Hewlett Packard Enterprise

Reference Architecture

HPE Reference Architecture for SAP HANA System Replication on HPE Superdome Flex with HPE 3PAR All-Flash Storage (Scale-out configuration built with Intel Cascade Lake)

Disaster recovery solution using SAP HANA system replication and dual purpose storage

Contents

Executive summary ... 3
Introduction ... 4
Solution overview ... 4
Solution components ... 5
Best practices and configuration guidance ... 6
SAP HANA system replication ... 6
Prerequisites for replication and Hewlett Packard Enterprise lab configuration ... 8
System replication with Active/Active (read enabled) production environment ... 8
Client/Application connectivity ... 10
Workload description ... 11
SAP HANA Cockpit ... 11
SAP HANA multi-SID ... 12
Summary ... 13
Appendix A: Configuration adjustments ... 14
Networking ... 14
Storage ... 14
Operating system ... 14
Workload ... 14
Management ... 14
Appendix B: Configuration of dual-purpose storage hardware for non-production use ... 15
Appendix C: Setting up network-based IP redirection for client failover ... 15
Appendix D: Multi-tier system replication setup ... 17
Appendix E: Initialize the secondary with storage copy from the primary ... 19
Appendix F: Multi-SID replication ... 20
Appendix G: Multitarget system replication ... 21
Appendix H: Secondary time travel ... 22
Appendix I: Useful commands ... 23
Glossary ... 25
Resources and additional links ... 26

Executive summary

High availability (HA) and disaster recovery (DR) capabilities are crucial for the SAP® in-memory computing platform in the event that an entire computing site is lost or there is system failure with its mission-critical applications. As a result, the organization may not be able to afford the time required to rebuild the SAP HANA® database and may not be able to afford the hit to the business.

In designing a disaster recovery solution for SAP HANA databases, the designer must focus on two key factors to specify the recovery parameters: recovery time objective (RTO) and recovery point objective (RPO). In response to these two factors, Hewlett Packard Enterprise has developed a high availability and disaster recovery solution for HPE Superdome Flex with HPE 3PAR All-Flash Storage scale-out configuration using kernel-managed and IP network-based SAP HANA system replication. This document describes Hewlett Packard Enterprise validated HA/DR solution that provides real-time replication for all the changes within primary to secondary and/or tertiary site.

In the event of a disaster, the primary site fails over to the secondary site resuming normal operations with minimum downtime and very little or no loss of data. A human decision initiates the failover of a production site to a secondary site.

Another solution to better utilization of the secondary system for non-production instances is to enable a dual-purpose disaster recovery site on the secondary host for Test, QA or Dev like configurations until the active state is restored in the event of a disaster.

What's new?

This Reference Architecture showcases the following offerings introduced in SAP HANA 2.0 SPS 04 or later:

Multi-target system replication: Automatic re-registration of all secondary sites to the new primary site in case the new secondary has taken over or the primary site is unavailable. One use case involves "Follow the sun" approach (where the active primary re-registers itself across geographies to optimize workforce latency during write transactions). Another case involves offloading the read transactions from primary to secondary which remained unproductive before serving the sole purpose of "acting as a secondary" for disaster recovery solution.

- Secondary Time Travel: Able to prepare the secondary system to quickly access data that was deleted in the original system using internal snapshots. This feature (started in SAP HANA 2.0 SPS 03) has been enabled to take place while replication is in progress in the latest version.
- Heterogeneous OS support: Replication is now possible across nodes having different base OS. For this Reference Architecture, replication is tested across RHEL7.6 and SLES 12 SP4 versions.

Target audience: This document is intended to assist solution architects, database and SAP Basis administrators, storage administrators, or IT professional services who are involved in planning and deploying high availability and disaster recovery solutions for HPE Superdome Flex with HPE 3PAR All-Flash Storage for scale-out configurations using SAP HANA system replication. This Reference Architecture assumes that the reader has experience with the SAP HANA database and familiarization with the HPE Superdome Flex configuration setup.

Document purpose: The purpose of this document is to describe a recommended replication architecture/solution for the target technical audiences and customers, highlighting the recognizable benefits of using Hewlett Packard Enterprise and SAP technologies.

This Reference Architecture describes solution testing performed in August 2019.

Introduction

The disaster recovery solution using SAP HANA system replication is built with the industry-leading high-density and highly scalable HPE Superdome Flex Server with HPE 3PAR All-Flash Storage featuring Intel® Cascade Lake processor architecture. The scale-out configuration provides high availability with a small recovery time objective (RTO) including management and automation of processes. The key to achieving high availability is by maintaining features like failover, failback, replication performance, and hardware-level redundancy (on components like power, switches, storage, network, and server). SAP HANA solutions and deployment services from Hewlett Packard Enterprise provide several levels of defense against failure. To learn more about SAP HANA – high availability, refer to this SAP HANA documentation, sap.com/documents/2016/05/f8e5eeba-737c-0010-82c7-eda71af511fa.html.

This Reference Architecture is divided into several sections with detailed configuration steps provided in the Appendix sections.

To learn more about general information on the SAP HANA system replication solutions, requirements, objectives, and the working model, refer to the Solution overview section.

Several sets of base tests were performed on the HPE Superdome Flex Server with HPE 3PAR All-Flash Storage scale-out configurations for SAP HANA replication environment to validate the configurations and to make sure the replication solution is performed as outlined in SAP documentation. These tests were performed with all the sites located in the same data center. A set of scripts were used to simulate transactions on the SAP HANA database by creating tables, adding/modifying data, and verifying data resiliency on the secondary site after a failover/failback happens. SAP HANA data replication performance depends on various factors such as the size of the database, amount of data being replicated, bandwidth between the replication sites, other traffic on the same link, and network latency.

Solution overview

Hewlett Packard Enterprise and SAP have collaborated to offer HPE Superdome Flex Server with HPE 3PAR All-Flash Storage for SAP HANA scale-out configurations as a portfolio of optimally pre-configured hardware. Hewlett Packard Enterprise offers a full range of services that include design, factory integration, on-site installation, custom services, and proactive support with a single point of contact.

SAP HANA system replication is a HA/DR solution for the SAP HANA database providing a zero RPO and shorter RTO with HPE Superdome Flex Server. Today, system replication employs a primary SAP HANA database on an HPE Superdome Flex Server with HPE 3PAR All-Flash Storage in a scale-out configuration with a secondary and/or tertiary tier standby that has an identical primary System Identifier (Instance name or SID), instance number and storage and memory greater than or equal to the primary system. Additional storage can be added to accommodate non-production instances (e.g., DEV/TEST) on the secondary and/or tertiary site while sharing the resources with the replicated SAP HANA production database instance (dual-purpose configurations).

In the case of SAP HANA Active/Active (read-only scenario) and syncmem operation mode, non-production instances cannot be run on the secondary site, as it is not possible to share resources with replicated production instance.

Figure 1 shows two double racks, each used as a scaled-out cluster in the Hewlett Packard Enterprise lab used for multiple testing scenarios. This Reference Architecture was tested with primary - secondary replication environment using sync, syncmem, and async modes with dual-purpose non-production on secondary/tertiary replication environment. When configuring a two cluster solution, both clusters can use sync, syncmem, or async modes.

The HPE Superdome Flex Server for SAP HANA scale-out at the secondary site could have a larger memory footprint than the primary site by using one of the certified HPE Superdome Flex Servers as shown in Figure 1.

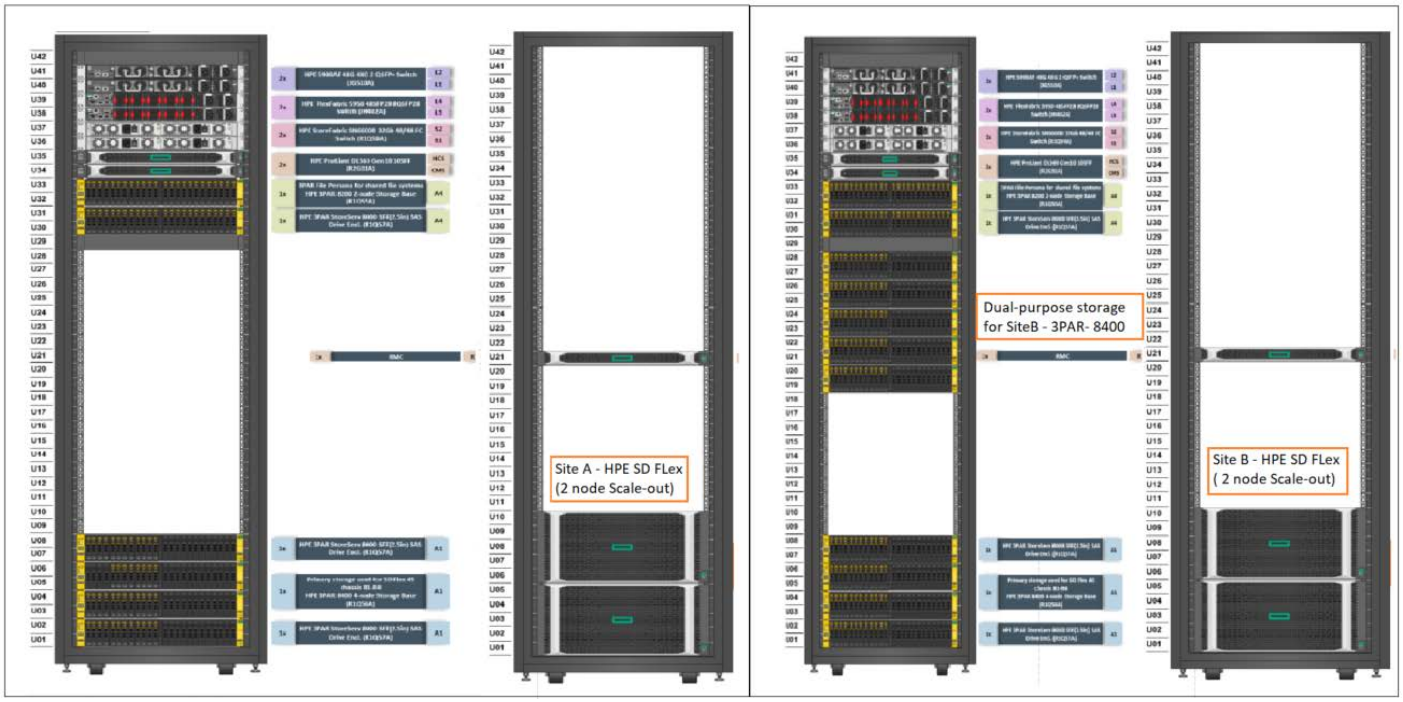

Figure 1. SAP HANA system replication (lab environment)

Solution components

Hewlett Packard Enterprise offers configurations that support large data volumes on SAP HANA databases. This is an enterprise-class real-time analytics or Suite on HANA (SoH) solution optimized to run SAP business application workloads while offering HA and DR architecture. It includes a semi-automated failover mechanism to a secondary system located in the same or different data center. This configuration uses industry-leading HPE Superdome Flex Servers.

Hewlett Packard Enterprise and SAP have teamed up to offer pre-configurations that greatly speed up the analysis of business data so more value can be driven from the information. HPE hardware platform is certified by SAP to run SAP HANA database workload. Refer to this support matrix for the latest product availability configurations for SUSE Linux Enterprise Server (SLES) and Red Hat® Enterprise Linux® (RHEL).

Table 1 shows the solution components.

Table 1. Solution components

Component	Details
Hardware	• Compute node – HPE Superdome Flex Server with 4 x Intel Xeon® Platinum 8280L CPU processors • Memory – 3TB • Additional storage for dual-purpose – HPE 3PAR 8400
Software	• SAP HANA Studio • SAP HANA Cockpit

Component	Details
Operating system	• SUSE Linux Enterprise Server (SLES) 12 SP4 for SAP HANA • Red Hat Enterprise Linux Server (RHEL) 7.6 for SAP HANA
Database	SAP HANA 2.0 database SPS 04 (2.00.040.00.1553674765)

Best practices and configuration guidance

SAP HANA system replication

SAP HANA offers different kinds of high availability mechanisms, supporting a broad range of recovery scenarios from various faults. One of those is system replication, in which SAP HANA replicates all the data to a secondary SAP HANA system (standard SAP HANA feature). Each persisted redo log in the primary system is sent to the secondary. Data is constantly pre-loaded on the secondary system to minimize the recovery time objective (RTO).

This Reference Architecture focuses on supporting decision making on SAP HANA system replication including setting up, testing, and maintaining such a system. Of course, a comprehensive high availability solution offers more design choices and requires the discussion of more details than can be covered in a short Reference Architecture; thus, additional consultation may be required.

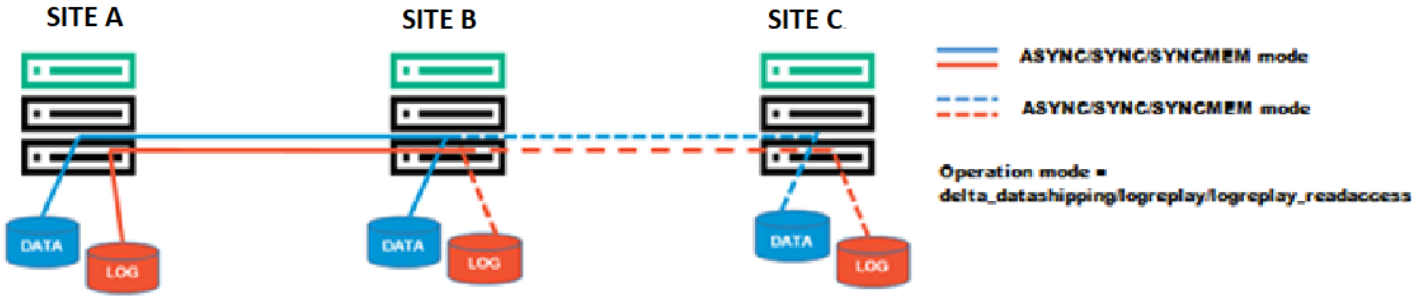

Figure 2. Replication between different sites with respect to modes

Figure 2 explains the replication between different sites. When the secondary system is brought up in live replication mode, each service component establishes a connection with its primary system counterpart and requests a snapshot of the data. Once the initial data synchronization is complete with the primary system, all logged changes in the primary system are replicated. When using a tertiary site, the secondary system acts as a source to the tertiary system. So, data shipping to the tertiary site can only start after the initial data/log transport has been written to the secondary system.

SAP HANA provides the following data and log replication options depending on the recovery time objective (RTO) and recovery point objective (RPO) requirement. SAP HANA customers could choose one of the following replication modes between the two sites:

Replication modes

- Synchronous (SYNC): Secondary system sends acknowledgment back to primary as soon as data is received and persisted to disk.
- Synchronous in-memory (SYNCMEM): Secondary system sends acknowledgment to primary as soon as data is received (this might lead to performance increase depending on disk speed).
- Asynchronous (ASYNC): As per the design of asynchronous replication, the primary does not wait until the secondary system sends an acknowledgment.

Note

In SYNC replication mode, the full sync option can be enabled. When the full sync option is enabled, the log buffer is shipped to the secondary system before any commit takes place on the primary site.

Operation modes

In a multi-tier system replication scenario, the following replication mode combinations are supported. From SAP HANA 2.0 SPS 01 onwards, system replication can be run in three different operation modes:

- **delta_datashipping:** In addition to the continuous redo log shipping taking place, the secondary system requests a delta data shipping from time to time (default every 10 minutes). During the takeover, the redo log needs to be replayed up to the last arrived delta data shipment. (This is the "classical" operation mode of SAP HANA system replication).
- **logreplay:** In this operation mode, pure redo log shipping is done after the system replication was initially set up with one full data shipping. The redo log is replayed on the secondary immediately after arrival making this step superfluous during a takeover, which shortens the RTO. Additionally, the amount of data that needs to be transferred to the secondary site is reduced dramatically, because no delta data shipping is required anymore.
- **logreplay_readaccess:** This operation mode is similar to the logreplay only. The difference here is that the secondary system is read-only.

Table 2 shows the operation modes between the multi-tier site.

Table 2. Operation modes between multi-tier site

Operation mode	Primary to secondary	Secondary to tertiary	Comments
delta_datashipping	SYNC	SYNC	The third system can be registered immediately after secondary is registered
	SYNC	SYNCMEM	
	SYNC	ASYNC	
	SYNCMEM	SYNC	
	SYNCMEM	SYNCMEM	
	SYNCMEM	ASYNC	
	ASYNC	ASYNC	
Logreplay	SYNC	SYNC	Register the tertiary system, only after initial data load is synchronized between primary & secondary
	SYNC	SYNCMEM	
	SYNC	ASYNC	
	SYNCMEM	SYNC	
	SYNCMEM	SYNCMEM	
	SYNCMEM	ASYNC	
	ASYNC	ASYNC	

To perform the system replication for SAP HANA, refer to How to perform System Replication for SAP HANA guide.

For more information on supported replication modes on SAP HANA SPS 12, refer to https://help.sap.com/viewer/6b94445c94ae495c83a19646e7c3fd56/2.0.02/en-US/c3fe0a3c263c49dc9404143306455e16.html.

For more information on system replication configuration parameters on SAP HANA SPS 12, refer to https://help.sap.com/viewer/6b94445c94ae495c83a19646e7c3fd56/2.0.02/en-US/0cd257970d514abd8ddf9ee1f45f3bca.html.

Prerequisites for replication and Hewlett Packard Enterprise lab configuration

There are several prerequisites[1] that need to be fulfilled before configuring system replication.

Table 3 depicts the requirement checklist. To learn more about the basics of SAP HANA system replication and requirements, view this SAP video.

Table 3. Prerequisites for replication and Hewlett Packard Enterprise lab configuration

	SITEA	SITEB	SITEC	Comments
RAM	3TB	3TB	3TB	
DATA	9TB	9TB 9TB	9TB	9TB of external SSD disks (HPE 3PAR-8400) used for dual purpose storage
LOG	1.5TB	1.5TB 1.5TB	1.5TB	Secondary and tertiary system would require more storage for dual purpose storage
System ID	HN1	HN1	HN1	SAP HANA database SID with instance number 01 same on all 3 servers
.ini files	Identical on all hosts	Identical on all hosts	Identical on all hosts	
Replication user	hn1adm	hn1adm	hn1adm	This user is created as part of SAP HANA installation
OS	SUSE/RHEL	SUSE/RHEL	SUSE/RHEL	This solution is tested on both OS
vlan network for replication	Vlan20	Vlan20	Vlan20	Network switch is configured with vlan 20 having speed of 10Gb/s
Vlan network for Administration	Vlan28	Vlan28	Vlan28	Network switch is configured with vlan 28 having speed of 1Gb/s
Vlan network for Shared NFS	Vlan30	Vlan30	Vlan30	Network switch is configured with vlan 30 having speed of 1Gb/s
Reserved instance number	02	02	02	Instance number +1 must free on all system. The user can use the different instance number.
SAP HANA software version	Identical on all hosts	Identical on all hosts	Identical on all hosts	2.00.040.00.1553674765[2]
Multitenant container database SID		NP1, NP2, and NP3		A new tenant database NP1, NP2, and NP3 has been created on SITEB.

System replication with Active/Active (read enabled) production environment

With SAP HANA 2.0, SAP has enabled Active/Active (read enabled) secondary system scenario for the system replication environment. This enables read access on the secondary site while the system replication is in place. The Active/Active (read enabled) feature is supported only for the secondary system and not the tertiary system in a 3-tier system replication scenario.

To enable Active/Active replication, set up SAP HANA system replication with `logreplay_readaccess` operation mode. For this operation mode, the primary and the secondary should have the same SAP HANA version. `logreplay_readaccess` is based on `logreplay` operation mode, so it inherits `logreplay` mode characteristics.

Active/Active (read enabled) provides read-only access on column tables of data on secondary with a delay compared to the primary. From SAP HANA 2.0 database SPS 02 onwards, row store read access is also available.

[1] The mentioned prerequisites are tested in the lab. Secondary and tertiary systems should have more memory than the primary system for the non-production database.
[2] No SAP applications were installed during the test phase.

Memory considerations

- When using Active/Active (read enabled), secondary system total statement memory is limited to 50% of the global allocation limit, because 50% of storage is reserved for logreplay.
- When Active/Active (read enabled) secondary system configuration is in place, secondary needs a separate virtual IP for read access on the secondary system, apart from virtual IP address to access the primary system. During the takeover, the secondary virtual IP address will be used for read access until the takeover completes. After the takeover, the secondary becomes primary and will have two virtual IP addresses. During failback, the secondary virtual IP address moves back again to the primary system.

Note

For the list of applications that are enabled to read from Active/Active (read enabled) HANA node, refer to the SAP Note 2405182 for scale-out configurations with dual-purpose storage (non-production environment).

With SAP HANA system replication, the servers on the secondary and/or tertiary system can be used for non-production SAP HANA testing under the following conditions:

- Additional independent disk volume is needed for DEV/QA (non-production database instances).
- The SIDs and instance numbers have to be different for DEV/QA. The <instance number>+1 of the production system must not be used. This must be free on both sites because this port range is used for system replication communication.
- Pre-load of tables must be switched off on the secondary by setting global.ini / [system_replication] -> preload_column_tables=false.
- The takeover process will take longer as no data is preloaded (sync or async modes) to memory on the secondary site (could still meet SLAs for disaster recovery).
- The DEV/QA (non-production database instances) need to be shut down in case of a takeover.
- Additionally, the global allocation limit on the secondary must be set in a way that the available memory covers the memory needed by the production as well as non-production instances using the following parameter: global.ini / [memorymanager] -> global_allocation_limit.

Table 4 shows the operation modes[3] which influence the memory size required on the secondary.

Table 4. Operation modes and the memory needed for the modes

Operation mode	Memory needed on secondary
delta_datashipping	• Minimum 64GB or • Row store size[4] + 20GB (if the sum is higher[5])
Logreplay	Size of loaded column tables (in-memory[6]) + row store size + 50GB

Note

In an Active/Active production environment, dual-purpose is not possible.

[3] These modes are supported with SAP HANA SPS 11 onwards.
[4] The row store size can be determined with this SQL statement: select host, round(sum(page_size*USED_BLOCK_COUNT)/1024/1024/1024,2) as "RowStore Size GB" from m_data_volume_page_statistics where page_sizeclass = '16k-RowStore' group by host;
[5] If this limit is not set, the SAP HANA database on the secondary site uses as much memory as it can get and possibly takes it away from the DEV/QA systems, which could run into out-of-memory.
[6] The size of loaded column tables (in-memory) can be found with this SQL statement: select round(sum(memory_size_in_total)/1024/1024/1024) size_GB from m_cs_tables.

Key point

1. For "logreplay" and "logreplay_readaccess" modes on the secondary site, log retention is required to do a failback with optimized data synchronization. The primary site periodically creates persistent snapshots during replication and provides the log position information to the secondary. The new primary site has to keep the log until a new secondary site is registered and has synced the missing log.
2. Because syncing can take some time, this behavior has to be explicitly turned on by setting this parameter on the new primary global.ini / [system_replication] / enable_log_retention = on.
3. In case of frequent failbacks between two sites, it is recommended to set the following parameter on both sites to simplify configuration: global.ini / [system_replication] / enable_log_retention = on.

The above configurations are the Hewlett Packard Enterprise recommended/HA/DR architectures for HPE Superdome Flex Server in scale-out configurations using SAP HANA with dual-purpose configurations running non-production instances on the secondary or tertiary system. In a 3-tier scenario, when primary and secondary are replicating in sync and async operation modes, non-production instances can be run on both secondary and tertiary sites. When primary and secondary are in Active/Active or in syncmem operation mode, non-production instances can be run on tertiary sites only. Hewlett Packard Enterprise recommends having reliable high bandwidth links with low latency between the primary and secondary sites.

Note

It is recommended to have a dedicated server network communication of 10Gbit/s or higher between the primary and the secondary system for efficient data replication. Refer to SAP HANA Hardware and Software requirements.

Client/Application connectivity

It is important for high availability to make sure the client users and applications can continue to communicate with the SAP HANA database whenever there is a failover or failback activity. Connection recovery after disaster recovery can be done with network-based IP redirection. The disaster recovery planning process needs to consider how IP addresses used by the clients accessing the database systems can be moved between primary and secondary systems. There are different possibilities for enabling client connection recovery.

For network-based IP redirections, refer to Appendix D, which outlines how to set up virtual IP addresses to handle client reconnection.

IP redirection requires that both the primary and failover hosts are on the same subnet. If the standby system is in a completely separate subnet network, then DNS redirection is the preferred alternative solution. The DNS redirection has an advantage over IP redirection as no client-specific configurations are required in the former. Further, it supports disaster recovery configurations, where the primary and secondary standby systems may be in two completely different network domains (separated by routers). Due to DNS caching in nodes (both clients and intermediate network equipment), it may take a while before the DNS changes are actually propagated, causing clients to experience downtime despite the recovery of the system. One way to reduce the length of the time out is to modify the Time to Live (TTL) value in the DNS servers.

Testing and Usage tests were carried out using a script that simulated table creation and insertion of data into a set of columnar tables. Failover and failback tests were conducted before, during, and after all tables were created or records were modified. The data created was verified before failover occurred (i.e., size of the database, number of tables, and row count were the same at both primary and replicated secondary). Tests included both sync, async, and syncmem[7] scenarios. Since the test was a run at a local HA site, the data was replicated and the failover was performed on the secondary system in the same rack in the same data center. SAP HANA database transactions executed during the failure on the affected node ("in-flight-transactions") were rolled back. The purpose of the test is to ensure and validate that the replication was working to the secondary system and tertiary system. If the primary system fails, then the secondary system (SITE B) would failover with the third DR system (SITE C) continuing to replicate data. In the case of an entire site failure, the third system (DR) would failover as production.

[7] For Syncmem replication, primary and secondary site should have the same size of memory allocated to the database.

The sites referred to in examples throughout this Reference Architecture are shown below:

- Primary: SITE A (10.xx.xx.22 – xxxxxx-22)
- Secondary: SITE B (10.xx.xx.20 – xxxxxx-20)
- Tertiary: SITE C (10.xx.xx.16 – xxxxxx-16)

Note
The testing was performed on HPE Superdome Flex with HPE 3PAR All-Flash Storage for scale-out configurations (3TB system memory) with SAP HANA database version: 2.00.040.00.1553674765. The size of the database is approximately 1.1TB.

Workload description

On SITE A, as user SYSTEM, a new user called USER1 was created with a new schema USER1. Test data was created by creating multiple tables and inserting records into each table in the SAP HANA database, using Hewlett Packard Enterprise internal tool from the SAP HANA Linux® Client. The size of the database is ~1.1TB.

SAP HANA Cockpit

The SAP HANA Cockpit is a web-based HTML5 user interface that provides a single point-of-access to a range of web-based tools for the online administration and monitoring of SAP HANA. From SAP HANA 2.0 onwards, SAP HANA Cockpit has to be installed on a separate server other than SAP HANA nodes.

In general, applications that provide core system and database administration features are available by default, for example, database monitoring, user management, and data backup. SAP HANA systems can be registered as resources at SAP HANA Cockpit manager (https://<SAP HANA cockpit hostname>:51023) by logging in with COCKPIT_ADMIN user, which is created during Cockpit installation. After registering the SAP HANA system as a resource at Cockpit manager, access SAP HANA Cockpit using the URL (https://<SAP HANA cockpit hostname>:510231) with COCKPIT_ADMIN user. The latest SAP HANA Cockpit can manage multiple SAP HANA systems.

For detailed information about accessing SAP HANA Cockpit, refer to,
https://help.sap.com/viewer/6b94445c94ae495c83a19646e7c3fd56/2.0.02/en-US/3188e43e196c46e5a26989f22e18f130.html.

Figure 3 shows the system replication dashboard in SAP HANA Cockpit.

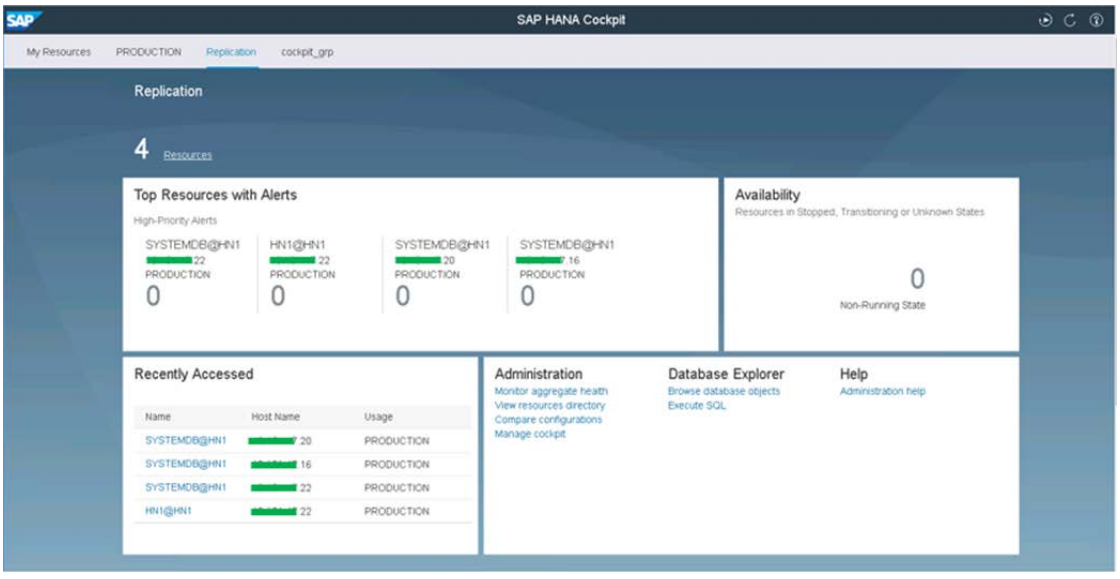

Figure 3. SAP HANA Cockpit view

Figure 4 shows the 2-tier system replication. This shows all the services are in “SYNC” replication mode and replication status is “Active” between the sites.

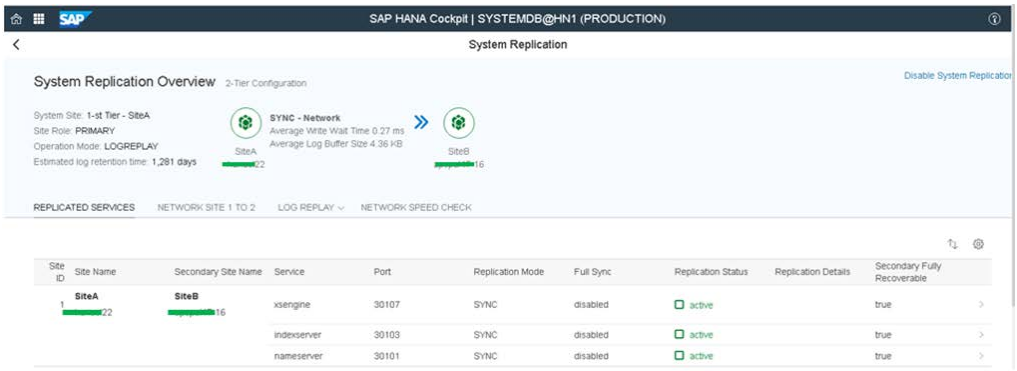

Figure 4. SAP HANA Cockpit view – System replication overview

SAP HANA multi-SID

Multiple SAP HANA databases can also run in an SAP HANA system replication configuration. Independent SIDs of the primary system are replicated to the secondary site. In a system, the SAP HANA database SIDs share the same storage area. This configuration is also said to be Multiple Components on One System (MCOS).

Figure 5 shows the SAP HANA multi-SID replication deployment between the primary site and secondary site. Refer to Appendix F section for more information. Refer to SAP note 1681092 for multiple SAP HANA SIDs on the SAP HANA system.

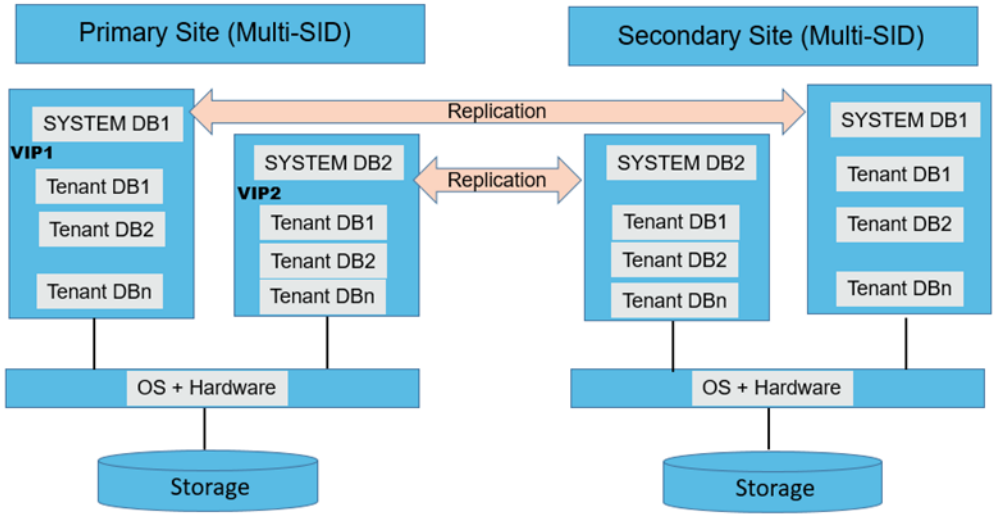

Figure 5. SAP HANA multi-SID deployment

While the system replication is running, the secondary system, which is configured identically to the primary, will be on standby until a takeover. Each SAP HANA database SID has its own VIP.

Summary

With database sizes growing continually and substantially, many organizations are considering efficient and cost-effective disaster recovery solutions over long distances. Hence, they designed the SAP HANA database environment which is suitable for their requirement of data availability and a small RTO in the event of any disaster.

In order to protect customer data investment, the primary system should be replicated to a secondary and/or tertiary site. The secondary and tertiary sites can also be used for non-production (i.e., development, test, etc.) purposes at the DR site, while the secondary or tertiary system is receiving replicated data snapshot and log data from the primary site. This effectively takes advantage of available resources and is a cost-effective solution for additional testing and development activities. It is, however, possible that a disaster recovery implementation can create challenges in different areas (e.g., in storage utilization, SAP HANA database performance, management, and availability). Hewlett Packard Enterprise addresses these challenges through a robust, easy-to-use solution.

Hewlett Packard Enterprise conducted a series of tests[8] to demonstrate that this HA/DR solution can ensure the SAP HANA database is available at the alternate remote secondary site during typical disaster scenarios by supporting synchronous or asynchronous system replication. Also, this solution is ideal for organizations that require a local HA solution for a scale-out system to protect the production site and a DR site, to prevent local disasters. Since the first version of system replication, the delta_datashipping operation mode has been the default replication method. With the logreplay operation mode, no delta-datashipping are necessary and the takeover time has been reduced and more components are already initialized at replication time.

HPE Superdome Flex with HPE 3PAR All-Flash Storage for SAP HANA scale-out configuration is designed to satisfy the requirements of a demanding high availability and disaster recovery solution considering cost and complexity. This solution is recommended for reducing the SAP HANA database downtime due to planned maintenance, faults, and disasters.

This solution is also helpful for customers to perform a technology refresh by applying updates at a local site as well as remote sites from their current HPE Superdome Flex with HPE 3PAR All-Flash Storage for SAP HANA Scale-out configuration setup.

Implementing a proof-of-concept

As a matter of best practice for all deployments, Hewlett Packard Enterprise recommends implementing a proof-of-concept using a test environment that matches closely to the planned production environment. In this way, appropriate performance and scalability characterizations can be obtained. For help with a proof-of-concept, contact Hewlett Packard Enterprise service representative (hpe.com/us/en/services/consulting.html) or your Hewlett Packard Enterprise partner.

[8] Testing performed in the Hewlett Packard Enterprise lab between adjacent HPE Superdome Flex with HPE 3PAR All-Flash Storage for scale-out configurations.

Appendix A: Configuration adjustments

Networking

A bonded network needs to be set up between two different Ethernet card ports to provide high availability for the dedicated SAP HANA replication network with a minimum of 10Gbps bandwidth.

To enable communication over a particular network, define a mapping from the replication IP address to the hostname for SAP HANA system replication, which is used exclusively for SAP HANA system replication. A dedicated separate replication network is configured to isolate all replication traffic from all other network traffic. An overview of the configuration of system replication over a separate network is shown in Figure A1.

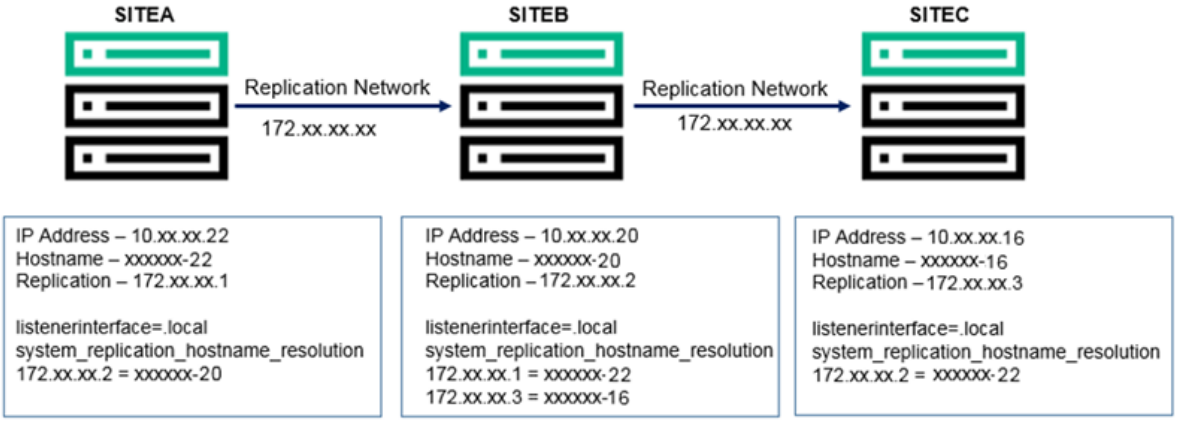

Figure A1. SAP HANA Cockpit view – System replication overview

It is important to map IP addresses to hostnames by editing the section "system_replication_hostname_resolution" in the global.ini file. In the case of multi-tier system replication, only the direct neighbors have to be specified in the mapping as shown in Figure D2. Now each system can resolve the hostnames correctly from the other system. This configuration must be done before the secondary or tertiary system is registered with the primary system because the command "hdbnsutil -sr_register" uses the mapping.

The separate network configurations have to be done for the NFS file persona for the Scale-out cluster. Refer to HPE HANA Solutions SD Flex 3PAR Sample file persona setup for setting up an NFS file share for the cluster nodes.

Storage

For production instance, HPE 3PAR All-Flash Storage has been used. For non-production instance, an additional HPE 3PAR 8400 storage has been used. Refer to the HPE Superdome Flex configuration guide for storage requirements.

Operating system

SUSE Linux Enterprise Server and Red Hat Enterprise Linux Server have been used in this setup. Refer to the HPE HANA Solutions SD Flex 3PAR Installation Guide for SUSE/RHEL installation.

Workload

The workload was generated using an HPE internal tool. The size of the database was ~1.1TB.

Management

SAP HANA Studio and SAP HANA Cockpit were used to monitor the SAP HANA database status and replication info, etc.

Appendix B: Configuration of dual-purpose storage hardware for non-production use

It is possible to make use of the secondary or tertiary system for DEV/QA environments while the primary system is in production when Active/Active is not enabled between primary and secondary.

The following prerequisites need to be reviewed prior to setting up the secondary system for dual-purpose DEV/Test/QA environments while the primary system is in live replication relationship with the production instances on the secondary site and still replicating data:

- Additional independent disk storage is needed for the DEV/QA, LOG, and DATA file systems for non-production tests. Cannot share production storage.
- The failover process will take longer as no data is preloaded in the RAM of the secondary system (could still meet SLAs for disaster recovery).
- Non-production DEV/QA must be shut down before executing a failover operation on the secondary system (which adds time to the RTO as tables are not preloaded).
- The SIDs and instance numbers have to be different for all non-production DEV/QA (start with at least +1 from the production instance number).

Table B1. Production and non-production instance numbers

INSTANCE TYPE	SID	INSTANCE NUMBER
Production instance	HN1	01
Non-production instance	NP1, NP2, and NP3	03, 05, and 07

Note

For “delta_datashipping” operation mode, in the Hewlett Packard Enterprise lab environment we have tested three non-production SAP HANA database instances in the secondary and tertiary systems. The test was carried out by setting memory values in the “global_allocation_limit” for production instance 65536 (64GB) and non-production instance 131072 (128GB).

For “logreplay” operation mode, in the Hewlett Packard Enterprise lab environment we have tested two non-production SAP HANA database instances in the secondary and tertiary systems. The test was carried out by setting memory values in the “global_allocation_limit” for production instance 247808 (242GB) and non-production instance 65556 (64GB).

For more information on memory calculation, refer to How to Perform System Replication for SAP HANA under the section 4.3, “Use Secondary site for DEV/QA system”.

Appendix C: Setting up network-based IP redirection for client failover

Client reconnection after disaster recovery can be enabled with network-based IP redirection[9]. As part of a complete disaster recovery plan, you need to define how user client connections and applications will access IP addresses on the failover system after a primary site failure.

Virtual IP addresses or VIPs are commonly used to enable database high availability for the client connectivity. Applications use them for database connections instead of the normal host IP address; however, VIPs can lead to very severe split-brain situations, if used incorrectly. To deploy virtual IP addresses without problems, you must first of all understand how they work, and secondly, use a sound cluster management solution that avoids split-brain and minimizes the impact if it does occur.

IP redirection requires that both the primary and secondary hosts are on the same network.

For our test environment, we used a public IP address that the HANA Studio or clients can use to reach the HANA database servers.

Edit the /etc/hosts file for the client systems and all SAP HANA database systems and add the IP address for the VIP as follows:

[9] Network-based DNS redirection is not tested in this setup.

```
 hanaSL23:- # cat /etc/hosts
# Host Network
xx.xx.xx.22   xxxx23        xxxx23
xx.xx.xx.22   xxxx22        xxxx22
xx.xx.xx.16   xxxx16        xxxx16
```

```
#Replication Network
xx.xx.xx.22   xxxx23-rep    xxxx23-rep
xx.xx.xx.22   xxxx22-rep    xxxx22-rep
xx.xx.xx.16   xxxx16-rep    xxxx16-rep
```

```
#vip network
xx.xx.xx.230  xxxx23-vip    xxxx23-vip
```

Below is a script that can be used to set up the Virtual IP on the SAP HANA servers. To start the Virtual IP on "xxxx-vip", run the script "/sbin/setupHanaVIP.sh[10] UP". To stop the Virtual IP on "xxxx-vip", run the script "/sbin/setupHanaVIP.sh DOWN".

```
#!/bin/bash
# This shell script configures the Virtual IP/Hostname for Network based client redirection using
(VIP/Hostname) during SAP HANA System Replication
NIC=bond4
VIRTUAL_HOSTNAME=xxxx-vip
NMASK=255.255.255.0
if [ $# -ne 1 ]
then
echo "Usage: setupHanaVIP.sh UP/DOWN"
exit
fi
if [ $1 = "UP" ]
then
/usr/sap/hostctrl/exe/saphostctrl -function ACOSPrepare -op ifup -iface $NIC -vhost
${VIRTUAL_HOSTNAME} -nmask ${NMASK}
else
if [ $1 = "DOWN" ]
```

[10] The SID and netmask values in the script has to be edited as per the customer environment.

```
then
/usr/sap/hostctrl/exe/saphostctrl -function ACOSPrepare -op ifdown -vhost ${VIRTUAL_HOSTNAME}
fi
fi
```

The lab test configuration uses the public IP address 10.xx.xx.22 for bond4 on the primary site, 10.xx.xx.20 on the secondary site, and 10.xx.xx.16 on the tertiary site. The bond4 interface is allowed to accept the traffic for another address 10.xx.xx.230 (VIP) using the following script on the SAP HANA system that is primary and secondary or the tertiary site after failover.

```
/sbin/setupHanaVIP.sh UP
```

The command in this script tells the TCP/IP stack to accept packets directed to IP address 10.xx.xx.230, as well as the original address 10.xx.xx.22. This means that user client software can connect to the primary, secondary, or tertiary site using the VIP address depending on which system is acting as primary.

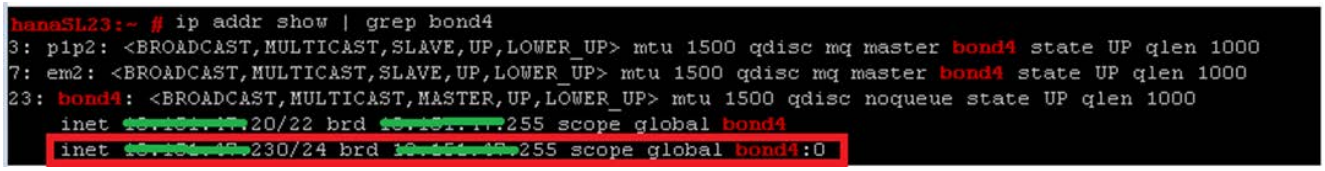

Figure C1. Verifying the management bond4 (VIP) status

Appendix D: Multi-tier system replication setup

To configure 3-tier replication, it is best to set up the replication SITE B and SITE C at the same time. If a third tier is being added to an existing two-tier system replication environment, then replication at secondary SITE B will need to be unregistered from the primary site to set up the network and any dual-purpose non-production configurations.

SAP HANA 2.0 replication connection uses certificate-based authentication. The authentication is done using the certificates in the system PKI SSFS store. To enable the authentication between primary, secondary, and tertiary sites, system PKI SSFS files should be the same on all sites. Copy the following system PKI SSFS .key (SSFS_<SID>.KEY) and .dat (SSFS_<SID>.DAT) files from the primary site to the same location on the secondary and tertiary sites.

```
$DIR_INSTANCE/../global/security/rsecssfs/data/SSFS_<SID>.DAT
$DIR_INSTANCE/../global/security/rsecssfs/key/SSFS_<SID>.KEY
```

Ensure to copy above files before registering secondary/tertiary system.

Following is the procedure to configure the system replication in a 3-tier environment:

```
@Site A:
su - su1adm
sapcontrol -nr 01 -function StartSystem HDB
hdbnsutil -sr_enable --name=SITEA
```

```
@Site B:
su - su1adm
```

```
sapcontrol -nr 01 -function StopSystem HDB

hdbnsutil -sr_register --remoteHost=xxxxx23 --remoteInstance=01 --replicationMode=syncmem --
operationMode=logreplay --name=SITEB

sapcontrol -nr 01 -function StartSystem HDB

hdbnsutil -sr_enable
```

```
@Site B - To stop non-production database:

su - su3adm

sapcontrol -nr 03 -function StopSystem HDB

su - su5adm

sapcontrol -nr 05 -function StopSystem HDB
```

```
@Site C:

su - su1adm

sapcontrol -nr 01 -function StopSystem HDB

hdbnsutil -sr_register --remoteHost=xxxxx22 --remoteInstance=01 --replicationMode=async --
operationMode=logreplay --name=SITEC

sapcontrol -nr 01 -function StartSystem HDB
```

```
@Site C - To stop non-production database:

su - su3adm

sapcontrol -nr 03 -function StopSystem HDB

su - su5adm

sapcontrol -nr 05 -function StopSystem HDB
```

Figure D1 shows the 3-tier replication view on SAP HANA Studio.

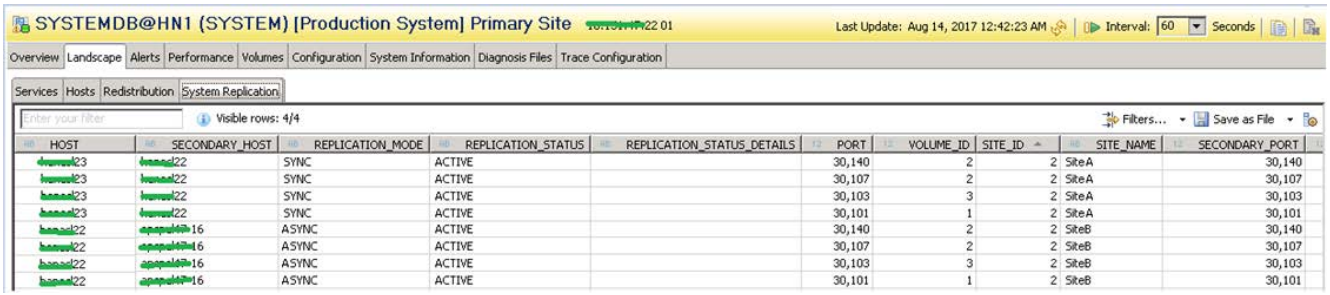

Figure D1. 3-tier replication view on SAP HANA Studio

Key point

For “logreplay” mode, configure the third site (SITE C) after the full replication is completed between SITE A and SITE B. For “delta_datashipping” mode, all three sites’ replication can be configured at the same time.

In a multi-tier system replication environment, if SITE A is taken over by SITE B, the failback to SITE A can be achieved without interrupting the high availability between sites as shown in Figure D2.

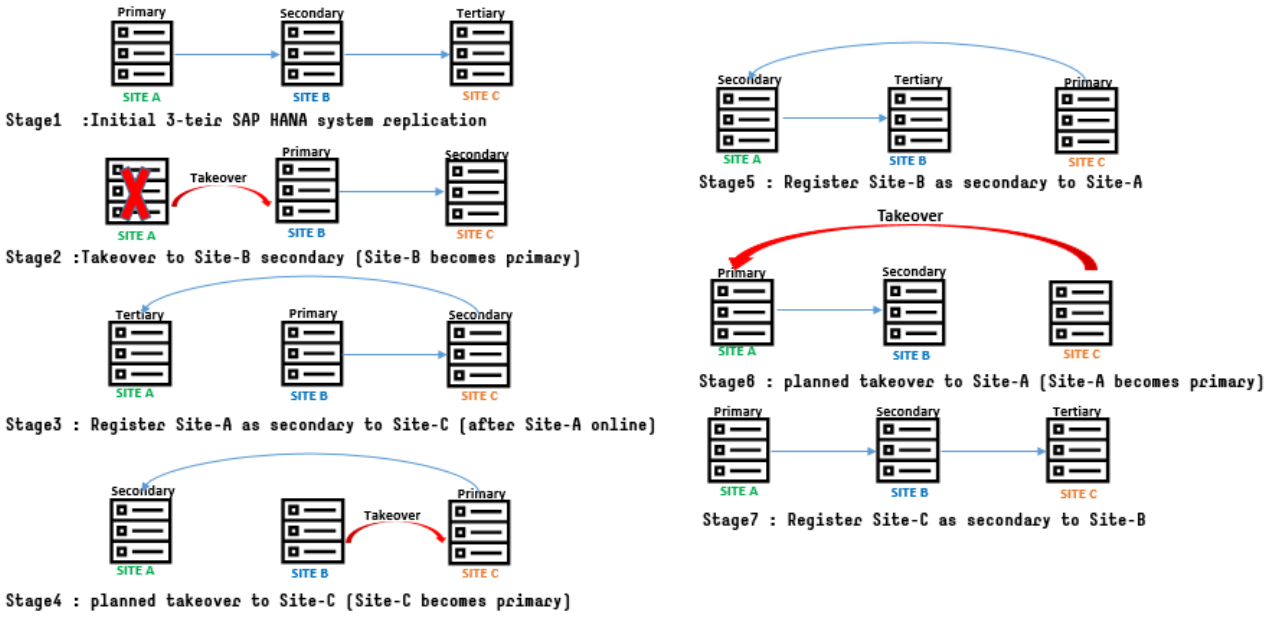

Figure D2. 3-Tier replication failback scenario

In the above diagram, in Stage1, the 3-tier SAP HANA system replication setup is configured. In Stage2, SITE A failed due to some hardware issue, so takeover to SITE B is done, and SITE B becomes primary and SITE C becomes secondary. Once SITE A becomes available, as shown in Stage3, register SITE A as a secondary to SITE C. Now SITE B is primary, SITE C is secondary to SITE B and primary to SITE A, and SITE A is secondary to SITE C. In this scenario, if we initiate a takeover to SITE C and then to SITE A, then SITE A becomes primary but without any high availability at the site level. So to maintain site-level high availability, failback SITE A as primary and initiate planned takeover to SITE C.

In Stage4 perform a planned takeover to SITE C, now SITE C becomes primary and SITE A is secondary to SITE C. In Stage5, register SITE B as secondary to SITE A. Take a planned takeover to SITE A from SITE C in Stage6, now SITE A becomes primary with SITE B as secondary. In the Final stage, Stage7, registering SITE C as secondary to SITE B, achieves failback to SITE A as primary, with SITE B and SITE C as secondary and tertiary.

Appendix E: Initialize the secondary with storage copy from the primary

After successful registration of the secondary site with the primary site, the data from the primary site is initialized with the secondary site. This data initialization can be prevented by initializing the secondary site with the binary storage copy of the primary site persistence.

Using storage snapshot technology creates an I/O consistent persistence copy of the primary site. Otherwise, create a consistent OS copy of persistence by stopping the primary site.

To configure initialization of the secondary site with storage copy on HPE Superdome Flex with HPE 3PAR All-Flash Storage configurations, perform the following procedure:

1. Stop the primary site. Create a consistent OS copy of the primary site.

2. Shut down the secondary site and transfer the full consistent copy of the primary to the secondary. Copy only the data, log is not required.
3. Replace the persistence of the secondary site with the full consistent copy of the primary site.
4. Register the secondary site without [--force_full_replica] parameter using hdbnsutil or uncheck the “initialize full datashipping” checkbox while registering with SAP HANA Cockpit.
5. Start the secondary site. At this stage, if the persistence of the secondary site is compatible with the primary site, the secondary system requests only delta_datashipping instead of full data initialization.

Appendix F: Multi-SID replication

SAP HANA multi-SID deployments are characterized by multiple SAP HANA systems on one host. This configuration is approved for production environments as of SAP HANA Support Package Stack (SPS) 09. This is restricted to single host/scale-up scenarios only. SAP HANA multi-SID requires significant attention to various detailed tasks related to system administration and performance management. Figure F1 shows the multi-SID database replication between SITE A and SITE B with two production instances (HN1 and DB1). The non-production database NP1 and NP2 are configured on SITE B.

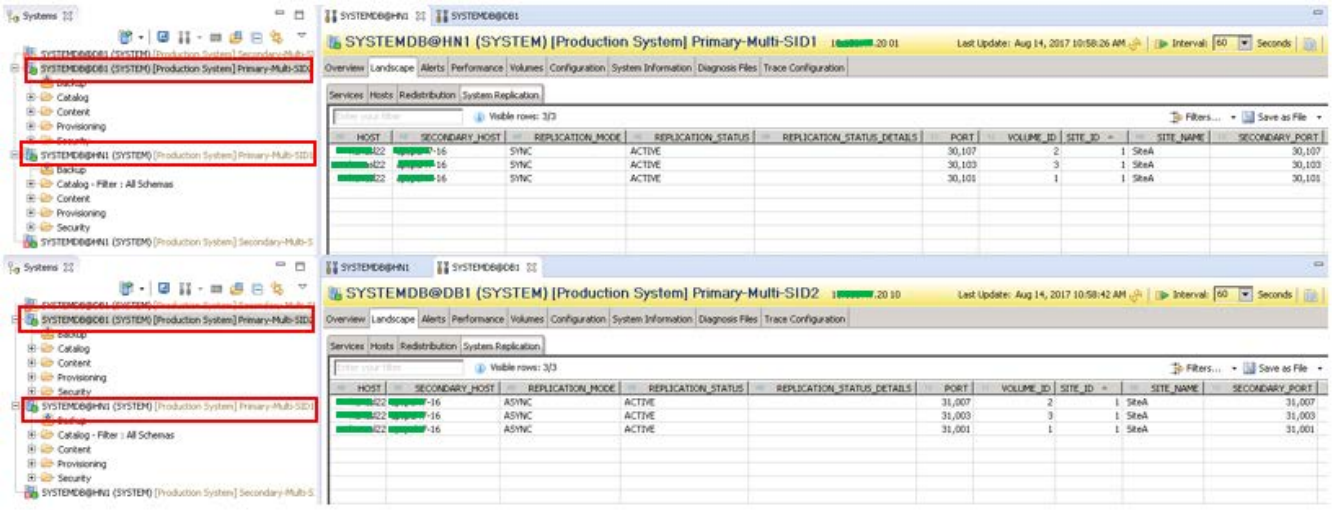

Figure F1. Multi-SID database replication

Note

In the SAP HANA multi-SID environment[11] in the Hewlett Packard Environment lab, we have tested two production and two non-production SAP HANA database instances in the secondary systems[12]. The test was carried out by setting memory values in the "global_allocation_limit" for production instance 307200 (300GB) and non-production instance 65536 (64GB). Each SAP HANA database instance was assigned with one Virtual IP (VIP). The scripts setupHanaVIP.sh (SID1) and setupHanaVIP1.sh (SID2) was created.

For more information on memory calculation (delta_datashipping and logreplay mode), refer to How to Perform System Replication for SAP HANA under the Section 4.3 use the secondary site for DEV/QA system.

Appendix G: Multitarget system replication

SAP HANA supports multitarget system replication from SAP HANA 2.0 SPS03 onwards. In system replication, the primary system replicates all the data to only one secondary system. In multitarget system replication, the primary system replicates all the data to more than one secondary system. Multitarget system replication will be useful in update scenarios, for increasing higher availability, rearrangement of system multitier chains, etc.

From SAP HAN2.0 SPS04 onwards, automatic re-registration of secondary sites has been enabled in case of takeover by the secondary site or when the initial primary source is unavailable.

For automatic re-registration on a secondary site using a takeover, a new parameter `global.ini/[system_replication]/register_secondaries_on_takeover = [true/false]`, has to be declared on the secondary site which has to act as a new primary. SAP HANA will re-register all the secondary systems of the former primary site to the new system that took over as primary.

A couple of new parameters in case of "unavailability of the source site" on a specific secondary was also introduced: `global.ini/[system_replication]/alternative_sources = <SiteName1>:<ReplicationMode1>,<SiteName2>:<ReplicationMode2>` has to be defined in secondary to select the alternative primaries in case the current primary is unreachable.

`Global.ini/[system_replicaiton]/retries_before_register_to_alternative_source=<time in secs>` has to be defined to check alternative sources if the current source site is unreachable for a given period of time.

Figure G1 shows an example environment for multitarget system replication. The Primary system in SITE A replicates the data to the secondary system in SITE A and it also replicates the data to the secondary system in the SITE B. This secondary system in SITE B is a primary for another secondary system in SITE B.

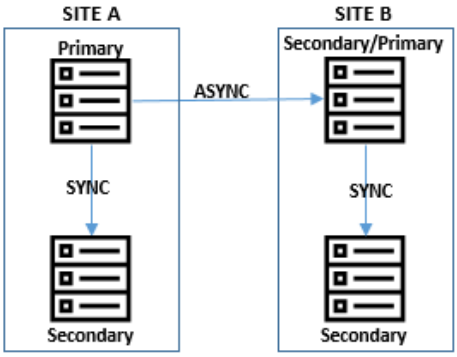

Figure G1. Multitarget system replication environment

[11] SAP HANA Multi-SID environment is tested only on RHEL operating environment (delta_datashipping mode). It may work on SLES operating environment as well.
[12] Primary system has two production SAP HANA database instances.

Configuring multitarget system replication is similar to the regular system replication. Figure G2 explains the configuration of multitarget system replication.

HOST	SECONDARY_HOST	REPLICATION_MODE	REPLICATION_STATUS	REPLICATION_STATUS_DETAILS	PORT	VOLUME_ID	SITE_ID	SITE_NAME	SECONDARY_PORT	SECONDARY_SITE_ID	SECOND.
[redacted]17	[redacted]15	SYNC	ACTIVE		30,007	2	2	SITEB	30,007	1	SITEA
[redacted]17	[redacted]15	SYNC	ACTIVE		30,003	3	2	SITEB	30,003	1	SITEA
[redacted]17	[redacted]15	SYNC	ACTIVE		30,001	1	2	SITEB	30,001	1	SITEA
[redacted]17	[redacted]20	ASYNC	ACTIVE		30,007	2	2	SITEB	30,007	3	SITEC
[redacted]17	[redacted]20	ASYNC	ACTIVE		30,003	3	2	SITEB	30,003	3	SITEC
[redacted]17	[redacted]20	ASYNC	ACTIVE		30,001	1	2	SITEB	30,001	3	SITEC

Figure G2. Configuration of multitarget system replication

Figure G3 (a) shows the initial setup of multi-target system replication. In the next Figure G3 (b) Primary system 1 on SITE A failed and takeover will happen to secondary system 2 on SITE A. At this stage replication from Site A to SITE B secondary system 3 will be stopped. In Figure G3 (c), after the takeover, secondary system 2 on SITE A becomes Primary and automatically register secondary system 3 on SITE B to new Primary system 2 on SITE A. Once system 1 is available, automatically register that as secondary to primary system 2 on SITE A as shown in Figure G3 (d). Now primary system 2 replicates the data to secondary system 1 on SITE A and secondary system 3 on SITE B using multi-target configuration.

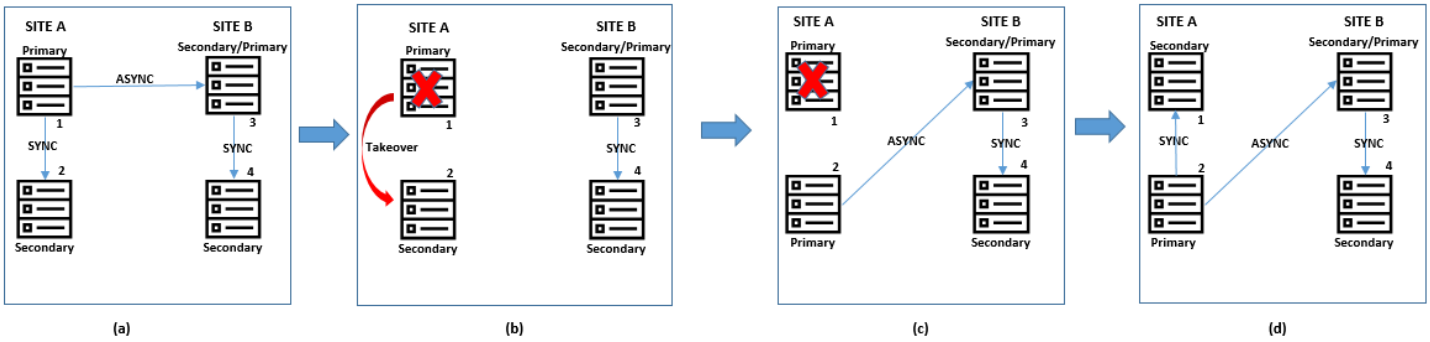

Figure G3. Failback scenario in multitarget replication

Appendix H: Secondary time travel

SAP HANA supports secondary time travel from HANA2SPS03 rev 36 onwards (logreplay and logreplay_readaccess modes only). It is a special version of takeover that can be used to access the data which was deleted in the original system or to read older data while the secondary keeps replicating.

The secondary system can be started to an older point in time by using hdbnsutil –sr_timetravel command. During the execution of hdbnsutil -sr_timetravel, the specified time and location are stored internally in a particular file. When invoking the command, takeover hooks can be called explicitly if not defined in the configuration parameter timetravel_call_takeover_hooks. The secondary system will enter in online mode on the specified point in time during restart. After a restart, the other services read the requested point in time and open their persistence using this information. If the requested point in time cannot be reached, then time travel will be aborted.

The start time and other statistics for the snapshots can be viewed by invoking the query “select * from M_SYSTEM_REPLICATION_TIMETRAVEL” from hdbsql prompt. The same will be available for view using HANA Cockpit from SPS05 onwards.

To invoke a secondary time travel manually:

1. Set `global.ini/[system replication]/ timetravel_logreplay_mode = manual` on the secondary system.
2. Stop the secondary system.
3. Execute `hdbnsutil -sr_timetravel --startTime=<startTime> --callTakeoverHooks=on --comment=<comment> command. Use format UTC: dd.mm.yyyy - hh.mm.ss` for starttime.

4. Start the secondary system.
5. Monitor the start time on the primary using `select * from M_SYSTEM_REPLICATION_TIMETRAVEL` query.
6. Use `hdbnsutil -sr_recoveruntil {--endTime=<timestamp>|max} [--nowait]. Use format UTC: dd.mm.yyyy - hh.mm.ss` for endtime. The endtime max is used to trigger logreplay to the newest possible time.

The manual replay mode can be stopped by setting the timetravel_logreplay_mode parameter back to auto or using hdbnsutil -sr_replaymode --mode=auto.

Appendix I: Useful commands

To enable the system replication:

```
hdbnsutil -sr_enable -name=SITEA
```

Register from the secondary site:

```
Sapcontrol -nr 01 -function StopSystem HDB

hdbnsutil -sr_register --name=SITEB --remoteHost=xxxxxx22 --remoteInstance=01 --mode=async
```

Perform failover from secondary site:

```
Sapcontrol -nr 01 -function StopSystem HDB (stop database instance first from primary site or make
sure not connected)

hdbnsutil -sr_takeover
```

Review status of replication (execute from both sites):

```
hdbnsutil -sr_state
```

From SQL prompt or from Studio:

```
select REPLICATION_STATUS from M_SERVICE_REPLICATION; or

select HOST, SITE_NAME, SECONDARY_HOST, SECONDARY_SITE_NAME, REPLICATION_MODE, REPLICATION_STATUS,
REPLICATION_STATUS_DETAILS, SECONDARY_ACTIVE_STATUS from PUBLIC.M_SERVICE_REPLICATION;
```

To get a detailed overview of the replication environment:

```
hdbcons -e hdbindexserver ”replication info”
```

Failback: old primary HDB is shut down; make it the secondary: HBD stop (old primary site)

```
hdbnsutil -sr_register --remoteHost=xxxxxx23 --remoteInstance=01 --mode=async --name=SITEB
```

To clean up the replication configuration:

```
hdbnsutil -sr_cleanup
```

To un-register a SITE from the replication configuration:

```
hdbnsutil -sr_unregister --name=SITEA
```

To start secondary time travel:

```
hdbnsutil -sr_timetravel --startTime=<starttime in format UTC: dd.mm.yyyy - hh.mm.ss > --
startMode=replicate --callTakeoverHooks=on --comment= <comment>
```

To monitor secondary time travel stats on primary:

```
select * from M_SYSTEM_REPLICATION_TIMETRAVEL
```

To check the overall status of the system replication using as <sid>adm OS user the script systemReplicationStatus.py:

```
su1adm# python $DIR_INSTANCE/exe/python_support/systemReplicationStatus.py
```

Check the overall status of the primary system using as <sid>adm OS user the script:

```
su1adm# python $DIR_INSTANCE/exe/python_support/landscapeHostConfiguration.py
```

Glossary

Table 5. Keywords and descriptions

Keyword	Description
DR	Disaster Recovery
HA	High Availability
RPO	Recovery Point Objective
RTO	Recovery Time Objective
Disaster	The failure of an entire compute site or data center
Failover	Switching to an alternate standby system/site in the event of the disaster at primary system/site
Failback	The process of restoring a site/system to its original state
Fault	A failure of a system or one of its components / sub-systems
Downtime	Inverse of availability: the duration of time that a system is not operational
HDD	Hard Disk Drives
SID	System Identifier (used as the database ID for HANA, not the name of a system)
VIP	Virtual IP Address

Resources and additional links

HPE Reference Architectures, http://www.hpe.com/info/ra

HPE Superdome Flex Server, QuickSpecs link

SAP HANA Administration Guide, http://help.sap.com/hana/SAP_HANA_Administration_Guide_en.pdf

SAP HANA System Replication FAQ - SAP note 1999880, http://service.sap.com/sap/support/notes/1999880

HPE & SAP Alliance, hpe.com/partners/sap

HPE IT Infrastructure Consulting Services for SAP HANA, hpe.com/us/en/services/platform-consulting-services.html

HPE Servers, hpe.com/servers

HPE Storage, hpe.com/storage

HPE Networking, hpe.com/networking

HPE Technology Consulting Services, hpe.com/us/en/services/consulting.html

To help us improve our documents, please provide feedback at hpe.com/contact/feedback.

a50000496enw, October 2019