COMMVAULT

Microsoft Azure

# SAP HANA in Azure: Data Protection and Snapshot Management with Commvault

# Contents


Introduction ..... 3

Commvault/Microsoft Partnership ..... 3

Commvault Architecture ..... 4

Integration with Azure ..... 5

Commvault Public Cloud Architecture Guide for Microsoft Azure ..... 6

Commvault Solutions for Azure SAP workloads ..... 7

Streaming versus Intellisnap ..... 7

Taking care of all SAP Databases ..... 8

SAP Landscape Management ..... 9

Setting up a test environment ..... 9

Test Environment ..... 9

SAP HANA VM configuration ..... 9

Cloud Libraries ..... 10

Storage Policies ..... 10

SAP HANA Commvault setup ..... 10

Testing Results ..... 12

Streaming Backup/Restore with Deduplication and Compression ..... 12

Streaming Backup/Restore with Compression only ..... 13

Snapshot Backup and Restore ..... 13

Summary of testing results ..... 14

Reference Architecture ..... 15

Storage requirements ..... 16

Basic considerations ..... 16

Production Systems ..... 16

Test and Development Systems ..... 17

Commvault infrastructure ..... 18

Single Azure Region ..... 18

Disaster Recovery across Azure Regions ..... 18

Summary and Conclusion ..... 19

# Introduction

SAP customers who are migrating their SAP applications to Azure need a solution which can meet or exceed their requirements for data management, protection, cost reduction, efficiency, as well as remaining compliant with their service level agreements. Commvault and Microsoft are leveraging their 20 year partnership and have teamed together to test and document a comprehensive solution for data management and protection for SAP on Azure.

# Commvault/Microsoft Partnership

Microsoft and Commvault's relationship spans nearly two decades, delivering industry-leading data management solutions (Commvault) on a powerful, secure infrastructure (Windows Server & Microsoft Azure). Commvault has built a single platform that unifies and automates data management and protection across all the operating systems and applications your organization relies on every day — simplifying the way you manage, protect, access and share data.

Commvault is a leader in data storage and management and is also the first to enable agentless backup and recovery of Azure instances. Not only have Commvault and Microsoft teamed up to create this whitepaper , we're also leveraging the two-decade partnership - which includes each company heavily using the others' technology - to help run our day-to-day businesses. By leveraging our combined vast experience and scale, as well as our exposure to early-access technology, leads to proven best-practice guidance for our customers.

Commvault has supported Microsoft Azure since 2008, delivering a deep depth and breadth of support for Azure compute and storage that uniquely positions our joint customers to benefit from the best practices of thousands of clients. Commvault is used extensively by Microsoft within the Office Products Group (since 1999), Xbox and Xbox Live (since 2002), as well as the Azure Data Protection Group (since 2013). Commvault is also a large Azure customer (since 2008), which included having Azure as the exclusive host for Commvault's SaaS offering. With support for hot, cool, page and cold (archive) Blob storage, as well as Express Route, Import/Export, Data Box, Geo-Replication and Azure Stack, we can help deliver an extremely optimized and flexible Azure solution - agile enough to move with the high demands of digital transformation.

Commvault data management software contains a full range of data management and data protection capabilities to/from the Azure cloud and Azure Stack, including best-of-breed data management capabilities for Microsoft products such as Exchange, SharePoint, SQL Server, Active Directory and Office 365.

In addition to the support of these Microsoft products, Commvault can also be used to manage and protect SAP HANA or any other mainstream database (Oracle, IBM DB2, Microsoft SQL Server, etc) within Microsoft Azure, on-premises or in a hybrid cloud scenario.

Running SAP HANA based applications on Azure is becoming more and more popular and requires an enterprise-grade data management solution. Commvault is both "Azure Certified" and "SAP-certified" for the integration with SAP HANA, making the Commvault data management solution the right choice to manage, move and protect SAP HANA workloads in Microsoft Azure.

Commvault provides a SAP HANA data protection solution that is tightly SAP-integrated and is designed for universal protection. Not only standalone SAP HANA deployments, but also SAP S/4HANA, SAP Business Suite on HANA, SAP BW/4HANA and SAP C/4HANA applications can be protected effectively. SAP HANA can run in scale-up and scale-out configurations and high availability provisions with SAP HANA System Replication are also supported.

Along with safe-guarding SAP data, Commvault provides a SAP archiving solution and also integrates with SAP Landscape Management (LaMa) tool for efficient, automated system refreshes of SAP test and development systems with production data. This allows for effective management of SAP landscapes running on SAP HANA and other databases supported by SAP.

As a market leader, Commvault has been named a leader in Gartner's Magic Quadrant for Data Center Backup and Recovery Solutions for 7 years in a row and has similar positioning from Forrester's Data Resiliency reporting.

# SAP on Azure offering

The partnership of Microsoft and SAP began more than 20 years ago when SAP certified Microsoft Windows and Microsoft SQL Server for running SAP R/3. The deep integration of SAP with Microsoft Office (Office 365) and the possibility to connect the core SAP system with Microsoft Azure PaaS offerings such as Azure Active Directory for security and identify management, LogicApps for simplified application integration and Data and Analytics (i.e DataLake and PowerBI) emphasizes the strong integration capabilities of SAP and Microsoft technologies. Moreover, SAP SE themselves are now running several of their internal SAP systems on Azure.

Microsoft Azure is the eminent public cloud for running SAP applications. With the largest portfolio of SAP HANA certified IaaS Cloud offerings customers can run their SAP HANA Production scale-up applications on certified virtual machines ranging from 192GB to 12TB of memory. In addition, customers can run SAP HANA scale-out applications such as BW on HANA and BW/4HANA on virtual machines with 2TB up to 16 nodes (32TB). In Q4CY19, Azure will offer SAP HANA scale-out with 6TB up to 16 nodes (96TB) where an option will also be provided for host auto-failover capability (N+M) with Azure NetApp Files. For customers that require extreme scale, today Azure offers bare-metal HANA Large Instances for SAP HANA scale-up to 20TB (24TB with TDIv5) and SAP HANA scale-out to 60TB (120TB with TDIv5).

As customers embark on their SAP to Azure journey, it is recommended to dive into the SAP on Azure documentation to deepen your understanding of using Azure for hosting and running your SAP applications. Use the SAP Workload on Azure Planning and Deployment Checklist as a compass to navigate through the various phases of your SAP migration project. The checklist will steer you in the right direction for a quality SAP deployment on Azure.

SAP and Microsoft provide joint support for SAP applications using Microsoft technologies; this provides customer confidence that they can run their SAP workloads in the most trusted and secured cloud environment. Microsoft is not only the best cloud for running SAP workloads, Microsoft also knows how to run SAP on Azure ; SAP ECC is the core ERP for running Microsoft's line of businesses and one of the largest SAP systems worldwide. Microsoft's transformation project to S/4HANA is also now underway. SAP also provides SaaS services to Microsoft such as Ariba and Successfactors and in many cases these services are provided out of Azure datacenters to SAP customers worldwide.

# Commvault solution for Microsoft Azure

## Commvault Architecture

Before explaining how Commvault can be used to protect SAP HANA workloads in Microsof Azure, here is a short overview of the main Commvault components.

A **CommCell environment** is the logical grouping of all software components that protect, move, store, and manage data and information. A CommCell environment contains one CommServe host, one or more MediaAgents, and one or more clients.

The **CommServe** host is the central management component of the CommCell environment. It coordinates and executes all CommCell operations, maintaining a Microsoft SQL Server database that contain all configuration, security, and operational history for the CommCell environment. There can be only one CommServe host in a CommCell environment.

The **MediaAgent** is the data transmission manager in the CommCell environment. It provides high performance data movement and manages the data storage libraries. The CommServe server coordinates MediaAgent tasks. For scalability, there can be more multiple MediaAgents in a CommCell environment.

A **client** is a logical grouping of the software agents that facilitate the protection, management, and movement of data associated with the client.

An **agent** is a software module that is installed on a client computer to protect a specific type of data. Different agent software is available to manage different data types on a client, for example, Windows file system data and SAP HANA databases.

The Virtual Server Agent or VSA is a specialised agent that protects hypervisor and cloud resources.

## Integration with Azure

The primary integration point with Azure is through Commvault's Virtual Server Agent (VSA). You can use VSA to perform the following tasks for Azure VMs:

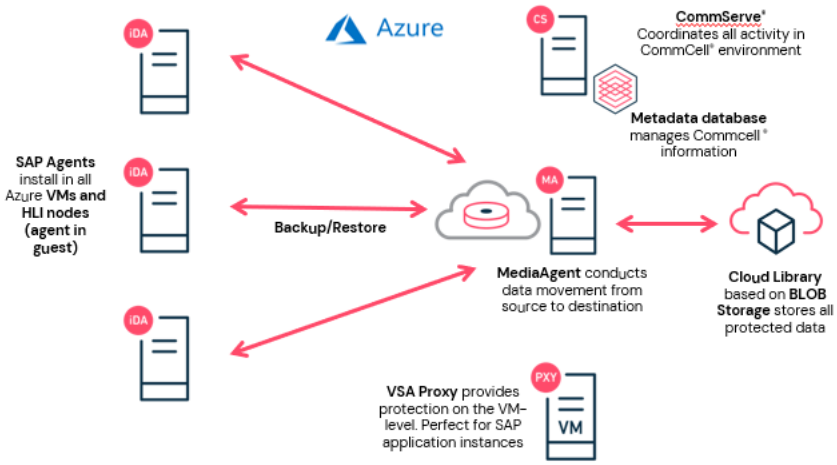
- Backup and recover Azure virtual machines using either the Azure Classic or the Azure Resource Manager deployment model. You can restore full virtual machines or guest files and folders.
- Seamlessly convert backups of Amazon, Hyper-V and VMware virtual machines to Azure virtual machines (Azure Classic or Azure Resource Manager).
- When performing a restore from a backup of an Azure VM, you can choose to restore a VM disk and attach it to a different VM that already exists.

Commvault data management and protection operations are managed through the CommCell Console. Use the CommCell Console to configure a virtualization client and other entities that are used to support operations.

A virtualization client instance is the access point for an Azure subscription and is used to backup full Azure virtual machines. You must define a VSA agent instance for each Azure subscription. If a database or application runs inside the VM on the Linux OS, the VSA approach is not appropriate, as it guarantees crash consistent backups only. For database workloads we need application consistent backups. Therefore we are using the "agent in guest" approach with the respective Commvault database agent installed inside the VM. This applies specifically for VMs running SAP databases. VMs running SAP application instances (PAS, AAS, ASCS, etc.) can be protected using the agentless VSA approach.

When you create a virtualization client instance (a so-called proxy), the Commvault software automatically creates an Azure instance, a so-called backup set, and a default subclient that can be used to protect all virtual machines. You can create additional subclients to perform separate protection operations for different groups of virtual machines. For example, you can create a different subclient for different guest operating systems and use the default subclient to protect any remaining virtual machines that are not covered by user-defined subclients.

With Commvault's VSA agent you can perform full, incremental, or synthetic full backups of virtual machines and perform full virtual machine, disk or guest files and folders at granular level.




Another important integration with Azure is available on the storage level. Commvault has the ability to leverage Azure Blob Storage as a secondary storage tier. So-called cloud libraries are built using Azure Blob storage containers. It is strongly recommended to use multiple containers in a single cloud library – this approach enables Commvault's load-balancing capabilities and provides the ability to easily increase storage capacity as and when it is required. Azure Blob Storage provides centralized data access, better failover capabilities and reduces the day-to-day storage administration tasks. As the data gets transferred over the network, protecting the integrity of data is an important aspect of any cloud storage implementation. Azure Blob Storage protects the integrity of the data using the following features:

- By default, data is transferred through secured channels using HTTPS protocol.
- Data encryption further encrypts the data providing data protection during network transfer as well as storage.

Commvault's Deduplication feature identifies and eliminates redundant data in the backup, thereby reducing not only the volume of data stored in cloud, but also the bandwidth required for data transfer. Enabling compression reduces the data footprint even further.

## Commvault Public Cloud Architecture Guide for Microsoft Azure

Commvault and Microsoft are providing this document as an architecture guide for solutions architects and Commvault customers who are building data protection and management solutions utilizing the Microsoft Azure public cloud and the Commvault platform.

It includes cloud concepts, typical use cases, architectural considerations, and sizing recommendations to support Commvault's platform in Azure. The approach defined in this guide extends existing functionality into easily scalable, re-usable architecture patterns to cover migration to the cloud, disaster recovery to the cloud as well as protecting operational SAP Applications already running in Azure.

For additional reading, we recommend Commvault's Cloud Architecture Guide for Azure. For sizing, we would like to refer you to the "Architecture sizing" section of the document. The full guide can be downloaded here: http://bit.ly/2v1IBN4

# Commvault Solutions for Azure SAP workloads

## Streaming versus Intellisnap

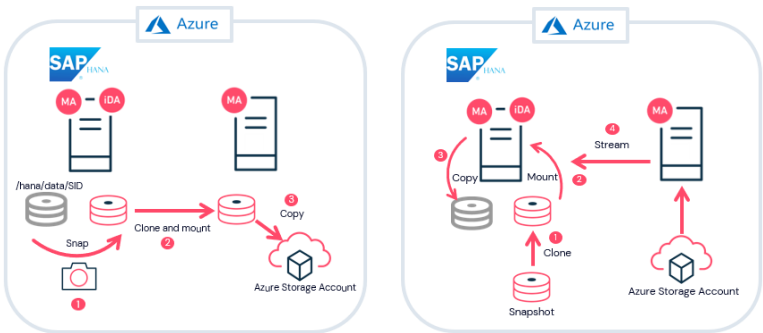
One option for protecting SAP environments in Microsoft Azure is to leverage streaming backups for SAP application instances and for SAP database VMs. This is a network-based data transfer where backup data is sent to a MediaAgent which will store the data in a cloud library. Alternatively, it is also possible to install a MediaAgent directly into the VM hosting the SAP database instance and share the cloud library. This way, all database nodes can backup their local data in parallel to blob storage, enabling high-speed backups and restores.

A more advanced way for protecting SAP databases is by using Storage-level Snapshots. That provides a number of advantages

- Minimal SAP application impact as the backup takes just seconds compared to hours in the streaming case
- Ability to drive backup windows to zero which is a huge benefit for large mission-critical production systems
- Fast restores allowing to meet even aggressive RTO targets

Commvault's integration with storage-level snapshots is called Intellisnap. Intellisnap operates fully application-aware and is very versatile as it currently supports a large number of on-prem storage products alongside with cloud storage. It also provides fully automated snapshot lifecycle management. Snapshots get a retention time assigned (like any other backup) for controlling its lifetime. Once its lifetime has expired, both backup metadata and the snapshot itself are automatically deleted. Snapshots can also be copied to Blob Storage or other media for creating secondary copies with different retention times. Those copies can automatically be created and placed on different Blob types (i.e. block blob, managed disks, etc.) and in different Azure regions serving disaster recovery or regulatory purposes. For some storage vendors, Intellisnap can also manage snapshot replication. One great example here is Netapp Storage, where Intellisnap can control SnapVault and Snapmirror relationships alongside with providing the ability to restore for all copies. The combination of both streaming and snapshot backups in one tool and under a single user interface is a key functionality of Commvault's solution. For customers this primarily means flexibility and support for a wide range of SLAs. For instance, if I'm running out of my backup window because of data growth or my SAP S/4HANA system transitions into production mode with RTO shortening dramatically, I can turn on snapshots with a couple if mouse clicks. This applies to all Commvault agents supporting Intellisnap in Azure.

Commvault Intellisnap for Azure is based on Azure Managed Disks. Simple disk configurations are supported alongside with Logical Volume Manager (LVM) configurations. At backup time, a snapshot is automatically created for each disk that's hosting the database volume. This usually takes a couple of seconds after the database was placed in backup mode. At restore time, temporary disks (clones) are created from all snapshots, mapped to the destination and data is read from the temporary disks and written to the original disks in the VM. In case of a revert restore, the original disks are replaced with the new disks that are created from the snapshots. As a result, this reduces recovery time of the database from hours to minutes.



## Taking care of all SAP Databases

The center of each SAP system is the database. It needs to be properly protected to avoid data loss and keep costly downtimes to an absolute minimum. Database backups need to include all data files along with an uninterrupted logfile chain to make sure the database can be recovered to any point in time.

When it comes to SAP environments, the most important database these days is certainly SAP HANA. SAP HANA uses an In-Memory Database technology that processes massive amounts of real-time data in a short span of time. Using the In-Memory computing engine, HANA processes data stored in RAM as opposed to reading it from a disk. This leads to nearly instantaneous results from customer transactions and data analytics.

Commvault software provides a SAP-certified and –integrated end-to-end backup and recovery solution for SAP HANA scale-up and scale-out environments using the SAP Backint for HANA interface. The agent combines a full implementation of the Backint for HANA standard including differential and incremental backups with a modern, easy to use web interface and a scalable logfile management architecture. The agent also connects the SAP HANA snapshot interface with Intellisnap and supports Azure managed disk snapshots – both for SAP HANA Scale-up and Scale-out configurations. High availability setups which include SAP HANA System Replication are also supported in Azure Availability Zone deployment models. Commvault's SAP HANA solution therefore is extremely flexible and universal. It enables you to meet a wide range of SLAs in combination with all important SAP configurations. From sandbox to production and from S/4HANA Scale-up systems to large BW/4HANA Scale-out deployments. That makes Commvault the perfect partner for Azure's SAP HANA virtual machines offering.

Another important offering is SAP HANA on Azure Large Instances (HLI). HLI, is a bare-metal offering with Netapp storage which offers scale up to 480 CPUs and 24 TB of memory; HLI address the needs of customers running very large SAP HANA databases. Commvault has full support for HLI including Intellisnap based on Netapp SVMs alongside with the ability to control SnapVault and Snapmirror relationship which can be part of an HLI configuration.

If you run other SAP database platforms in your SAP environment, Commvault can certainly also help. For SAP on Oracle, there is an automated solution available which is integrated with and certified on SAP's Backint for Oracle interface. This agent is based on SAP BR*Tools but eliminates the need for writing scripts via GUI-based management and on-the-fly generation of all required commands and scripts. The agent supports Intellisnap in Azure alongside with threshold-based archive log management. Further agents are available for SAP MaxDB, SAP Sybase ASE, Microsoft SQL Server and IBM DB2 ensuring the all SAP deployments can be protected appropriately. Below's table summarizes the capabilities of all these agents. All agents support backup/restore to/from blob storage, premium disk and to UltraSSD.

| Commvault iDataAgent | SAP certified | SAP Integration | SAP Interface | Intellisnap for Azure | GUI-based Database Copy | Granular Restore | Table-level Restore | Auto Log Backup | Auto Discovery for DBs |
|---|---|---|---|---|---|---|---|---|---|
| **SAP on Oracle** | **Y** | **BRTOOLS<br>DB13<br>DBA Cockpit** | **Backint for Oracle** | **Y** | **Y** | **Y** | **N** | **Y** | **N** |
| Oracle | N/A | N | N/A | Y | Y | Y | Y | Y | Y |
| **SAP on MaxDB** | **Y** | **DBMCLI<br>MaxDB Studio<br>DB13<br>DBA Cockpit** | **Backint for MaxDB** | **N** | **N** | **N** | **N** | **Y** | **N** |
| DB2 | N/A | DB13<br>DBA Cockpit | N/A | N | Y | Y | Y | Y | Y |
| MSSQL | N/A | N | N/A | Y | Y | Y | Y | Y | Y |
| SYBASE | N/A | N | N/A | N | Y | N | N | N | Y |
| **SAP HANA** | **Y** | **HANA Studio<br>HANA Cockpit<br>DBA Cockpit** | **Backint for HANA** | **Y** | **Y** | **N** | **N** | **Y** | **Y** |

## SAP Landscape Management

SAP applications are typically deployed in so-called landscapes. Each landscape typically consists of a production system plus development and test systems. During SAP development projects, feature, service pack or application rollouts changes are developed and/or tested in cylces before go-live in the production system. In order to be able to test on current data sets it is required to periodically refresh test and development systems from production data. The SAP process for this is called System Refresh. This process consists of 2 phases:

- Database Copy
- Post Copy Activities

Commvault has a graphical dialog available for making a database copy simple. You can pick source and destination server or Azure VM and chose from different options like full vs. Point-in-Time recovery. The rest happens in an automated way. Post Copy Activties (like RFC adjustments, etc) usually need to be handled by an SAP Basis administrator in a manual way. This can be very time consuming.

The larger your SAP environment becomes, the more time you need to spend on System Refreshes which drives the need for automation. SAP's Landscape Management (LaMa) tool is the answer to these challenges and there is an. SAP LaMa supports powerful orchestration and automation for for System Copies, Refreshes and SAP HANA management tasks; moreover SAP LaMa can automate the tedious and time consuming post copy activities. Commvault provides an add-on which integrates with SAP LaMa which can trigger a database copy at the right point in time during the SAP LaMa system refresh workflow. This gives you an additional synergy as you can leverage your existing backups and the data protection infrastructure for management of SAP landscapes. The integration of Commvault with SAP LaMa delivers fully automated SAP System refreshes on Azure which can bring down the required time from days to hours. Additional benefits are integration of backup management functionality in LaMa's GUI and with custom processes. Microsoft also offers a SAP LaMa Connector for Azure which can be leveraged in combination with Commvault's LaMa integration.

# Setting up a test environment

## Test Environment

For testing and benchmarking, a test environment was built. This environment consisted of the following Azure resources:

- 1x Standard_D8s_v3 VM, for running the Commserve, Windows 2016
- 1x  Standard_D16s_v3 VM, for running a Media Agent, Windows 2016
- 2x 200 GB Premium Disks on Media Agent VM for holding the IndexCache and Deduplication databases
- 3x M128s VMs, for running a 3 node Scale-out SAP HANA 2.0 setup, SuSE SLES 15 for SAP applications
- 1x M128s VM, for running a Scale-up SAP HANA 2.0 setup, SuSE SLES 15 for SAP applications
- 4 x Cool Blob Containers and 4 x Hot Blob Containers serving as backup targets

#### SAP HANA VM configuration

With 128 vCPUs, 2 TB of memory and a maximum I/O throughput of 2000 MB/s Azure's M128s VM is a powerful production platform for all SAP applications requiring SAP HANA Scale-up or Scale-out configurations. In order to be able to run realistic tests we configured the VMs for production use. After deploying the M128s VMs we installed SAP HANA 2.0 SPS04 according to SAP and Azure documentation. In order to meet the requirements of a production system with maximum I/O throughput we used the following disk layout on each VM:

With 128 vCPUs, 2 TB of memory and a maximum I/O throughput of 2000 MB/s Azure's M128s VM is a powerful production platform for SAP applications requiring SAP HANA scale-up or scale-out configurations.

After deploying the M128s VMs we installed SAP HANA 2.0 SPS04 according to SAP and Azure documentation. And, to meet the requirements of a production SAP HANA system with maximum I/O throughput we used the following disk layout on each VM:

- 3 x P30 (1 TB) Premium Disk for /hana/data, configured as a striped LVM volume
- 2 x P20 (512 GB) Premium Disk for /hana/log, configured as a striped LVM volume
- 1 x P30 (1 TB) Premium Disk for /hana/shared
- 1 x P6 (64 GB) Premium Disk for /usr/sap

For /hana/data we verified 600 MB/s read and write throughput performance using Linux "dd" tool. And, for /hana/log the performance test showed 300 MB/sec.

On the scale-up system we created HANA instance TD3 with a single tenant named WB1 where we loaded 1.5 TB of test data. The scale-out system was created with HANA instance TD4 and single tenant WB3, which was configured to be distributed across all 3 nodes. 1.5 TB of test data was loaded.

### Cloud Libraries

All Azure Blob storage containers were configured into Commvault cloud libraries consisting of four mount paths each. All test backups landed in these libraries. In order to achieve maximum throughput and evenly distributed load balancing, the "Spill and Fill" option was activated.

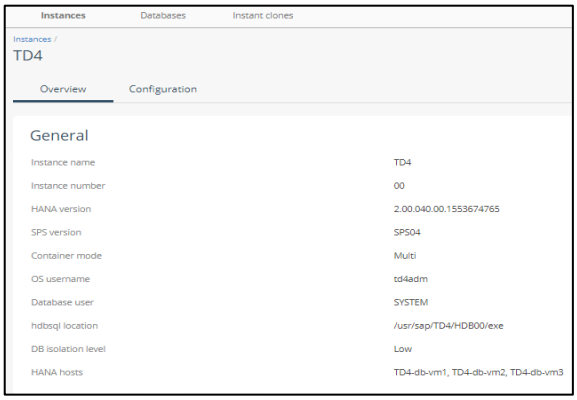
### Storage Policies

Using Commvault Storage Policies we can define the data management rules which will be applied to the SAP HANA backup data. Storage Policies determine how the data is backed up (deduplicated, compressed or without data reduction), how many copies will be created, where these copies are stored (on the cloud libraries in our case) and how long each copy will be retained. We created Storage Policies for serving 2 test cases:

- HANA Backup and Restore with deduplication and compression
- HANA Backup and Restore with compression only

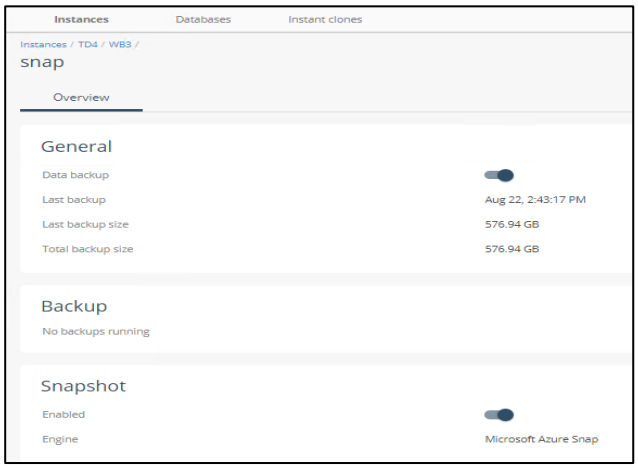
### SAP HANA Commvault setup

To perform the setup in Commvault, we started with HANA instance TD4. It is required to specify the instance details alongside either an hdbuserstore key or a database user to enable Commvault to connect to HANA. In addition, all HANA nodes (in this case three) were added and Storage Policies for HANA logs and data were specified. Deduplication and compression settings can be fine-tuned using this GUI dialog as well.

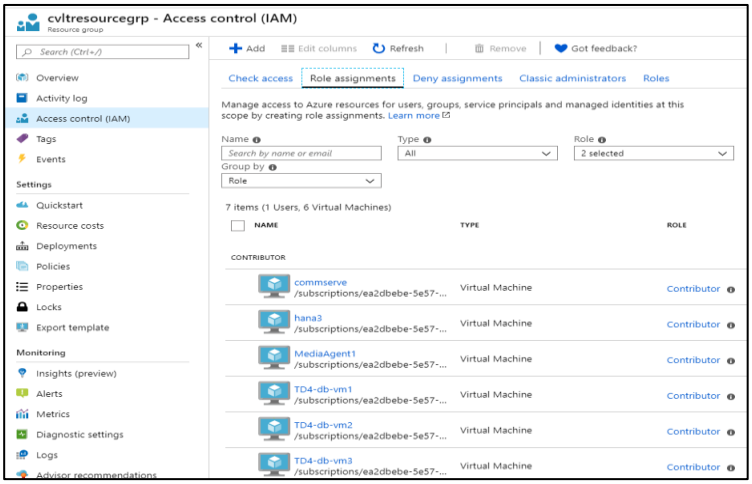


While the instance is generated in the Commvault Command Center GUI, the HANA agent will also auto-discover all tenants which in turn inherit all settings of the embracing HANA instance. All SAP HANA side configuration settings (parameter file, creation of symbolic links, switch logfile backup to Commvault) will be taken care of automatically at the time of the first backup job and can be verified through SAP HANA Studio.

As we wanted to test both streaming and snapshot backup for both HANA instances, we configured 2 so-called subclients per tenant (WB1 and WB3). For turning on Azure snapshots, we enabled the GUI switch and selected "Microsoft Azure Snap" as snap engine. See below for a screenshot of the "snap" subclient.




Some Azure side configuration is also required. On the level of the resource group which contains all disks being snapshot targets you need to assign the "contributor" role to all VMs managing the snapshots (create, copy, delete).




Finally, the Azure subscription needs to be configured on the Commvault side as a VSA proxy. An important aspect for the testing effort, as well as for customer environments, is the ability to tune streaming backup and restore performance by taking advantage of SAP HANA's multi-streaming capabilities. SAP HANA databases consist of up to four internal services (name server, index server, XS engine, etc.). Each of these services stores its own data and generates its own log files. However, the vast majority of data always resides in the index server. By default, all SAP HANA services are backed up in parallel, each using a single stream. While all other HANA services typically finish their data transfer within minutes, index server backups can run for many hours.

SAP HANA supports multi-streaming capabilities on the level of index server. This can be configured by specifiying the desired number of streams using the global.ini parameter "parallel_data_backup_backint_channels" in combination with the parameter "data_backup_buffer_size", which needs to be set to number of streams multiplied by 512.

Please refer to Commvault's Best Practice Guide for SAP HANA for further information and agent installation, configuration and operation.

## Testing Results

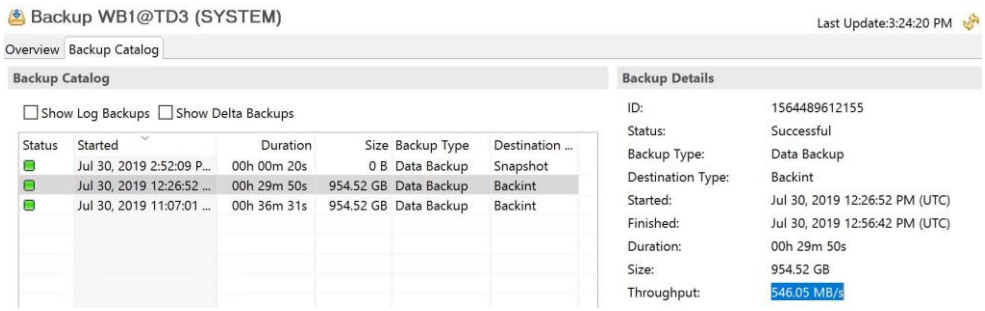
All testing was done based on SAP HANA full backups.

**Streaming Backup/Restore with Deduplication and Compression**

### SAP HANA Scale-up

For the first round of tests we enabled deduplication and compression at the client side. Deduplication provides an efficient method to store data by identifying and eliminating duplicate blocks of data during backups. We enabled deduplication on the client side to make sure the least amount of data will be sent over the network and to ensure we consume only the smallest amount of object storage, since only unique chunks will be stored. We also configured multiple streams for the SAP HANA backup. Default stream count is 4. We also used 8.

The tests immediately delivered backup throughput beyond 2.5 TB/h resulting in backup job runtimes around 30 minutes for our 1.5 TB test system. 16 streams proved to be a good setting here. Client-side CPU load was a bit elevated, which should be fine given the quick job runtime. If this is still an issue for a critical production system, a backup window needs to be defined at a time with minimal SAP end user and batch system load.




We found that restores from deduplicated backups didn't run at the speed of the corresponding backups. This is because of their non-linear read pattern from the backup media. Customers need to be aware of this in the context of RTO. We also want to make clear that we measured full RTO time, which not only includes the actual data restore, but also the time needed to roll-forward the database based on the logs and for database shutdown and startup.

### SAP HANA Scale-out

For Scale-out, we added a MediaAgent to each HANA node, shared the cloud library with all nodes and enabled local backups (LAN-Free). Backup performance with 4 streams easily exceeded 5 TB/h. For restores we achieved around 1 TB/h.

Pros:

- Fastest backup option
- Best data reduction at around 60 – 70%
- Best option for non-production systems

Cons:

- Restores don't achieve backup performance

### Streaming Backup/Restore with Compression only

#### SAP HANA Scale-up

For this test, all deduplication was disabled (client-side and media agent) by using a storage policy without deduplication. Client-side compression remained in place. CPU load on the SAP HANA side was lower compared to the the deduplication test. However, more CPU resources were consumed on the media agent. Using 8 streams, the backup performance was still beyond 2 TB/h. Restore performance was above 1.5 TB/h with 8 streams.

#### SAP HANA Scale-out

Backup performance was not as good as for deduplicated backups, but still above 5 TB/h. On the restore side we observed a nice improvement to more than 1.5 TB/h. 4 streams turned out to be the best option, 8 stream didn't scale anymore.

Pros:

- High restore performance
- Good option for production systems

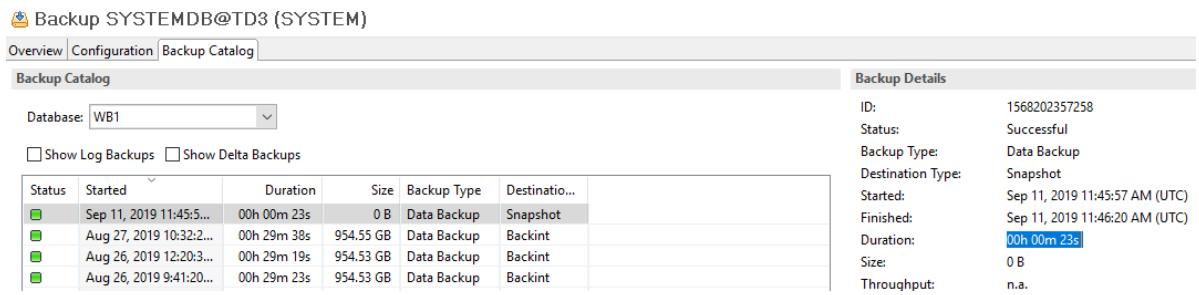
Cons:

- Data reduction around 30 - 35%
- Higher cost for secondary BLOB storage

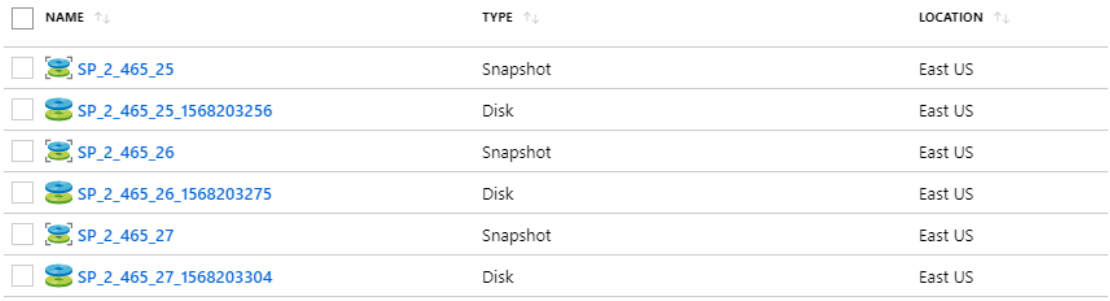
### Snapshot Backup and Restore

#### SAP HANA Scale-up

We tested snapshot backups and revert restores. Results were very good. Backup just took around 20 seconds.




As /hana/data consists of 3 P30 disks in stripe LVM volume, we can see 3 snapshots which were created by backup job 465:



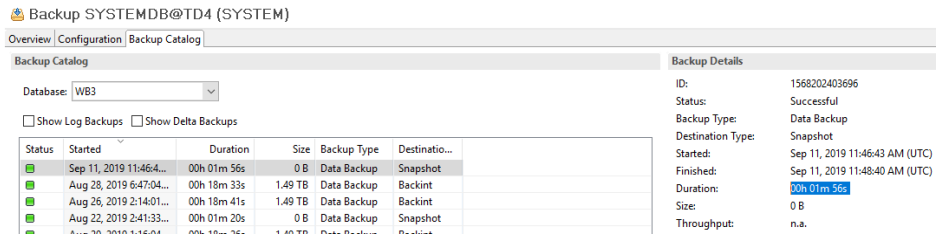
For revert restore, 3 new P30 disks were allocated (see above), mapped and the LVM volume was rebuilt. Restore including database recovery time took in total around 12 minutes (aggregated time over 3 jobs, see below). Further log analysis showed that database shutdown, recovery and startup consumed most of the restore time. The actual revert is taking just a couple of minutes (not shown in the screenshot).

Restore Jobs

| Server/Destination... | Agent | Instance | BackupSet | Job ID | Status | Initiating User | Start Time | End Time |
|---|---|---|---|---|---|---|---|---|
| HANA_TD3/hana3 | SAP HANA | TD3 | WB1 | 472 | Completed | admin | Sep 11, 2019, 12:04:18 PM | Sep 11, 2019, 12:10:22 PM |
| HANA_TD3/hana3 | SAP HANA | TD3 | SYSTEMDB | 471 | Completed | admin | Sep 11, 2019, 12:03:15 PM | Sep 11, 2019, 12:09:19 PM |
| HANA_TD3/hana3 | SAP HANA | TD3 | WB1 | 467 | Completed | admin | Sep 11, 2019, 11:58:27 AM | Sep 11, 2019, 12:04:34 PM |

### SAP HANA Scale-out

For Scale-out, backup time was around 2 minutes. As for backup resource consumption, 9 snapshots were created as tenant WB4 runs on all 3 nodes having 3 /hana/data disks each.




For the revert restore, 9 new P30 disks were allocated (3 per node), mapped and used as revert snapshot target. Restore and recovery running time was in total around 20 minutes in total.

Restore Jobs

| Server/Destination ... | Agent | Instance | BackupSet | Job ID | Status | Initiating User | Start Time | End Time |
|---|---|---|---|---|---|---|---|---|
| HANA_TD4/TD4-db-vm2 | SAP HANA | TD4 | WB3 | 475 | Completed | admin | Sep 11, 2019, 12:34:24 PM | Sep 11, 2019, 12:40:26 PM |
| HANA_TD4/TD4-db-vm1 | SAP HANA | TD4 | SYSTEMDB | 474 | Completed | admin | Sep 11, 2019, 12:33:09 PM | Sep 11, 2019, 12:39:12 PM |
| HANA_TD4/TD4-db-vm1 | SAP HANA | TD4 | WB3 | 473 | Completed | admin | Sep 11, 2019, 12:21:58 PM | Sep 11, 2019, 12:34:47 PM |

Pros:

- Ideal option for production systems
- Very fast backups and restore
- No application impact at all
- Blob storage for backup copies only

Cons:

- Cost for snapshots to be retained

#### Summary of testing results

As the tests have shown, using deduplication in combination with compression primarily helps to reduce Blob secondary storage consumption (around 60 - 70% in practice, depending on data change rate and database content) and is certainly a fast backup option at well over 2.5 TB/h for Scale-up and more than 5 TB/h for Scale-out. However, restores were much slower, making this option less than ideal for crtitical production systems.

In cases where we disabled deduplication and just used compression, we could observe backup performance going down a bit, but still saving around 30 - 35% of Blob storage space. This test case had good restore performance, meaning that a RTO of an hour is achievable for our test databases.

The best option for large and/or critical production systems is certainly to enable Intellisnap for Azure. Remember, only /hana/data will get snapped. Just a couple of seconds backup time for a 1.5 TB Scale-up database – it's very hard to beat that. As for restores we achieved 12 minutes for 1.5 TB Scale-up and 20 minutes for Scale-out which is exceptional performance. In practice, it is expected that the recovery time will become a bit longer as more logs need to be consumed to recover the database. However, this effect can be limited by creating more than one recovery point, i.e. snapshot backup per day. Depending on the data retention time, a certain amount of snapshots need to be retained which is a cost factor that needs need to be taken into consideration. Blob consumption is driven here by snapshot copies you may want to create and the streaming data consistency backups.

To summarize, the recommendation is to look at RTO before making a decision on which of the three backup methods should be used. For a SAP production system, one option is to use streaming protection with client-side compression and without deduplication. Backups can run for longer times as CPU resource impact is still moderate. For production systems being very large and/or with aggressive RTO requirements we recommend to leverage snapshot backups. For test, development and sandbox systems, RTO requirements are usually lower, so deduplication can be used for minimzing the Blob storage footprint.

Using deduplication along with the previously introduced Commvault VSA agent is also the recommended approach for protection all standalone SAP application instance VMs.

# Reference Architecture

Now it's time to build a reference architecture for a typical customer environment. Let's consider this example SAP environment consisting of four application landscapes. Depending on the workload use case (PROD, DEV, TEST) we will have different Restore Time Objectives (RTO) and retention requirements to complete the example.

| | SAP S/4HANA | SAP BW/4HANA | SAP CRM | SAP PLM | RTO |
|---|---|---|---|---|---|
| Production | 1500 GB | 3000 GB | 800 GB | 500 GB | 1h |
| Development | 1500 GB | 3000 GB | 800 GB | 500 GB | 5h |
| Test | 1500 GB | 3000 GB | 800 GB | 500 GB | 5h |

Total source data volume is approximately 17.4 TB. Data retention time for database and log backups is 14 days. S4/HANA and CRM production systems are deployed as HANA scale-up systems and have a RTO SLA of 1 hour. SAP BW/4HANA is deployed as a 3-node scale-out architecture. BW and PLM production has a RTO SLA of 3 hours. All test and development systems have a RTO of 5 hours.

As S/4HANA and CRM production systems are critical, we need to select a data protection approach which not only solves today's challenges but will continue to meet the SLAs in a year or two as databases grow over time. Leveraging snapshot backups and restores will meet this requirement best. However, in order to ensure SAP HANA data consistency, we need to plan for at least one full streaming backup per week. Combining a full backup with daily incrementals would be ideal.

For protecting test and development systems we recommend enabling deduplication and compression, which will not only meet the SLA, but also help in reducing the Azure Blob storage footprint.

Additionally, disaster recovery for the backup solution between two Azure regions or two Azure Availability Zones needs to be considered. All backup data from region A (or Zone1) needs to be available in region B (or Zone2). In case of an Azure Region of Availability Zones experiencing an outage, all SAP applications need to be restorable in the surviving region.

## Storage requirements

### Basic considerations

The SAP HANA agent backs up the following:

- All database files on each node
- Log files on each node
- SAP HANA backup catalog

For SAP HANA data the following backup types are available:

- Full (snapshot or streaming)
- Incremental
- Differential

In order to protect the SAP application instance footprint as well as /hana/shared, the Commvault file server agent which is part of the installation is recommended to be used.

A typical customer with a retention requirement of 2 weeks and running all backups at a full level, would have 14 full backups stored in their Azure Blob storage based cloud library. If the customer decides to run only one full backup per week and incremental or differential backups on the other days, less storage would be consumed. Incremental backups would store all changes since the last backup and the differential stores all changes since the last full backup.

So if performing one full backup per week, a differential backup half way through the week and incremental on all other days, we would see less storage consumed, but an additional cycle will be retained because the incremental and differential jobs rely on their parent backups. With a 14 day retention time and a full backup happening on day 1, the subsequent full backup of day 8 will be expired on day 20 because of the dependent delta backups. When using the full-only backup cycle, there will be fewer jobs retained, but more Blob storage would be consumed anyway.

Snapshot backups take only minutes, so they can performed multiple times per day. This gives you multiple daily recovery points and reduces the number of log files to be processed during a recovery operation.

### Production Systems

In this example we consider a larger daily database change rate (churn) of 10%. For the BW/4HANA and PLM production systems we plan streaming protection with daily fulls for which we assume an average compression ratio of 30% (compression only).

S/4HANA and and CRM are considered as the most critical production systems and require snapshot-based protection. As it is anticipated that production restores usually happen within a 7 days timeframe, all snapshots are retained for this time only. In addition we need to plan for a consistency backup. A weekly full combined with daily incrementals will fulfill this task. We can also restore from these jobs if snapshots have expired already. This allows to save Blob storage as we don't

need snapshot copies. Running 3 snapshot backups per day should be sufficient for minimizing log replay time and keeping within RTO. This schedule can be further optimized for saving cost. For cost related to Azure Manged Disk snapshots you can refer to this link: https://azure.microsoft.com/en-us/pricing/details/managed-disks/.

So how much Snapshot and Blob storage do we need for implementing this 14 day backup stategy? The following table summarizes the storage amount needed for the database backups.

| Production Systems | S/4HANA, CRM Consistency Backup | BW/4HANA, PLM Daily Full | S/4HANA, CRM Snapshots |
|---|---|---|---|
| Full | 2.07 TB | 34.3 TB | 126 (6 x 3 x 7) |
| Incremental | 3.91 TB | | |
| Differential | - | | |
| Total backup storage required | 5.98 TB Blob | 34.3 TB Blob | 126 Premium Disk Snaps |

#### Test and Development Systems

For the test and development systems, we are assuming a daily change rate of 5%. With deduplication and compression enabled we can assume a data reduction rate of around 70%. As with BW/4HANA and PLM production we can run full backups only. Another option would be to run just one full per week, combined with daily incrementals and differential which capture the changes. This leads to slightly longer restore times, but helps to limit Azure blob storage footprint. To put things in numbers; running fulls only requires 48.7 TB vs. 19.7 TB with the levelled backup approach (based on 5% daily change rate). This represents a factor of 2.5.

| Test and Development Systems | Only Full | Levelled backups |
|---|---|---|
| Full | 48.7 TB | 10.4 TB |
| Incremental | - | 5.7 TB |
| Differential | - | 3.6 TB |
| Total TB Azure Blob storage consumed | 48.7 TB | 19.7 TB |

Incremental and differential backups need to be managed by a non-deduplicated storage policy, as those data is considered unique. The same applies to SAP HANA log files whereas deduplication needs to be enabled for protecting SAP application instance data and /hana/shared via the file system or VSA agent. In terms of storage capacity, we recommend adding another 1TB of Blob storage capacity for protecting SAP HANA logs and SAP application instance files.

Finally, all storage polices will use the same Commvault cloud library. It consists of four Azure Cool Blob storage containers as mountpoints where backup data is randomly written to each storage account to load balance for read/write performance, and to overcome any storage account throughput limitations.

## Commvault infrastructure

### Single Azure Region

At a minimum, you need a Commvault Commserve and one MediaAgent to control a cloud library. We suggest to set up two MediaAgents to share the cloud library and to configure them to handle redundancy/failover and load balancing. This can be achieved via Commvault's GridStor technology. One of the Media Agents can be powered down during times with limited backup and restore job activity (e.g. outside of the main backup window), to help save costs. Restores are still possible using the other MediaAgent which remains in operation for handling the SAP HANA logs. Please refer to the Commvault Public Cloud Architecture Guide for Mircosoft Azure for sizing recommendations for Commserve and MediaAgents.

Please note that you need another Windows VM for running a so-called VSA proxy for protecting SAP application instances via the VSA method. As for storage accounts, we recommend to use General-Purpose V2 and Blob Cool storage containers for building libraries.

### Disaster Recovery across Azure Regions

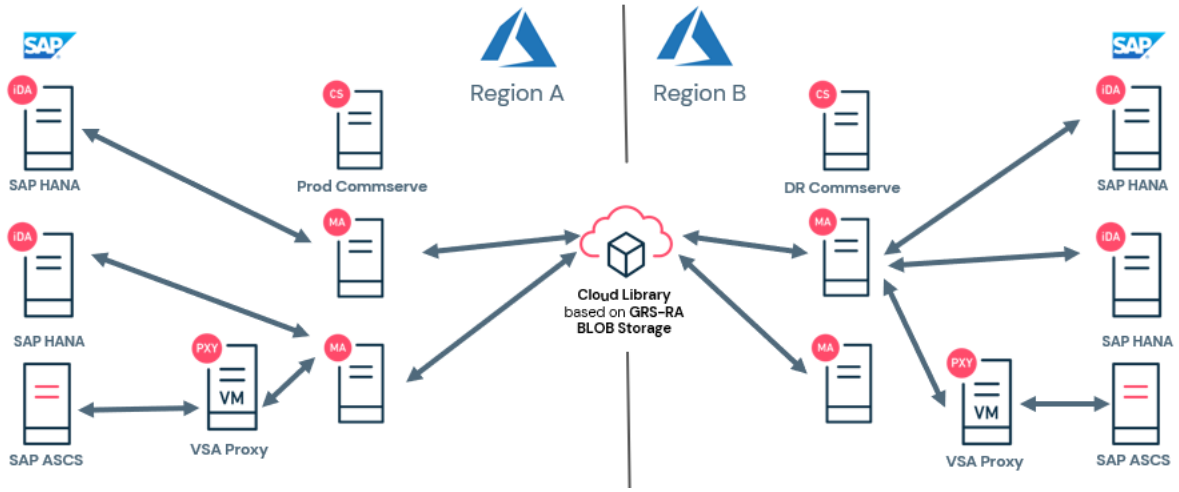
In case a disaster recovery solution is required, the Commvault setup needs to be designed to span two Azure regions. As a first step, the Commserve needs to be installed in a high availability setup. For instance, a standby Commserve VM can be set up in the DR region in conjunction with Commserve LiveSync for keeping the DR Commserve in sync ready for failover. For building cloud libraries, Azure GRS-RA Blob containers are recommended to be provisioned so we can leverage Azure storage-level replication. That means all backup data from primary region A will be automatically and asynchronously replicated to DR region B. We further require at least one pair of MediaAgents in each region. All cloud libraries need to be shared within the region and also cross region.

During normal operation, backups and snapshots run as scheduled in the primary region. For all snapshots, a backup copy process can to be scheduled to make these backups available in the DR region. Another option would be to rely on the SAP HANA data consistency backups for DR purposes. All other backups go to Azure Blob anyway.

At any time, all SAP data can be restored in the DR region, while backups in DR to the same Azure Blob containers are not. After failover, any production system from the primary region can immediately be restored in DR. As the GRS-RA blob containers become writeable after failover the restored SAP systems can be backed up to the same containers making a failback operation simple.

It should be pointed out that this architecture is just one option for DR. An alternative architecture can, for instance, be designed using Commvault's built-in backup copy technology which is called Auxiliary Copy or auxcopy. In this case backup jobs will be copied between MediaAgents across the network. In case of deduplicated backups only changed blocks with be sent.

For very critical SAP production systems, SAP customers will mostly likely configure SAP HANA System Replication where the passive secondary node will be placed in the DR region. In this case, Commvault protects the primary either with streaming backups or via snapshots and has the ability to follow the primary role in case of takeovers.




# Summary and Conclusion

Many SAP customers are currently looking at migrating their SAP environments to run SAP HANA and in Microsoft Azure. These customers are experienced with professional, enterprise-level solutions like Commvault to protect their on-premises SAP environment and expect the same level of protection when they move to Azure.

Starting from an Azure test environment, which was built around the powerful M128s SAP HANA VM and Azure Cool Blob storage, this paper has discussed various approaches on how to protect SAP HANA in Azure and ensure recovery SLAs can be met. No matter which SAP workload type you are using, the right solution can be built. We showed that Commvault's data protection solution for SAP HANA in Microsoft Azure is not only tightly integrated but also market leading. We have validated multiple ways to bring back mission-critical SAP HANA production databases – if required within minutes using Intellisnap for Azure - and how test and development systems can be protected efficiently.

We've also put together a reference architecture for a typical customer environment (including a Disaster Recovery option), discussing various backup strategy options and their impact on Azure Blob storage consumption. Azure Premium Storage showed excellent performance both for running Commvault's internal Deduplication and Index databases.

Commvault and Microsoft have teamed up and demonstrated how Azure and Commvault together can be leveraged to provide an enterprise-level solution for data protection which is cost-efficient, automated and flexibel enough to meet a wide range of SLAs, supports all important SAP configurations, databases and use cases for safe-guarding your SAP HANA workloads in Azure.