Technical Paper

How to Implement High Availability for the SAS® Metadata Server Using High Availability Cluster Multi-Processing (HACMP)

Technical White Paper by SAS and IBM

Table of Contents

Abstract **1**
Introduction **1**
System configuration **2**
Scripts 3
Testing failover of the SAS Metadata Server running on an IBM HACMP cluster **3**
Conclusion **4**
References **5**
Appendix 1: System Configuration Details **6**
Appendix 2: HACMP control and monitor scripts **7**
Appendix 3: HACMP cluster SNMP configuration listing **9**
Appendix 4: Modifications to MetadataServer.sh (diff output) **10**

Abstract

SAS and IBM worked to configure and test high availability using the SAS® Metadata Server running under IBM® AIX 5L V5.3. This paper details the steps used to configure and test a high-availability configuration of the IBM High Availability Cluster Multi-Processing (HACMP) product and the SAS Metadata Server.

This implementation included the following:

- The IBM HACMP product was installed to establish a base high-availability environment.
- The SAS Metadata Server was installed into the high-availability environment.
- Three custom shell scripts were developed to interconnect SAS Metadata Server operationally with the high-availability environment.
- The environment was validated with a series of tests.

This material is intended for system administrators and engineers responsible for configuring, managing, and troubleshooting high availability issues involving HACMP clusters and the SAS Metadata Server.

This document assumes that the reader has the following:

- Experience with SAS® Foundation servers and the SAS Metadata Server installation and operation
- Knowledge of IBM eServer pSeries system components including disk devices, cabling, and network adapters
- Experience with the IBM AIX 5L operating system, including the Logical Volume Manager subsystem
- The System Management Interface Tool (SMIT)
- Communications, including the TCP/IP subsystem
- Reviewed the *IBM High Availability Cluster Multi-Processing for AIX Administration Guide*—available at www-03.ibm.com/systems/p/library/hacmp_docs.html.

Introduction

The SAS Metadata Server provides an open, central repository for all metadata that is created and required by an organization to support its enterprise intelligence strategy. System managers leverage high-availability solutions to monitor and protect the SAS Metadata Server ensuring that critical information is obtainable.

IBM defines high availability in *Implementing High Availability Cluster Multi-Processing (HACMP) Cookbook* as:

> *… one of the components that contributes to providing continuous service for the application clients, by masking or eliminating both planned and unplanned systems and application downtime. … A high availability solution will ensure that the failure of any component of the solution, either hardware, software, or system management, will not cause the application and its data to become permanently unavailable to the end use. (Lascu et al. 2005).*

System configuration

The SAS Metadata Server, on the SAS 9.1.3 platform, was deployed on an IBM eServer p570 running IBM AIX 5L V5.3 TL05. The IBM eServer p570 was configured as two logical partitions representing two system nodes: *hacmp0* and *hacmp2*. Figure 1 depicts the testing system configuration.

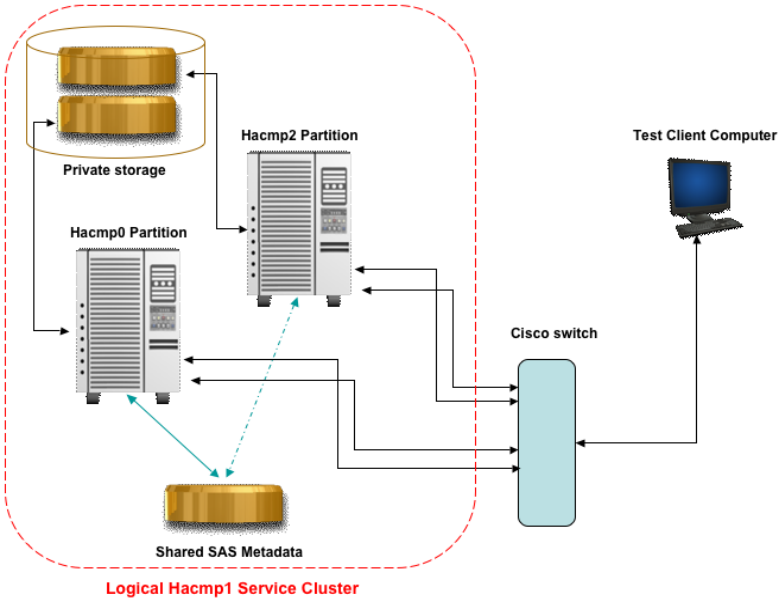

Figure 1. Testing system configuration

Each partition was configured with an independent root volume group that contained the operating system, HACMP software, private data, and shell scripts. Each root volume group was stored on a virtual SCSI disk that was private to its partition.

Each partition was connected to a Cisco network switch over two Ethernet adapters. The adapters were configured as two Virtual Local Area Networks (VLANs), designed to simulate a two-way redundant network.

The HACMP software was configured to create a high-availability cluster, *meta_cluster,* which contained the two nodes *hacmp0* and *hacmp2.*

The SAS Foundation software, including the SAS Metadata Server, was installed on a file system in a second, *non-root*, volume group. Each *meta_cluster* node was allowed concurrent access to this file system. This allowed access to the

SAS® System and the SAS Metadata Server data repository by the cluster node that was actively servicing SAS metadata requests. The cluster was run in an active/passive configuration with the SAS Metadata Server running on a single node at a time. The spare node acted as a failover resource for the SAS metadata service. Access to this shared file system was only allowed for the active node as coordinated by HACMP.

The HACMP resource group *meta_group* was configured containing the two nodes *hacmp0* and *hacmp2*, with a service IP label of *hacmp1*. External client applications used this service label to access the SAS metadata service running on the cluster *meta_cluster.*

A stand-alone SAS Metadata Server testing application was run on a separate Microsoft Windows XP PC connected to the Cisco switch.

See Appendix 1 for a detailed description of the system configuration.

Scripts

Three shell scripts, *hacmp_start, hacmp_stop*, and *sasms_hacmp_mon.sh*, were written to implement failover of the SAS metadata service through HACMP. Please refer to Appendices 2, 3, and 4 for samples of script code.

Copies of the scripts resided in local storage on each node of the *meta_group* resource group. These scripts were used by the HACMP system to programmatically start, stop, and monitor the SAS Metadata Server on the local node. To control execution of the SAS Metadata Server application, the start and stop scripts invoked the *MetadataServer.sh* control script that is supplied by SAS and installed as part of the standard SAS Metadata Server deployment.

Note that one modification was made to the standard *MetadataServer.sh script*: An additional option, *restart_no_error*, was added to the script.

The HACMP system can monitor the health of applications either with process application monitoring, which uses IBM Reliable Scalable Cluster Technology (RSCT) services, or with custom application monitoring, which uses a custom monitor method to check the application at user-specified intervals. For this implementation, the shell script *sasms_hacmp_mon.sh* was used as a custom application monitor to periodically check the SAS Metadata Server process. This script uses IBM AIX commands to check for the existence of the SAS Metadata Server process and its auxiliary *elssrv* process on the local node. The script returns a nonzero status if the check fails.

Note that an alternative approach is to issue service requests to the SAS Metadata Server directly to verify its continued functioning.

Testing failover of the SAS Metadata Server running on an IBM HACMP cluster

A system can be considered to have high availability when it can maintain an acceptable level of service, with or without a short outage, after a hardware or software failure, or human error that would normally make the system unavailable. Restarting or failover of the server software should occur rapidly, usually without operator intervention.

The SAS Metadata Server must appear to exist at a constant IP address, with access to related file storage at the same logical file paths even after a restart or failover to another system node. The SAS Metadata Server and related storage should appear to run on a fixed virtual machine, no matter which physical node the virtual machine is running on.

SAS has developed a series of steps to test various failure modes of a high-availability cluster. These tests are intended to verify that the high-availability system can efficiently and correctly respond to events that would necessitate restarting the SAS Metadata Server on the current or an alternative node of the cluster.

As part of this series of tests, an internal test application was used to rapidly and continuously query the SAS Metadata Server running on the cluster. The test application noted any interruptions in service, and then repeated attempts to reconnect to the SAS Metadata Server at the original service port and IP address. The test application noted if and when service was restored and resumed querying the server.

When testing the HACMP configuration, the following was verified:

1. An operator was able to use standard system tools and commands to programmatically start and stop the SAS Metadata Server on initial and alternative failover nodes in the cluster. The SAS Metadata Server ran on the initial or alternative node and was able to respond to queries from an external test application. The test application ran on an independent machine attached to the network.

2. While the SAS Metadata Server ran on a node in the cluster and responded to queries, the operator force-quit or terminated the active server process and its auxiliary *elssrv* process. The HACMP system noted that the service had failed, and then automatically restarted the SAS Metadata Server and began successfully responding to queries.

3. While the SAS Metadata Server ran on a node in the cluster and responded to queries, the operator physically disconnected the Ethernet network cable that was actively carrying the network traffic. The HACMP system noted the loss of connectivity, resulting in the programmatic shutdown and failover of the SAS Metadata Server to an alternative node in the cluster. The new instance of the SAS Metadata Server began responding to queries.

4. While the SAS Metadata Server ran on a node in the cluster and responded to queries, the operator was able to cause or simulate a catastrophic power failure on the active node. The HACMP system noted the loss of service, resulting in the programmatic failover of the SAS Metadata Server to an alternative node in the cluster. The new instance of the SAS Metadata Server began to respond to queries.

Conclusion

SAS Metadata Server on the SAS 9.1.3 platform successfully integrates with the IBM High Availability Cluster Multi-Processing product for IBM AIX 5L. In the event of service failure, HACMP successfully restarts the SAS Metadata Server on the current or an alternative cluster node. This enables the SAS Metadata Server to rapidly resume responding to service queries, improving the availability of access to data in the SAS metadata repository.

References

Lascu, Octavian, Shawn Bodily, et al. 2005. *Implementing High Availability Cluster Multi-Processing (HACMP) Cookbook*. Austin, TX: IBM Corporation. Page 4, section 1.1.1. Available www.redbooks.ibm.com/redbooks/pdfs/sg246769.pdf.

Appendix 1: System Configuration Details

The SAS 9.1.3 Metadata Server was deployed on an IBM eServer p570 running IBM AIX 5L V5.3 TL05. The system was configured with two logical partitions (LPARs) representing two system nodes: hacmp0 and hacmp2. Each partition had 4 GB of main memory, two Ethernet adapters, and two SSA adapters.

The four Ethernet adapters were connected to a Cisco 2950 switch. Two VLANs were defined to simulate a 2-way redundant network.

The four SSA adapters were connected to two 7133 SSA disk drawers using a fully redundant connection scheme. Each 7133 SSA disk drawer contained 16 disks for a total of 32 physical disks. Two RAID 5 logical disks were configured from the physical disks of each disk drawer. The two logical disks were used to construct an IBM AIX 2-way mirrored logical volume for an additional level of redundancy. Each RAID 5 array could tolerate a single disk failure without loss of function. Hot standby disks were also configured for automatic rebuilding and replacement in the array of the failed disk to allow the array to self heal to tolerate a future disk failure. The IBM AIX 2-way mirrored logical volume allowed the complete failure of one disk drawer to be tolerated.

The hacmp0 and hacmp2 partitions were each configured with a root volume group. The root volume group, rootvg, held the operating system, private data, and shell scripts. For each partition, this volume group contained two 33 GB virtual SCSI disks in a 2-way mirrored configuration. Each virtual SCSI disk was private to its partition. The only access to these disks was through the owning partition.

HACMP for IBM AIX 5L, V5.3 (5765-F62) was installed on partitions hacmp0 and hacmp2. HACMP was then configured to create an HA cluster called meta_cluster that contained two nodes named hacmp0 and hacmp2.

The SAS Foundation software, including the SAS Metadata Server, was installed on a file system on a 2-way mirrored logical volume in a second (non-root) volume group. Each cluster node was allowed concurrent access to this file system. This allowed access to the SAS System and the SAS Metadata Server data repository by the cluster node that was actively servicing SAS metadata requests. The cluster was run in an active/passive configuration with the SAS Metadata Server running on a single node at a time. The spare nodes acted as a failover resource for the SAS metadata service. Access to this shared file was only allowed for the active node as coordinated by HACMP.

The HACMP resource group meta_group was configured containing the two nodes hacmp0 and hacmp2, with a service IP label of hacmp1. External client applications used this service label to access the SAS metadata service running on the cluster meta_cluster.

Appendix 2: HACMP control and monitor scripts

```
hacmp_start
**********************************
#!/bin/sh
#
#hacmp_start – used by HACMP to start the SAS Metadata Server
#
# The referenced script is installed as part of the SAS Metadata Server deployment.
# The file path will vary with your installation.
#
/home/sas/SAS/MetadataServer/Lev1/SASMain/MetadataServer/MetadataServer.sh \
restart_no_error

**********************************
hacmp_stop
**********************************
#!/bin/sh
#
#hacmp_stop – used by HACMP to shutdown the SAS Metadata Server
#
# The referenced script is installed as part of the SAS Metadata Server deployment.
# The file path will vary with your installation.
#
/home/sas/SAS/MetadataServer/Lev1/SASMain/MetadataServer/MetadataServer.sh \
stop_no_error > /dev/null
slibclean
**********************************

sasms_hacmp_mon.sh
**********************************
#!/bin/sh
#
# Monitor the existence of the SAS Metadata Server and its aux processes.
#
# The MSDIR file path will vary with your installation.

#
MSDIR="/home/sas/SAS/MetadataServer/Lev1/SASMain/MetadataServer"

if [ -f $MSDIR/server.pid ];
then
  mspid=`cat $MSDIR/server.pid`
#  echo "MSPID = $mspid"
else
#  echo "Can't find pid file"
  exit 1   # can't find the pid file
fi

#
# see if Metadata process is alive
#
kill -0 "$mspid"

if [ $? -ne 0 ];
then
#  echo "No SASMS process = $mspid"
  exit 2   # process is gone
fi
```

```
#
# See if elssrv associated with THIS Metadata is alive.
#
ps -ef | grep elssrv | grep $mspid >/dev/null 2>&1

if [ $? -ne 0 ];
then
#  echo "No elssrv associated with this SASMS"
  exit 3   # no elssrv parented by this Metadata
else
  exit 0   # they're both up
fi
**********************************
```

Appendix 3: HACMP cluster SNMP configuration listing

```
_____________________________________________________________
Cluster Name: meta_cluster
Cluster State: UP
Cluster Substate: STABLE
_____________________________________________________________


Node Name: hacmp0               State: UP
  Network Name: net_diskhb_01      State: UP
    Address:                  Label: hacmp0_hdisk5_01   State: UP

  Network Name: net_ether_01       State: UP
    Address: xx.xx.127.111    Label: hacmp1             State: UP
    Address: xx.xx.150.110    Label: hacmp0             State: UP
    Address: xx.xx.9.159      Label: metahacmp2         State: UP

  Network Name: net_tmssa_01       State: UP
    Address:                  Label: hacmp0_tmssa2_01   State: UP


Node Name: hacmp2               State: UP
  Network Name: net_diskhb_01      State: UP
    Address:                  Label: hacmp2_hdisk5_01   State: UP

  Network Name: net_ether_01       State: UP
    Address: xx.xx.150.112    Label: hacmp2             State: UP
    Address: xx.xx.9.22       Label: metahacmp1         State: UP

  Network Name: net_tmssa_01       State: UP
    Address:                  Label: hacmp2_tmssa1_01   State: UP



Cluster Name: meta_cluster

Resource Group Name: meta_group
Startup Policy: Online On Home Node Only
Fallover Policy: Fallover To Next Priority Node In The List
Fallback Policy: Never Fallback
Site Policy: ignore
Priority Override Information:
    Primary Instance POL:
Node                          Group State
----------------------------- ---------------
hacmp0                        ONLINE
hacmp2                        OFFLINE
```

Appendix 4: Modifications to MetadataServer.sh (diff output)

This code fragment is inserted after the case entry for STOP_NO_ERROR and before the wildcard *) entry. The line numbers indicated in the following diff output might change relative to future releases of this script.

```
89,95d88    ???
<    restart_no_error)
<          $0 stop_no_error
<          if [ $? -eq 0 ]; then
<             sleep 5
<             $0 start
<          fi
<          ;;
```

.