Configuration Guide

SAP HANA Disaster Tolerance with System Replication on Cisco Unified Computing System

Contents

Overview

This document is not a step-by-step installation guide but instead provides guidance for setting up SAP HANA system replication. This document refers to specific vendor information, such as the SUSE SAP HANA replication white paper and other SAP HANA documentation listed at the end of this document.

This document is based on SUSE Linux Enterprise Server (SLES). Red Hat Enterprise Linux (RHEL) will be integrated later.

At this point, automatic failover is available only with scale-up (single-system) architecture. Cisco is working with SUSE to enable automatic failover for scale-out (multiple-system) designs.

SAP HANA is SAP's implementation of in-memory database technology.

The SAP HANA database takes advantage of the low cost of main memory (RAM), the data processing capabilities of multicore processors, and the fast data access of solid-state disk (SSD) drives relative to traditional hard-disk drives (HDDs) to deliver better performance for analytical and transactional applications. Its multiple-engine query processing environment allows it to support relational data (with both row- and column-oriented physical representations in a hybrid engine) as well as graph and text processing for management of semistructured and unstructured data within the same system. The SAP HANA database is 100 percent compliant with the atomicity, consistency, isolation, and durability (ACID) model.

For more information about SAP HANA, see the SAP help portal, at http://help.sap.com/hana/.

Audience

The intended audience for this document includes sales engineers, field consultants, professional services staff, IT managers, partner engineering staff, and customers deploying SAP HANA with NetApp clustered Data ONTAP.

Architecture

The solution presented in this document is based on the Cisco Unified Computing System™ (Cisco UCS®) and FlexPod, but it could also be configured with EMC VNX systems. The solution provides a general approach to a single system (scale-up design) with system replication and automatic failover.

The scenarios for this solution can use the following hardware:

- Cisco UCS C460 M4 Rack Server
- Cisco appliance for SAP HANA Tailored Datacenter Integration (TDI)
- Cisco appliance for NetApp
- Cisco appliance for EMC

Possible failover scenarios are:

- Standalone Cisco UCS C460 M4 > Standalone Cisco UCS C460 M4
- Standalone Cisco UCS C460 M4 > Cisco appliance for SAP HANA TDI
- Cisco appliance for NetApp > Cisco appliance for NetApp
- Cisco appliance for EMC > Cisco appliance for EMC
- Cisco appliance for NetApp or EMC > Cisco appliance for SAP HANA TDI
- Cisco appliance for SAP HANA TDI > Cisco appliance for SAP HANA TDI

This first version of this solution is based on SLES for SAP Applications with one SAP HANA scale-up system. SUSE has not yet released a scale-out version.

A Red Hat version also is not yet available.

Network Recommendations

In general, SAP does not give a network bandwidth recommendation for the replication network.

The bandwidth required between the two sites depends by the change rates of the database. Therefore, you should start with a 10-Gbps Layer 2 connection for one scale-up system. If you configure a scale-out system, the bandwidth can be increased to the number of nodes x 10 Gbps.

If the change rate is not high, less bandwidth may be required. In this case, even a switched (Layer 3) connection is possible.

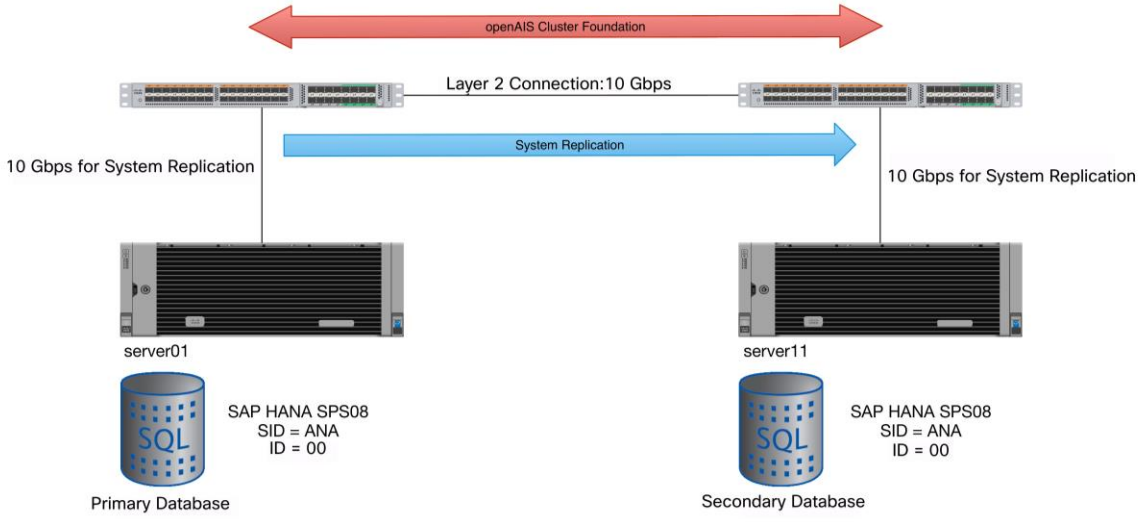
Single-Node Disaster-Tolerance Configuration

The single-node disaster-tolerance configuration is built on SLES for SAP Applications 11 SP3 with installed SUSE Linux Enterprise High Availability Extension (SLE HAE) packages.

SUSE provides a step-by-step description of how to set up the SLE HAE cluster. Register at https://www.suse.com/promo/saphana-replication.html to get the latest copy of it.

Figure 1 shows the design of the current version of the solution.

Figure 1. SAP HANA System Replication in the Cluster



Limitations of the First Resource Agent Release

For the first version of the SAP HANA system replication (SAPHanaSR) resource agent software package, support is limited to the following scenarios and parameters:

- Only two-node clusters are supported.
- Only scale-up (single device–to–single device) system replication is supported.
- Both nodes are in the same network segment (Layer 2).
- Technical users and groups such as sidadm are defined locally in the Linux system.
- Name resolution for the cluster nodes and the virtual IP address is performed locally on all cluster nodes.
- Time synchronization between the cluster nodes uses Network Time Protocol (NTP).
- There is no other SAP HANA system (such as a quality assurance [QA] system) on the replicating node that needs to be stopped during takeover.
- Only one system replication is performed for the SAP HANA database.
- Both SAP HANA instances have the same system identifier (SID) and instance number.
- If the cluster nodes are installed in different data centers or data center areas, the environment must meet the requirements of the SLE HAE cluster product. In particular, the environment must meet the requirements for the network latency between the nodes and the recommended maximum distance. Please review the product documentation for SLE HAE to learn more about these recommendations.
- As the starting configuration for a proof-of-concept (PoC) project, switch off automatic registration of a failed primary device. The configuration **AUTOMATED_REGISTER="false"** is the default as of Version 0.139.
- The SAP Host Agent must be installed on both nodes.

If you need to implement a different scenario, you should define a PoC with SUSE. This PoC will be the focus of testing.

The resource agent supports SAP HANA for system replication beginning with SAP HANA Version 1.0 SPS 7 Patch Level 70.

Source: SUSE AG

Prepare for Installation

Both nodes for system replication must be installed identically.

First update the nodes with the latest available kernel and OS patches for SLES for SAP 11 SP3.

```
server01:~ # uname -a
 Linux server01 3.0.101-0.46-default #1 SMP Wed Dec 17 11:04:10 UTC 2014
(8356111) x86_64 x86_64 x86_64 GNU/Linux


server11:~ # uname -a
 Linux server11 3.0.101-0.46-default #1 SMP Wed Dec 17 11:04:10 UTC 2014
(8356111) x86_64 x86_64 x86_64 GNU/Linux
```

Install SAP HANA

SAP HANA must be installed on both systems with the exact same SID and instance number.

server01 (Primary Database)

```
server11:/ #  su - anaadm
server01:/usr/sap/ANA/HDB00> sapcontrol -nr 00 -function GetProcessList
 name, description, dispstatus, textstatus, starttime, elapsedtime, pid
 hdbdaemon, HDB Daemon, GREEN, Running, 2015 01 14 18:40:39
 hdbnameserver, HDB Nameserver, GREEN, Running, 2015 01 14 18:40:45
 hdbpreprocessor, HDB Preprocessor, GREEN, Running, 2015 01 14 18:40:57
 hdbindexserver, HDB Indexserver, GREEN, Running, 2015 01 14 18:41:08
 hdbstatisticsserver, HDB Statisticsserver, GREEN, Running, 2015 01 14 18:41:08
 hdbxsengine, HDB XSEngine, GREEN, Running, 2015 01 14 18:41:08
 sapwebdisp_hdb, SAP WebDispatcher, GREEN, Running, 2015 01 14 18:42:31
 hdbcompileserver, HDB Compileserver, GREEN, Running, 2015 01 14 18:40:57


 server01:/usr/sap/ANA/HDB00> HDB info
 USER       PID  PPID %CPU    VSZ   RSS COMMAND
 anaadm   13025 13024  0.0  13884  2788 -sh
 anaadm   13119 13025  0.3  12900  1720  \_ /bin/sh /usr/sap/ANA/HDB00/HDB info
 anaadm   13142 13119  3.2   4944   876      \_ ps fx -U anaadm -o
user,pid,ppid,pcpu,vsz,rss,args
 anaadm   11858     1  0.0  22024  1508 sapstart
pf=/hana/shared/ANA/profile/ANA_HDB00_server01
 anaadm   11870 11858  0.0 825768 301404  \_
/usr/sap/ANA/HDB00/server01/trace/hdb.sapANA_HDB00 -d -nw -f /usr/s
 anaadm   11889 11870  1.1 23685688 2208920      \_ hdbnameserver
 anaadm   12068 11870 33.1 16865944 6288988      \_ hdbpreprocessor
 anaadm   12071 11870  0.0 12361768 295608      \_ hdbcompileserver
 anaadm   12101 11870  8.8 20706496 7320808      \_ hdbindexserver
 anaadm   12104 11870  2.7 15484116 2771180      \_ hdbstatisticsserver
 anaadm   12107 11870  2.1 16653840 2314620      \_ hdbxsengine
 anaadm   12394 11870  0.0 271080 72324      \_ sapwebdisp_hdb
pf=/usr/sap/ANA/HDB00/server01/wdisp/sapwebdisp.p
 anaadm   11770     1  0.0 152580 74932 /usr/sap/ANA/HDB00/exe/sapstartsrv
pf=/hana/shared/ANA/profile/ANA_HDB00
 server01:/usr/sap/ANA/HDB00>
```

server11 (Secondary Database)

```
server11:/ #  su - anaadm
 server11:/usr/sap/ANA/HDB00> sapcontrol -nr 00 -function GetProcessList
 name, description, dispstatus, textstatus, starttime, elapsedtime, pid
 hdbdaemon, HDB Daemon, GREEN, Running, 2015 01 15 03:11:34
 hdbnameserver, HDB Nameserver, GREEN, Running, 2015 01 15 03:11:41
 hdbpreprocessor, HDB Preprocessor, GREEN, Running, 2015 01 15 03:11:55
 hdbindexserver, HDB Indexserver, GREEN, Running, 2015 01 15 03:12:07
 hdbstatisticsserver, HDB Statisticsserver, GREEN, Running, 2015 01 15 03:12:07
 hdbxsengine, HDB XSEngine, GREEN, Running, 2015 01 15 03:12:07
 sapwebdisp_hdb, SAP WebDispatcher, GREEN, Running, 2015 01 15 03:13:57
 hdbcompileserver, HDB Compileserver, GREEN, Running, 2015 01 15 03:11:55


server11:/usr/sap/ANA/HDB00>  HDB info
 USER       PID  PPID %CPU    VSZ   RSS COMMAND
 anaadm   31848 31847  0.3  13884  2772 -sh
 anaadm   31907 31848  0.6  12900  1720  \_ /bin/sh /usr/sap/ANA/HDB00/HDB info
 anaadm   31930 31907 13.3   4944   876      \_ ps fx -U anaadm -o
user,pid,ppid,pcpu,vsz,rss,args
 anaadm   30631     1  0.0  22024  1508 sapstart
pf=/hana/shared/ANA/profile/ANA_HDB00_server11
 anaadm   30643 30631  0.0 825796 301416  \_
/usr/sap/ANA/HDB00/server11/trace/hdb.sapANA_HDB00 -d -nw -f /usr/s
 anaadm   30663 30643  1.9 17904260 2165336      \_ hdbnameserver
 anaadm   30867 30643  134 16770920 7402644      \_ hdbpreprocessor
 anaadm   30870 30643  0.1 10406304 293128      \_ hdbcompileserver
 anaadm   30899 30643 21.2 19714576 6921608      \_ hdbindexserver
 anaadm   30902 30643  5.6 16580252 2822236      \_ hdbstatisticsserver
 anaadm   30905 30643  4.5 16402576 2332228      \_ hdbxsengine
 anaadm   31246 30643  0.0 271080 72380      \_ sapwebdisp_hdb
pf=/usr/sap/ANA/HDB00/server11/wdisp/sapwebdisp.p
 anaadm   30541     1  0.0 152580 74936 /usr/sap/ANA/HDB00/exe/sapstartsrv
pf=/hana/shared/ANA/profile/ANA_HDB00
 server11:/usr/sap/ANA/HDB00>
```

Prepare SAP HANA for the Cluster User

Modify **hdb_add_user.sh** with your system information and credentials on the first node.

Call the script to create the user needed for the cluster. You need to implement this call only on the primary database because it will be replicated to the secondary database.

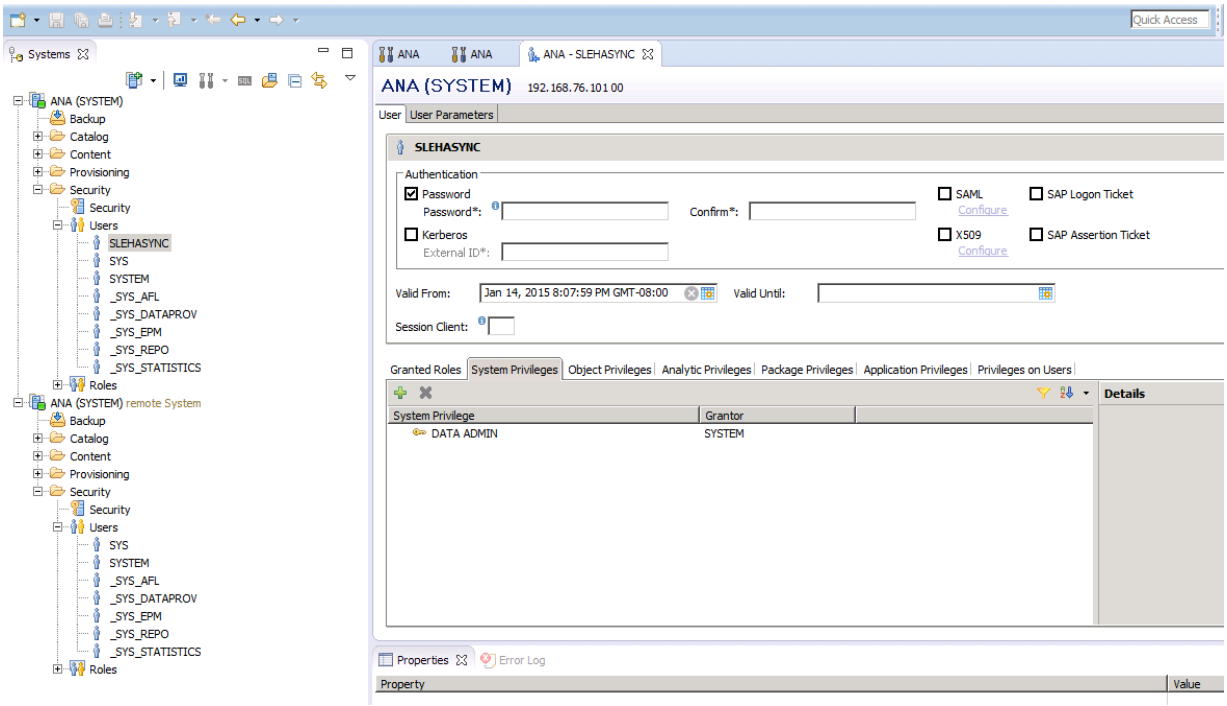
Copy the script from the management node, or re-create it.

```
mgmtsrv06:/opt/Cisco/scripts # scp hdb_add_user.sh server01:/root
 hdb_add_user.sh                                                    100%
```

Modify the script with your system information and user credentials.

```
server01:~ # vi hdb_add_user.sh
 #! /bin/bash
 echo "============================================================================="
 echo "Add User SLEHASYNC for Linux Cluster ...."
 echo "============================================================================="
 echo "\connect -i 00 -n server01 -u system -p manager
 create user slehasync password L1nuxlab
 " | /usr/sap/ANA/HDB00/exe/hdbsql
 echo "\connect -i 00 -n server01 -u system -p manager
 grant DATA ADMIN to slehasync
 " | /usr/sap/ANA/HDB00/exe/hdbsql

 server01:~ #
```

Alternatively, use SAP HANA Studio to create the user.



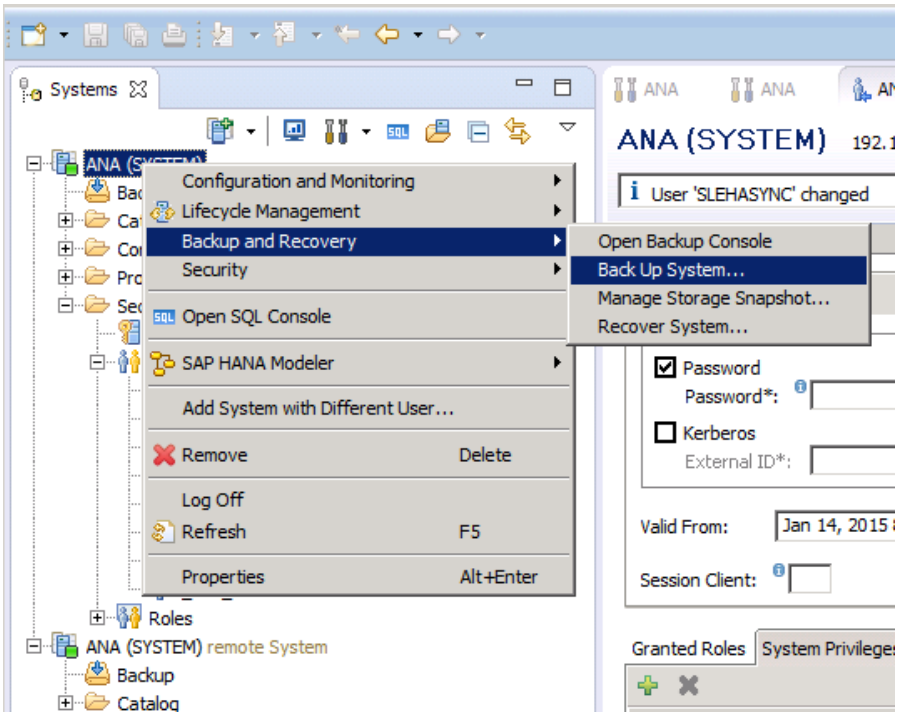
The user keys are stored in the file system, so you need to create the key on each node.

```
server01:~ # export PATH=$PATH:/usr/sap/ANA/HDB00/exe
server01:~ # hdbsql -c -m -i 00 -n server01:30015 -u SLEHASYNC -p L1nuxlab

 Welcome to the SAP HANA Database interactive terminal.

 Type:  \h for help with commands
        \q to quit

hdbsql ANA=> alter user SLEHASYNC password Cisco2015


server01:~ # export PATH=$PATH:/usr/sap/ANA/HDB00/exe
server01:~ # hdbuserstore SET slehaloc localhost:30015 slehasync Cisco2015


server11:~ #  export PATH=$PATH:/usr/sap/ANA/HDB00/exe
server11:~ #  hdbuserstore SET slehaloc localhost:30015 slehasync Cisco2015
```

Test the password for database access.

```
server01:~ # hdbsql -U slehaloc "select * from dummy"


DUMMY
 "X"
lines 1-2/2 (END)
```

Back up your database now using SQL or SAP HANA Studio.



Set Up Two-Node System Replication

Primary Node

Enable the primary node.

```
server01:~ # su - anaadm
 server01:/usr/sap/ANA/HDB00> hdbnsutil -sr_enable --name=SJC-Rack-1
 checking for active nameserver ...
 nameserver is active, proceeding ...
 successfully enabled system as system replication source site
 done.
 server01:/usr/sap/ANA/HDB00>
```

Verify the status.

```
server01:/usr/sap/ANA/HDB00> hdbnsutil -sr_state
 checking for active or inactive nameserver ...

 System Replication State
 ~~~~~~~~~~~~~~~~~~~~~~~~

 mode: primary
 site id: 1
 site name: SJC-Rack-1

 Host Mappings:
 ~~~~~~~~~~~~~~


 done.
 server01:/usr/sap/ANA/HDB00>
```

Secondary Node

The SAP HANA database instance on the secondary side must be stopped before the instance can be registered for system replication.

Enable the secondary node.

```
server11:/usr/sap/ANA/HDB00> sapcontrol -nr 00 -function StopSystem

 15.01.2015 06:17:03
 StopSystem
 OK
server11:/usr/sap/ANA/HDB00> hdbnsutil -sr_register --remoteHost=server01 --
remoteInstance=00 --mode=sync --name=SJC-Rack-2
 adding site ...
 checking for inactive nameserver ...
 nameserver server11:30001 not responding.
 collecting information ...
 updating local ini files ...
 done.
 server11:/usr/sap/ANA/HDB00>
```

Verify the status.

```
server01:/usr/sap/ANA/HDB00>  hdbnsutil -sr_state
 checking for active or inactive nameserver ...

 System Replication State
 ~~~~~~~~~~~~~~~~~~~~~~~~

 mode: primary
 site id: 1
 site name: SJC-Rack-1

 Host Mappings:
 ~~~~~~~~~~~~~~

 server01 -> [SJC-Rack-1] server01
 server01 -> [SJC-Rack-2] server11


 done.
 server01:/usr/sap/ANA/HDB00>
```

```
server01:/usr/sap/ANA/HDB00> hdbsql -U slehaloc 'select distinct
REPLICATION_STATUS from SYS.M_SERVICE_REPLICATION'


REPLICATION_STATUS
 "UNKNOWN"
```

Start the secondary site.

```
server11:/usr/sap/ANA/HDB00> sapcontrol -nr 00 -function StartSystem

 15.01.2015 06:35:17
 StartSystem
 OK
 server11:/usr/sap/ANA/HDB00>
```

Test the system replication status.

```
server01:/usr/sap/ANA/HDB00> hdbsql -U slehaloc 'select distinct
REPLICATION_STATUS from SYS.M_SERVICE_REPLICATION'
 1 row selected (overall time 6920 usec; server time 6493 usec)


REPLICATION_STATUS
 "ACTIVE"
```

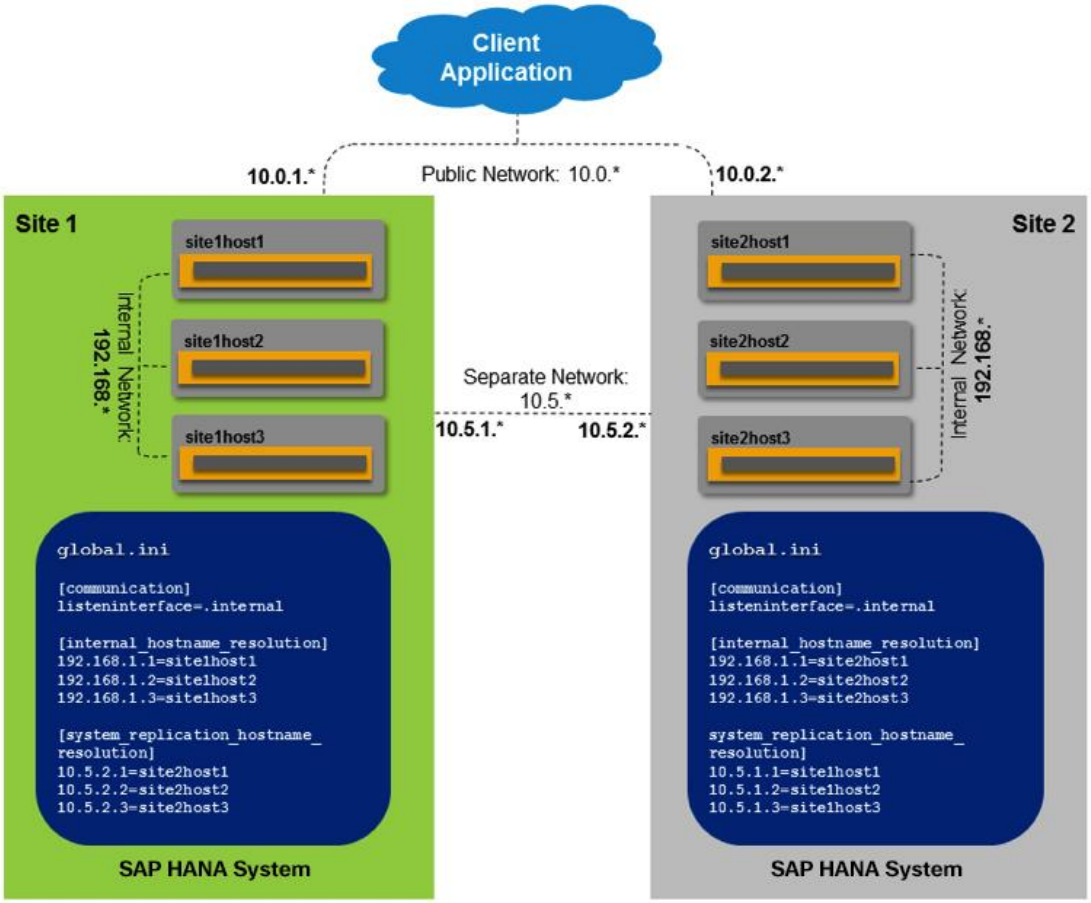
Set Up Network for System Replication

If nothing is configured, SAP HANA uses the access network to synchronize the systems. The solution in this document uses a separate network for system replication, so you should configure the network (Figure 2).

For more information, see:

- http://help.sap.com/saphelp_hanaplatform/helpdata/en/32/c22f81e8c14268a4e5de01cd033e8f/content.htm
- http://help.sap.com/saphelp_hanaplatform/helpdata/en/0c/6738ab85c64da1aed0fa91c25ed47c/content.htm

Figure 2. Network Configuration for System Replication



Source: SAP AG

Change the configuration as shown here.

```
server01:/usr/sap/ANA/HDB00> netstat -ia
 Kernel Interface table
 Iface   MTU Met    RX-OK RX-ERR RX-DRP RX-OVR    TX-OK TX-ERR TX-DRP TX-OVR Flg
 access   1500   0    17116      0  17085     0        7      0      0      0 BMRU
 appl     9000   0        0      0      0     0        0      0      0      0 BM
 backup   9000   0    17099      0  17085     0        7      0      0      0 BMRU
 datasrc  9000   0    17116      0  17085      0       8      0      0      0 BMRU
 eth0     1500   0 96907431      0  17085      0 115121556      0      0      0
BMRU
 lo      16436   0 48241697      0      0      0 48241697      0      0      0 LRU
 mgmt     1500   0    41529      0  17085      0    2043      0      0      0 BMRU
 nfsd     9000   0  9547248      0  17085      0 10164551      0      0      0 BMRU
 nfsl     9000   0  1088756      0  17087    0  1515500      0      0      0 BMRU
 server   9000   0  1296168      0  17085      0 10105637      0      0      0 BMRU
 sysrep   9000   0    17118      0  17085      0        8      0      0      0 BMRU
 server01:/usr/sap/ANA/HDB00>
```

server01

```
server01:/usr/sap/ANA/HDB00> cdglo
 server01:/usr/sap/ANA/SYS/global> cd hdb/custom/config
 server01:/usr/sap/ANA/SYS/global/hdb/custom/config> cat global.ini
 [persistence]
 basepath_datavolumes = /hana/data/ANA
 basepath_logvolumes = /hana/log/ANA

 [system_replication]
 mode = primary
 actual_mode = primary
 site_id = 1
 site_name = SJC-Rack-1
```

server11

```
server11:/usr/sap/ANA/HDB00> cdglo
 server11:/usr/sap/ANA/SYS/global> cd hdb/custom/config/
 server11:/usr/sap/ANA/SYS/global/hdb/custom/config> cat global.ini
 [persistence]
 basepath_datavolumes = /hana/data/ANA
 basepath_logvolumes = /hana/log/ANA

 [system_replication]
 site_id = 2
 mode = sync
 actual_mode = sync
 site_name = SJC-Rack-2

 [system_replication_site_masters]
 1 = server01:30001
```

Reroute the network traffic from the access network to the system replication network.

server01

```
server01:/usr/sap/ANA/SYS/global/hdb/custom/config> cat global.ini
 [persistence]
 basepath_datavolumes = /hana/data/ANA
 basepath_logvolumes = /hana/log/ANA

 [system_replication]
 mode = primary
 actual_mode = primary
 site_id = 1
 site_name = SJC-Rack-1

 [communication]
 listeninterface=.internal

 [internal_hostname_resolution]
 192.168.220.101 = server01
 192.168.220.111 = server11

 [system_replication_hostname_resolution]
 192.168.222.101 = server01
 192.168.222.111 = server11
```

server11

```
server11:/usr/sap/ANA/SYS/global/hdb/custom/config> cat global.ini
 [persistence]
 basepath_datavolumes = /hana/data/ANA
 basepath_logvolumes = /hana/log/ANA

 [system_replication]
 site_id = 2
 mode = sync
 actual_mode = sync
 site_name = SJC-Rack-2

 [system_replication_site_masters]
 1 = server01:30001

 [communication]
 listeninterface=.internal

 [internal_hostname_resolution]
 192.168.220.101 = server01
 192.168.220.111 = server11

 [system_replication_hostname_resolution]
 192.168.222.101 = server01
 192.168.222.111 = server11
```

Restart SAP HANA.

Choose a SAP HANA Synchronization Mode

Several log replication modes are available to send log information to the secondary instance. You need to decide which mode to use.

- Synchronous **(mode=sync)**: In this mode, the log write operation is considered successful when the log entry has been written to the log volume of the primary and secondary systems. If the connection to the secondary system is lost, the primary system continues transaction processing and writes the changes only to the local disk. No data loss occurs in this scenario as long as the secondary system is connected. Data loss can occur if takeover is performed while the secondary system is disconnected.
- Synchronous in memory **(mode=syncmem)**: In this mode, the log write operation is considered successful when the log entry has been written to the log volume of the primary system and transmission of the log has been acknowledged by the secondary system after the log has been copied to memory. If the connection to the secondary system is lost, the primary system continues transaction processing and writes only the changes to the local disk. Data loss can occur if the primary and secondary systems fail at the same time when the secondary system is connected or takeover is performed when the secondary system is disconnected. This option provides better performance, because it is not necessary to wait for disk I/O on the secondary system, but it is more vulnerable to data loss.
- Asynchronous **(mode=async)**: In this mode, the primary system sends a redo log buffer to the secondary system asynchronously. The primary system commits a transaction when it has been written to the log file of the primary system and sent to the secondary system through the network. It does not wait for confirmation from the secondary system. This option provides better performance because it is not necessary to wait for log I/O on the secondary system. Database consistency across all services on the secondary system is guaranteed. However, this option is more vulnerable to data loss. Data changes may be lost on takeover.

To set up system replication, you need to perform the configuration steps on the secondary system. You can complete this configuration using the hdbnsutil tool, which initializes the topology of the database during installation, or exports, imports, and converts the topology of an existing database. You also can use SAP HANA Studio.

Enable Full Synchronization for System Replication

When activated, the full synchronization (fullsync) option for system replication makes sure that a log buffer is shipped to the secondary system before a commit operation occurs on the local primary system.

As of SPS08, the fullsync option can be enabled for SYNC replication (that is, not for SYNCMEM). With the fullsync option activated, transaction processing on the primary blocks, when the secondary is currently not connected and newly created log buffers cannot be shipped to the secondary site. This behavior helps ensure that no transaction can be locally committed without shipping the log buffers to the secondary site.

The fullsync option can be switched on and off using the following command:

```
hdbnsutil -sr_fullsync --enable|--disable
```

Source: SAP HANA Administration Guide

Set Up SUSE Linux Enterprise High Availability Extension

For more information about SLE HAE, see the SUSE cluster documentation at https://www.suse.com/documentation/sle_ha/pdfdoc/book_sleha/book_sleha.pdf.

Test to determine whether the SLE HAE packages are installed.

Test whether the cluster package is installed on all nodes.

```
server01:~ # rpm -qa | grep openais
 openais-1.1.4-5.19.7
 libopenais3-1.1.4-5.19.7

server01:~ # rpm -qa | grep corosync
 libcorosync4-1.4.7-0.19.6
 corosync-1.4.7-0.19.6

server01:~ # rpm -qa | grep csync2
 csync2-1.34-0.8.1

server01:~ # rpm -qa | grep pacemaker
 pacemaker-mgmt-client-2.1.2-0.15.12
 drbd-pacemaker-8.4.4-0.22.9
 pacemaker-mgmt-2.1.2-0.15.12
 libpacemaker3-1.1.11-0.7.53
 pacemaker-1.1.11-0.7.53

server01:~ # rpm -qa | grep SAP
 SUSE_SLES_SAP-release-11.3-1.17
 SAPHanaSR-0.148-0.7.1
 SAPHanaSR-doc-0.148-0.7.1
 SUSE_SLES_SAP-release-DVD-11.3-1.17
 server01:~ #
```

Make sure that the password and hosts files are identical between the two nodes.

```
server11:/ # scp server01:/etc/passwd /etc
 passwd
100% 1536     1.5KB/s   00:00
 server11:/ # scp server01:/etc/group /etc
 group
100%  685     0.7KB/s   00:00
 server11:/ # scp server01:/etc/shadow /etc
 shadow
100%  799     0.8KB/s   00:00
 server11:/ # scp server01:/etc/hosts /etc
 hosts
100% 5692     5.6KB/s   00:00
 server11:/ #
```

Set Up the Cluster

For a step-by-step description of the cluster setup, see the SUSE SAP HANA HA guide at https://www.suse.com/promo/saphana-replication.html.

Perform Basic Cluster Setup

To set up a cluster, you first need to decide what cluster interconnect networks to use:

- eth0 – ring0
- mgmt – ring1 (optional)

This configuration does not include a shared block device (SBD), so you cannot use an SBD for the cluster. If a small device is available for use as an SBD, then an SBD is preferred.

- NetApp Small Computer System Interface over IP (iSCSI) or Fibre Channel (a separate license is required)
- EMC iSCSI or Fibre Channel

In this case, the Cisco Integrated Management Controller (IMC) interface is configured to enable STONITH.

To get a detailed overview of all possible settings for the Intelligent Platform Management Interface (IPMI) resource, enter the command shown here.

```
server11:~ # crm ra info stonith:external/ipmi
```

To test the IPMI connection and function, request the power status.

```
server01:~ # ipmitool -I lanplus -H server01-ipmi -U sapadm -P cisco power status
 Chassis Power is on
server01:~ # ipmitool -I lanplus -H server11-ipmi -U sapadm -P cisco power status
 Chassis Power is on
```

Set Up SAP Host Control

Before you configure anything on the cluster, make sure that the SAP Host Control (SAPHostCtrl) agent is working properly on all nodes. You should use the latest available SAP Host Agent you can download from the SAP marketplace.

Install the latest SAP Host Agent on all nodes.

```
server01:/sapcd # rpm -Uvh saphostagentrpm_203-20005731.rpm
 Preparing...                ########################################### [100%]
    1:SAPHostAgent           ########################################### [100%]
 sapinit                   0:off  1:off  2:off  3:on   4:off  5:on   6:off
 server01:/sapcd #
```

Restart the SAP Host Agent on all nodes.

```
server01:/usr/sap/hostctrl/exe # service sapinit restart
 saphostexec is already running (pid=18222). Stopping...Stopped
 start hostcontrol using profile /usr/sap/hostctrl/exe/host_profile
 Impromptu CCC initialization by 'rscpCInit'.
   See SAP note 1266393.
 Impromptu CCC initialization by 'rscpCInit'.
   See SAP note 1266393.
 server01:/usr/sap/hostctrl/exe #
```

Test the functions.

```
server01:/usr/sap/hostctrl/exe # ./saphostctrl -function ListInstances
  Inst Info : ANA - 00 - server01 - 740, patch 36, changelist 1444691
 server01:/usr/sap/hostctrl/exe #

server11:/usr/sap/hostctrl/exe #  ./saphostctrl -function ListInstances
  Inst Info : ANA - 00 - server11 - 740, patch 36, changelist 1444691
 server11:/usr/sap/hostctrl/exe #
```

Initialize the Cluster

Initialize the cluster.

```
server01:~ # sleha-init

WARNING: NTP is not configured to start at system boot.
   Do you want to continue anyway? [y/N] y
   Enabling sshd service
   /root/.ssh/id_rsa already exists - overwrite? [y/N] N
   Configuring csync2
   Generating csync2 shared key (this may take a while)...done
   Enabling csync2 service
   Enabling xinetd service
   csync2 checking files

 Configure Corosync:
   This will configure the cluster messaging layer.  You will need
   to specify a network address over which to communicate (default
   is mgmt's network, but you can use the network address of any
   active interface), a multicast address and multicast port.

   Network address to bind to (e.g.: 192.168.1.0) [192.168.76.0] 192.168.127.0
   Multicast address (e.g.: 239.x.x.x) [239.37.130.18]
   Multicast port [5405]

 Configure SBD:
   If you have shared storage, for example a SAN or iSCSI target,
   you can use it avoid split-brain scenarios by configuring SBD.
   This requires a 1 MB partition, accessible to all nodes in the
   cluster.  The device path must be persistent and consistent
   across all nodes in the cluster, so /dev/disk/by-id/* devices
   are a good choice.  Note that all data on the partition you
   specify here will be destroyed.

   Do you wish to use SBD? [y/N] N
 WARNING: Not configuring SBD - STONITH will be disabled.
   Enabling hawk service
     HA Web Konsole is now running, to see cluster status go to:
       https://192.168.76.101:7630/
     Log in with username 'hacluster', password 'linux'
 WARNING: You should change the hacluster password to something more secure!
   Enabling openais service
   Waiting for cluster........done
   Loading initial configuration
```

```
  Done (log saved to /var/log/sleha-bootstrap.log)
server01:~ #
```

The cluster should not be allowed to start during bootup. Make sure that the **chkconfig** setting for the cluster is disabled.

```
server01:~ # chkconfig |grep openais
 openais                  off
server01:~ # ssh server11 chkconfig |grep openais
 openais                  off
 server01:~ #
```

Configure the corosync.conf file for User Datagram Protocol (UDP).

```
server01:~ # vi /etc/corosync/corosync.conf
 compatibility: whitetank
 aisexec {
        user:           root
        group:          root
 }

 service {
        ver:            0
        name:           pacemaker
        use_mgmtd:      yes
        use_logd:       yes
 }

 totem {
        version: 2
        token:          5000
        token_retransmits_before_loss_const: 10
        join:           60
        consensus:      6000
        vsftype:        none
        max_messages:   20
        clear_node_high_bit: new
        secauth: off
        interface {
                ringnumber: 0
                bindnetaddr: 192.168.127.0
                member {
                        memberaddr: 192.168.127.101
                }
                member {
                        memberaddr: 192.168.127.111
```

```
                }
                mcastaddr: 239.37.130.18
                mcastport: 5405
                ttl: 1
        }
        transport: udpu
        threads: 4
}

logging {
        fileline: off
        to_stderr: no
        to_logfile: no
        to_syslog: yes
        syslog_facility: daemon
        debug: off
        timestamp: off
        logger_subsys {
                subsys: AMF
                debug: off
        }
}
server01:~ #
```

Synchronize the corosync.conf file.

```
server01: # scp /etc/corosync/corosync.conf server11:/etc/corosync/
```

The second node joins the cluster.

```
server11:~ # sleha-join
 WARNING: NTP is not configured to start at system boot.
   Do you want to continue anyway? [y/N] y

 Join This Node to Cluster:
   You will be asked for the IP address of an existing node, from which
   configuration will be copied.  If you have not already configured
   passwordless ssh between nodes, you will be prompted for the root
   password of the existing node.

   IP address or hostname of existing node (e.g.: 192.168.1.1) [] server01
   Enabling sshd service
   /root/.ssh/id_rsa already exists - overwrite? [y/N] N
   Configuring csync2
   Enabling csync2 service
   Enabling xinetd service
```

```
  Merging known_hosts
  Probing for new partitions......done
  Enabling hawk service
    HA Web Konsole is now running, to see cluster status go to:
      https://192.168.76.101:7630/
    Log in with username 'hacluster', password 'linux'
WARNING: You should change the hacluster password to something more secure!
  Enabling openais service
  Waiting for cluster...done
  Done (log saved to /var/log/sleha-bootstrap.log)
server11:~ #
```

Restart the cluster on both nodes.

```
server01:~ # service openais stop
 Stopping OpenAIS/corosync daemon (corosync): .done OK
 server01:~ #

server11:~ # service openais stop
 Stopping OpenAIS/corosync daemon (corosync): .done OK
 server11:~ #


server01:~ # service openais start
 Starting OpenAIS/Corosync daemon (corosync): starting... OK
 server01:~ #
  server11:~ # service openais start
 Starting OpenAIS/Corosync daemon (corosync): starting... OK
 server11:~ #
```

The actual status of the cluster is reported.

```
server01:~ # crm_mon -1 -r
 Last updated: Mon Jan 26 01:27:41 2015
 Last change: Mon Jan 26 01:20:10 2015 by root via crm_attribute on server11
 Stack: classic openais (with plugin)
 Current DC: server11 - partition with quorum
 Version: 1.1.11-3ca8c3b
 2 Nodes configured, 2 expected votes
 0 Resources configured


 Online: [ server01 server11 ]

 Full list of resources:
```

```
server01:~ #
```

Configure the STONITH IPMI.

```
server01:~ # crm configure
crm(live)configure# primitive STONITH-Server01 stonith:external/ipmi op monitor
interval="0" timeout="60s" op monitor interval="300s" timeout="60s" on-
fail="restart" op start interval="0" timeout="60s" on-fail="restart" params
hostname="server01" ipaddr="server01-ipmi" userid="sapadm" passwd="cisco"
interface="lanplus"

crm(live)configure# primitive STONITH-Server11 stonith:external/ipmi op monitor
interval="0" timeout="60s" op monitor interval="300s" timeout="60s" on-
fail="restart" op start interval="0" timeout="60s" on-fail="restart" params
hostname="server11" ipaddr="server11-ipmi" userid="sapadm" passwd="cisco"
interface="lanplus"

crm(live)configure# location LOC_STONITH_Server01 STONITH-Server01 inf: server11
crm(live)configure# location LOC_STONITH_Server11 STONITH-Server11 inf: server01

crm(live)#
```

With **inf:server0x**, you specify that the server runs only on this node. For IPMI, the service must always run on the other node: for instance, STONITH-Server01 must run on server11.

```
server01:~ # crm configure property no-quorum-policy="ignore"
server01:~ # crm configure property stonith-action="reboot"
server01:~ # crm configure property startup-fencing="false"
server01:~ # crm configure property stonith-timeout="30s"
```

Restart the cluster.

```
server01:~ # service openais stop
Stopping OpenAIS/corosync daemon (corosync): .done OK
Server11:~ # service openais stop
Stopping OpenAIS/corosync daemon (corosync): .done OK
server01:~ # service openais start
Starting OpenAIS/Corosync daemon (corosync): starting... OK
server11:~ # service openais start
Starting OpenAIS/Corosync daemon (corosync): starting... OK
```

The result is shown here.

```
server01:~ # crm_mon -1 -r
```

```
Last updated: Mon Jan 26 23:32:46 2015
Last change: Mon Jan 26 23:26:52 2015 by root via cibadmin on server11
Stack: classic openais (with plugin)
Current DC: server01 - partition with quorum
Version: 1.1.11-3ca8c3b
2 Nodes configured, 2 expected votes
2 Resources configured


Online: [ server01 server11 ]

Full list of resources:

 STONITH-Server01       (stonith:external/ipmi):        Started server01
 STONITH-Server11       (stonith:external/ipmi):        Started server11
server01:~ #
```

Configure the basic settings for the cluster.

```
server01:~ # vi crm-base.txt
```

```
property $id="cib-bootstrap-options" \
no-quorum-policy="ignore" \
stonith-enabled="true" \
stonith-action="reboot" \
stonith-timeout="150s"
rsc_defaults $id="rsc-options" \
resource-stickiness="1000" \
migration-threshold="5000"
op_defaults $id="op-options" \
record-pending="false" \
timeout="600"
```

```
 server01:~ # crm configure load update crm-base.txt
```

Install the SAP HANA–specific properties.

```
server01:~ #  vi saphanatop.txt
primitive rsc_SAPHanaTopology_ANA_HDB00 ocf:suse:SAPHanaTopology \
 operations $id="rsc_sap2_ANA_HDB00-operations" \
 op monitor interval="10" timeout="600" \
 op start interval="0" timeout="600" \
 op stop interval="0" timeout="300" \
 params SID="ANA" InstanceNumber="00"
 clone cln_SAPHanaTopology_ANA_HDB00 rsc_SAPHanaTopology_ANA_HDB00 \
 meta is-managed="true" clone-node-max="1" target-role="Started"
```

```
server01:~ # crm configure load update saphanatop.txt
```

Configure the SAP HANA–specific parameters.

```
server01:~ # vi saphana.txt
 primitive rsc_SAPHana_ANA_HDB00 ocf:suse:SAPHana \
 operations $id="rsc_sap_ANA_HDB00-operations" \
 op start interval="0" timeout="3600" \
 op stop interval="0" timeout="3600" \
 op promote interval="0" timeout="3600" \
 op monitor interval="60" role="Master" timeout="700" \
 op monitor interval="61" role="Slave" timeout="700" \
 params SID="ANA" InstanceNumber="00" PREFER_SITE_TAKEOVER="true" \
 DUPLICATE_PRIMARY_TIMEOUT="7200" AUTOMATED_REGISTER=”false”
 ms msl_SAPHana_ANA_HDB00 rsc_SAPHana_ANA_HDB00 \
 meta is-managed="true" notify="true" clone-max="2" clone-node-max”1”
 target-role="Started"
```

```
server01:~ # crm configure load update saphana.txt
```

Configure the virtual IP address of the customer access LAN or application LAN.

```
server01:~ # vi crm-ip.txt
 primitive rsc_ip_ANA_HDB00 ocf:heartbeat:IPaddr2 \
 meta target-role="Started" is-managed="true" \
 operations $id="rsc_ip_ANA_HDB00-operations" \
 op monitor interval="10s" timeout="20s" \
 params ip="192.168.220.200"
 server01:~ #
```

```
server01:~ # crm configure load update crm-ip.txt
```

Configure the co-location.

```
server01:~ # vi crm-colo.txt
 colocation col_saphana_ip_ANA_HDB00 2000: rsc_ip_ANA_HDB00:Started \
 msl_SAPHana_ANA_HDB00:Master
 order ord_SAPHana_ANA_HDB00 2000: cln_SAPHanaTopology_ANA_HDB00 \
 msl_SAPHana_ANA_HDB00
 server01:~ #
```

```
server01:~ # crm configure load update crm-colo.txt
```

The result is shown here.

```
Server01:/ # /usr/share/SAPHanaSR/tests/show_SAPHanaSR_attributes
 Host \ Attr clone_state remoteHost roles site srmode sync_state vhost
lpa_ana_lpt
--------------------------------------------------------------------------------
 server01 PROMOTED server11 4:P:master1:master:worker:master SJC-Rack-1 sync PRIM
server01 1422591580
 server11 DEMOTED server01 4:S:master1:master:worker:master SJC-Rack-2 sync  SOK
server11 30
 server11:/usr/sap/hostctrl/exe #
```

```
server01:/ # crm_mon -r -1
 Last updated: Thu Jan 29 20:23:10 2015
 Last change: Thu Jan 29 20:21:46 2015 by root via crm_attribute on server01
 Stack: classic openais (with plugin)
 Current DC: server11 - partition with quorum
 Version: 1.1.11-3ca8c3b
 2 Nodes configured, 2 expected votes
 7 Resources configured

 Online: [ server01 server11 ]

 Full list of resources:

  STONITH-Server01       (stonith:external/ipmi):        Started server01
  STONITH-Server11       (stonith:external/ipmi):        Started server11
  Clone Set: cln_SAPHanaTopology_ANA_HDB00 [rsc_SAPHanaTopology_ANA_HDB00]
```

```
      Started: [ server01 server11 ]
  Master/Slave Set: msl_SAPHana_ANA_HDB00 [rsc_SAPHana_ANA_HDB00]
      Masters: [ server01 ]
      Slaves: [ server11 ]
  rsc_ip_ANA_HDB00       (ocf::heartbeat:IPaddr2):       Started server01

server01:/ #
```

Configure the Cluster Resource Manager

Configure the cluster resource manager (CRM).

```
server01:~ # crm configure show
 node server01 \
        attributes hana_ana_remoteHost="server11" hana_ana_site="SJC-Rack-1"
hana_ana_srmode="sync" lpa_ana_lpt="1423192047" hana_ana_vhost="server01"
 node server11 \
        attributes hana_ana_remoteHost="server01" hana_ana_site="SJC-Rack-2"
hana_ana_srmode="sync" lpa_ana_lpt="30" hana_ana_vhost="server11"
 primitive STONITH-Server01 stonith:external/ipmi \
        op monitor interval="0" timeout="60s" \
        op monitor interval="300s" timeout="60s" on-fail="restart" \
        op start interval="0" timeout="60s" on-fail="restart" \
        params hostname="server01" ipaddr="server01-ipmi" userid="sapadm"
passwd="cisco" interface="lanplus" \
        meta target-role="Started"
 primitive STONITH-Server11 stonith:external/ipmi \
        op monitor interval="0" timeout="60s" \
        op monitor interval="300s" timeout="60s" on-fail="restart" \
        op start interval="0" timeout="60s" on-fail="restart" \
        params hostname="server11" ipaddr="server11-ipmi" userid="sapadm"
passwd="cisco" interface="lanplus" \
        meta target-role="Started"
 # enter the following to crm-saphanatop.txt
 primitive rsc_SAPHanaTopology_ANA_HDB00 ocf:suse:SAPHanaTopology \
        operations $id="rsc_sap2_ANA_HDB00-operations" \
        op monitor interval="10" timeout="600" \
        op start interval="0" timeout="600" \
        op stop interval="0" timeout="300" \
        params SID="ANA" InstanceNumber="00"
 primitive rsc_SAPHana_ANA_HDB00 ocf:suse:SAPHana \
        operations $id="rsc_sap_ANA_HDB00-operations" \
        op start interval="0" timeout="3600" \
        op stop interval="0" timeout="3600" \
        op promote interval="0" timeout="3600" \
        op monitor interval="60" role="Master" timeout="1400" \
        op monitor interval="61" role="Slave" timeout="1400" \
```

```
        params SID="ANA" InstanceNumber="00" PREFER_SITE_TAKEOVER="true"
DUPLICATE_PRIMARY_TIMEOUT="7200" AUTOMATED_REGISTER="false"
 primitive rsc_ip_ANA_HDB00 ocf:heartbeat:IPaddr2 \
        meta target-role="Started" is-managed="true" \
        operations $id="rsc_ip_ANA_HDB00-operations" \
        op monitor interval="10s" timeout="20s" \
        params ip="192.168.127.200"
 ms msl_SAPHana_ANA_HDB00 rsc_SAPHana_ANA_HDB00 \
        meta is-managed="true" notify="true" clone-max="2" clone-node-max="1"
target-role="Started"



clone cln_SAPHanaTopology_ANA_HDB00 rsc_SAPHanaTopology_ANA_HDB00 \
        meta is-managed="true" clone-node-max="1" target-role="Started"
 location LOC_STONITH_Server01 STONITH-Server01 inf: server11
 location LOC_STONITH_Server11 STONITH-Server11 inf: server01
 colocation col_saphana_ip_ANA_HDB00 2000: rsc_ip_ANA_HDB00:Started
msl_SAPHana_ANA_HDB00:Master
 order ord_SAPHana_ANA_HDB00 2000: cln_SAPHanaTopology_ANA_HDB00
msl_SAPHana_ANA_HDB00
 property $id="cib-bootstrap-options" \
        stonith-enabled="true" \
        placement-strategy="balanced" \
        dc-version="1.1.11-3ca8c3b" \
        cluster-infrastructure="classic openais (with plugin)" \
        expected-quorum-votes="2" \
        no-quorum-policy="ignore" \
        stonith-action="reboot" \
        startup-fencing="false" \
        stonith-timeout="150s" \
        last-lrm-refresh="1422587833"
 rsc_defaults $id="rsc-options" \
        resource-stickiness="1000" \
        migration-threshold="5000"
 op_defaults $id="op-options" \
        timeout="600" \
        record-pending="false"
 server01:~ #
```

In some cases, instance monitoring may return errors as shown here.

```
Failed actions:
     rsc_SAPHana_ANA_HDB00_monitor_60000 on server11 'ok' (0): call=47,
status=complete, last-rc-change='Thu Jan 29 20:11:35 2015', queued=0ms,
exec=27433ms
     rsc_SAPHana_ANA_HDB00_monitor_61000 on server01 'not running' (7): call=31,
status=complete, last-rc-change='Thu Jan 29 20:07:34 2015', queued=0ms, exec=0ms
```

If errors occur, tune the **op monitor** parameter at the primitive **rsc_SAPHana_ANA_HDB00**.

```
primitive rsc_SAPHana_ANA_HDB00 ocf:suse:SAPHana \
         operations $id="rsc_sap_ANA_HDB00-operations" \
         op start interval="0" timeout="3600" \
         op stop interval="0" timeout="3600" \
         op promote interval="0" timeout="3600" \
         op monitor interval="60" role="Master" timeout="1400" \
         op monitor interval="61" role="Slave" timeout="1400" \
   params SID="ANA" InstanceNumber="00" PREFER_SITE_TAKEOVER="true"
DUPLICATE_PRIMARY_TIMEOUT="7200" AUTOMATED_REGISTER="false"
```

Use 1400 or higher for the value.

Disabling System Replication

```
Before disabling system replication, verify that it is enabled.
hdbsql -U slehaloc 'select distinct REPLICATION_STATUS from
SYS.M_SERVICE_REPLICATION'

REPLICATION_STATUS
 "ACTIVE"
```

Now stop the secondary site.

```
server11:
 server11:/usr/sap/ANA > sapcontrol -nr 00 -function StopSystem

 25.01.2015 22:26:43
 StopSystem
 OK
 server11:/usr/sap/ANA/SYS/global/hdb/custom/config>
```

server01

```
hdbsql -U slehaloc 'select distinct REPLICATION_STATUS from
SYS.M_SERVICE_REPLICATION'

REPLICATION_STATUS
 "ERROR"
```

server11

```
server11:/usr/sap/ANA/SYS/global/hdb/custom/config> hdbnsutil -sr_unregister
 unregistering site ...
 nameserver server11:30001 not responding.
 nameserver server11:30001 not responding.
 checking for inactive nameserver ...
 nameserver server11:30001 not responding.
 nameserver is shut down, proceeding ...
 opening persistence ...
 run as transaction master
 updating topology for system replication takeover ...
 mapped host server01 to server11
 sending unregister request to primary site (2) ...


##########################################################################################
#
 # CAUTION: You must start the database in order to complete the unregistration!

##########################################################################################

 done.
 server11:/usr/sap/ANA/SYS/global/hdb/custom/config>
```

server01

```
server01:/usr/sap/ANA/SYS/global/hdb/custom/config> hdbnsutil -sr_disable
 checking local nameserver:
 checking for inactive nameserver ...
 nameserver is running, proceeding ...
 done.
 server01:/usr/sap/ANA/SYS/global/hdb/custom/config> hdbnsutil -sr_state
 checking for active or inactive nameserver ...

 System Replication State
 ~~~~~~~~~~~~~~~~~~~~~~~~

 mode: none

 done.
 server01:/usr/sap/ANA/SYS/global/hdb/custom/config>
```

For More Information

- SAP HANA takeover process: http://scn.sap.com/docs/DOC-52345
- Introduction to SAP HANA High Availability: http://scn.sap.com/docs/DOC-60334
- SAP guide to performing system replication: https://scn.sap.com/docs/DOC-47702
- SAP HANA administration guide: http://help.sap.com/hana/SAP_HANA_Administration_Guide_en.pdf
- System replication blog: http://scn.sap.com/community/hana-in-memory/blog/2013/12/16/sap-hana-system-replication--using-hdbnsutil-sr
- SLE HAE installation documentation: https://www.suse.com/promo/saphana-replication.html
- Cluster SBD fencing: http://www.linux-ha.org/wiki/SBD_Fencing
- SAP HANA system replication: http://help.sap.com/saphelp_hanaplatform/helpdata/en/09/9caa1959ce4b3fa1144562fa09e163/content.htm
- SAP HANA multinode network configuration: http://help.sap.com/saphelp_hanaplatform/helpdata/en/0c/6738ab85c64da1aed0fa91c25ed47c/content.htm
- Takeover process: http://scn.sap.com/docs/DOC-52345
- Network setup for system replication:
  - http://help.sap.com/saphelp_hanaplatform/helpdata/en/32/c22f81e8c14268a4e5de01cd033e8f/content.htm
  - http://help.sap.com/saphelp_hanaplatform/helpdata/en/0c/6738ab85c64da1aed0fa91c25ed47c/content.htm

Trademarks and Disclaimers

Cisco's trademarks signify Cisco's high quality products and services, and they are valuable assets of the company. Cisco's trademark and brand policies can be found at http://www.cisco.com/go/logo. The following is a list of Cisco's trademarks and registered trademarks in the United States and certain other countries. Please note, however, that this listing is not all-inclusive and the absence of any mark from this list does not mean that it is not a Cisco trademark.

http://www.cisco.com/web/siteassets/legal/trademark.html

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C11-734028-00 2/15