On node niveviocl4
1. Change the reservation policy on any disks to be shared as well as the cluster repository disk
 # chdev -l hdiskpower4 -a reserve_policy=no_reserve
```
hdiskpower4 changed
```

2. Determine a free major number:
 # lvlstmajor
```
41...
```
So in the example output below major numbers 41 upwards are free. You will need to run this same command on the other node (juraviocl4) to find a similar free number. For this example we shall assume "41" is free on both nodes.

3. Create an enhanced concurrent volume group
```
# mkvg –y vgA –s 32 –n –V 41 –C hdiskpower4
0516-1254 mkvg: Changing the PVID in the ODM.
vgA
mkvg: This concurrent capable volume group must be varied on manually.
```

4. Varyon the created volume group, create a logical volume, add a filesystem and check you can mount it
```
# varyonvg vgA
# mklv -y lvA -t jfs2 vgA 100
lvA
# crfs -v jfs2 -m /fsA -d lvA -a logname=INLINE -a logsize=70
File system created successfully.
3276496 kilobytes total disk space.
New File System size is 6553600
# mount /fsA
```

5. Unmount and varyoff the volume group
```
# umount /fsA
# varyoffvg vgA
```

On node juraviocl4
Next check we can access the disk from the other node:

1. Change the reservation policy on any disks to be accessed
```
# chdev –l hdiskpower4 -a reserve_policy=no_reserve
hdiskpower0 changed
```

2. Check you can import the volume group to the node. In this case the major number being used is "41" (see above). Note you may need to run **cfgmgr** first to pick up the pvid on the disk.

```
# cfgmgr
# importvg –V 41 –y vgA hdiskpower4
```

3. Check you can varyon the volume group and access the shared filesystem
```
# varyonvg vgA
# mount /fsA
# umount /fsA
# varyoffvg vgA
```

Repeat the above procedure to create a volume group called vgB on hdiskpower1 for appB

15.4 Creating and Configuring the Cluster

HACMP 7.1 introduces a CLI interface which can be used to configure a cluster. Hence this is used in the procedures below. Smit menus are still available.

15.4.1   Define the Cluster Nodes, Networks and Heartbeats

1. To ease administration add the following lines to the root prfile on both nodes by editing /root/paths.conf:

```
ADD_PATH /usr/lpp/mmfs/bin /usr/es/sbin/cluster /usr/es/sbin/cluster/utilities /
usr/es/sbin/cluster/etc /usr/es/sbin/cluster/diag
```

2. Disable snmpd as this is not required by HACMP and can lead to security vulnerabilities.

```
# echo "foo" > /etc/snmpd.boots
# stopsrc –s snmpd
```

3. Edit /etc/cluster/rhosts on both nodes to include the two nodes. Note the FQDN should be used:

```
# cat /etc/cluster/rhosts
niveviocl4.eulab.nsroot.net
juraviocl4.eulab.nsroot.net
```

4. If using the LPAR network is provided via the VIO server, the network interfaces are virtual and as such a failure a CEC level network failure may not show up on VIO clients as the internal virtual network is still operational. To ensure that such a failure is detected by HACMP, it is required to add a line to the file "/usr/es/sbin/cluster/netmon.cf" with an entry of the format:

```
!REQD <owner> <target>
```

where <owner> is the originating interface name or IP and <target> is the ping address to be used to test the connectivity. Note this ping address should be on the external network such as the HSRP or gateway address. An example is:

```
!REQD en2 10.154.216.1
```

5. Create the cluster:

```
clmgr add cluster cluslab repository=hdiskpower3
nodes=niveviocl4.eulab.nsroot.net,juraviocl4.eulab.nsroot.net
```

6. Add the application service IP

```
clmgr add service gironde-appa network=net_ether_01
```

7. Define the application appA

```
clmgr add application_controller appA startscript="/usr/local/bin/appa start"
stopscript="/usr/local/bin/appa stop"
```

8. Create the resource group assigning a service IP, volume group and application.

```
clmgr add rg rgA nodes=niveviocl4,juraviocl4 startup=ohn fallback=nfb
service_label=gironde-appa volume_group=vgA application=appA
```

9. Synchronize and verify the cluster

```
clmgr sync cl fix
```

15.4.2   Procedure to disable Firstalias option

PowerHA 7.1 automatically configures the service IP as an alias with firstalias option regardless of the user's setting. However in certain scenarios such as where products use host IP for authentication/validation, the default firstalias option can cause issues. Set the "Disable Firstalias" option in PowerHA with following procedure.

1. smitty sysmirror --> Cluster Applications and Resources --> Resources --> configure Service IP Labels/Addresses --> Configure Service IP Label/Address Distribution Preferences --> Select the network

   Change Distribution Preference to "Disable Firstalias"

                      Configure Service IP Labels/Address Distribution Preference
   Type or select values in entry fields.

Press Enter AFTER making all desired changes.

```
                                    [Entry Fields]
* Network Name                      net_ether_01
* Distribution Preference           Disable Firstalias +
Source IP Label for outgoing packets +
```

2. Verify and sync the cluster

3. Bring down resource group (if already running) and bring up again.

15.4.3   Application Monitoring

It is advisable to set up an application monitor which can, for example, send a notification of an application failure or try to restart an application on a node. The number of retries is a definable parameter. This can either be in the form of a process monitor or a Custom Monitor. As an example, to configure a custom monitor:

```
clmgr add mon appa_mon TYPE=Custom APPLICATION=appA MONITORINTERVAL=60 FAILUREACTION=fallover STABILIZATION=300\
RESTARTINTERVAL=1200 CLEANUPMETHOD=/usr/local/bin/appa_cleanup.sh RESTARTMETHOD=/usr/local/bin/appa_restart.sh \
RESTARTCOUNT=3 MONITORMETHOD=/usr/local/bin/appa_mon.sh
```

Note: Although application monitors can be used to invoke a failover to another node, this should be done with extreme care. In particular if the issue is application related, failing over may just have the effect of reproducing the same failure on the failover node. This not only fails to resolve the issue, but can make the root cause determination more difficult.

16 Configuring PowerHA SystemMirror using Disk Resources Provided via the VIO Servers

16.1 Shared Application Disks Provisioned via NPIV

Disks Provisioned to the LPARs via NPIV

Figure3 below shows how NPIV is used to provide SAN disks (LUNs) to the VIO client.
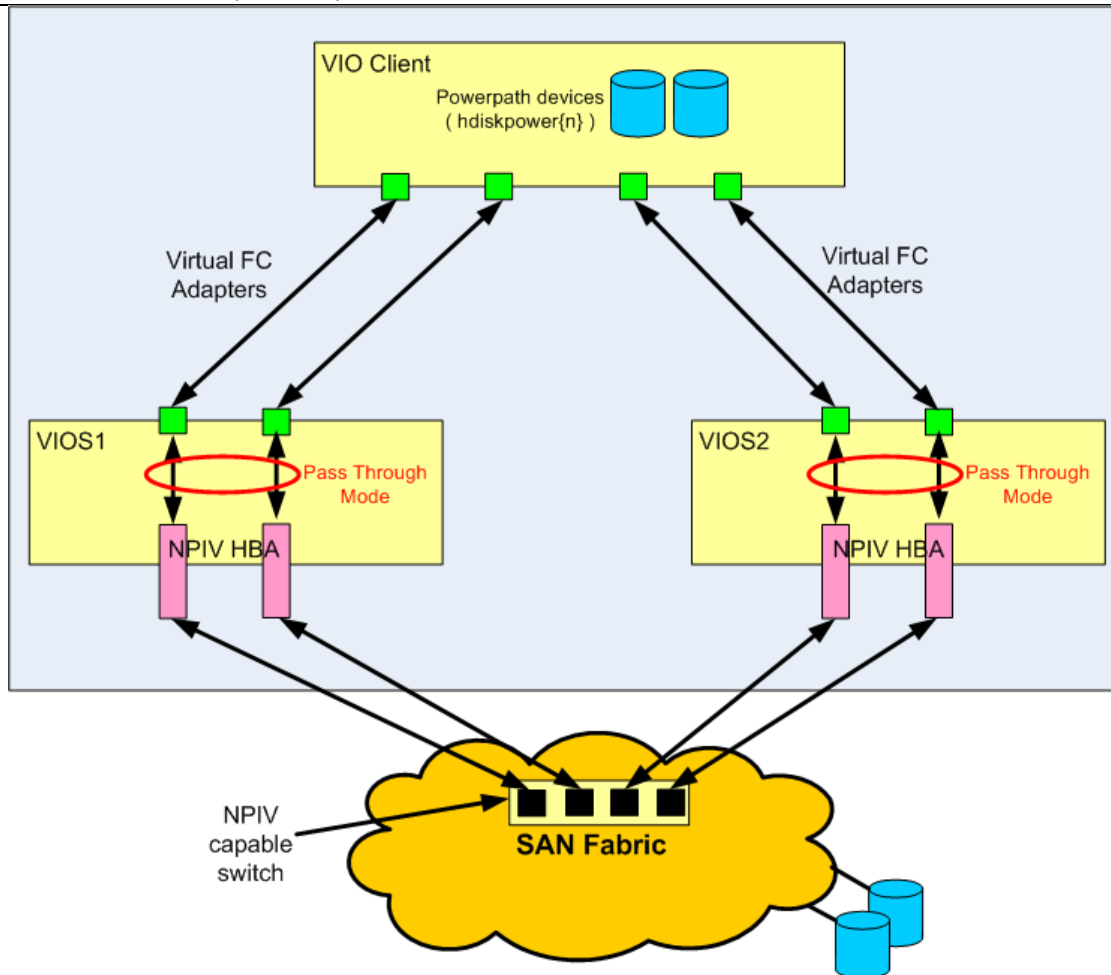
Figure3: NPIV Configuration of HACMP application LUNs

Configuration of the shared application SAN disks is identical to section 15.2. Please note:

- For details on how to configure NPIV see the VIO server Release Notes and Best Practices Guides (https://catecollaboration.citigroup.net/domains/platstor/osunix/stdsrelateddocs/Forms/AllItems.aspx)and SAN stack release notes (https://catecollaboration.citigroup.net/domains/platstor/gdse/products/techstdshmpgs/SANStacks.aspx)

- The O.S. disks are either local disks or provided by the VIO servers as vscsi LUNs.

- The application data disks are provided through the VIO servers via virtual fibre adapters. Four of these should be visible (two from each VIO server) on the VIO client (HACMP node). For example:

```
# lsdev -Cc adapter | grep fcs
fcs0   Available 55-T1 Virtual Fibre Channel Client Adapter
fcs1   Available 56-T1 Virtual Fibre Channel Client Adapter
fcs2   Available 57-T1 Virtual Fibre Channel Client Adapter
fcs3   Available 58-T1 Virtual Fibre Channel Client Adapter
```

- The application disks need to be visible to all HACMP cluster nodes.

- The application disks themselves are still managed by Powerpath in an identical fashion to direct attached LUNs.

16.1.1   Shared Application Disks Provisioned to the LPAR as VSCSI Devices

Note – no new PowerHA clusters should use vscsi disks. Only NPIV or direct attached will be supported going forward. This is due to the added complexity of using vscsi data disks, and in particular the disk reservation settings.

Figure 4 below shows how SAN disks are presented to the VIO client LPARs as vscsi devices.
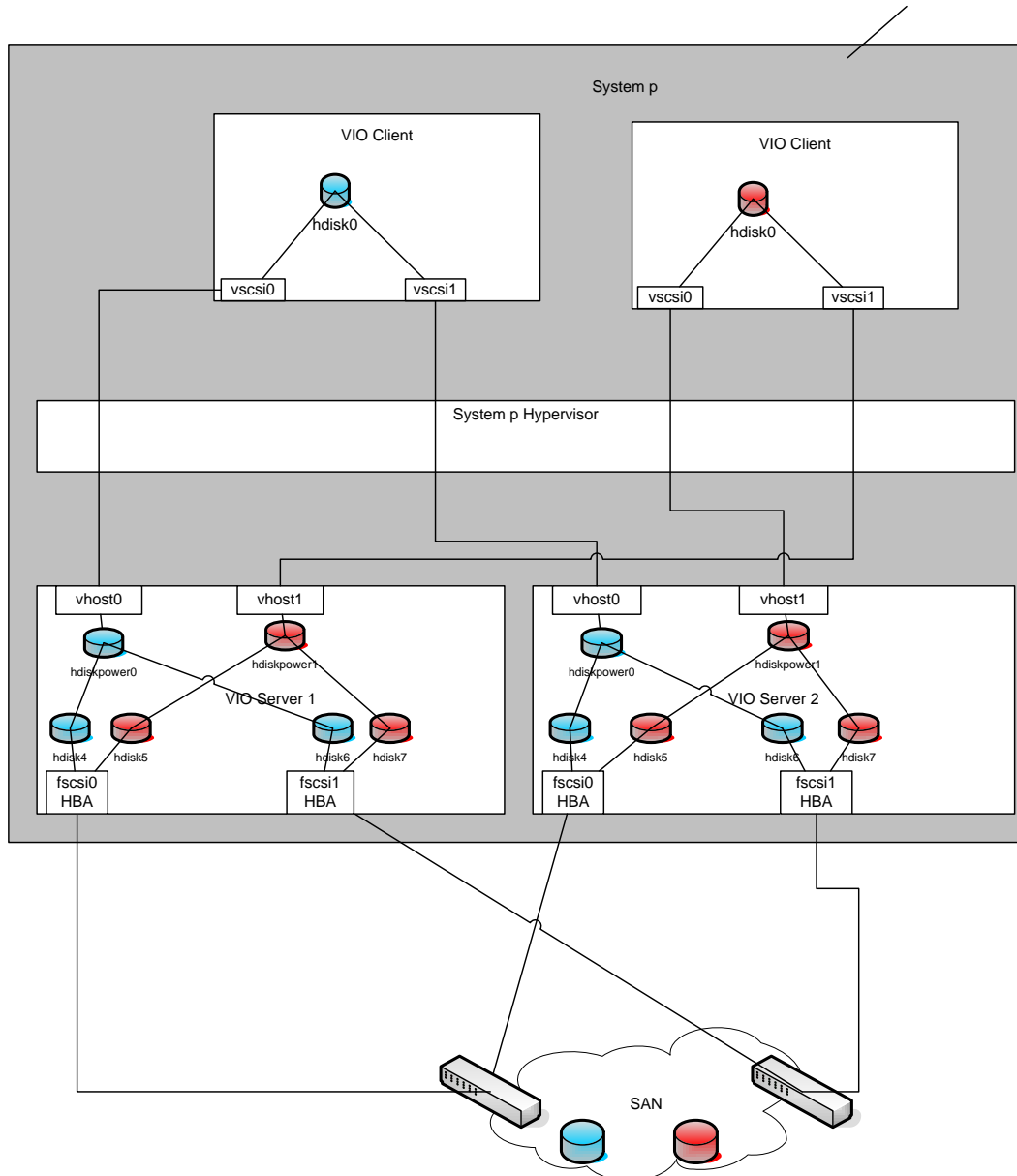
Figure 4: Disks provisioned to LPARs as VSCSI Devices

Configuration of the shared application SAN disks is identical to section 15.3. Please note:

- For details on how to configure vscsi disks see the VIO Server Release Notes and Best Practices Guides (https://catecollaboration.citigroup.net/domains/platstor/osunix/wpages/ProductDetail.aspx?Name=AIX&Publish=1)

- The PowerHA data disks are provided through the VIO servers as vscsi devices and will appear as hdiskX. They are managed by MPIO and two paths to each disk should be visible.

- The PowerHA data disks need to be visible to all PowerHA cluster nodes.

17 Live Partition Mobility

Live Partition Mobility for HACMP clusters is supported. CATE Engineering will only support this for cluster data disks provisioned via NPIV.

When performing validation for an LPM operation, the following error will show which can be safely ignored:

```
"HSCLA27A The management console was not able to locate the source virtual I/O server
partition's hosting server adapter for vlan 3358 in its internal database."
```

18 Cluster Commands and Logfiles

Cluster related commands:
- lssrc –ls clstrmgrES   - To show the PowerHA cluster status
- clmgr query cluster   - To show CAA cluster status
- clRGinfo –p   - To show status of the resource groups
- lscluster –i   - To view cluster interface information
- lscluster –m   - To view the cluster node information
- cltopinfo / clmgr view report basic   - To view the cluster topology
- smitty clstart / clmgr start clutser   - To start the cluster
- smitty clstop / clmgr stop cluster   - To stop the cluster
- clmgr   - CLI for cluster configuration.
- smitty sysmirror   - equivalent to "smitty hacmp"

Cluster log files
- /var/hacmp/log/hacmp.out   - To see cluster logs during startup/failover
- /var/hacmp/log/clutils.log   - To see the logs generated by cluster util commands
- /var/hacmp/adm/cluster.log   - Output from cluster script actions
- /var/adm/ras/syslog.caa   - CAA cluster logs
- errpt   - AIX error log

19 Testing a Cluster

Prior to go-live it is essential to test any clustered applications will failover. The tests below were conducted on the test cluster created in section 15. It is strongly recommended that these same tests be conducted on any cluster at installation time. In addition it is also essential that applications are periodically manually failed over to the standby node(s).

Test Environment
Servers: 2 x Power7 p720 servers
Network: 100Mb connections throughout
SAN: EMC VMAX-1SE, HDS Tagmaster
O.S.: AIX 6.1 patchset 13, AIX 7.1 patchset6
SAN stack: AIX SAN stack version 7.1 was used.

Initial AIX machine State
HACMP down on both nodes to start

Steps	HACMP status		Pass/Fail
	niveviocl4	juraviocl4	
AIX installed and configured HACMP Installed and configured Machines rebooted	Down	Down	✓
Check to see that persistent IPs available	Down	Down	✓

Start HACMP – check resource groups brought on-line on home nodes only
HACMP down to start

Steps	HACMP status		Pass/Fail
	Niveviocl4	Juraviocl4	
Start HACMP on niveviocl4	appA on-line only	Down	✓
Start HACMP on juraviocl4	appA on-line only	appB on-line only	✓

Graceful shutdowns
HACMP up and running as of end of previous test

Steps	HACMP status		Pass/Fail
	Niveviocl4	juraviocl4	
Stop HACMP on niveviocl4	Down	appB on-line only	✓
Stop HACMP on juraviocl4	Down	Down	✓

Graceful shutdown and move resource groups
HACMP down as of end of previous test

Steps	HACMP status		Pass/Fail
	niveviocl4	juraviocl4	
Start HACMP on juraviocl4	Down	appB on-line only	✓
Start HACMP on Niveviocl4	appA on-line only	appB on-line only	✓
Stop HACMP on Juraviocl4 and move resource group	appA on-line appB on-line	Down	✓
Start HACMP on Juraviocl4	appA on-line appB on-line	Up – no resource groups	✓

Test resource group move
HACMP up with both RGs on Niveviocl4 as of end of last test

Steps	HACMP status		Pass/Fail
	niveviocl4	juraviocl4	
Confirm RG status	appA on-line appB on-line	Up – no resource groups	✓
Perform RG move of appB back to juraviocl4 node	appA on-line only	appB on-line only	✓

Graceful shutdown and move resource group
HACMP up as of end of previous test

Steps	HACMP status		Pass/Fail
	niveviocl4	juraviocl4	
Stop HACMP on niveviocl4 and move resource group	Down	appB on-line appA on-line	✓
Start HACMP on Niveviocl4	Up – no resource groups	appB on-line appA on-line	✓

Test resource group move
HACMP up with both RGs on juraviocl4 as of end of last test

Steps	HACMP status		Pass/Fail
	niveviocl4	juraviocl4	
Confirm RG status	Up – no resource groups	appB on-line appA on-line	✓
Perform RG move of appB back to niveviocl4 node	appA on-line only	appB on-line only	✓

Complete node failure testing
HACMP up and both RGs on their home nodes as of end of last test

Steps	HACMP status		Pass/Fail
	niveviocl4	juraviocl4	
Confirm RG status	appA on-line only	appB on-line only	✓
Shutdown niveviocl4 LPAR using HMC and immediate halt or pull out power.	Down	appB on-line appA on-line	✓
Restart niveviocl4 and HACMP	Up – no resource groups	appB on-line appA on-line	✓
Perform RG move to move appA back to home node	appA on-line only	appB on-line only	✓

Steps	HACMP status		Pass/Fail
	niveviocl4	juraviocl4	
Confirm RG status	appA on-line only	appB on-line only	✓
Shutdown juraviocl4 LPAR using HMC and immediate halt or pull out power.	appA on-line appB on-line	Down	✓
Restart juraviocl4 and HACMP	appA on-line appB on-line	Up – no resource groups	✓
Perform RG move to move appB back to home node	appA on-line	appB on-line only	✓

		only		

VIO Server Failure Tests

Steps	HACMP status			Pass/Fail
	Niveviocl4		Juraviocl4	
Shutdown VIO1 on nive	appA on-line network and disk resources provided by VIO1 failed over to VIO2		appB on-line	✓
Shutdown VIO2 on nive (all network/disk resource to niveviocl4 failed)			appB on-line appA on-line only	✓

VIO Server Network Failure Tests

HACMP restored back to the "Initial AIX machine State" as listed above:

Steps	HACMP status		Pass/Fail
	niveviocl4	juraviocl4	
	appA on-line	appB on-line	✓

Steps	HACMP status			Pass/Fail
	Niveviocl4		Juraviocl4	
Remove single network cable from VIO1 on niveviocl4	appA on-line VIO1 etherchannel failover to alternate adapter		appB on-line	✓
Remove second network cable from VIO1 (no network connectivity to VIO1	appA on-line only niveviocl4 switches to using VIO2 for network		appB on-line	✓
Remove network cables from VIOS2	Cluster detects complete network failure at CEC level and fails over appA		appA on-line appB on-line	✓

VIO Server SAN Connectivity Failure Tests for NPIV disks

HACMP restored back to the "Initial AIX machine State" as listed above:

Steps	HACMP status		Pass/Fail
	niveviocl4	juraviocl4	
	appA on-line	appB on-line	✓

Steps	HACMP status			Pass/Fail
	Niveviocl4		Juraviocl4	
Remove fibre cable from NPIV slot on HBA on VIO1 (niveviocl4)	appA on-line link error on single adapter		appB on-line	✓
Remove final fibre cable from NPIV slot on HBA on VIO2 (total NPIV connectivity lost to niveviocl4)			appA on-line appB on-line	✓
On the SAN remove access to a disk being used by appB	appB on-line		appA on-line	✓

Document Classification: Internal

20 References

- **Product Documentation**   http://www.redbooks.ibm.com/redbooks.nsf/RedpieceAbstracts/sg248167.html?Open
- **VIO Standards**   VIOS standards
- **HA Support Matrix**   HA_Suppport_Matrix_v2.0

Server Platform Engineering   RFP-AIX-POWERHA-7.1-d7.31.doc, Version 7.4   Page 38 of 38
Document Classification: Internal