Business Continuity requires that the operation of business critical systems remain highly available at all times, even in the presence of failures. This paper discusses the functionality of SAP HANA in support of High Availability and Disaster Recovery.
1 Introduction

SAP HANA

SAP HANA™ is an innovative in-memory database and data management platform, specifically developed to take full advantage of the capabilities provided by modern hardware to increase application performance. By keeping all relevant data in main memory (RAM), data processing operations are significantly accelerated.

The key performance indicators of SAP HANA appeal to many of our customers, and hundreds of deployments are in progress. SAP HANA has become the fastest growing product in SAP’s 40-year history.

About this Document

Loss of business critical system resources and services, like SAP HANA, translate directly into lost revenue. The goal therefore is Business Continuity, using systems designed for continuous operation even in the presence of inevitable failures. Mission critical systems require High Availability, this is no longer optional.

SAP HANA is fully designed for High Availability, supporting a broad range of recovery scenarios from various faults, from simple software errors, to disasters that decommission an entire site.

This paper describes SAP HANA’s High Availability support for Fault and Disaster Recovery. 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 paper, and may therefore require additional consultations.

2 What is High Availability?

Availability, the measure of a system’s operational continuity, is expressed as a percentage of time, inversely proportional to downtime. For example, if a given system is designed to be available for 99.9% of the time (sometimes called "three nines"), its downtime per year must be less than 0.1%, or 9 hours.

Downtime is the consequence of outages, which may be intentional (e.g. for system upgrades) or caused by unplanned faults. A fault can be due to equipment malfunction, software or network failures, or due to a major disaster such as a fire, a regional power loss or a construction accident, which may decommission the entire data-center.

High Availability is a set of techniques, engineering practices and design principles for Business Continuity. This is achieved by eliminating single points of failure (fault tolerance), and providing the ability to rapidly resume operations after a system outage with minimal business loss (fault resilience).

Fault Recovery is the process of recovering and resuming operations after an outage due to a fault. Disaster Recovery is the process of recovering operations after an outage due to a prolonged datacenter or site failure. Preparing for disasters may require backing up data across longer distances, and may thus be more complex and costly.

Recovery - Key Performance Indicators

Customers commonly use two key measures to specify the recovery parameters of a system following an outage: the Recovery Period Objective (RPO) and the Recovery Time Objective (RTO). The RPO and RTO of a system are illustrated below:

- The RPO is the maximal permissible period of time during which operational data may be lost without ability to recover (time between the last backup and the crash)
- The RTO is the maximal permissible time it takes to recover the system, so that its operations can resume.
3 Eliminating Single Points of Failure

The key to achieving fault tolerance is to eliminate single points of failure by introducing redundancy. SAP HANA Appliance vendors deliver several levels of redundancy to avoid outage due to component failure, which are briefly discussed here. Generally speaking, these techniques are "transparent" to SAP HANA’s operation, but they form a crucial line of defense against avoidable system outage, and therefore greatly contribute to Business Continuity.

Hardware Redundancy

SAP HANA appliance hardware vendors design multiple layers of redundant hardware components and subsystems. These include redundant and hot-swappable power supply units (PSUs), fans, network interface cards and enterprise-grade error-correcting protected memories. These subsystems are designed such that the redundant component can sustain the operation of the system if the other component fails.

Particularly critical is the storage system. Enterprise-grade storage systems combine multiple physical drives into logical units, with built-in standard (RAID) techniques for redundancy and error recovery. These include mirroring, the writing of the same data to two different drives in parallel, and parity, extra bits written to allow the detection and automatic correction of errors.

Network Redundancy

Redundant networks, network equipment and network connectivity is required to avoid network failures from affecting system availability. This is typically accomplished by deploying a completely redundant switch topology, using the Spanning Tree Protocol to avoid loops. Routers can be configured with the Hot Standby Router Protocol (HSRP) for automatic failover. BGP is commonly used to manage dual WAN connections.

Data Center Redundancy

Data centers that host SAP HANA solutions are equipped with Uninterrupted Power Supply (UPS) and backup power generators, redundant cooling systems and multi-sourced providers of network connectivity and electricity, achieving operational availability in the presence of individual failures, and significantly reducing the probability of a business-impacting outage.

Some enterprises operate fully duplicated data centers, providing a high level of disaster tolerance.

---

1 The SAP HANA software itself is a single point of failure, as it can cease to operate due to software errors or extreme out-of-memory situations. Fault Recovery support is discussed in the next section.
2 An example of high availability hardware design can be found here: [http://www.redbooks.ibm.com/redpapers/pdfs/redp4664.pdf](http://www.redbooks.ibm.com/redpapers/pdfs/redp4664.pdf)
3 Read further: [http://download.intel.com/support/motherboards/server/sb/enterprise_class_versus_desktop_class_hard_drives_.pdf](http://download.intel.com/support/motherboards/server/sb/enterprise_class_versus_desktop_class_hard_drives_.pdf)
4 SAP HANA High Availability Support

As an in-memory database, SAP HANA must not only concern itself with maintaining the reliability of its data in the event of failures, but also with resuming operations with most of that data loaded back in memory as quickly as possible.

The following figure shows the phases of High Availability. The first phase is readiness, being prepared for the inevitable fault. During this time, data is backed up and standby systems are ready to take over. A fault must be detected, either automatically or administratively (to avoid false positives), and a recovery process is put in action. Finally, the fault must be repaired, and the system may need to be reverted to the original configuration (failed back), to be ready again for the next fault.

Different RPO/RTO values can be associated with different kinds of faults. Business critical systems are expected to operate with an RPO of zero data loss in the case of local faults, and often even in the case of a disaster. The challenges of disaster recovery are different from locally recoverable faults; to achieve zero RPO and low RTO, data must be replicated synchronously over longer distances, which impacts regular system performance and may require more expensive standby and failover solutions.

All of this leads to tradeoff decisions around the attributes of fault recovery functionality, cost and complexity. SAP accordingly offers complementary design options, including three levels of Disaster Recovery support and two automatic Fault Recovery support features, summarized in the following table and further discussed in the sections below.

<table>
<thead>
<tr>
<th>DISASTER RECOVERY SUPPORT</th>
<th>Cost</th>
<th>RPO</th>
<th>RTO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Backups</td>
<td>$</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>2. Storage Replication</td>
<td>$$</td>
<td>~0</td>
<td>high</td>
</tr>
<tr>
<td>3. System Replication</td>
<td>$$$$</td>
<td>0</td>
<td>low</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FAULT RECOVERY SUPPORT</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Service Auto-Restart</td>
<td>-</td>
<td>0</td>
<td>med</td>
</tr>
<tr>
<td>2. Host Auto-Failover</td>
<td>$$</td>
<td>0</td>
<td>med</td>
</tr>
</tbody>
</table>

**Backups**

SAP HANA uses in-memory technology, but of course it fully persists any transaction that changes the data, such as row insertions, deletions and updates, so it can resume from a power-outage without loss of data. SAP HANA persists two types of data to storage: transaction logs, and data changes in the form of savepoints.

A transaction redo log is used to record a change. To make a transaction durable, it is not required to persist the complete data when the transaction is committed; instead it is sufficient to persist the redo log. Upon an outage, the most recent consistent state of the database can be restored by replaying the changes recorded in the log, redoing completed transactions and rolling back incomplete ones.

A savepoint is a periodic point in time, when all the changed data is written to storage, in the form of pages. One goal of performing savepoints is to speed up restart: when starting up the system, logs need not be processed from the beginning, but only from the last savepoint position. Savepoints are coordinated across all processes (called SAP HANA services) and instances of the database to ensure transaction consistency. By default, savepoints are performed every five minutes, but this can be configured.

Savepoints normally overwrite older savepoints, but it is possible to freeze a savepoint for future use; this is called a snapshot. Snapshots can be replicated in the form of full data backups, which can be used to restore a database to a specific point in time. This can be useful in the event of data corruption, for instance. In addition to data backups, smaller periodic log backups ensure the ability to recover from fatal storage faults with minimal loss of data.
Local Persistence and Backups

The above figure shows the savepoints, saved to local storage, and the additional backups, saved to backup storage. Local recovery from the crash uses the latest savepoint, and then replays the last logs, to recover the database without any data loss. If the local storage was corrupted by the crash, it is still possible to recover the database from the data and log backups, possibly with loss of some data.

Regularly shipping backups to a remote location over a network or via couriers can be a simple and relatively inexpensive way to prepare for a disaster. Depending on the frequency and shipping method, this approach may have an RPO of hours to days.

Backup Shipping

Storage Replication

One drawback of backups is the potential loss of data from the time of the last backup to the time of the failure. A preferred solution therefore, is to provide continuous replication of all persisted data. Several SAP HANA hardware partners offer a storage-level replication solution, which delivers a backup of the volumes or file-system to a remote, networked storage system. In some of these vendor-specific solutions, which are certified by SAP, the SAP HANA transaction only completes when the locally persisted transaction log has been replicated remotely. This is called synchronous storage replication. Synchronous storage replication can be used only where the distance between the primary and backup site is up to 100 kilometers (no or few hops, with no more than ~5 μsec latency per kilometer), allowing for sub-millisecond round-trip latencies.

Due to its continuous nature, storage replication (sometimes also called remote storage mirroring) offers a more attractive RPO than backups, but this solution of course requires a reliable, high bandwidth and low latency connection between the primary site and the secondary site.
**System Replication**

**SAP HANA System Replication** is a new SAP HANA feature, compatible with all SAP HANA hardware partner solutions. System replication employs an "N+N" approach, with a secondary standby SAP HANA system that is identically configured as the active, primary system. Each service and instance of the primary SAP HANA system communicates pair-wise with a counterpart in the secondary system.

The secondary system can be located near the primary system to serve as a rapid failover solution for planned downtime, or to handle storage corruption or other local faults, or, it can be installed in a remote site to be used in a disaster recovery scenario. Like Storage Replication, this Disaster Recovery option requires a reliable link between the primary and secondary sites.

The instances in the secondary system operate in live replication mode. In this mode, all secondary system services constantly communicate with their primary counterparts, replicate and persist data and logs, and load data to memory. The main difference is that the secondary system does not accept requests or queries. Here is how this works. When the secondary system is brought up in live replication mode, each service component establishes a connection with its counterpart, and requests a snapshot of the data in the primary system. From then on, all logged changes in the primary system are replicated. Whenever logs are persisted in the primary system, they are also sent to the secondary system. A transaction in the primary system is not committed until the logs are replicated. What this means in detail, can be configured by choosing one of the log replication options:

- **Synchronous**: The primary system waits with committing the transaction until it receives a reply that the log is persisted in the secondary system. This mode guarantees immediate consistency between both systems, at a cost of delaying the transaction by the time for data transmission and persisting in the secondary system.

- **Synchronous in-memory**: The primary system commits the transaction after it receives a reply that the log was received by the secondary system, but before it was persisted. The transaction delay in the primary system is shorter, because it only includes the data transmission time.

Note: asynchronous replication, where the primary system commits the transaction after sending the log without waiting for a response, will be supported in a subsequent SAP HANA release. This mode is useful when the secondary site is hundreds of kilometers away from the primary site.

If the connection to the secondary system is lost, or the secondary system crashes, the primary system (after a brief, configurable, timeout) will resume operations without the backup protection.

Currently, the secondary system persists, but does not immediately replay the received log. To avoid an ever growing list of logs, incremental data snapshots are transmitted asynchronously from time to time from the primary system to the secondary system. If the secondary system has to take over, only that part of the log needs to be replayed that represents changes that were made after the most recent data snapshot.

In addition to snapshots, the primary system also transfers status information regarding which table columns are currently loaded into memory. The secondary system correspondingly preloads these columns.

In the event of a failure that justifies full system failover, an administrator instructs the secondary system to switch from live replication mode to full operation. The secondary system, which already preloaded the

---

4 As a result, the primary system and secondary system might get out of sync. Such a situation is detected by the secondary system when it resumes, reestablishes the connection, and receives the next set of log entries. In such case, the secondary system requests a data backup delta based on which the log replication can be restarted.
same column data as the primary system, becomes the primary system by merely replaying the last transaction logs, and then starts to accept queries\(^5\).

In future versions of SAP HANA, the log entries will be replayed in the secondary system continuously, immediately after they have been received. This means that the secondary system can take over with virtually no delay, if the primary system fails.

One topic that requires some attention is how to handle database clients that were configured to reach the primary system, and need to be “diverted” to the secondary system after failover. The most common technique is to use a DNS failover configuration. Many DNS products support failover configuration by using short (few minutes or less) TTL response fields, and can be set up with watchdog functionality and automatically triggered switchover. Another simple and attractive technique that can be considered is the use of Virtual IP Addresses (VIP)\(^6\). The precise implementation may differ from customer to customer, typically depending on network and cluster management capabilities.

### Service Auto-Restart

In the event of a software failure (or an intentional intervention by an administrator), that disables one of the configured SAP HANA services (Index Server, Name Server, etc.), the service will be restarted by the SAP HANA Service Auto-Restart watchdog function, which automatically detects the failure and restarts the stopped service process. Upon restart, the service loads data into memory and resumes its function. While all data remains safe (RPO=0), the service recovery takes some time.

### Host Auto-Failover

**Host Auto-Failover** is a local "N+m" (\(m\) is often 1) Fault Recovery solution that can be used as a supplemental or alternative measure to the system replication solution described earlier. One (or more) standby hosts are added to a SAP HANA system, and configured to work in standby mode. As long as they are in standby mode the databases on these hosts do not contain any data and do not accept requests or queries.

When an active (worker) host fails, a standby host automatically takes its place. Since the standby host may take over operation from any of the primary hosts, it needs shared access to all the database volumes. This can be accomplished by a shared, networked storage server, by using a distributed file system, or with vendor-specific solutions that use a SAP HANA programmatic interface to dynamically detach and attach (mount) networked storage (e.g. using block storage via Fiber Channel) upon failover.

---

\(^5\) When the original system can be restored to service, it can be configured as the new secondary system, or, reverted to the original configuration by “falling back”.

\(^6\) E.g. see here: [http://scale-out-blog.blogspot.de/2011/01/virtual-ip-addresses-and-their.html](http://scale-out-blog.blogspot.de/2011/01/virtual-ip-addresses-and-their.html)
Once repaired, the failed host can rejoin the system as the new standby host, to reestablish the failure recovery capability.

In support of host auto-failover, database clients can be configured with the connection information of multiple hosts, optionally including the standby host. The client connection code (ODBC, JDBC, etc.) will try to connect to one of these, and upon successful connection receives the updated connection configurations. This ensures that clients can continue to reach the SAP HANA database, even after failover. Note that this solution does not apply to clients of the SAP HANA web services. DNS-based, or Virtual IP Address-based redirection as part of an overall network/cluster management solution can be deployed to handle these cases as well.

5 Design for High Availability

The following table summarizes the main advantages and limitations of the SAP HANA High Availability support options.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
</table>
| **Backups** | • Allows Disaster Recovery  
• Lowest cost, simplest  
• Supports point-in-time recovery  
• Can also be used to "clone" or copy systems | • RPO of hours or days, depending on frequency of backup and shipping method.  
• In case of disaster, need to acquire and configure secondary system (hours-days)  
• Cold start – longer RTO (~hour)  
• Extra time (up to hours) to load column data and return to full performance |
| **Storage Replication** | • Allows Disaster Recovery  
• RPO=0 with synchronous replication; RPO of a few seconds otherwise  
• Secondary system can be used for other purposes, until needed | • In case of disaster, need to possibly free up, boot up and re-configure secondary system (hours)  
• Cold start – longer RTO (~hour)  
• Not yet offered by all SAP HANA hardware partners  
• Extra time (up to hours) to return to full performance  
• Requires networked storage systems and efficient inter-site link  
• Synchronous replication only supports distances of up to 100 kilometers  
• Doesn’t protect against storage corruption  
• More bandwidth wasteful than System Replication |
| **System Replication** | • Allows Disaster Recovery, and can be used as main HA failover for planned downtime or failures  
• RPO=0 (synchronous)  
• RTO of only a few minutes  
• Full performance right after failover  
• Compatible with all partner solutions  
• Supports single-host systems with local storage, no need for external network storage appliances | • Requires dedicated live standby system and efficient inter-site link  
• Synchronous replication supports distances of up to 100 kilometers  
• Requires a custom solution for client redirection upon failover (e.g. DNS or Virtual IP address based) |
| **Host Auto-Failover** | • Can be used to complement System Replication or by itself  
• Automatic detection and failover | • Requires access to database storage by the standby host (shared network storage or other partner-specific solution) |

In addition to the aforementioned SAP HANA High Availability options, one other approach deserves to be mentioned, for analytic "data mart" applications where the data in SAP HANA is the result of using SAP Landscape Transformation (SLT) replication from another data source. In such a situation, High Availability through redundancy can be achieved by setting up concurrent SLT replication streams from the common data-source to two separate SAP HANA systems. Both systems can actively operate independently; in the case of a failure or disaster, the other system remains available.
Planning for Failure

Failures are inevitable. Planning a comprehensive High Availability solution for SAP HANA requires an evaluation of the impact of potential failures, the company’s tolerance and requirements for different RPO and RTO values in the presence of common transient local failures vs. extremely rare disasters, and an understanding of the benefits and cost of the different alternatives offered.

To recap, here is a brief summary of main faults and how SAP HANA addresses them:

<table>
<thead>
<tr>
<th>Fault</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service down (software fault)</td>
<td>• Service Auto-Restart. System Replication can also be used to fail over.</td>
</tr>
<tr>
<td>Power outage</td>
<td>• Persistence of savepoints and transaction logs guarantees recovery without data loss</td>
</tr>
<tr>
<td>Host crash (hardware fault)</td>
<td>• Host Auto-Failover. Alternatively, System Replication can be used to fail over.</td>
</tr>
<tr>
<td>Storage or Data Corruption</td>
<td>• Backups and snapshots allow point-in-time recovery.</td>
</tr>
<tr>
<td>Data center out (disaster)</td>
<td>• System Replication supports rapid resumption of operation. Alternatively, Storage Replication or Backups can be used to bring up the system in an alternate datacenter</td>
</tr>
</tbody>
</table>

Besides the high-level consideration of RPO/RTO in the different scenarios, other aspects will need to be evaluated as well: the size of the system and database, the frequency and size of the logs and data files that need to be replicated, the bandwidth availability, reliability and latency of the links between the systems, the nature of the landscape management and availability solutions used for other non-SAP HANA systems, and other considerations.

Small RTO requirements lead to the preferred system replication solution, which can also be used for rapid failover in case of planned and unplanned outages. Tradeoffs may lead to other alternatives. The following decision tree summarizes the main design choices:

Realistically, the above decision process will be further influenced by considerations like timelines, costs, budgets and customer paradigm-preferences which are outside the scope of this short paper.
6 In Summary

SAP HANA supports a comprehensive range of High Availability options, designed to satisfy tradeoffs between demanding High Availability and Disaster Recovery requirements, while also considering cost and complexity.

In particular, the SAP HANA System Replication solution supports an RPO of zero seconds, and an RTO measured in minutes, and is SAP's recommended configuration for addressing SAP HANA outage reduction due to planned maintenance, faults and disasters.


Glossary

Industry Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fault</td>
<td>A failure of a system or one of its components / sub-systems (hardware, network, software...)</td>
</tr>
<tr>
<td>Disaster</td>
<td>Major fault: the failure of an entire data center/site</td>
</tr>
<tr>
<td>Outage</td>
<td>A system’s inability to operate (due to failure or planned downtime)</td>
</tr>
<tr>
<td>Availability</td>
<td>The measure of a system’s operational continuity, expressed as a percentage of time</td>
</tr>
<tr>
<td>Downtime</td>
<td>Inverse of availability: the duration of time that a system is not operational</td>
</tr>
<tr>
<td>High Availability (HA)</td>
<td>A framework of design principles, techniques and best practices to reduce downtime</td>
</tr>
<tr>
<td>Fault Recovery (FR)</td>
<td>Recovery of system operations after outage due to a local fault</td>
</tr>
<tr>
<td>Disaster Recovery (DR)</td>
<td>Recovery of system operations after outage due to a disaster</td>
</tr>
<tr>
<td>Failover</td>
<td>Switching to a backup (standby) system/host, upon failure of the primary system/host</td>
</tr>
<tr>
<td>Failback</td>
<td>Process of restoring a system to its original state</td>
</tr>
<tr>
<td>Recovery Point Objective (RPO)</td>
<td>the maximal permissible period of time during which operational data may be lost without ability to recover (time between the last backup and the crash)</td>
</tr>
<tr>
<td>Recovery Time Objective (RTO)</td>
<td>The maximal permissible time it takes to recover the system, so that its operations can resume</td>
</tr>
</tbody>
</table>

SAP HANA Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP HANA System</td>
<td>A SAP HANA system is identified by a system id (SID). It is perceived as one unit from the perspective of the administrator, who can install, update, start up, shut down, or backup the system as a whole. A distributed SAP HANA system is a system which is installed on more than one host. The collection of elements of the system on each host are referred to as an instance.</td>
</tr>
<tr>
<td>SAP HANA Service</td>
<td>A SAP HANA service is an independent functional component of a SAP HANA System, such as the Index Server, the Name Server, etc. They appear as separate processes from an Operating System perspective.</td>
</tr>
</tbody>
</table>

Disaster Recovery Support

<table>
<thead>
<tr>
<th>Backup</th>
<th>Periodic saving of database copies in safe place</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Replication</td>
<td>Continuous replication (mirroring) between primary storage and backup storage over a network (may be synchronous)</td>
</tr>
<tr>
<td>System Replication</td>
<td>Continuous (currently: synchronous) update of secondary system by primary system, including in-memory table loading</td>
</tr>
</tbody>
</table>

Fault Recovery Support

<table>
<thead>
<tr>
<th>Service Auto-Restart</th>
<th>Automatic restart of stopped services on host (watchdog)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host Auto-Failover</td>
<td>Automatic failover from crashed host to standby host in the same system</td>
</tr>
</tbody>
</table>