EMC Data Domain Deduplication Storage Systems - SAP HANA Data Protection

Abstract
This white paper provides technical considerations and specific configuration guidance for successfully deploying EMC Data Domain Systems in an SAP HANA environment.

September 2013
Copyright © 2013 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is.” EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.
Table of Contents

Executive Summary ............................................................................................................. 4
The Solution Overview ......................................................................................................... 4
Audience................................................................................................................................. 4
Technology overview ............................................................................................................. 5
EMC Data Domain System ..................................................................................................... 5
SAP HANA Appliances ........................................................................................................... 6
SAP HANA Backup and Recovery .......................................................................................... 7
SAP HANA Columnar Data Storage Compression ................................................................. 8
Solution Test Scenarios ......................................................................................................... 9
Methodology .......................................................................................................................... 10
Results ..................................................................................................................................... 11
One Step Backup using EMC NetWorker ............................................................................... 12
Simple Disaster Recovery ...................................................................................................... 24
Conclusion .............................................................................................................................. 24
Executive Summary

SAP HANA is an in-memory data platform that is deployable as an on-premise appliance, or in the cloud. It is a ground-breaking platform that’s best suited for performing real-time analytics, and developing and deploying real-time applications. At the core of this real-time data platform is the SAP HANA database that is fundamentally different than any other database engine in the market today.

Many IT organizations perform SAP HANA database backups on a nightly basis. To meet their backup and recovery window requirements, most businesses store these backups for thirty days or more. Unfortunately this leads to rapid growth in backup storage requirements, which has kept some users stuck with legacy tape systems as the default solution for database backups. However, this reliance on tape can limit the number of backups that can be performed, impacting recovery point objectives (RPOs).

In addition, SAP administrators are constantly challenged to improve recovery time objectives (RTOs). Recovering HANA databases from previous backups, then rolling the archive/redo logs forward is time consuming and complex. However, restoring the database in the shortest possible time is essential to business operations.

The Solution Overview

EMC Data Domain systems deliver industry-leading data duplication and performance for the protection of HANA databases. The solution described in this paper covers SAP HANA backups to Data Domain systems over NFS using SAP HANA Studio and single-step NetWorker-based SAP HANA backups to Data Domain or any other disk or tape device supported by EMC NetWorker. EMC NetWorker uses standard scripting functionality and SAP HANA’s hdbsql backup utility.
Audience

This white paper is intended for SAP Basis administrators, Backup administrators, technical consultants, partners, and members of the EMC and partner professional services community who are looking for faster, more efficient SAP HANA database backup and recovery with complete control over the SAP HANA disaster recovery procedures.

Technology overview

The following section provides an overview of components as listed in table 1.

<table>
<thead>
<tr>
<th>System</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMC Data Domain DD640</td>
<td>• DDOS 5.2.2.4</td>
</tr>
<tr>
<td>SAP HANA Appliance (3 active node)</td>
<td>• SAP HANA DB 1.0 SP5</td>
</tr>
<tr>
<td></td>
<td>• SAP HANA Studio</td>
</tr>
</tbody>
</table>

Table 1. SAP HANA Solution Components

EMC Data Domain System

Data Domain deduplication storage systems offer a cost-effective alternative to tape that allows users to provide the retention and recovery benefits of inline deduplication.

Data Domain deduplication storage systems dramatically reduce the amount of disk storage needed to retain and protect enterprise data. With the industry’s fastest deduplication storage controller, Data Domain systems allow more backups to complete faster while putting less pressure on limited backup windows.

All Data Domain systems are built as the data store of last resort. The data in each Data Domain system is protected by the EMC Data Domain Data Invulnerability Architecture – an end-to-end data verification, continuous fault detection and self-healing mechanisms coupled with other resiliency features transparent to the application.

Data Domain systems also integrate easily into your existing environment, enabling you to take advantage of the benefits of deduplication across workloads, infrastructure, and backup and archiving applications. For example, it enables efficient and non-disruptive backup of large databases such as data warehouses and data marts.

**SAP HANA Appliances**

The SAP HANA database is an in-memory database that combines transactional data processing, analytical data processing, and application logic processing functionality in memory. SAP HANA removes the limits of traditional database architecture that have severely constrained how business applications can be developed to support real-time business.

SAP HANA’s database design thus enables one to perform real-time online application processing (OLAP) analysis on an online transaction processing (OLTP) data structure. As a result, one can address today’s demand for real-time business insights by creating business applications that previously were neither feasible nor cost-effective.

SAP HANA was delivered as an appliance with factory pre-installed hardware, operating system and SAP HANA database software, either single node or multi-node cluster, on proven hardware provided by SAP’s hardware partners. It can also be delivered under SAP’s Enterprise Shared Storage and Tailored Datacenter Integration program, which allows customer to use existing infrastructure.

With the appliance model, the hardware components include dedicated servers using the latest Intel CPU architecture to run the SAP HANA database software as well as dedicated network and storage components.

The SAP HANA architecture has two deployment options:
- Single host configuration
- Multi-node cluster configuration

![Figure 1 - SAP HANA System Architecture](image-url)
**Single host configuration (Scale-Up)**
In a SAP HANA single host configuration, all SAP HANA components run on a single server. The persistence layer resides on server internal flash and SAS disks. The scalability of the database is limited by the amount of RAM that can be installed in the server. A Single host configuration does not provide HA (High Availability) functionality.

**Multi-node cluster configuration (Scale-Out)**
In a multi-node cluster configuration, the SAP HANA database is distributed across multiple servers. The scalability of the database is achieved by adding servers to the cluster. Shared storage is required for a multi-node cluster configuration to provide HA functionality. Each node uses its own persistent devices on the shared storage layer. In case of a node failure, the SAP HANA HA feature reassigns the storage persistence to a standby node, which recreates the in-memory database part from the storage persistence.
In a distributed scale-out environment, one node (usually the first installed node) is the master node. It handles the workload of the NetWeaver stack, the statistics and the tables located in the row store. It is strongly recommended to have a minimum of three nodes for a HANA scale-out database available. In two nodes landscapes, the table distribution and the parallel process degree would not be beneficial.

**Figure 2 - SAP HANA Persistent Layer**

**SAP HANA Persistence**
The SAP HANA in-memory database holds the bulk of its data in memory for maximum performance but still uses persistent storage to provide a fallback in case of a failure. The SAP HANA persistence resides on either internal disks (single node) or on shared storage (multi-node cluster) and uses two volumes: Data and Log.
The primary SQL data is kept in memory but asynchronously written to the persistence as a savepoint every 5 minutes (default). Data also contains undo log information. Information about data changes are directly saved to the persistence whenever a transaction is committed or the internal log buffer (1MB) is full.

**SAP HANA Backup and Recovery**
Even though SAP HANA database is an in-memory database, it maintains its own persistency. Data is always saved from memory to disk at save points and data changes are written into redo log files. In case of failure, this persistency can be leveraged for recovery.

Backups can be triggered using the SAP HANA studio, the DBA Cockpit in Business Warehouse (BW), SQL script commands or 3rd party tools. Since these are not automatically run by the SAP HANA system, each HANA database administrator will have to design a backup strategy with the hardware vendor before purchasing SAP HANA to ensure that they have the appropriate hardware to support the backup process.

When performing a backup of the SAP HANA system the tables, views, undo logs, packages, information views and metadata are all saved to a configurable persistent disk location. In summary, the benefit is that all of the data that is stored in SAP HANA will be backed up to a data path that you specify. The default location of the data backup is configured as $(DIR_INSTANCE)/backup/data.

**SAP HANA Columnar Data Storage Compression**

Column-stores are well suited for analytical queries on massive amounts of data. For high read performance using the SAP HANA DB’s column-store, it leverages a very efficient compression scheme in combination with cache-aware and parallel-tasking algorithms. Every column is compressed using a sorted dictionary, i.e., each value is mapped to an integer value (the valueID). These valueIDs are further bit-packed and compressed. By re-sorting the rows in a table, the most beneficial compression (e.g., run-length encoding (RLE), sparse coding, or cluster coding) for the columns of this table can be determined and used. Compressing data not only allows for retaining more data on a single node, but it also allows for faster query processing, e.g., by exploiting the RLE to compute aggregates. Scans are accelerated by using SIMD algorithms working directly on the compressed data. For more information on Columnar Data Storage Compression, this link provides more details:  
http://sites.computer.org/debull/A12mar/hana.pdf

---

![Figure 3: Data Domain NFS for SAP HANA backup solution](image-url)
Solution Test Scenarios

This document presents a test scenario that demonstrates how Data Domain systems complement SAP HANA environments as shown in Figure 3 above.
Methodology

The SAP HANA Studio was the backup application used to perform backup and restore testing. A proprietary synthetic data generation tool was used to generate globally unique data within a controlled change rate environment. Three SAP HANA nodes were connected directly to a Data Domain DD640 system over two 10 Gbps data links.

Backup workload scenarios
- Initial database backup.
- The subsequent six days full backup with 5% change rates were based upon the last backup database size. (NOTE: HANA does not support incremental backup in the current release)

Data set size and type
- Initial database size was 500 GB.
- Database tables were created and partitioned among three HANA nodes - all the datasets were backed up in parallel.

The NFS Protocol was used in all tests. The “Destination Type” used was File for NFS backups.
Results

Figure 4 shows seven days of full backups with HANA where column compression is enabled (this is the default setting on the HANA configuration). The initial backup took 41 min for 500+ GB datasets. The balance of all six subsequent backups with a 5% controlled change rate ranged from 23 - 27 minutes completion.

As shown in Figure 3 below, the deduplication ratios show that the local compression factor is 1.2x across the seven days of full backups. The global compression ratios started at 1.4x and and then increase to 6.9x after seven days of backups.

The global compression values show the amount of storage required for the file after the deduplication process finishes, but before local compression occurs.

The local compression value provides the storage requirement after local compression (LZ [default], gzfast, gz) is applied to the deduplicated segments.
Figure 5. Global-Comp Factor = Pre-Comp / (Size after de-dupe) Local-Comp Factor = (Size after de-dupe) / Post-Comp. Total-Comp Factor = Pre-Comp / Post-Comp

In Figure 6 below, the total compression ratio is varies from 1.6 to 7.6 over the seven day range of full backups.

The pre-compression value is the storage requirement for the contents of data to be backup.

The post-compression values is the actual space required to store the deduplicated data

Figure 6. Total-Comp Factor = Pre-Comp / Post-Comp

One Step Backup using EMC NetWorker
**Native SAP HANA Backup and Restore**

SAP HANA comes with native tools and commands to perform backups of database data and redo logs to disk, as well as restoring the database from these backups.

There are a few ways to perform SAP HANA database data backup, such as using SAP HANA Studio, using SQL commands, etc. In this document, and specifically, in the script, we use SQL commands to perform data backups with SAP HANA client tool “hdbsql”.

The syntax of using hdbsql to perform a database data backup:

```
> hdbsql -i <instance number> -u <username> -p <password> "backup data using file
<(path>/<prefix>)"
```

Using explicitly username and password with hdbsql is not recommended by SAP HANA, because the password will be visible in the process list. SAP HANA provides a tool, “hdbuserstore”, which can be used to create a user key. The user key encapsulates the credential information, including username, password, and the instance number. Using user key instead of plain username and password with hdbsql looks like:

```
> hdbsql -U <user_key> "backup data using file (path)/<prefix>)"
```

Refer to section 13.9.3 in SAP HANA Administration Guide for details on creating user key.

/path> is a user-provided destination for the backup.
<prefix> is a user-provided prefix for the backup name.

For details on how to perform backup and recovery with SAP HANA native tools, refer to Chapter 13 in SAP HANA Administration Guide.

**Introduction to NSRSAPHANA**

nsrsaphana is a shell script written to perform SAP HANA 1-step backup to NetWorker. It can be used to configure a scheduled backup with NMC, and it can also be used on its own to perform a manual backup. The scheduled backups can be further replicated to remote site using NetWorker catalog-aware Clone Controlled Replication if desired.

The script consists of three sections: pre-command section, backup to NetWorker section, and post-command section.

The primary functionality of the pre-command section is to perform the first stage of the 1-step backup process: SAP HANA database backup to disk. It uses the “hdbsql” client tool of SAP HANA to perform this backup operation. The syntax of hdbsql performing a database data backup:

```
> hdbsql -i <instance number> -u <username> -p <password> "backup data using file
<(path>/<prefix>)"
```

Or, if using the user key:

```
> hdbsql -U <user_key> "backup data using file (path)/<prefix>)"
```
nsrsaphana accepts either user key or username/password/instance number combination for connection authentication.

<path> would be a user-provided destination for the backup, as mentioned in section 1. In the script, it is the staging area to which the intermediate backup files will be generated. This is provided by the user.

<prefix> would be a user-provided prefix for the backup name, as mentioned in section 1. In the script, it is the timestamp of when the backup started. This is internally provided by the script, and not by the user.

Once the pre-command issues hdbsql and successfully performs a database backup to the staging area (we refer to backups made in the staging area “backup files”), the script invokes NetWorker “save” command to backup the backup files to NetWorker. All the backup files in the staging area will be saved to one saveset. The name of the saveset has the following format:

SAPHANA_<Data/Log>_<SID>_HANA_Instance_<Timestamp>_<Path>

The saveset name identifies whether the backup is data backup or logs backup. “SID” (System ID) identifies to which SAP HANA system the backup belongs. “HANA_Instance” identifies from which instance in the SAP HANA system the backup came. “Timestamp” identifies when the backup was performed. “Path” is the staging area where the backup files in the saveset were from.

The last section, post-command, performs cleanup operation to the staging area. The user can configure that files older than a certain number of days can be deleted from this staging area.

Scheduled Backup via NMC
This section describes how to use nsrsaphana and NMC to configure scheduled backup of SAP HANA database data or redo logs.

Data Backup:
The user can follow the following steps to configure a SAP HANA scheduled backup of database data to NetWorker:

1. Firstly, we need to set some parameters in nsrsaphana. The script can be found in the Appendix, and you shall see that towards the beginning of the script, three parameters are required to be set by the user:
   - hdbsql_cmd: Specify the full path to hdbsql.
   - hana_SID: Specify the SAP HANA System ID (SID).
   - hana_instance: Specify the SAP HANA instance number.
   There are examples in the script to show how to set these parameters.

The user also needs to set credential parameters. If the user has already created a user key, then set the following parameter in the script:
   - hana_key: Specify the SAP HANA user key for backup.
Otherwise, the user needs to specify a database username and password that has privilege BACKUP ADMIN:

- `hana_username`: Specify the database username that performs the backup.
- `hana_password`: Specify the corresponding password for the username.

If all of the three parameters above are provided, the user key takes the higher precedence and will be used, because it is more secure than the plain username/password (password can be leaked through the process list).

Above are all the parameters that the user needs to set in the `nsrsaphana` script. The `nsrsaphana` script needs to be deployed under the same directory where NetWorker “save” is, such as `/usr/sbin` on Linux systems.

Configure a client resource with NMC (The NMC version used in this document is 8.0.1). Under the “Configuration” tab in NMC, click on “File” menu, and then click “New” to start creating a client resource. Under the “General” tab, fill in the information required for the client resource. In the following example (Figure 1), I provided the following three piece of information (circled in red): Client resource name, “Saveset”, and “Group”.

“Saveset” is special. It specifies the staging area. SAP HANA will backup the data to this staging area during the first stage of the backup, and then NetWorker will backup the backup files from this area. The user needs to make sure that the staging area is big enough to hold the intermediate backup files (at least large enough to hold one backup of SAP HANA database data).
Under “Apps & Modules”, fill in the “Backup command” textbox, with the name of the script: nsrsaphana (Figure 2). You can optionally put a number after nsrsaphana. This number specifies that any files that are older than this number of days will be deleted from the staging area at the end of the backup. If no number is specified, all files are removed immediately after the backup to NetWorker is done. This number affects the size of the staging area: The longer the backup files stay, the more backup files accrue, and the bigger this area has to be.
Figure 8: NMC GUI

Click “OK”, and the client resource for SAP HANA database data backup is created.

2. Perform a scheduled backup with this newly created client resource. After the backup is finished you can see that a saveset containing the backup files from the staging area is created (Figure 3).

Figure 9: NMC GUI
The saveset contains all backup files in the staging area. If the backup files are erased from the staging area every time after the data backup, then the saveset only contains backup files corresponding to one database backup. If the user configures the client resource in such a way that backup files stay in the staging area for a period of time, then the saveset might contain backup files from a few database backups.

For more information on configuring scheduled backups with NMC, refer to EMC NetWorker Administration Guide.

**Redo Logs Backup:**
The backup of redo logs from SAP HANA database to the staging area is initiated periodically by the SAP HANA database server, and not initiated by any backup commands as the case for data backup. For more detail on configuring redo logs backup, refer to section 13.4.2 in SAP HANA Administration Guide. nsrsaphana does not issue any pre-command for redo logs backup. It just backs up the redo log backup files from the staging area to NetWorker, and then perform cleanup on the staging area. It is not recommended to use the same staging area for both data backup and redo logs backup.

To configure a scheduled backup for redo logs is very similar to that of database data backup.

Step 1 in the previous data backup section applies the same to redo logs backup. Step 2 for redo logs backup is almost the same as that for data backup in the previous section, except that the value for the “Backup Command” textbox is different. As shown in Figure 4, the user needs to put “log” argument right after the nsrsaphana, in order for nsrsaphana to know that this is for redo logs backup. Again, the user can optionally put a number at the end of the command to specify that any files that are older than this number of days will be deleted from the staging area at the end of the backup. If no number is specified, then by default the number is 1, which implies that the intermediate redo log backup files are retained for 1 day. This behavior is different from that of data backup. (As mentioned in the previous section, data backup by default uses value 0, i.e., all files are deleted from the staging area at the end of the backup.) The reason for retaining redo log backup files for 1 day is that redo logs backup to disk is not controlled by the nsrsaphana script, and therefore the script could be backing up the log files from the staging area to NetWorker at the same time that the SAP HANA database server is backing up redo logs to the staging area. In this situation, we could backup incomplete redo log backup files. If we erase the backup files immediately after the backup to NetWorker is finished, then we would have incomplete redo logs backed up on NetWorker, and the complete ones would have been deleted.
After the client resource for log backup is configured, it can be scheduled to perform redo log backups to NetWorker. An example saveset produced by SAP HANA redo logs backup is shown in Figure 5.

Same as data backup, every redo logs backup backs up all the files in the staging area. These files may represent many redo log backups from the SAP HANA server, depending on how often
the SAP HAHA performs a redo log backup, and how long nsrsaphana retains the backup files in the staging area.

**Manual Backup with nsrsaphana**

This section describes how to use nsrsaphana to perform manual backup of SAP HANA database data and redo logs.

Firstly, parameters in nsrsaphana need to be provided by following step 1 in section 3.2.

**Data Backup:**
The syntax of using nsrsaphana to perform SAP HANA database data backup:

```
> nsrsaphana [num_of_days] [CLI options for NetWorker save command] <staging area>
```

- **num_of_days:** Specify an integer representing the number of days. Any file in the staging area that is older than this number of days will be deleted at the end of the 1-step backup. If this option is not specified, all files in the staging area will be deleted at the end of the 1-step backup.
- **CLI options for NetWorker save command:** Specify any CLI options for NetWorker save. These options will be passed on to the save command verbatim by nsrsaphana. Refer to EMC NetWorker Command Reference Guide for all CLI options for the save command. Note that the “-N” switch will be ignored, because nsrsaphana creates its own symbolic names for savesets.
- **staging_area:** Specify the directory to which SAP HANA will perform backup, and from which NetWorker will then backup the intermediate backup files.

Example:

```
> nsrsaphana 3 –s bu-saphana1.emc.com /usr/sap/AAA/HDB01/backup/data
```

This command performs SAP HANA database data backup to NetWorker server bu-saphana1.emc.com, with /usr/sap/AAA/HDB01/backup/data as the staging area. Any file in the staging area that is older than 3 days will be deleted at the end of the backup.

**Redo Log Backup:**
The syntax of using nsrsaphana to perform SAP HANA database redo logs backup:

```
> nsrsaphana log [num_of_days] [CLI options for NetWorker save command] <staging area>
```

This is almost the same as that of data backup, except that right after the nsrsaphana command, the user needs to add the “log” argument to tell nsrsaphana that this is a redo logs backup operation. Another difference is that if num_of_days is not provided, then by default it is 1.

Example:

```
> nsrsaphana 5 –s bu-saphana1.emc.com /usr/sap/AAA/HDB01/backup/log
```

This command performs SAP HANA redo logs backup to NetWorker server bu-saphana1.emc.com, with /usr/sap/AAA/HDB01/backup/log as the staging area. Any file in the staging area that is older than 5 days will be deleted at the end of the backup.
With this command-line backup option, the user can create scheduled backup with tools other than NMC, such as cron.

**SAP HANA Database Recovery**

The recovery of SAP HANA Database is performed through SAP HANA Studio graphical user interface. SAP HANA supports the following recovery types: (1). Recover the database to its most recent state, (2). recover the database to a point-in-time, and (3) recover the database to a specific data backup. For (1) and (2), both data backup and redo log backup will be used. For (3), only data backup will be used. (There is another not-commonly-used recovery option which is to recover the database to a log position. Both data backup and redo log backup will be used for this option.)

The recovery process involves two manual steps: 1). Restore the desired backup from NetWorker to the staging area. 2). Restore the backup in the staging area to SAP HANA database server. The script, nsrsaphana, does not perform any recovery operation.

The saveset names produced from the 1-step backup with the script can aid the user in preparing for database recovery by helping the user to identify the following information about the backup: (1) whether the backup contains database data or redo logs, (2) the SAP HANA system that the backup belongs to, (3) the instance in the system that the backup belongs to, and (4) the time corresponding to when the backup was made. We will, in the following paragraphs, expand on how to leverage these benefits, especially the timestamp, for restoring the data backups as well as redo log backups from NetWorker to the staging area, regardless which one of the three types of SAP HANA recovery operation (mentioned in section 1) is being carried out.

The user first needs to figure out what data backup and redo log backups are needed to perform a SAP HANA recovery. The user can use SAP HANA backup catalog or the SAP HANA backup log for this task. We will not elaborate on how to do so, but to ask the user to refer to chapter 13 of SAP HANA Administration Guide on how to identify what backups to use for any type of SAP HANA recovery operation.

Once the user figures out which data and redo log backups are needed, the user then needs to identify which savesets on NetWorker contain the desired data and redo log backups for the SAP HANA database recovery operation, and restores these savesets from NetWorker to the staging area.

Each SAP HANA saveset has the following format:
```
SAPHANA_<Data/Log>_ <SID>_<HANA_Instance>_<Timestamp>_ <Path>
```
<Data/Log> helps to identify whether the backup is a data backup or a logs backup.
<SID> helps to identify the SAP HANA system where the backup came from.
<HANA Instance> helps to identify the instance in the SAP HANA system where the backup came from.
<Timestamp> helps to identify when the 1-step backup operation was started.
<Path> is the staging area where the backup files in the saveset were from.

The user can use NetWorker “mminfo” command to query for the list of savesets stored on the NetWorker server. For details on “mminfo” command, refer to EMC NetWorker Command Reference Guide. For example:

```
> mminfo

volume client date size level name
bu_saphana1.001 bu-saphana1 05/02/2013 1086 MB manual
SAPHANA_Data_AAA_01_2013y_05m_02d_11h_08m_/hanamnt/AAA/HDB01/backup/data/
bu_saphana1.001 bu-saphana1 05/02/2013 2173 MB manual
SAPHANA_Data_AAA_01_2013y_05m_02d_11h_11m_/hanamnt/AAA/HDB01/backup/data/
bu_saphana1.001 bu-saphana1 05/02/2013 3268 MB full
SAPHANA_Data_AAA_01_2013y_05m_02d_13h_02m_/usr/sap/AAA/HDB01/backup/data/
bu_saphana1.001 bu-saphana1 05/01/2013 1022 MB full
SAPHANA_Log_AAA_01_2013y_05m_02d_12h_57m_/usr/sap/AAA/HDB01/backup/log/
bu_saphana1.001 bu-saphana1 05/02/2013 1177 MB full
SAPHANA_Log_AAA_01_2013y_05m_02d_12h_49m_/usr/sap/AAA/HDB01/backup/log/
```

In the above, there are 5 savesets. Three are data backups and two are redo log backups. “AAA” is the SID, and “01” is the instance number. The timestamp and staging area form the last part of the saveset name.

Here is an example on how one can identify the saveset that contains the desired data backup: SAP HANA maintains a history of all backups in a log file called backup.log (usually under directory $DIR_INSTANCE/<SAPLOCALHOST>/trace/). Whenever a backup is started, the start time of the backup as well as the content of backup, among other information, are logged into this file. The user can leverage this history file and their recovery requirement to figure out which data backup is needed for the restore. Once the data backup is identified, the user can use the start time in the backup.log for that backup to cross-reference the corresponding saveset on NetWorker, as the timestamp in the saveset name corresponds closely to the start time in backup.log (there may be a deviation of a few minutes).

Even if backup.log is not available, the user shall be able to figure out which data backup is needed by using the timestamp information in the saveset name, as it corresponds to the start time of the SAP HANA database backup.

Once the saveset that needs to be restored is identified, the user can perform NetWorker filesystem files recovery operation to restore backup files in the saveset to the staging area. For details on how to restore flat files with NetWorker, refer to EMC NetWorker Administration Guide. Here is a simple example on how to restore saveset based on SSID (Save Set ID): 1). Use “mminfo –v” to get the corresponding SSID for the saveset that you want to restore. Let’s say that the SSID is: 12345678
2). Use “recover –S 12345678” to recover the files in the saveset to the directory where they were backed up.

Regarding restoring redo logs, the user needs to identify the saveset(s) that contains the necessary redo logs that will be used to recover the SAP HANA database. For example, if the user wants to perform database recovery to the most recent time, then the redo logs needed are usually the ones generated from the last full database backup till the most recent redo log. Differing from data backup, the timestamp in the redo log saveset name represents when the redo log backup to NetWorker happened, not when the redo logs in the saveset were generated by SAP HANA (Recall that redo logs backup to disk is always initiated by SAP HANA database server, in user-configurable time intervals.). Therefore, the timestamp tells that this saveset contains redo logs generated up to the point-in-time represented by this timestamp. How many days' worth of redo logs a saveset contains depends on how long the user keeps the redo log files in the staging area (via the “num_of_days” argument to nsrsaphana).

If the user cannot decide which savesets contain all necessary redo logs based on just the timestamp information in the saveset name, then the user can use NetWorker “nsrinfo” command to look into the saveset, and get information on each redo log file contained in the saveset. Specifically, the “-V” switch for “nsrinfo” shows the creation time (ctime) of each redo log file. For details on how to use “nsrinfo”, refer to EMC NetWorker Command Reference Guide.

Once the necessary data and redo log backup files are restored from NetWorker to the staging area in the file system, the user can perform the second step, which is to restore and recover the SAP HANA database. This document will not expand on how to perform restore and recovery of SAP HANA database from backup files. Chapter 13 of SAP HANA Administration Guide contains details on how to do so.

**Every NetWorker backup of the staging area is a full backup**
With the current nsrsaphana script, every NetWorker backup (with the save command) of the backup files from the staging area is a full backup, because every backup has a different saveset name, due to the reason that the timestamp in the saveset name changes every time. Note that this only pertains to the second stage of the 1-step backup when NetWorker is backing up the backup files. For the first stage where SAP HANA backs up the database data to the staging area, SAP HANA always performs full whole database data backup.

If the user would like to have incremental NetWorker backups, the user can modify the script such that the timestamp is not a part of the saveset name (taking off $(bk_start_time) from nw_sym_name in the script).

Without the timestamp in the saveset name, the user can use “nsrinfo” command to look into the saveset, and the names of the savefiles in the saveset bear the timestamp representing the start time of the SAP HANA backup (This is only for data backups. For log backups, use “nsrinfo –V” to obtain the redo log creation time).

NetWorker incremental backup gives the functionality that if files in the staging area have been backed up before by NetWorker under the same saveset name, then the backup only includes new files created under the staging area. This does not affect SAP HANA database backup taking place during the first stage of the 1-step backup.
**Simple Disaster Recovery**

Integration with Data Domain Replicator software provides streamlined, network-efficient and cost effective disaster recovery protection by sending only unique, compressed data segments across the WAN to the remote Data Domain system.

Data Domain Replicator software provides automated, fast, and reliable replication of data for disaster recovery (DR), remote office data protection, and multiple site tape consolidation.

**Conclusion**

The backup test results presented here confirm that using an NFS setting on Data Domain storage systems is one of the simplest and most effective methodologies to backup a HANA database. HANA delivers leading performance and scalability while enabling businesses to backup in-memory databases in an aggressive production OLAP environment. The test results should be used only as a general guideline since these results will also vary from implementation to implementation due to a number of factors such as the workload of application datasets.

Backups were done online while the database was running, the impact on HANA system resources were negligible, and users can continue to work normally.
Appendix

One Step Back up With Networker

The following shell script is provided as an example for customers to use EMC NetWorker for their SAP HANA FileSystem Backups. The script allows the customer to do both manual and scheduled one step backup using a staging area.

Shell script: nsrsaphana

```
#!/bin/bash

# Specify the full path to hdbsql, for example:
# hdbsql_cmd=/hanamnt/<SID>/HDB01/exe/hdbsql
#
# Specify the SAP HANA System ID, SID. For example:
# hana_SID=ABC
#
# Specify the SAP HANA instance number. For example:
# hana_instance=01

# Specify the SAP HANA user key for backup. This key
# is created by the hdbuserstore tool. For example:
# hana_key=BACKUPKEY
```

IMPORTANT NOTE:
# FEEDBACK IS ALWAYS WELCOME, HOWEVER, IT SHOULD BE CLEARLY UNDERSTOOD THAT
# USE OF THIS SCRIPT SHOULD ONLY BE MADE AFTER AN EXPERIENCED SAP HANA admin
# CLEARLY UNDERTANDS ALL STEPS IN THIS SCRIPT AND HAS MADE NECESSARY
# MODIFICATIONS AND PERFORMNED SUFFICIENT TESTING TO BE ASSURED OF ITS
# PROPER OPERATION IN THEIR PARTICULAR ENVIRONMENT

```
# If no user key created, then specify the username and password. However, this is not a preferred authentication method, as SAP HANA user key has better security feature. See section 13.9.3 in SAP HANA Administration Guide for detail.

# hana_username=backupadmin
# hana_password=abc123

#################################################
# hana_key=
# hana_username=
# hana_password=

#################################################
# User provided value verification

if [ "$hdbsql_cmd" = "" ]
then
  echo "Path to hdbsql is not specified."
  exit 1
fi

if [ "$hana_SID" = "" ]
then
  echo "SAP HANA SID is not specified."
  exit 1
fi

if [ "$hana_instance" = "" ]
then
  echo "SAP HANA instance number is not specified."
  exit 1
fi

if [ "$hana_key" = "" ]
then
  echo "SAP HANA user key is not specified."
  echo "Use username/password for authentication instead."
  if [ "$hana_username" = "" ]
    then
      echo "username for SAP HANA backup is not specified."
exit 1
elif [ "$hana_password" == "" ]
then
    echo "password for user \"${hana_username}\" is not specified."
exit 1
fi
fi

################################################
# pre-cmd section
################################################

# Determine if the operation is data backup or log backup
bk_type=
if [ "$1" = "log" ]
then
    bk_type=Log
    shift
else
    bk_type=Data
fi

# Determine what is the data preservation period (in days).
# This is used during post-cmd cleanup operation.
time_to_live=
if [[ "$1" =~ ^[0-9]+$ ]]
then
    time_to_live=$1
    shift
fi

# Construct the credential for backup
if [ "$hana_key" == "" ]
then
    hana_credential="-U $hana_key"
else
    hana_credential="-i $hana_instance -u $hana_username -p $hana_password"
fi

# Make sure that the path to the
# staging area ends with "/".
# This is needed later for constructing
# the backup command.
eval savepath='$'{$#}
if [ ${savepath:(-1)} != "/" ]
then
    savepath=${savepath}/
fi

# Construct the backup name for HANA backup.
bk_start_time=$(date +%Yy_%mm_%dd_%Hh_%Mm)
data_bk_cmd="backup data using file ("${savepath}${bk_start_time}")"

# Construct the SAP HANA backup command.
run_bk="$hdbsql_cmd $hana_credential $data_bk_cmd"

#NW save section
nwsave_cmd=save

# Just in case the environment cannot find save.
# nsrsaphana needs to be in the same directory
# where save is (this is a core NW requirement).
case $0 in
    /* ) PATH=$PATH:/bin:/sbin:/usr/bin/:$(dirname $0)
    ;;
    * ) PATH=$PATH:/bin:/sbin:/usr/bin:/usr/sbin
    ;;
esac
export PATH

# Construct the save options and
# skip symbolic name passed to save because
# we already have one.
save_options=
while [ "$1" != "" ]
do
    if [ "$1" = "-N" ]
    then
        shift; shift
        continue
    fi

    save_options="${save_options} "$1"
    shift

if [ 

# Refresh the backup start time for NW save.
# This string will be used for saveset construction.
# NOTE: We decided not to refresh the backup start time.
# Therefore, the timestamps on the backup name and on
# the saveset name are the same. This will help the user
# to identify which saveset to use for restore.
#
# bk_start_time=$(date +%Yy_%mm_%dd_%Hh_%Mm)

# Construct the symbolic name for the NW saveset
nw_sym_name="SAPHANA_${bk_type}_${hana_SID}_${hana_instance}_${bk_start_time}_${savepath}"

nw_sym_name_opt="-N $nw_sym_name"

run_save="$nwsave_cmd $nw_sym_name_opt $save_options"

# start SAP HANA backup
if [ "$bk_type" != "Log" ]
  then
echo $run_bk
  sleep 70
  $run_bk

  if [ "$?" != "0" ]
    then
echo "SAP HANA database backup failed."
    exit 1
  fi
fi

# start NetWorker save
echo $run_save
$run_save

if [ "$?" != "0" ]
  then
echo "NetWorker backup of "$savepath" failed."
    exit 1
fi

########################################################################
# post-cmd section

# Remove data in the staging area, based on user's
# "time_to_live" specification.
if [ "$time_to_live" = "" ] || [ "$time_to_live" -lt 1 ]
then
    if [ "$bk_type" != "Log" ]
    then
        rm -rf "${savepath}"*
    else
        # For log backup, we always save them for at least 1 day.
        find "${savepath}" -mtime +1 -print -exec rm -rf {} \;
    fi
else
    find "${savepath}" -mtime +${time_to_live} -print -exec rm -rf {} \;
fi

exit 0

END

END

END