

SAP on Google Cloud: Backup strategies and solutions

Overview



Contents

About this document	4
Introduction	4
Services for storage and backup	5
Overview of storage options	5
Persistent disks	6
Cloud Storage buckets	6
Filestore	6
Elastifile	6
NetApp Cloud Volumes	6
Backup mechanisms	7
Persistent disk snapshots	7
Application-specific agents	7
Back up with managed NFS	7
Back up with partner solutions	8
Backup strategies and options	9
Persistent disk snapshots	10
Shared file storage	11
HANA Backint agent for Cloud Storage	12
Third-party network storage	12
Third-party backup agents and managed services	13
SAP HANA data snapshot	13
SAP HANA stop/start snapshot of secondary HANA instance	14
Snapshot and disk deallocation	14
Database backup strategies with other RDBMS	15
Automation	16
Comparison summary	17
Recommendations	18
Summary	19
Further reading	19

About this document

This document is part of a series about working with SAP on Google Cloud. The series includes the following documents:

- [High availability](#)
- [Migration strategies](#)
- Backup strategies and solutions (this document)
- [Disaster-recovery strategies](#)

Introduction

Backup and storage design is a key part of overall business continuity planning for SAP systems. Backups are intended to protect against data loss by making copies of data (replication). You can lose data due to physical reasons (for example, data center failure or a disk failure) or logical reasons (for example, accidental deletion of data). The strategies and solutions discussed in this document address protection against data loss due to physical reasons. To prevent data loss due to logical reasons, you must do extra steps and set up mechanisms like access control, which is out of scope for this document.

The *Recovery Point Objective (RPO)* and *Recovery Time Objective (RTO)* [disaster recovery requirements](#) of an SAP deployment directly influence the design of a backup strategy and the choice of associated Google Cloud and third-party solutions. However, fulfillment of RPO/RTO requirements also involves a broader disaster recovery architecture definition. Within that architecture definition, disaster recovery requirements, like the need to replicate backups across regions for RTO purposes, can influence the choice of backup strategy.

The key consideration for your backup strategy and associated backup solution is the trade-off between the overall cost of backup operations and the cost of time lost due to lost data. When reviewing the overall cost of backup operations, consider the following technical capabilities of cloud-based backup options:

- Availability of storage locations
- Time for backup and restore
- [Full, incremental](#), and [differential](#) backup capabilities
- Tiering of data to storage classes based on frequency of access
- [Compression and backup deduplication](#)

Services for storage and backup

The most relevant resources for SAP workloads on Google Cloud are virtual machines (VMs) on [Compute Engine](#) and [Persistent Disk](#). A typical SAP workload consists of a set of virtual machines that run database servers and application servers. By default, each of these VMs has a boot disk for the operating system (OS) and the application (either database server or SAP NetWeaver application server) relevant binaries. The VMs have more disks (or volumes) for application files. We recommend allocating a separate disk for all files that are not part of the OS disk. This configuration means that the OS volume is replaceable, and helps to ensure that a system administrator can easily relocate the instance to another zone or region during failures. The following table summarizes the primary and secondary storage options for SAP systems in Google Cloud. For more information on various storage options, see [overview of storage options](#) in Google Cloud.

SAP system component / layer	Primary storage options	Secondary storage options (backup locations)
Instance boot disk and application binaries	Persistent Disk	Persistent Disk snapshots
SAP HANA database	Persistent Disk	Persistent Disk snapshots, Cloud Storage buckets
Any database (non-HANA)	Persistent Disk	Persistent Disk snapshots, Cloud Storage buckets
SAP NetWeaver Application Server	Persistent Disk , Managed NFS, like Elastifile or Cloud Volumes (NetApp)	Filestore , Cloud Storage buckets , Persistent Disk snapshots (for managed NFS)

Overview of storage options

For SAP applications, the primary storage option in Google Cloud is Persistent Disk. To help ensure that the SAP environment is secure and always available you can also configure a cloud-based SAP environment with a Network File Storage (NFS) that you can secure across [Google Cloud zones](#) (for example, for high availability at the application layer), and regions (disaster recovery). NFS storage options are available from services like [Cloud Filestore](#) (zonal resource), [Elastifile](#), and partner solutions like [NetApp Cloud Volumes Service](#).

Persistent disks

If your SAP system is deployed on Google Cloud, we recommend using [persistent disks](#) for storage. You can use standard [hard disk drives \(HDD\)](#) or [solid-state drives \(SSD\)](#). Persistent disks can be [zonal](#) or [regional](#), which have different throughput and latency characteristics. Zonal persistent disks are the primary block-level storage for storing SAP application and database files because of the latency and throughput considerations required by the SAP two-tier stack. Particularly for SAP HANA, persistent disk SSDs are currently the only certified storage solution for data and log volumes. You can use regional persistent disks to give you block-level replication for zonal resilience with your SAP deployments, but there are [limitations](#).

Cloud Storage buckets

[Cloud Storage buckets](#) are the most flexible, scalable, and durable storage option in Google Cloud. You can retrieve any amount of data around the world from Cloud Storage buckets at any time for a range of scenarios, including storing data backups for archival and disaster recovery. Buckets can be [regional, dual region, or multi-regional](#). Cloud Storage buckets have different [storage classes](#) designed for varying frequency of access, but all with a single API access, with access latency, and with identical eleven 9's of durability. You can manage the lifecycle of data—for example, TTL—in a storage bucket with [rule-based configuration](#) of a bucket's lifecycle.

Filestore

[Filestore](#) is a zonal resource that provides a fully managed NFS endpoint. It is intended for large-scale, shared file storage. Filestore is useful for the following:

- SAP NetWeaver configuration, transport, and interface file shares in a distributed NetWeaver setup in the same zone.
- Shared file volumes for SAP HANA scale-out deployments, because these volumes are zonal only (due to latency requirements for the HANA scale-out nodes).

Elastifile

Elastifile offers scalable, cloud-native enterprise file storage. Elastifile works for multi-zones and multi-regions in Google Cloud, which addresses the [HA/DR requirements of SAP installations](#).

NetApp Cloud Volumes

[Cloud Volumes Service](#) is a solution from Google Cloud partner [NetApp](#) for file storage available in select Google Cloud regions. Cloud Volumes Service for Google Cloud is a fully managed file storage service that provides a robust [set of capabilities for migrating or deploying new SAP workloads](#). Currently, Cloud Volumes Service provides a multiple-zone availability feature, but not multi-region disaster recovery.

Backup mechanisms

One option for creating backups is taking a snapshot of a disk, which is a block-level capture of the state of a disk. Another option is using software-managed systems like database-specific backup tools (for example, SAP HANA Studio tools for SAP HANA databases) that generate specific files as backup. Depending on the storage service you choose, you can set up snapshot backups manually or on a schedule using Cloud Console, Cloud SDK, or API-based access to snapshots. If you use a shared file storage solution like Cloud Volumes, you can use snapshots for backup in the management console of the shared file solution. You can also configure database backup tools and interfaces at the application level.

Persistent disk snapshots

The primary mechanism for backup of persistent disks in Google Cloud is by the periodic [creation of persistent disk snapshots](#). Persistent disk snapshots are a straight-forward, software-independent, and cost-effective way to back up data. Snapshots are global resources and the storage location is different from Cloud Storage. By default, the storage location for snapshots is the multi-region that is geographically closest to the source persistent disk. For example, if your persistent disk is stored in the `us-central1` region, your snapshot is stored in the `us` multi-region by default. However, if the location of the persistent disk is not multi-region, then if you create or recover snapshots and recovery, you might incur [network costs](#). For example, if your persistent disk is stored in the `australia-southeast1` region, the nearest multi-region is the `asia` region. You can define the storage location for snapshots as regional. You can also schedule snapshots depending on the frequencies dictated by RPO/RTO requirements. For more information, see [Creating persistent disk snapshots](#).

Application-specific agents

In addition to persistent disk snapshots, you can also use Cloud Storage directly for backups and recoveries of applications like the [Cloud Storage Backint agent for SAP HANA](#). The Backint agent for SAP HANA is certified by SAP and supports multistreaming for backing up large datasets to increase throughput. The usage of the Backint agent lets you back up directly from the HANA database to a Cloud Storage bucket, bypassing the persistent disk. With this configuration, you avoid the costs associated with a secondary disk and benefit from incurring no egress charge to offload the backup data into a separate, highly durable storage layer.

Back up with managed NFS

Secondary storage options come with bundled backup options that typically rely on snapshots and storing these snapshots in Cloud Storage, but the managed service determines the setup. Typically, you set up the backup settings of these file system services during deployment, but you can configure them further because some, like NetApp, are integrated with the Google Cloud Console.

Back up with partner solutions

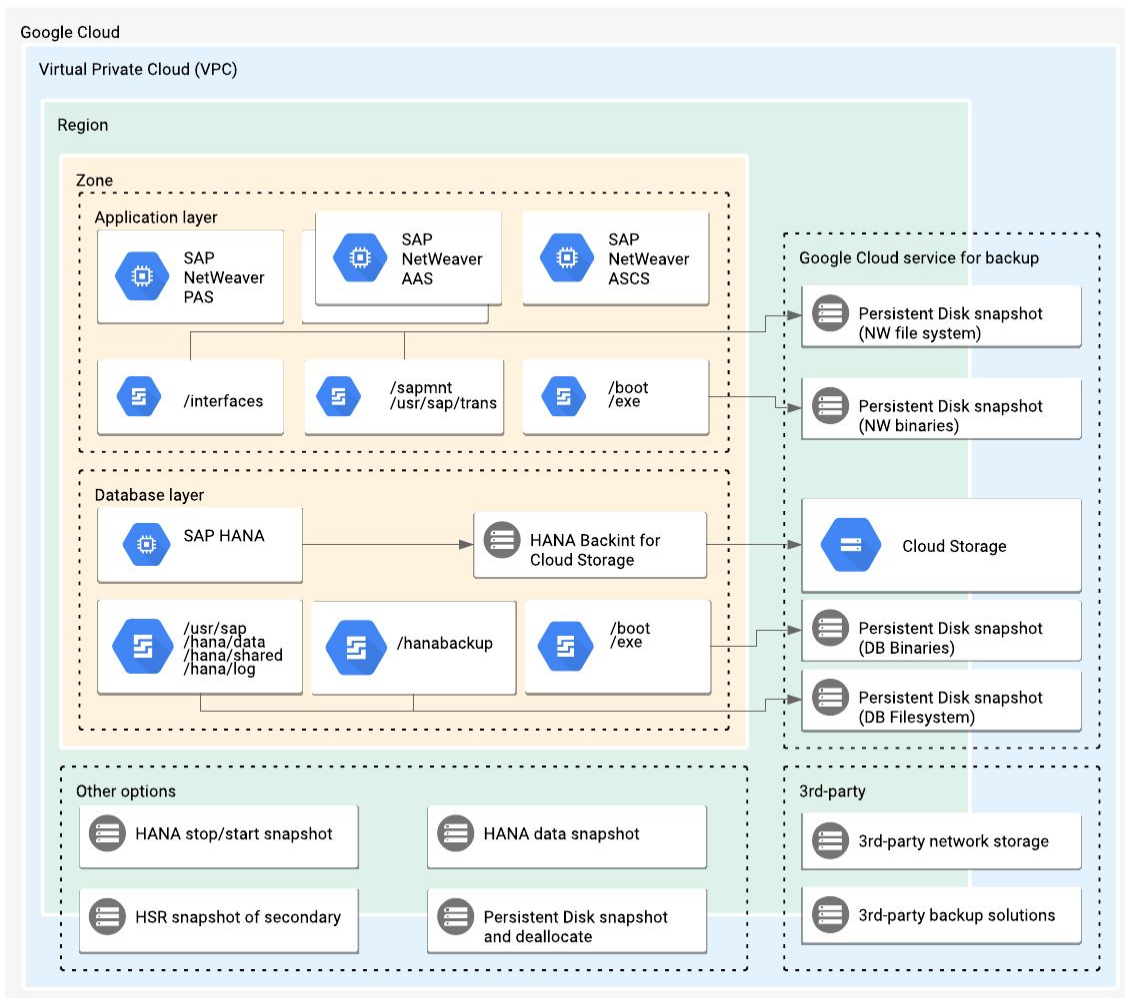
Google Cloud also works with industry-leading partners that offer state-of-the-art solutions for backup, recovery, and migration of large-scale SAP workloads with specific RTO/RPO, data retention, and compliance requirements. Third-party backup solutions might be the right choice for you for the following reasons:

- They're an existing toolset that is part of your landscape and operational process.
- They provide technical functionality like deduplication or encryption technologies.
- They satisfy a compliance requirement needed in your particular industry, such as long-term data retention.

Backup strategies and options

The core principle for backup solutions is to segregate backup data copies from the primary storage location. Conventionally, the backup of data in on-premises environments involved storing copies of data to in-house storage devices, such as network-attached storage, storage servers, and tape or disk backups. When workloads migrate to the cloud, the backup solutions also have multiple cloud-based options to choose from. Backups in the cloud are flexible and scalable and can be self-managed by enterprise-internal IT workforces or consumed as managed services from cloud providers or third-party vendors.

The backup of SAP workloads is classified into application layer and database layer options. Application-layer backups involve the backup of the root disk with SAP NetWeaver binaries and other relevant directories in the SAP NetWeaver filesystem. The following diagram gives an overview of all options for SAP workloads (full stack) that run on the SAP HANA database



Persistent disk snapshots

Persistent disk snapshots are fast and cost-effective. You can specify the storage location for snapshots as regional or multi-regional. Compute Engine stores multiple copies of each snapshot across multiple locations with automatic checksums to ensure the integrity of the disk data. By storing copies in multiple locations, these snapshots of resources of the same Google Cloud project are globally addressable (though not globally redundant), thus eliminating the need for extra replication across regions. You can also use [Cloud IAM roles](#) to share snapshots across Google Cloud projects.

One consideration for backing up with persistent disk snapshots is the allocation of separate disks for organizing parts of the file system that belong together and are needed as a unit for recovery. For example, in an SAP HANA database server, you can store the backup folder on a separate persistent disk so you can capture it and recreate the database server independently. For more information, see [Best practices for persistent disk snapshots](#).

To enable full recovery of the application layer from a failed virtual machine or disk, you can schedule persistent disk snapshots to back up root disks containing OS and required binaries. You can also schedule snapshots for volumes containing the following:

- Set of SAP NetWeaver binaries and configuration files.
- Common files that are used to share (or transport) customer developments or other transports between systems in a single SAP landscape.
- Folders that are required by a highly available system (for example, to restore locking table if an instance fails over).
- Folders containing interfaces to surrounding systems.

For the database layer of an SAP HANA DB installation, you can schedule snapshots for volumes containing the following:

- Application binary directories.
- Database configuration files.
- Directories containing data or database transactions.
- Shared files such as binaries, logs, and configurations. For a multiple-host SAP HANA system, this must be an NFS-shared file system.
- Recent redo logs. For a multiple-host SAP HANA system, a shared NFS volume is required to enable SAP HANA multiple-host cluster functionality.
- Files or file systems containing database, file interfaces, and application configuration backups.

For the database layer, although persistent disk snapshots (snapshot of complete disks) are crash-consistent, they do not ensure file-level or application-level consistency, which is

important for a database recovery. An alternative is to snapshot the HANA backup volume, which contains a local database dump from the database. You can configure HANA backups in the HANA Administration Cockpit. Full backups at shorter frequencies mean mass [block-change](#) for the volume snapshot. For a snapshot, a full backup means that on the disk level a set of new blocks need to be written, which increases backup time. Lower frequency, full backups increase the RTO. You can achieve a balance between time-consuming full backups and the need to have the latest data backed up by configuring less frequent full HANA backups. Instead, you can do incremental or differential backups between the full backups. The RTO/RPO requirements of the business drive the persistent disk snapshot schedule frequency and the HANA database full, incremental, and differential backup frequencies.

When scheduling snapshots of HANA backup volumes consider the following:

- Configure snapshots at the disk level. In contrast, some database snapshots offered by database administration tools are at the application level
- Configure subsequent snapshots to be incremental, so the changed blocks determine the data that needs to be captured with each snapshot. For example, if a full backup file is dumped to a disk location that is deleted and recreated, all blocks are considered to be changed.
- Flush write caches and freeze file system interactions. Because snapshots focus only on a specific volume (the backup volume), you can control these actions.
- Unable to age snapshots like Cloud Storage buckets with [rule-based configuration](#) for lifecycle management.

Shared file storage

SAP systems can use shared file storage like Filestore or Elastifile to fulfill the high availability and disaster recovery requirements. Filestore is a zonal resource that provides a fully managed NFS endpoint and is intended for large-scale shared file storage. You can use Filestore for file shares in a distributed SAP NetWeaver setup in the same zone. You can use Elastifile to help ensure that the associated storage is [highly available across zones](#) in a region with synchronous replication and a disaster recovery copy is stored in a separate region by using asynchronous replication.

HANA Backint agent for Cloud Storage

Beta

This feature is in a pre-release state and might change or have limited support. For more information, see the [product launch stages](#).

For SAP HANA database backup, Google Cloud offers a free and certified [Cloud Storage Backint agent for SAP HANA](#) (Backint agent). The main advantage of using the Backint agent is the elimination of the persistent disks for the backup volume.

Backups using the Backint agent are application-aware because the streaming happens from APIs integrated with SAP HANA database tools. Also, the streaming checks the physical consistency of the database backup.

When using the Backint agent, consider the following:

- Use SAP HANA DB administration tools to schedule backups.
- Use native OS compression because backups might not be efficiently compressed, but at the cost of CPU cycles and throughput.
- Unable to do deduplication even though it is an agent integrated to SAP HANA DB tool, but the application awareness is better than a volume snapshot.
- Take extra manual steps to do system copies using backups.

Third-party network storage

Third-party NFS solutions also offer a backup of all relevant file system volumes of an SAP installation (both application server and database) with scheduled snapshots. These snapshots are stored in Cloud Storage. These NFS solutions offer equivalent features required for a HA/DR architecture, such as cross-zonal resilience for high-availability failover and file system replication to other regions, or Cloud Storage-based snapshots for cross-region disaster recovery.

Consider the following with third-party NFS solutions:

- Extra license costs
- Slower throughputs than Persistent Disk
- Egress charges for cross-regional replications

For an SAP HANA database, these solutions are only suitable for hosting backup and share volumes. They are not able to support data or log volumes due to higher throughput requirements.

Third-party backup agents and managed services

Third-party backup and storage solutions offer advanced technical features that enable rapid backup and fast recovery times. These solutions use block-level change capture (that is not reliant on the file system) for backups that do not have an adverse impact on disk-write times. Recoveries are also faster than file-system-based backups because they do not rely on database-level incremental backups. This recovery time enables fast playback of increments. For enterprise-scale SAP landscape deployments, the storage sizes are reduced. For HANA database backups to be fast, these solutions rely on extraction mechanisms other than the HANA Backint interface. We recommend that you only consider third-party solutions that are certified SAP HANA backup solutions.

As managed services, these solutions also offer features like central management and de-duplication that significantly reduce the manual and automatic tasks of the IT teams in an enterprise organization.

For third-party backup agents and managed services, consider the following:

- Tradeoff between licensing for a managed service and enterprise-internal IT costs and efforts.
- Egress costs for cross-regional data replications.

Despite these considerations, such solutions might be a viable and cost-effective option for SAP installations for industries like banking and financial services, where retention policies and compliance requirements drive your backup strategy.

SAP HANA data snapshot

For SAP HANA databases, you can create a [data snapshot](#) using native SQL without extra certification. However, this technique is somewhat complex. For example, if you use [XFS](#), you need to suspend access to the XFS file system before you create a snapshot. A data snapshot is normally application-consistent and is executed by the SAP HANA database tooling, so you don't need to separately back up multiple volumes and devices. Because the data snapshot tooling is at the database level, there is no influence of Google Cloud infrastructure or tools like Persistent Disk. You can also take data snapshots of multi-tenant SAP HANA databases for SAP HANA 2.0 SPS4. Use SAP HANA data snapshots with regular backups for large production SAP HANA Databases (more than 6 TB).

For data snapshots, consider the following:

- A data snapshot for a database with more than one tenant is not supported.
- Data snapshots rely on previously created internal snapshots. If the database or a database service is restarted, the snapshot is lost.

SAP HANA stop/start snapshot of secondary HANA instance

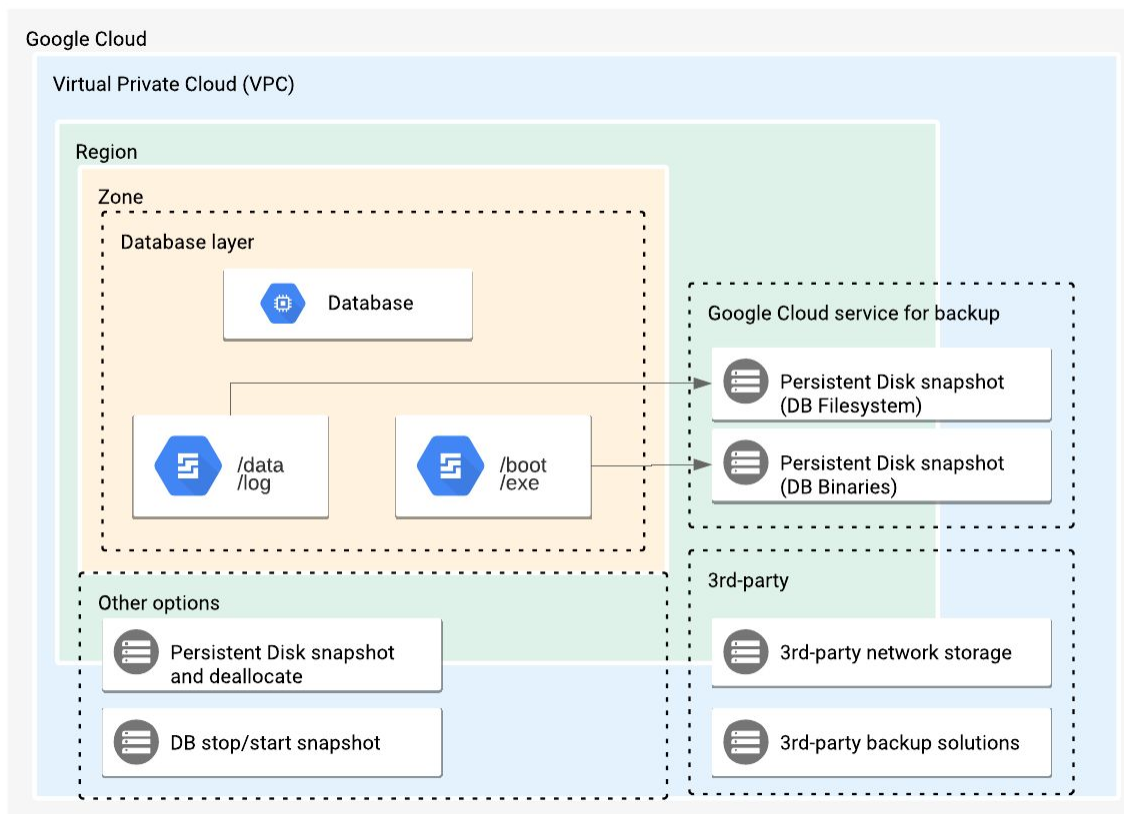
Creating a snapshot of a secondary database instance means using a smaller standby instance in an SAP HANA system replication (HSR) setup. For more information, see, [SAP HANA Secondary System Usage](#). In addition to backing up its disks (file system), you can take this instance offline and make a complete virtual machine snapshot for point-in-time recoverability. Because this method involves stopping and starting the entire virtual machine, it is only suitable for non-production use cases, where cost considerations supersede RPO requirements.

Snapshot and disk deallocation

The flexibility of standing up resources only when you need them is a core feature of cloud-based infrastructure. By using this flexibility, you can create persistent disks before a backup for a volume is scheduled in a backup cycle. The backup can then move to the new disk. After the backup is complete, you can take a snapshot. Then you can delete the disk to stop incurring costs for that disk. The snapshot storage is [less expensive](#) than running the disk. For scenarios where cost considerations are highly prioritized, you can allocate a persistent disk in time for a snapshot and then deallocate when the backup is completed.

Database backup strategies with other RDBMS

The backup options with other relational database management systems (RDBMS) underlying SAP installations are similar to those options with SAP HANA databases, except for the options that require the SAP HANA application-level tools. The relevant volumes are database application binaries, boot folders, and the folders that store transactions and log information, which are referred to in the following diagram as `/data` and `/log`. The following diagram shows database backup strategies with non-SAP HANA databases.



The following backup options apply for non-SAP HANA RDBMS:

- Persistent disk snapshots
- Third-party network storage
- Third-party backup agents and managed services
- Stop and start snapshots
- Snapshot and disk deallocation

Automation

Every backup strategy relies on the automation of backup tasks to reduce the workload of an IT administrator. For SAP workloads running on Google Cloud, you can use automation features like scheduling snapshots of persistent disks or rule-based configuration of TTL for Cloud Storage buckets.

You can also use [Cloud Scheduler](#) to run advanced automation tasks. Cloud Scheduler is a fully managed enterprise-grade cron job scheduler that lets you schedule virtually any job, including batch jobs, big data jobs, and cloud infrastructure operations. You can automate everything, including retries during a failure, to reduce manual toil and intervention. Cloud Scheduler even acts as a single control interface, letting you manage all of your automation tasks from one place. To schedule customized backup tasks, you can use Cloud Scheduler.

Beyond the capabilities offered at the infrastructure-level in Google Cloud, there are automation possibilities offered by database-specific tools (for example, SAP HANA Cockpit). Also, managed services like third-party NFS and third-party backup solutions offer configurable automation capabilities.

Comparison summary

Strategy	Features	Considerations	Environment Suitability	Recommendations
Persistent disk snapshots (complete disks)	Straight-forward tooling, crash-consistent, globally addressable	File and application-level consistency not ensured	Production and non-production	For system clones. For production to complement the other options.
Volume snapshots (database backup volumes)	Straight-forward tooling, crash-consistent, reasonably fast, globally addressable	Not application-aware, quiesce file system for correct snapshots, no rule-based life-cycle management	Production and non-production	For production
SAP HANA Backint Agent for Cloud Storage	No egress, application-aware, free, streaming with inherent consistency checks	Compression not effective, extra steps for system copy	Production and non-production	Specific to SAP HANA database
third-party NFS	Multi-zone availability	Cost / managed trade-off, slower than native storage, egress charges	Production and non-production	For SAP NetWeaver HA installations
third-party Backup Agent and managed service	Enterprise-grade, centralized management, industry-specific retention and compliance requirements	Cost versus managed service trade-off, egress charges	Production	For specific customers and industries that require long retention periods
SAP HANA data snapshot	SQL based, application consistent, database-level tooling	Not supported for multi-tenant database, extra configuration required	Non-production only	Specific to SAP HANA database. Not recommended for multi-tenant SAP HANA.
Stop and start Snapshot	Trivial way to help ensure consistency	VM restarts required	Non-production only	Suitable for quick non-production clones
Snapshot and deallocate	Allocate storage only for taking snapshot	Extra scripting required	Non-production only	Use cases with stringent cost considerations

Recommendations

When you back up SAP installations on Google Cloud, choose the same backup technique across production and non-production environments to simplify your operations. You can cover the differences in RTO/RPO requirements between these environments by adjusting the backup schedules and the related automation.

Our recommended best practice is to allocate separate disks for volumes that are relevant for backup. Then we recommend backing up SAP systems on Google Cloud with persistent disk snapshots of the relevant volumes with automation configured to suit RTO/RPO requirements. For the application layer, the relevant volumes are the boot and application binaries, and the application server directories. You can directly configure persistent disk snapshots for single servers and store them globally as addressable snapshot storage. For multi-server installations using NFS (like Elastifile), you configure backups by using the Cloud Console for the NFS service, and you store the backups in multi-regional Cloud Storage buckets (for cross-region disaster recovery). For the database layer, the relevant volumes are the data and log in addition to the database application boot and binaries.

In the context of the overall disaster-recovery architecture, allocating separate disks for non-OS volumes eliminates needing to replicate redundant data across regions as multi-regional or global resources. When NFS solutions are used in the application layer, you need to configure the backup options in those solutions to comply with the overall high-availability and disaster recovery architecture.

Overall, Google recommends that critical systems use at least two redundant means of protection. Using two means of protection can help to ensure the integrity and recoverability of the data during a disaster. We recommend that you use database technology and persistent disk snapshots.

Summary

The backup strategy and the selection of suitable services for backup are an important part of the overall business continuity planning for SAP systems. Google Cloud offers multiple native capabilities for cost-effective backup of SAP systems. These capabilities also come with automation features. You can also use the backup options and interfaces that are offered by applications (for example SAP HANA Backint) and managed backup solutions by third parties.

Allocation of separate persistent disks for virtual machine and application relevant volumes is a recommended best practice that helps you achieve robust backups of SAP systems. Persistent Disk snapshots of virtual machine, application, and database relevant volumes is a recommended technique for backup. Scheduling snapshots and using Cloud Scheduler for advanced job scheduling enable the required automation for the backup of SAP systems.

The recommendations in this document are considered a baseline for starting an assessment for backup of SAP systems on Google Cloud.

Further reading

- [Backup and recovery SAP HANA on Google Cloud - operations guide](#)
- [SAP on Google Cloud - SAP NetWeaver - operations guide](#)
- [Create snapshots](#)
- [Schedule snapshots](#)
- [Restore and delete snapshots](#)
- [Snapshot best practices](#)