COMMVAULT®

Public Cloud Architecture Guide for Microsoft Azure

COMMVAULT® VERSION 11 SP14

Notices

This document is provided for informational purposes only. It represents Commvault's current product offerings and practices as of the date of issue of this document, of which are subject to change without notice. The responsibilities and liabilities of Commvault to its customers are controlled by Commvault® agreements, and this document is not part of, nor does it modify, any agreement between Commvault and its customers.

Revision History

VERSION	DATA	CHANGES
1.0	March 2015	• Initial Version
1.1	May 2015	• Updates
1.2	June 2015	• Added new Architecture Consideration sections - Networking (Azure AVN), Infrastructure Access, Performance / Storage • Added new Installation sections - Bring Your Own License / Software sections (Installation), Video Tutorial links • Added new Additional Architecture Resources section • Updated document & layout for new Commvault® branding • Updated core cloud concepts and technology and Azure Sizing Recommendations and Security (Architecture Considerations) section • Modified section layout • Removed Data Aging caveats with SP11 Micro-pruning for Cloud Storage release, replaced text to refer to this only for pre-SP11 sites
1.3	July 2015	• Updated with new trademark guidelines
1.4	August 2015	• Minor reformatting • Added new links to video content
1.5	September 2015	• Added Selecting the right Storage Class section • Minor reformatting
1.6	November 2015	• New logo style • Updated requirements for disaster recovery to the Cloud • Updated cloud use case diagrams • Updated documentation links for VM Recovery to Azure • Added Unsupported Cloud Configurations section
2.0	March 2016	• Updated to reflect new Virtual Server Agent methodologies, deployment and changes to use cases • Updated Backup to the Cloud, DR to the Cloud and Protection in the Cloud use case scenarios and requirements • Updated micro pruning section • Updated Drive Shipping to add note about Snowball support arriving 2016 • Updated all BOL links to use Commvault® Version 11 documentation • Added new Documentation section to Additional Resources • Added Automating Deployment with Puppet/Chef section • Added Pre-packaging Commvault® within a VM Template section • Removed Azure Marketplace component • Minor reformatting changes

VERSION	DATA	CHANGES
2.1	June 2016	• Updated Backup to the Cloud use case for more clear language around DDB requirements • Added VSA for Azure
2.2	September 2016	• Added Migration to the Cloud use case and Application Migration section • Fixed error in data seeding table • Minor updates to Backup/Archive to the Cloud use case verbiage • Updated links to 2 Clicks to the Cloud videos, added new Backup videos for Azure • Updated micro pruning statement on Azure page blobs • Updated all original "Agent-in-Guest" references to "Agent-in-Guest (Streaming)" • Added Azure Reference Architectures • Updated verbiage on Selecting the right Storage Class
2.3	March 2017	• Revised cloud library storage deduplication performance recommendations • Added Live Sync DR for Azure and revised DR structure • Added Blob storage backup feature • Added GET/PUT storage cost considerations • Added partitioned dedupe recommendation • Updated Azure VM Names / Sizes to align with recent changes • Added information about leveraging multiple mount points for cloud libraries • Added EC2 to Azure conversion feature • Added Azure CBT feature • Added Azure VSA Commvault IntelliSnap feature • Updated Azure VSA features and recommendations (Advanced file system restore, CBT) • General grammatical fixes
2.4	April 2017	• Updated Cloud pricing • Updated Azure throughput and timing diagram • Changed a few areas to Commvault® where Simpana was referenced
2.5	August 2017	• Added to Azure Architecture Requirements • Changed Azure Proxy recommendations • Changed CS & MA sizing wording • Added Windows 2016(SP7) to CS and MA
2.6	September 2017	• Added support for Azure Managed Disk VMs (Beta) • Added Azure Archive Storage • Added Extra Large specs for Azure media agents • Added recommendations on dedupe block size to use in hybrid environments
2.7	February 2018	• Added Azure Cloud diagram • Updated Azure Managed Disk information
2.8	May 2018	• Review entire document and update with new content • Updated MediaAgent VM sizing for all Clouds • Included Move, Use, Manage use cases • Updated document to make specific for Microsoft Azure
2.9	October 2018	• Added Back Up and Restore Azure SQL Databases • Fixed hyperlinks • Added Live Sync Replication • Added Azure SQL protection • Added performance measurements for VSA and Agents • Added Azure File share protection
3.0	December 2018	• Added IntelliSnap support for Azure Managed Disks • Azure snapshot support for large disk (<8TB)

Table of Contents

ABSTRACT

This document serves as an architecture guide for solutions architects and Commvault customers who are building data protection and management solutions utilizing public cloud environments and Commvault® software.

It includes public cloud concepts, architectural considerations, and sizing recommendations to support Commvault® software in public cloud. The approach defined in this guide applies to both running Commvault solely in public cloud environments and extending existing on-premises Commvault® functionality into hybrid cloud architectures. The guide covers several common use cases for public cloud including moving data to public cloud, disaster recovery to public cloud, as well as protecting workloads in public cloud.

Currently this guide delivers architecture considerations and sizing recommendations for the Microsoft Azure public cloud platform. Guides for other public cloud environments are available as well.

THE CLOUD ADVANTAGE

The public cloud megatrend is one of the most disruptive and challenging forces impacting customers' applications and infrastructure, requiring new business models and new architecture decisions. This impacts the decisions about solutions for the protection and management of data in public cloud.

In general, Commvault believes the public cloud contains these attributes that upon which its value proposition focuses.

INFRASTRUCTURE AS PROGRAMMABLE AND ADDRESSABLE RESOURCES

In a traditional on-premises, non-cloud, environment: (i) infrastructure assets must be manually configured, (ii) capacity requires manual tracking, (iii) capacity predictions are based on the guess of a theoretical maximum peak, and (iv) deployment can be measured in days to weeks.

Within cloud, these infrastructure building blocks are not only provisioned on-demand as required driven by actual usage, but can also be programmed and addressed by code allowing for a cost-effective pay-as-you-go model. This greatly enhances flexibility for both production, non-production environments for scenarios such as development, testing, and disaster recovery.

Resources in cloud environments can be provisioned as temporary, disposable units, freeing users from the inflexibility and constraints of a fixed and finite IT infrastructure. Infrastructure provisioning is automated through code, allowing for greater self-service and more agile delivery of desired business and technical outcomes.

Intelligently provisioned and managed resource consumption in cloud environments is measured by what you consume, not what you could consume, drastically changing the cost model challenges experienced today in traditional on-premises architectures that typically operate on a three-year to five-year technology refresh cycle.

This represents a major, disruptive reset for the way in which you approach common infrastructure for data usage such as secondary backup storage, long-term archiving, disaster recovery, new application development and testing, reliability and capacity planning for bursting production workloads. Commvault utilizes this attribute of public cloud to enable cost effective on-demand use cases for both data protection and data management both to and in public cloud platforms.

GLOBAL, FLEXIBLE AND UNLIMITED RESOURCES

Public cloud providers offer globally distributed infrastructures available to customers on a pay-as-you-go model, allowing for more flexibility in meeting requirements for both geographically distributed workloads and recoveries.

Cloud resources, bandwidth, and their availability, are often localized via massive regional presence to proximity of on-premises corporate assets and human resources, allowing for an easy on-ramp to public cloud. The cost model implications of pay-as-you-go do not just extend to only production workloads, but also to the ever-present challenge of providing a flexible, agile, yet capable, recovery solution for your applications and data. Today, many recovery environments have less compute and storage capacity than their production counterparts, resulting in degraded service in the event of a disaster. Even more so, hardware is often re-purposed to fulfill both the recovery requirements as well as non-production uses, resulting in higher than expected maintenance costs and slowed recovery times.

With the public cloud model, the infrastructure availability and refresh aspect are disrupted by removing the need to maintain a hardware fleet that can meet both your recovery requirements and sustain your service level agreements. Public cloud VMs can be rapidly provisioned to meet the needs tied to business requirements, rather than purchasing cycles. For specific recovery events – both real and simulated – the underpinning hardware is maintained and upgraded by the cloud provider without any need for technical input, and no upgrade costs are incurred by the organization.

This dynamic shift allows you to begin costing per recovery event, instead of paying for availability, improving your level of disaster recovery preparedness through the application of flexible, unlimited resources to stage both recovery tests and execute actual recovery events – all without requiring pre-purchased hardware or disrupting production operations. While the recovery use case is the most common foray into a public cloud architecture, many other use cases such as application testing and development, business intelligence and analytics, and production bursting all benefit from the public cloud model.

Commvault® software is designed as an orchestrated, hardware and cloud agnostic, highly modular, distributed solution that conforms with this new architecture reality, allowing data protection and management solutions to be built to support and remain flexible with a highly distributed infrastructure built on-top of cloud architecture – public, private or hybrid.

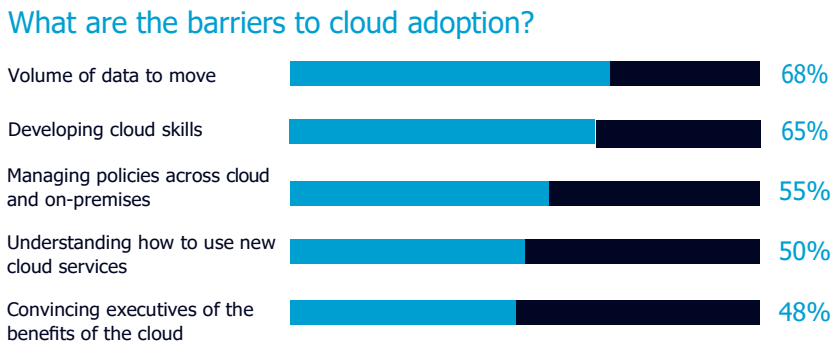
BARRIERS TO CLOUD ADOPTION

In this section, we outline some of the challenges that have been observed when it comes to on-ramp to public cloud platforms.

Due to the existence of a variety of public cloud platforms, many organizations are deploying multi-cloud architecture is being deployed out of need to overcome their technological challenges and to optimize costs for specific types of services. In some cases, this is being coupled with in-house developed private cloud environments that operate with the agility and flexibility of public cloud for specific data types. The resulting reality for most environments is a hybrid multi-cloud architecture, which offers optimal flexibility and control of the costs.

With the formation of a hybrid multi-cloud environments, some common challenges surface with respect to data management and protection. The human element is fundamental to cloud transformation. Having the appropriate skillsets available in a timely manner to create, administer, and manage the data across cloud environments becomes a top priority and can quickly result in a costly offset to the benefits of cloud being originally sought. The number of different toolsets and techniques to perform the same operation can become an overwhelming decision point and can result in a steep learning curve coupled with considerable custom automation and added cloud cost if not done correctly. Lastly, the risks associated with the volume of data movement to, from, and across clouds and their associated regulatory and compliance implication on your data being placed in public cloud often warrants a deeper examination, before embarking on cloud adoption.

An executive survey around cloud adoption produced the following data points:



Source: 2017 Executive Cloud Survey, CITO Research and Commvault

Regardless of cloud preferences, the data is yours and needs to adhere to the standards set forth by your organization with respect its management, compliance, and protection. With the Commvault® platform, several barriers discussed above get mitigated by having a single software-defined solution that provides the following capabilities:

- native integration with multiple cloud platforms compute, application, and storage layers
- single pane of glass for managing all data in hybrid environments – public and private clouds and multiple hybrid clouds
- cost optimization in public cloud by usage-based scaling infrastructure
- integrated global deduplication and encryption
- data portability through cross platform conversion and recoveries

COMMVAULT PLATFORM DESIGN PRINCIPLES FOR CLOUD

In this section, we provide design principles and architecture principles that have been employed within the Commvault® platform to provide an optimal cloud experience for organizations planning to leverage the cloud as part of their data protection and management strategy.

NATIVE CLOUD CONNECTIVITY

The Cloud Library feature of a MediaAgent is the native integration within the Commvault® platform that directly communicates with object storage such as Azure Blob Storage - Hot, Cool, and Archive storage tiers, and many others* including on-premises object storage without requiring translation devices, gateways, hardware appliances or Virtual Tape Libraries (VTLs).

This Cloud Library works by communicating directly with object storage's REST API interface over HTTP or HTTPS, allowing for Commvault platform deployments on both virtual and physical compute layers to perform read/write operations directly against cloud storage targets, reducing the TCO of the data management solution. The Cloud Library is part of the native code of the Commvault platform, and it optimizes the data exchange with cloud object storage platform to maximize the transfer speed while minimizing recall needs and costs.

Since the Cloud Library essentially treats cloud storage akin to a disk target, data management functions such as compression, encryption, deduplication, and data life-cycling can be performed against cloud storage targets to ensure that both costs and risks are managed effectively. This also allows the data to be retained independent to the cloud format thereby enabling optimized recall and movement of data across different cloud platforms for future use cases.

*For more information on all the supported vendors, please refer to this comprehensive list located in Commvault online documentation.

SCALABILITY

Application environments and the data and VMs that service those environments grow over time, and a data protection and management solution needs to adapt with the change rate to protect the dataset quickly and efficiently, while maintaining an economy of scale that continues to generate business value out of that system.

Commvault addresses scalability in cloud architectures by providing these key constructs:

DEDUPLICATION BUILDING BLOCKS

Commvault® software maintains a scale-up or scale-out "building block" approach for protecting datasets, regardless of the origin or type of data. These blocks are sized based on the front-end data they will ingest, prior to compression and deduplication. This provides clear scale-out and scale-up guideline for the capabilities and requirements for each Commvault MediaAgent that will perform the data movement (both ingestion and storage), compression and deduplication.

Furthermore, these deduplication MediaAgent building blocks may be logically grouped together in a grid formation, providing further global deduplication scale, load balancing, and redundancy across all nodes within the grid.

This software on architecture, with scale-up and scale-out enables cloud adoption to start with a cost-conscious approach however scales to meet SLAs quickly without locking the architecture into a specific unit of operation.

CLIENT-SIDE DEDUPLICATION

As is the nature of deduplication operations, each data block must be hashed to determine if it is a duplicate block, or unique; and then must be captured. While this is a way to improve the ingest performance of the data mover (MediaAgent), it has the secondary effect of reducing the network traffic stemming from each client communicating through to the data mover.

In public cloud environments where network performance can vary, the use of client-side deduplication can reduce backup windows and drive higher scale, freeing up bandwidth for both production and backup network traffic. By utilizing client-side deduplication, the workload of backup can be distributed across all the VMs, compared to building a larger data protection architecture in cloud. This can also help reduce the recovery points for critical application by enabling more frequency of protection.

VIRTUAL SERVER AGENT (VSA) OR PROXY FOR CLOUD PLATFORMS

Utilizing agents in each cloud operating VM is an approach that distributes the overall workload and cost for data protection across all the VMs. However, in many cases with large scale deployments, management of each VM can become an overhead. The Commvault platform automates the management of such operations from initial deployment to upgrading and disposition. When this approach is deemed insufficient, the Commvault Virtual Server Agent (VSA) proxy software capability can be loaded into a public cloud VM to perform agent-less operations.

Akin to proxy based protection for on-premises hypervisors, the Commvault VSA* proxy, interfaces directly with APIs available from the hypervisor layer of public cloud platforms to perform protection and management operations of VMs within the public cloud platform. The VSA not only manages operations such as snapshot creation and orchestration, but can also perform automatic VM identification and selective VM data reads (Change Block Tracking) from cloud platforms that support this capability. The VSA further performs any data format conversions and enables disaster recovery operations for VMs to cloud platforms. Working together with the MediaAgent (data mover), the VSA offers enhanced protection and management of cloud workloads.

*For a complete list of supported VSAs and updated VSA capabilities please review the online VSA Feature Comparison Matrix.

DESIGN FOR RECOVERY

Using native cloud provider tools, such as creating a snapshot of a cloud-based VM may be easy to orchestrate, but does not always deliver the application-consistency required by an application or database such as Microsoft SQL Server or Oracle Database residing within the VM. The general approach requires database and application-specific scripting or manual handling to deliver a successful application recovery. Across a large enterprise estate, this bespoke manual management becomes time-consuming and subject to human error.

As part of any data protection and management solution, it is important to ensure that you design for recovery in order to maintain and honor the recovery time objective (RPO) and recovery point objective (RTO) requirements identified for your individual applications groups.

CRASH CONSISTENCY VERSUS APPLICATION CONSISTENCY

While crash-consistency within a recovery point may be sufficient for a file-based dataset or cloud VMs such as Azure VMs it is not generally appropriate for an application such as Microsoft SQL Server or Oracle Database where the database instance needs to be quiesced to ensure the database is valid at the time of backup.

Commvault® software supports both crash and application consistent backups, providing flexibility in your design while assuring VM recoverability coupled with application recovery to a specific point in time. Not only are the most common types of applications covered, but a wide variety of classic applications and cloud applications are supported. For a complete list of updated application support please review the online documentation.

STORAGE-LEVEL REPLICATION VERSUS DISCRETE INDEPENDENT COPIES

Many cloud providers support replication at the object storage layer from one region to another, however, in the circumstance that bad or corrupted blocks are replicated to the secondary region, your recovery points are invalid. Further network and storage costs continue to accumulate regardless of the validity of both "sides" of the data.

While Commvault® software can support a replicated cloud library model, in which the secondary storage location for backups is replicated using the Cloud vendors storage-based replication tools, we recommend that you consider Commvault® software to create an independent copy of your data, either to another region, or cloud provider, or back to an on-premises infrastructure to address broader risks. Deduplication is also vital as part of the latter option and this ensures that Commvault® software can minimize the cross-region and cross- provider copy time and costs by ensuring only the unique changed blocks are transferred over the network. This recommendation not only ensures recoverability to multiple points in time, it further manages the cost and risk through the assurance that the data is independent of the platform and ensures that different SLAs for protection and retention can be maintained for different classes of data.

DECIDING WHAT TO PROTECT

Not all workloads within the cloud need protection – for example, with micro services architectures, or any architecture that involves worker nodes that write out the valued data to an alternate location, presents no value in protecting the worker nodes. Instead, the protection of the gold images and the output of those nodes provides the best value for the business. However, it is important to note that data stored in ephemeral locations may need to be protected prior to termination operations against those VMs to ensure that any valuable data is not lost.

DESIGNED FOR CLOUD EFFICIENCY

As already discussed, the ability to provide compression and deduplication for both data to and data in the cloud begins to provide initial cost savings for many of the common use cases for secondary data. However, deduplication savings are closely tied to the type of data being managed and additional methods can result in even more overall cloud efficiency.

A common consideration is to utilize multiple tiers of storage for data as the service life of that data reduces. This has been a common practice on-premises and the Commvault platform extends this capability to cloud platforms. By having not only native integration to primary object storage targets such as Azure Blob Hot storage tier, but also having native access to more cost-effective tiers such as Azure Blob Cool storage tier and Azure Blob Archive storage tier, data lifecycle management can be performed within the cloud. For example, it is not uncommon to see Azure Hot storage tier being used as the primary or secondary copy for short-term retention followed by Cool storage tier or Archive storage tier. Having a data management platform that can utilize SLA policies to orchestrate the data movement and be aware of the location of data for recall and disposition becomes a valuable quality in gaining cloud efficiency.

CLOUD POWER MANAGEMENT

Shutdown of VMs in an on-premises data center is a very uncommon practice or design concept, however in cloud environments this type of operation is welcomed by those paying the cloud bills. By having the ability to create policies which monitor resource usage of cloud VMs and can both alert and act by terminating such VMs. However, the risk of data loss is mitigated since it is ensured a copy of ephemeral data is protected before such an operation is performed.

The ability to shutdown VMs is extended to Commvault platform components running in the public cloud. Referring to the MediaAgent (data movers) referenced above, these VMs can be both shutdown and powered- up via a policy operating on a proxy running in public cloud or on-premises. The trigger events are around data protection operations. For example, shutting down the Cloud MediaAgent after all protection operations have ceased and restarting prior to the next SLA window can help further reduce operational costs within public cloud environments.

AUTOMATION

The cloud encourages automation, not just because the infrastructure is programmable, but the benefits in having repeatable actions reduces operational overheads, bolsters resilience through known good configurations and allows for greater levels of scale. Commvault® software provides this capability through three key tenets:

PROGRAMMATIC DATA MANAGEMENT

Commvault® software provides a robust Application Programming Interface (API) that allows for automated control over deployment, configuration, and backup and restore activities within the solution.

Whether you are designing a continuous delivery model that requires automated deployment of applications, or automating the refresh of a disaster recovery copy, data warehouse or development/testing environment that leverages data from a protection copy, Commvault® software provides the controls necessary to reduce administrative overhead and integrate with your toolset of choice.

Beyond API access, the most common use cases for data protection and management are built into the Commvault user interface. Simply enter the cloud credentials and necessary permission and the Commvault platform will query the cloud environment accounts and present wizards with necessary attributes to create VMs and populate with the data required to support the above uses discussed. Since format conversions are handle by the VSA, the entire operation is orchestrated even if the source of data is an on-premises hypervisor. This reduces the operational overhead and unique skillsets required to on-board cloud usage.

WORKLOAD AUTO-DETECTION AND AUTO-PROTECTION

The Commvault® Intelligent Data Agents (iDA), whether via the Virtual Server Agent for the various cloud platforms, or the multitude of application and database iDAs, provide auto-detection capabilities to reduce administrative load.

Fresh instances, new volumes recently attached to cloud instances and virtual machines, or databases imported and created into a database instance are some examples of how Commvault® software automatically detects new datasets for inclusion in the next data protection SLA window, all without manual intervention. Even agent-in-guest deployments can be auto-detected by Commvault® software and included in the next data protection schedule through intelligent Client Computer Groups. This capability is especially valuable in the assurance of data protected in large scale cloud environments where many users can provision workloads in the cloud, but may have little or no consideration for the protection of those workloads.

This auto-detection and auto-protection level removes the requirement for a backup or cloud administrator to manually update the solution to protect the newly created datasets. This results in improving your operational excellence, improving resiliency within your cloud infrastructure, and ensuring new data protected SLAs are maintained.

SELF-SERVICE ACCESS AND RESTORE

A common task performed by system administrators is facilitating access to recovery points for end-users and application owners, shifting their attention away from other day-to-day operations and strategic projects.

The Commvault self-service interfaces empower users to access their datasets through a web-based interface, allowing security mapped access to individual files and folders within the protected dataset, freeing up administrators to work on critical tasks. Commvault's robust role-based security function provides assurance that self-servicing users have access to only their data assets, while bespoke auditory reporting capabilities capture how these users are accessing those data assets.

MOST COMMON CLOUD USE CASES WITH COMMVAULT® SOFTWARE

The most common use cases observed at most customer environments by Commvault related to cloud, fall into three categories depending on the maturity level of initiatives around cloud adoption:

- Move data to the cloud – typically involves using public cloud object storage as a target for backups and archive data and moving certain types of VM workload into cloud VMs.
- Manage data in and across clouds – protecting and life-cycling data and VMs in cloud, moving data across clouds and back to on-premises in some cases.
- Use data in the cloud – utilizing the data stored in public cloud for use cases such as disaster recovery, dev/test, and other production and non-production use cases.

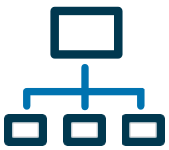
These three primary use cases can be visualized as follows:

Move data

Seamlessly extend the datacenter to the cloud.

Manage data

Unlock IT agility with a comprehensive view of data.



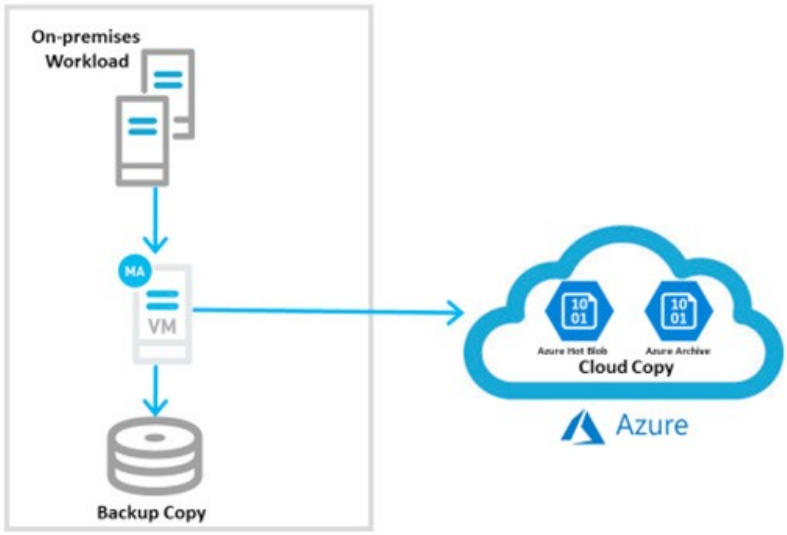
Use data

Enable a more strategic, customer-focused business.

Each use case can have multiple phases and types of data associated. For example, movement could involve simple backup data, but can graduate to workloads being moved back and forth for agility as an extension to on- premises. Management of data can start with basic snapshot management and graduate to complete data lifecycle management with cloud snapshots, operational recovery deduplicated copies, and archive of data coupled with searching and indexing for compliance. The use of data can involve uses such as disaster recovery that eliminate the need to sustain secondary on-premises sites and utilize the agility of the cloud to on-ramp recovery testing and real recoveries.

MOVE DATA – BACKUP AND ARCHIVE TO THE CLOUD

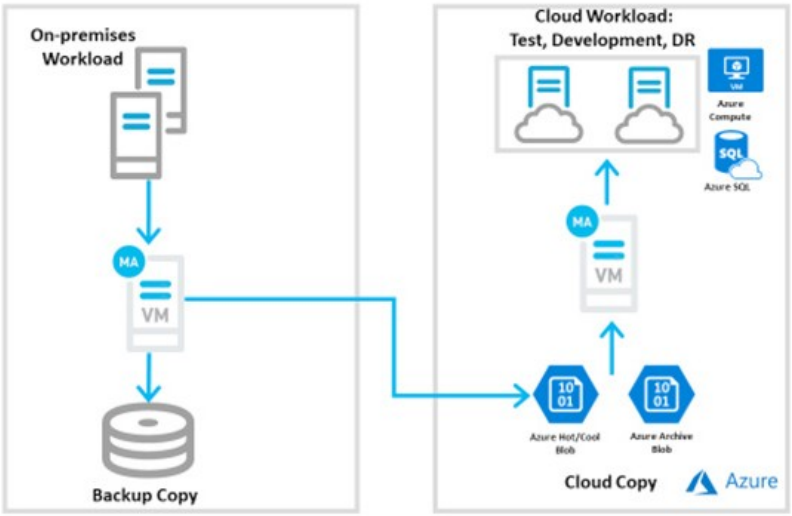
Business Value: Protecting data at the primary on-premises location by writing directly to an external cloud provider's storage solution, or retaining a local copy and replicating the backup and archive data copy (either in full, or only selective portions of that data) into an external cloud provider's storage service suitable for both short and long-term retention configurations.



SCENARIO / SUITABILITY	REQUIREMENTS
• Offsite Storage and "Tape Replacement" Scenario – replace long-term on-site retention with cloud storage. • Native, direct connectivity to 40+ object storage endpoints – no requirement for translation, gateway, and hardware deduplication devices. • Avoid point solution on a per application basis. Any data (physical or virtual) that can be backed up by Commvault® on-premises can be moved to cloud. • Cloud object storage target can be provided by either public IaaS provider (Microsoft Azure) or via a Managed Service Provider.	• Minimum 1x MediaAgent on-premises with no VMs in cloud required for backup to the cloud • Can use direct internet connection, or a dedicated network to cloud provider for optimized data transport performance in a secure manner (e.g. Azure ExpressRoute) • In-cloud MediaAgent can be created to support DR solution to cloud using the data that is placed in cloud. This can be done at time of DR or DR Test, as required.

MOVE DATA - MIGRATION OF VMS AND APPLICATIONS TO THE CLOUD

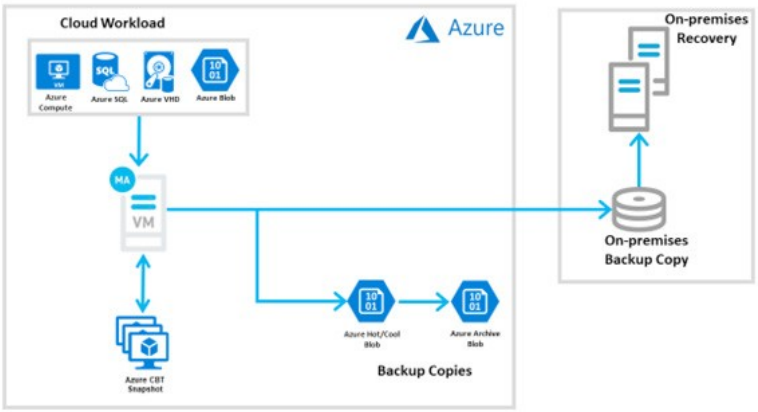
Business Value: Upon protecting VM and application data at the primary on-premises location, Commvault® software orchestrates the migration of application workloads into the cloud, either at the VM container level or the application level. While providing the migration lifecycle and workloads are in a transition phase between on- premises and public cloud, data is still protected on-premises.



SCENARIO / SUITABILITY	REQUIREMENTS
• Lift & Shift of virtual machines – Application-consistent VM backups are used to restore and convert VMware and Hyper-V VMs into Azure as part of a migration with a phased cut-over strategy reducing on-premises downtime. • Application Restore Out-of-Place –Leverage Commvault® iDataAgents for your supported workload to restore the target application out-of-place to a warm VM residing in cloud.	• Minimum 1x MediaAgent on-premises to protect and capture workloads • Minimum 1x MediaAgent (& DDB) in cloud to protect workloads post-migration in-cloud, and for optimal migration performance. • Highly recommended to use dedicated network to cloud provider for best performance (e.g. Azure ExpressRoute).

MANAGE DATA - PROTECTION IN THE CLOUD

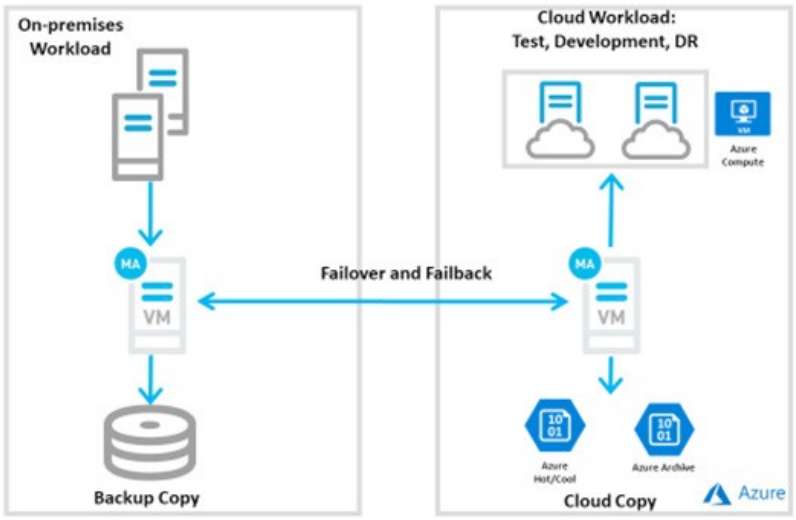
Business Value: Providing operational recovery for active workloads and data within an external provider's cloud. Provide the ability to lifecycle data and cloud VMs to meet SLA and cost requirements.



SCENARIO / SUITABILITY	REQUIREMENTS
• Data protection for Cloud based workloads – protecting active workloads within an existing IaaS Cloud (Production, Dev/Test, etc.). • Azure Agentless VM Protection – Protect VMs with an agentless and script-less protection mechanism through the Virtual Server Agent. • DASH Copy data to another region, cloud, or back to on-premises – Complete data mobility by replicating to another geographical region within IaaS provider, a different IaaS provider, or back to on-premises sites. • Protect Azure Blob storage – Backup object storage repositories with data created by other 3rd party applications either in cloud, to an alternative provider, or back to on-premises sites.	• Azure Virtual Server Agent and MediaAgent deployed on a proxy within IaaS provider for agentless backup. Applications will require agent- in-guest deployed in VM. • Applications requiring application-level consistency, and all other cloud providers can be protected via agents deployed in each VM within Azure. • Minimum 1x MediaAgent in cloud, and (optional) minimum 1x MediaAgent at secondary site (whether cloud or on-premises) for receiving replicated copy of data. Recommended to use a dedicated network from cloud provider to on-premises for best performance when replicating back to on-premises (Azure ExpressRoute).

USE DATA - DISASTER RECOVERY TO THE CLOUD

Business Value: Providing operational recovery of primary site applications to a secondary site from an external cloud provider.



SCENARIO / SUITABILITY	REQUIREMENTS
• Off-site storage & cold DR site in the Cloud – Only use the cloud compute infrastructure when a DR event occurs, saving time & money via the elimination of asset allocation with long idle periods between DR operations. • Live Sync data replication for Warm Recovery in cloud – Automate the creation of cloud VMs and replication of on-premises VMs to Azure on a periodic cycle basis more frequently than backups. Reduces recovery time to the cloud. • VM Restore & Convert – Convert VMware and Hyper-V VMs into Azure VMs on-demand with data intact. This data transformation automation reduces time & complexity costs. • Automate Failover and Failback of VMs - From on-premises VMware to Azure.	• Database/Files – Restore out-of-place, whether on- demand or scheduled, to refresh DR targets. When combined with job-based reporting, this scheduled operation is of benefit to enterprises that must maintain audit and compliance reporting associated with business continuity reporting. • Minimum 1x MediaAgent on-premises, and minimum 1x MediaAgent in cloud • MediaAgent in cloud only needs to be powered on for recovery operations • Highly recommended to use dedicated network to cloud provider for best performance (Azure ExpressRoute).

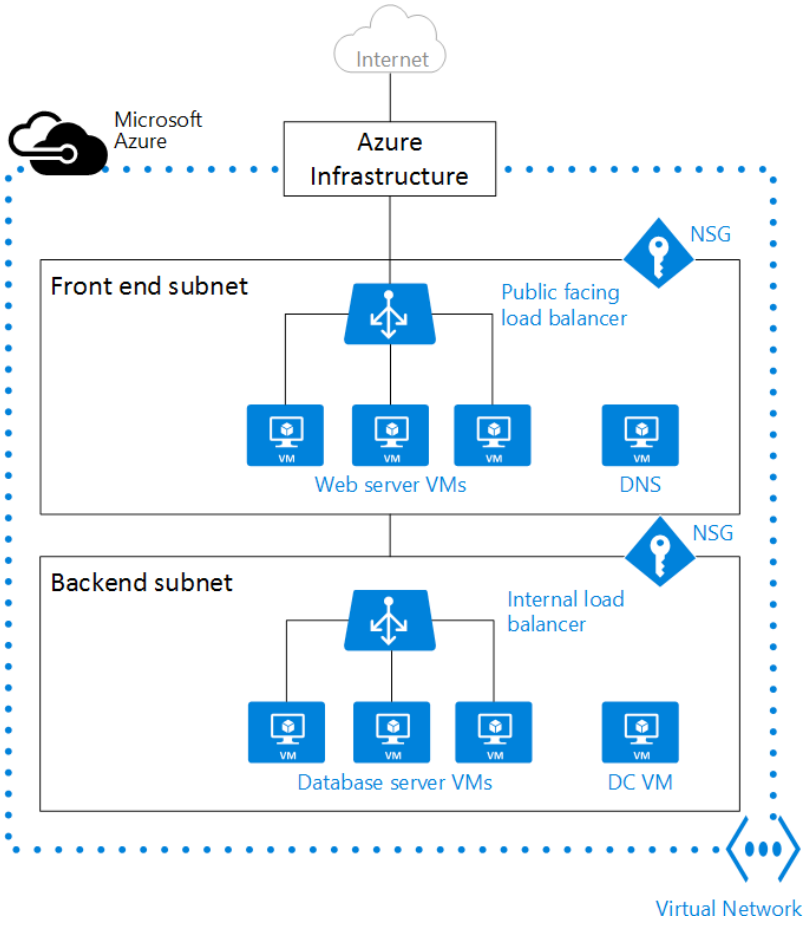
ARCHITECTURE CONSIDERATIONS

NETWORKING

VIRTUAL PRIVATE CLOUD/NETWORKING

Microsoft Azure (Azure) has the capability to establish an isolated logical network. This is referred to within Azure as an Azure Virtual Network (AVN).

Instances / virtual machines deployed within an AVN, by default, have no access to the Public Internet, and utilize a subnet of the customer's choice. Typically, AVN's are used when creating a backbone between virtual machines, and also when establishing a dedicated network route from a customer's existing on-premises network directly into the public cloud provider via Azure ExpressRoute.



BRIDGING ON-PREMISES INFRASTRUCTURE – VPN & EXPRESSROUTE

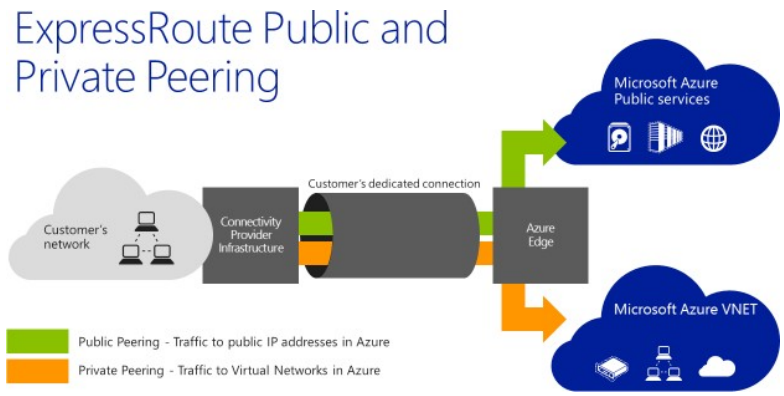
Customers may find a need to bridge their existing on-premises infrastructure to their public cloud provider, or bridge systems and workloads running between different cloud providers to ensure a common network layer between compute nodes and storage endpoints.

This is particularly relevant to solutions where you wish to Backup/Archive directly to the cloud, or create deduplicated secondary data copies (DASH Copy) of existing backup/archive data to object storage within a cloud provider.

To utilize these features there are two primary choices available:

- VPN Connection – network traffic is routed between network segments over Public Internet, encapsulated in a secure, encrypted tunnel over the customer's existing Internet Connection. As the connection is shared, bandwidth is limited and regular data transfer fees apply as per the customer's current contract with their ISP.
- Azure ExpressRoute – a dedicated network link is provided at the customer's edge network at an existing on- premises location that provides secure routing into an Azure Virtual Network.

Typically, these links are less expensive when compared to a customer's regular internet connection, as pricing is charged on a monthly dual-port fee, with all inbound and outbound data transfers included free of charge, with bandwidth from 10Mbit/s to 10Gbit/s.



INFRASTRUCTURE ACCESS

HYPERVISOR ACCESS IN PUBLIC CLOUD

Public Cloud providers do not allow direct access to the underlying hypervisor, instead access to functionality such as VM power on/off, Console access are provided through a REST API.

AZURE VIRTUAL NETWORK

Azure Virtual Network (VNets) enables many types of Azure resources, such as Azure virtual machines, to securely communicate with each other, the internet, and on-premises networks.

Azure Virtual Network provides the following key capabilities:

- Isolation and segmentation
- Communicate with the internet
- Communicate between Azure resources
- Communicate with on-premises resources
- Filter network traffic
- Route network traffic
- Connect virtual networks

For more information on Azure Virtual Networks, please refer to this Azure documentation.

DATA SECURITY

IN-FLIGHT

By default, all communication with Cloud Libraries utilize HTTPS which ensures that all traffic is encrypted while in-flight between the MediaAgent and the Cloud Library end-point, but traffic between Commvault® nodes is not encrypted by default. We recommend that any network communications between Commvault® modules routing over public internet space be encrypted to ensure data security. This is employed by using standard Commvault® firewall configurations (Two-Way & One-Way).

AT-REST

Data stored in a public cloud is usually on shared infrastructure logically segmented to ensure security. Commvault recommends adding an extra layer of protection by encrypting all data at-rest. Most cloud providers require that any seeded data is shipped in an encrypted format. Examples of seeding data is with the use of Azure Data Box.

HTTPS PROXIES

Please take note of any HTTP(S) proxies between MediaAgents and endpoints, whether via public Internet or private space, as this may have a performance impact upon any backup/restore operations to/from an object storage endpoint. Where possible, Commvault® software should be configured to have direct access to an object storage endpoint.

DATA SEEDING

Data Seeding is the process of moving the initial set of data from its current location to a cloud provider in a method or process that is different from regular or normal operations. For seeding data to an external cloud provider, there are two primary methods:

"OVER-THE-WIRE"

This is typically performed in a small logical grouping of systems to maximize network utilization in order to more quickly complete the data movement per system. Some organizations will purchase "burst" bandwidth from their network providers for the seeding process to expedite the transfer process.

Major cloud providers offer a direct network connection service option for dedicated network bandwidth from your site to their cloud such as Azure ExpressRoute.

Please see the chart below for estimated payload transfer time for various data sizes and speeds.

LINK SIZE	DATA SET SIZE							
	1 GB	10 GB	100 GB	1 TB	10 TB	100 TB	1 PB	10 PB
10Mbit	14m	2.2 hrs	22.2 hrs	9.2 days	92.6 days	-	-	-
100Mbit	1m 20s	13m 20s	2.2 hrs	22.2 hrs	9.2 days	92.6 days	-	-
1Gbit	8s	1m 20s	13m 20s	2.2 hrs	22.2 hrs	9.2 days	92.6 days	-
10Gbit	0.8s	8s	1m 20s	13m 20s	2.2 hrs	22.2 hrs	9.2 days	92.6 days

DRIVE SEEDING

If the data set is too large to copy over the network, or transport over network is too costly, then physical drive seeding is a valid alternative option. Drive seeding is copying the initial data set to external physical media and then shipping it directly to the external cloud provider for local data ingestion.

Please refer to the Commvault's Online documentation for the Seeding the Cloud Library procedure for more information.

In addition to this, Azure has their own process for drive seeding:

Azure

- Import Services
- Databox

COST / CONSUMPTION

NETWORK EGRESS

Moving data into a cloud provider, in most cases, has no provider cost, however moving data outside the cloud provider, virtual machine instance, or cloud provider region, usually has a cost associated with it. Restoring data from the cloud provider to an external site or replicating data between provider regions are examples of activities that are classified as Network Egress and usually have additional charges. Pay special attention to the tier of storage. Some storage tiers cost more for egress and others are free. This may impact your storage costs enough to decide to choose a higher tier of storage like Hot storage tier instead of Cool or Archive storage tier.

STORAGE I/O

The input and output operations to storage attached to the virtual machine. Cloud storage is usually metered with a fixed allowance included per month and per unit "overage" charges beyond the allowance. Frequent restores, active data, and active databases may go beyond a cloud provider's Storage I/O monthly allowance, which would result in additional charges.

GET/PUT TRANSACTION COSTS

Azure customers usually incur a cost for GET/PUT transactions to cloud object storage. These costs are primarily to enforce good practices for applications when retrieving and placing data in the cloud. As such, the cost when using the Commvault® solution is minimal.

When Commvault® writes data to a Cloud Library, the Cloud Library splits the data up into a sub-chunk size of 32 MB. Each 32 MB chunk write or read will incur a GET or PUT request. As of January 2018, Azure charges $0.05 per 10,000 requests for its LRS Hot storage tier.

A baseline of 200 GB with a saving of 40% at 32MB sub-chunk size would result in an approximately 3840 PUT requests. At a charge of $0.05 per 10,000 requests, Azure LRS Hot storage tier ($0.05/10,000) the net charge would be 5 cents.

Note: All cost figures are referenced in USD.

DATA RECALL

Low-cost cloud storage solutions may have a cost associated with accessing data or deleting data earlier than an agreed time. Storing infrequently accessed data on a low-cost cloud storage solution may be attractive upfront, however Commvault recommends modeling realistic data recall scenarios. In some cases, the data recall charges may be more than the potential cost savings vs. an active cloud storage offering.

As a best practice, Commvault recommends developing realistic use case scenarios and modeling cost against the identified scenarios to ensure the cloud solution meets your organization's SLAs, as well as cost objectives.

Visit the Azure cost calculator for supported cloud vendors.

PERFORMANCE / STORAGE

MULTI-STREAMING WITH OBJECT STORAGE

Object storage performs best with concurrency, and as such with any Cloud Libraries configured within the Commvault® environment, best performance is achieved when configured for multiple readers / streams.

CLOUD CONNECTOR BEST PRACTICES

There are additional Data Path settings and additional settings used to adjust and fine-tune the performance of the Cloud Library. The exact settings for best performance may vary between Cloud vendors.

The following combined settings are recommended to increase data read performance from cloud libraries utilizing deduplication.

1 Increase deduplication block size to either 256 KB or 512 KB for maximum performance

   Cloud object storage is subject to 50% or higher latencies than traditional storage. When requesting blocks from object storage, this delay may reduce overall read performance. To counteract the delay, increase the deduplication block size to allow larger data retrievals for each request. Note that changing existing storage policies will initially cause an increase in storage as new deduplication data re-baselines. Only new data written with the higher block size will benefit from retrieval performance improvements.
   If the requirement is to keep one copy on-premises and another in the cloud, recommendation is to use 256 KB block size for on-premises and cloud copy. Otherwise, if one or all copies involved will be using cloud storage, recommendation is to use 512 KB block size. The reason for this is, you cannot choose a different deduplication block size for multiple copies within a Storage Policy, allowing this will unnecessarily increase the overhead in creating the secondary copies as data will now have to be rehydrated and re-deduplicated with the new block size. As of April 2018 (Commvault® v11 SP11), the block size for a deduplication Cloud Library will automatically be created to use 512KB.

2 Configure SILookAheadAsyncIOBlockSizeKB to 1024 KB

   When used in conjunction with either 256 KB or 512 KB block size, this setting will further increase read performance from a cloud library utilizing deduplication.

For more tunable settings and information, please refer to Cloud Connection Performance Tuning online documentation.

COMPRESSION VS. DEDUPLICATION

Deduplication is recommended to be used where possible, with the exception of environments where there are significant bandwidth concerns for re-baselining operations, or for Archive-only use cases where the data pattern spread generates no benefit from deduplication operations.

While additional compute resources are required to provide the necessary foundation for optimal deduplication performance, using deduplication in a cloud context can still achieve greater than a 10:1 reduction.

Even with sealing of the deduplication database (DDB), stored data results can achieve a 7:1 reduction in footprint, providing significant network savings and reduced backup/replication windows (DASH Copy).

In comparison, software compression can only achieve 2:1 reduction on average, and will constantly consume the same bandwidth when in-flight between endpoints (no DASH Copy).

LEVERAGING MULTIPLE MOUNT PATHS FOR A CLOUD LIBRARY

Just like regular disk libraries, Cloud Libraries have the option to leverage multiple mount paths. The benefit of using multiple mount paths depends on the cloud storage vendor.

In Azure, creating additional mount points overcomes the management complexities of the limit of 500 TB per storage account (as of May 2018).

BLOCK STORAGE VERSUS OBJECT STORAGE

While Public IaaS environments allow block-based storage to be provisioned and leveraged as Disk Libraries, the overall cost of those volumes can quickly exceed that of object storage.

With the inclusion of Commvault® micro pruning, and its benefit of reducing cost of data stored in object storage, it is highly recommended that object storage be the primary choice for writing data to the cloud, and other forms of storage by exception.

With Azure, be aware that both object blobs as well as page blobs, are potential options. While Commvault® software supports and can consume page blobs, this option incurs a higher cost compared to standard Block blobs, and this cost should be evaluated before making this choice.

If you are unsure as to which offering to use, you should consume regular object storage blobs instead.

PARTITIONED DEDUPLICATION

Like on-premises configurations, making use of partitioned deduplication can provide several benefits. When possible, make use of partitioned deduplication to increase scale, load balancing, and failover. Version 11 allows for the addition of two extra nodes (up to 4) to an existing deduplication store dynamically, allowing for rapid scale-out configurations. Learn more.

MICRO PRUNING

The micro pruning support for object storage is effective for any new data written into the active store.

For customers who have upgraded from Version 10 of Commvault, but have not yet enabled micro pruning support, macro pruning rules will still apply to existing data within the active store until the store has been sealed. Once the active store is sealed, there will no longer be a need for continued periodic sealing against that store.

SELECTING THE RIGHT STORAGE CLASS FOR BACKUP AND ARCHIVE DATA

Depending on the provider, there may be different tiers of object storage available that offer different levels of cost, performance and access/SLA's. This can have a significant impact on both the cost and the user experience for the datasets within.

For example, storing infrequently accessed backups within an intermediate storage tier (Cool blob) can significantly lower the cost of your cloud bill, while storing archives in a deep archive storage tier (Archive blob) may greatly impact accessibility for end-users to the archived data while reducing storage costs further.

To delve into further detail, these storage classes can be broken into three primary categories:

- Standard storage – this storage class represents the base offering of any object storage platform – inexpensive, instant access to storage on-demand. Offerings in this category include Azure Blob Hot storage tier.

  Typically, this tier would be used for backup and archive workloads in a short-term retention configuration.

- Intermediate storage – this is a storage tier that addresses a gap between the standard Hot storage tier offering and deep Archive storage tier, in that it is offered at a lower price point than Hot storage, but is aimed at scenarios where data is infrequently accessed.

  While the storage is always accessible, similar to the Hot offering, the cost model is structured to enforce an infrequent access use case by charging $0.01/GB for any retrieval from this storage tier. Offerings in this category include Azure Blob Cool storage tier.

  This tier would be leveraged for Backup workloads in a medium to long-term retention configuration, and for Archive workloads that require instant access to the archived data.

- Deep archive storage – sometimes referred to as "cold storage", this tier is intended for data that will probably not be accessed again, but must be retained in the event of compliance, legal action, or another business reason, Azure Archive Blob storage tier is an example of archive storage which Commvault® software supports.

  The cost of this storage class is the lowest compared to all three offerings – $0.002/GB/month, but as with the intermediate class, the deep archive class's cost model is also structured with the expectation that retrievals are infrequent and unusual, and data will be stored for an extended period of time. In addition to the per-GB, per- month charge, any data that is moved to Archive is subject to an Archive early deletion period of 180 days. The charge is prorated based on the amount of time it has been stored in archive storage. You can think of this class of storage as equivalent to tape and is therefore recommended not to use deduplication.

It is highly recommended that you review the cost options and considerations of each of these storage classes against the use case for your architecture in order to gain the best value for your cost model. Commvault® Professional Services can assist in necessary service class / data class valuations in designing the correct cost value model for your enterprise.

INFREQUENT ACCESS STORAGE CLASS SUPPORT

Support for the following Infrequent Access storage classes are available in Commvault v10 Service Pack 12 and Version 11:

- Microsoft Azure Blob Cool storage tier
- Others listed at Commvault.com

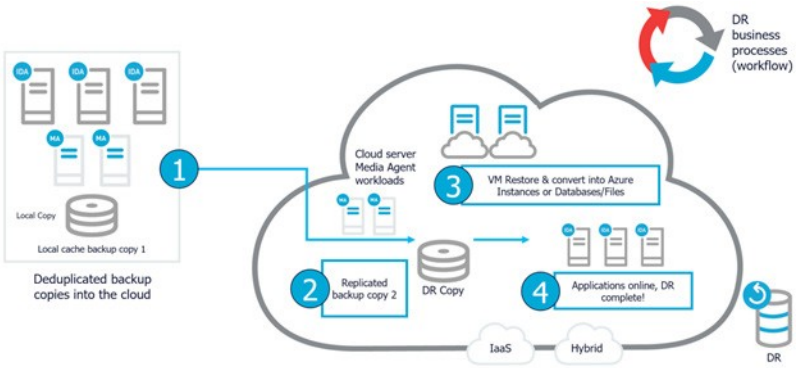
STORAGE ACCELERATOR TO AZURE BLOB STORAGE

For remote office locations, small cloud environments, roaming devices such as laptops, and any architecture that proves unfeasible or cost prohibitive to implement a traditional or cloud-based MediaAgent, backups can be done directly from the source to a cloud target such as Azure Blob storage, completely bypassing the MediaAgent. This is achieved by installing and enabling the Storage Accelerator feature on the client for direct communication to a storage target and will speed up the backup and reduce costs in these instances. When deduplication is used, the client communicates with the Media Agent to reference and update the dedupe database, and then sends only the unique blocks directly from the client to the cloud storage blob.

Get additional information on the Storage Accelerator.

PERFORMING DISASTER RECOVERY TO THE CLOUD

This section will cover the steps required to perform disaster recovery into the Azure public cloud platform. We examine recovery methods available for both image and agent based protection. This also addresses different recovery scenarios that may be needed to meet short recovery time objectives.



RESTORING APPLICATIONS (AUTOMATED OR ON-DEMAND)

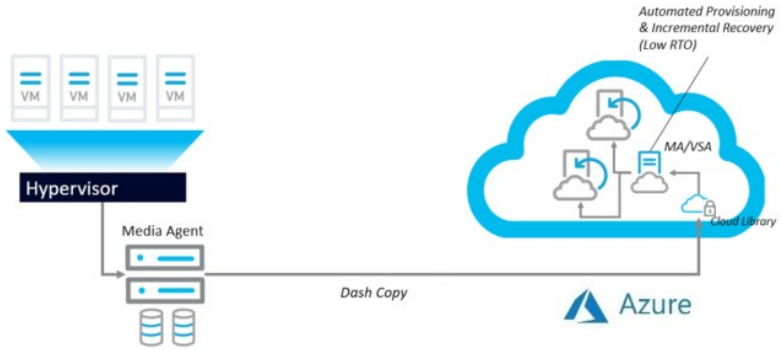
An agent-in-guest approach allows for the recovery of a wide variety of operating systems and applications. These can be captured at the primary site and replicated to the cloud-based Media Agent in a deduplicated efficient manner. Once replicated, the data can be held and restored in the event of a DR scenario or automatically recovered to existing VMs for more critical workloads.

REPLICATING VM WORKLOADS WITH LIVE SYNC

Live Sync also allows you to replicate VMs to public cloud infrastructure. As of December 2016, Azure is a supported cloud infrastructure vendors for Commvault® v11. Live Sync combines the VM conversion feature with incremental replication to provide a DR solution utilizing on-demand cloud infrastructure. As Live Sync to cloud integrates with Dash Copy, highly efficient WAN replication is possible, by reducing the amount of data being replicated. You can replicate to the same Azure subscription or a different subscription, and to the same region or a different region. Live sync for Azure is supported from streaming backups or IntelliSnap backup copies.

Azure provides greater flexibility to important virtual machines and offers superior achievable RPO and RTO targets when used in conjunction with Live Sync. Live Sync can also be used as an alternative to limited bandwidth scenarios where large data transfer isn't achievable. The following hypervisors are supported for Live Sync replication to Azure: VMware, Hyper-V, and Azure VM.

As such, a good strategy is to identify multiple approaches depending on the business RTO/RPO requirements and implement them accordingly, while also considering the limitations and requirements specific to the cloud vendor. For example, Tier 1 applications may be a better fit for near-continuous replication using Commvault's CDR technology, while Tier 2 applications could make use of Live Sync (VMs, Files, DB), and Tier 3 apps could use on- demand VM conversion from cloud storage when needed.



Get additional information on the Live Sync feature.

REPLICATING OTHER WORKLOADS

Commvault® Continuous Data Replicator (CDR) allows near time continuous data replication for critical workloads that must be recovered in adherence to Service Level Agreements that exceed the capabilities associated with Live Sync operations. These VMs require similarly sized Azure VMs to receive any replicated data. In order for CDR to operate, an Azure VM must be running at all times to receive application changes. Get additional information on CDR.

VIRTUAL MACHINE RECOVERY FROM AMAZON EC2 TO AZURE

With Commvault® v11 SP7, you can now recover Amazon EC2 instances protected with the Virtual Server Agent to an Azure VM for disaster recovery or migration purposes. Currently streaming backups are supported for recovery to both Azure Resource Manager and Azure Classic.

Get additional information on the Conversion feature.

AZURE-SPECIFIC WORKLOADS

VIRTUAL MACHINE RECOVERY INTO AZURE VM INSTANCES

The Commvault® Virtual Server Agent provides the ability to easily perform direct conversion of protected VMware, Hyper-V, or Amazon instances into Azure VMs, from backups stored either within Azure Blob storage, another Cloud Library or from an on-premises Disk Library.

This process is used as part of a disaster recovery strategy using Azure as a Cold DR site, or as a migration strategy (Lift-and-Shift). Additional details can be found here.

USING COMMVAULT® WORKFLOWS TO AUTOMATE DR

The Commvault® Workflow engine provides a framework in which the DR runbook process, covering the deployment of new VMs, recovery of data and applications, and validation aspects of a DR operation can be automated to deliver a simplified, end-to-end GUI-driven DR process. This can be developed and maintained by your administrators, or with the assistance of the Commvault® Personalization Services team.

For more information on Commvault's Personalization Services team, please contact Commvault or Commvault® Partner Account team.

For more information on the Workflow engine, please refer to the Workflow Overview link.

PROTECTING AND RECOVERING ACTIVE WORKLOADS IN THE CLOUD

This section outlines the basics on protecting active workloads running in Microsoft Azure. This portion of the document outlines the various protection approaches as well as replication and recovery to different geographic regions. This section also reviews cross platform recovery as well as recovery to onsite locations.

AZURE

AGENT-LESS VM PROTECTION (VIRTUAL SERVER AGENT FOR AZURE)

Introduced in Version 11 Service Pack 4, the Virtual Server Agent for Azure (VSA for Azure) delivers an agent-less, block-level capture of Azure VMs and their attached block volumes. Restoration options include Full virtual machine recovery, attaching disks to an existing virtual machine, and granular-level file recovery. Azure VSA optionally includes Changed Block Tracking, included with v11 SP5, which helps accelerate incremental backup performance.

There are two types of storage accounts that can be used to provision blob, table, queue, file storage, and virtual machine hard disks. You can create a virtual disk in the Azure cloud by working directly with a storage account or you can let Azure manage the storage account for you with Managed Disks. Azure Managed Disks simplifies disk management for Azure IaaS VMs by managing the storage accounts associated with the VM disks. You only have to specify the type (Premium or Standard) and the size of disk you need, and Azure creates and manages the disk for you.

Commvault® software has had agent-less VSA protection for Non-Managed Disks since v11 SP4. As of v11 SP10, we now have the ability to protect Azure Managed Disks.

WHEN TO USE THE VSA FOR AZURE

- Agent-less protection approach for Azure VMs & file-level data – no agents are required in-guest to perform a block-level backup to provide VM and File-level recovery

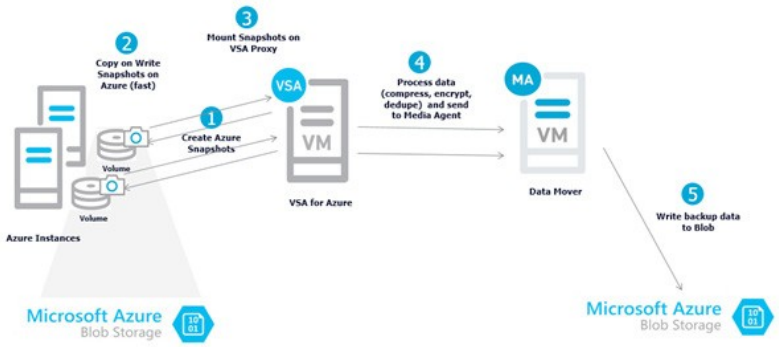
WHEN NOT TO USE THE VSA FOR AZURE

- When you require application-consistent backups – the VSA for Azure approach creates a Crash-consistent image of the source VM and its block volumes. If you require application consistency, use an agent-in-guest either standalone or in conjunction with the VSA for Azure backup schedule.
- Protecting worker/stateless VMs – Worker nodes may generate valued data that is moved to another centralized repository and the nodes themselves do not require protection. It is recommended to instead target that centralized repository for Data Protection instead of the individual worker nodes, whether with VSA for Azure or agent-in-guest, depending on the required level of backup (crash vs. application consistent).

HOW VMS ARE QUALIFIED FOR PROTECTION

- Each VSA can be configured with one or more subclients. Each subclient defines a rule set on which to Auto- Detect and Protect Azure VMs, based on a user-defined criteria of VM Name or Resource Groups.
- During the Discovery phase of the backup job, the VSA will use the subclient rule to qualify VMs to add/remove for protection within that job.

Commvault® software does not require access to the Azure hypervisor-level, instead using the REST APIs to create snapshots of each block volume, attaching the snapshot to a nominated proxy (Azure VM-based VSA / Media Agent) to read and de-duplicate the blocks before writing out to an Azure Hot, Cool, or Archive storage tiers.



Commvault IntelliSnap® functionality introduced in Version 11 Service Pack 7 for the VSA for Azure. Commvault IntelliSnap allows snapshots to be retained on the VM configured via Storage Policy's Snap Primary retention setting.

Commvault IntelliSnap backups enable reducing backup windows considerably, providing fast, snapshot-based restoration capability. Azure also allows for the use of CBT (Changed Block Tracking) which can accelerate incremental backup performance considerably.

Note: Azure changed block tracking (CBT) is currently only available for Azure unmanaged disks. CBT support for Azure managed disks is on the Azure roadmap and coming soon.

Get additional details, here.

ARCHITECTURE REQUIREMENTS FOR THE VSA FOR AZURE

- Minimum 1x VSA/MA per Region, Recommended 1x VSA/MA per Availability Set.
- Each 1x "VSA/MA" node represents a single Windows Server Azure VM with the Virtual Server Agent and MediaAgent modules deployed. The Azure VM specifications should match the MediaAgent specifications within this Architecture Guide.
- If the environment contains more than >25-50 VM's within a single Availability Set, it is recommended to scale out with additional VSA proxies (1 or more) placed inside of the source Availability Set to improve performance and reduce cross-zone traffic by containing it within the fault/update domain.
- When considering the VM instance type for the Proxy, the Commvault platform does not actually mount the guest VM snap disks on the proxy for backup. This means that if you are protecting VMs with Premium Disks, you don't necessarily need to select a Proxy VM type that is capable of mounting a Premium Disk. You can choose a cheaper VM type if it suits your environment better. Although keep in mind that there can be a performance hit if the proxy VM type is not able to keep up with the required compute and IOPS demanded of it.

ARCHITECTURE RECOMMENDATIONS

- While the readers count can be increased to improve concurrency per VSA/MA node, consider scaling out with multiple VSA proxies. Azure recommendations state that for optimal performance, you will want to limit the number of highly utilized disks attached per worker node to avoid possible throttling.[1]
- Use of Premium Storage for the deduplication database and index cache used by the MediaAgent module is highly recommended for optimal performance.
- (Backup Copy / Streaming) To restore from advanced Linux files systems such as EXT4, XFS and others, you can deploy a file recovery enabler by converting an existing Linux media agent to a FREL. When browsing data on advanced file systems, Commvault® software will leverage the FREL to access the file system.
- (Backup Copy / Streaming) Enable CBT to improve incremental backup performance.

AGENT-IN-GUEST (STREAMING)

An agent-in-guest approach can be used to protect a wide variety of operating systems and applications. These can be captured on the production workload and protected to the MediaAgent residing in Azure, using client-side deduplication to reduce the network consumption within the cloud. These can also be replicated to a secondary MediaAgent residing in a different geographic region. Once replicated the data can be held and restored in the event of a DR scenario or automatically recovered to existing VMs for the more critical workloads.

WHEN TO USE AGENT-IN-GUEST APPROACH:

- When you require application-consistent backups – Deployment of agents can either be pushed/installed by Commvault® software, or baked into an Azure template using de-coupled installation, or deployed as part of a continuous deployment method (i.e. Puppet/Chef/Ansible).
- When you require granular-level protection and restoration features for applications – the Commvault® iDataAgents can deliver granular-level protection for supported application workloads, such as SQL Server or Oracle Database, in comparison to a Full VM or File-level approach.

PROTECT AZURE SQL DATABASES

Commvault provides a complete data protection solution for Azure SQL databases by automating backup operations and by providing the following recovery methods:

- Restore from Azure Cloud SQL to Azure Cloud SQL
- Restore from Azure Cloud SQL to an on-premises SQL server
- Restore from an on-premises SQL server to Azure Cloud SQL

Best Practice: The SQL server on the proxy client should run the latest SQL server release to ensure that the proxy server is in sync with the Azure instance (Azure always has the latest SQL version). Add the SQL server instance to the CommCell Console or Admin Console. When you restore an on-premises database to the Azure cloud, the restored database uses the standard tier model.

Get additional details, here.

ARCHITECTURE REQUIREMENTS FOR AGENT-IN-GUEST:

- Minimum 1x iDataAgent per VM for the intended dataset (i.e. SQL, File). Multiple iDataAgents can be deployed on the same VM.
- Minimum 1x MediaAgent per region. MediaAgents connect to the target object storage, and can either be deployed on the same VM, or on a dedicated host for a fan-in configuration. The Azure VM specifications of the MediaAgent should match the MediaAgent specifications within this Architecture Guide.
- Check the Systems Requirements section in Online Documentation to determine if the iDataAgent supports your application.

1 "Virtual Machine disk limits", from Azure Documentation.

ARCHITECTURE RECOMMENDATIONS

- Use of multiple readers to increase concurrency to the Azure Blob endpoint is recommended

PERFORMANCE TESTS IN AZURE

Below are the results of tests that were performed to compare the performance of the different backup methods within Azure. The tests were done to compare the Storage Accelerator, VSA backups with Azure managed Disks, and VSA backups with Azure unmanaged disks for a file based data-set.

Backup Statistics

App size : 1096 GB for Full, 234 GB for Incremental

Backup Method Used	Time taken for backups		
	CV Storage Accelerator	VSA for Azure Managed Disk	VSA for Azure Un-Managed Disk
Full	3 hrs 36 min	10 hrs 31 min	10 hrs 8 min
Incr 2 40GB change	34 min	3 hrs 51 min	2 hrs 40 mins
DASH Full	24 mins	16 mins	15 mins

Backend Storage Consumed

App size : 1096 GB for Full, 234 GB for Incremental

Backup Method Used	Backend storage consumed		
	CV Storage Accelerator	VSA for Azure Managed Disk	VSA for Azure Un-Managed Disk
Full	1020 GB	1001 GB	1000 GB
Incr 2 40GB change	201 GB	217 GB	213 GB
DASH Full	5.5 GB	0.5 GB	0.7 GB

Restore Statistics

Granular and Full Volume restores

Backup Method Used	Time taken to restore data		
	CV Storage Accelerator	VSA for Azure Managed Disk	VSA for Azure Un-Managed Disk
Granular restore of 100 files (1GB)	1 min	8 mins	13 mins
Granular Restore of 40GB	23 mins	1 hr 9 mins	1 hr 9 mins
Whole VM Restore	4 hrs 32 mins	1 hr 39 mins	1 hr 39 mins

Snapshots

Backup Method Used	Time taken to create a snapshot		
	CV Storage Accelerator	VSA for Azure Managed Disk	VSA for Azure Un-Managed Disk
Snapshot Incremental	X	2 mins	3 mins
Snapshot Full	X	2 mins	3 mins

TEST INFRASTRUCTURE AND DATA SET:

Five Azure instances as source clients, each with 200 GB volume that contained a mix of files varying in sizes from 64K to 100MB. Instance type used was Standard D4s.v3 (4 CPU core, 16 GB RAM) and Premium storage (SSD) for OS and Data disks. Before incremental backup, 20% (40 GB) of data was deleted and re-created with random, non-compressible data. The Media Agent, CommServe®, and VSA Proxy (for VSA backups only) components were configured to run on separate instances in Azure, configuration of which was as follows: instance type Standard D4s v3 (8 CPUs, 32 GB RAM) with 128 GB Premium SSD disk for the DDB and 128 GB standard SSD disk for the Index Cache.

Note that incremental backups via VSA for Azure unmanaged disks ran fast as Azure supports CBT (Change Block Tracking), although it's currently not yet supported by Azure for managed disks.

The storage accelerator tests were done by running backups of all five clients at the same time in parallel.

The storage footprint consumed for hosting the backups is similar for all methods. From a cost perspective, since the time taken by all methods is nearly the same, costs will be in the same ballpark.

AZURE SNAPSHOTS

Azure snapshots allow for a crash consistent point-in-time copy of an Azure disk, and can be automated with the use of Workflows. Snapshots are Copy-On-Write (COW) so it is recommended to keep more than one snapshot if you require the ability to restore from snap in the case of corruption or loss of data on the production VM. Commvault IntelliSnap currently supports Azure managed disks and unmanaged disks.

With Commvault IntelliSnap support for Microsoft Azure, you can:

- Perform full and incremental snap backups of virtual machines.
- Perform backup copy jobs from snap backups.
- Perform backup copy jobs with Changed Block Tracking (CBT). Currently only supported for Azure Unmanaged Disks.
- Perform full VM restores from snap backups, and full VM and file level restores (using Live Browse) from backup copy jobs.

Get additional details, here.

MACHINE EXPORT FROM AZURE

Azure offers the ability to export running machines in VHD format. These allow for import into the most commonly used hypervisors.

AZURE BLOB STORAGE BACKUP

Introduced in Version 11 Service Pack 6, is the ability to perform backups of Azure Blob storage created by 3rd party applications. This capability will allow Commvault® software to capture the data contained inside an Azure blob container, allowing for restore back to a blob container or file system client.

Get additional information on Azure blob storage backups here.

WHEN TO USE BLOB BACKUP:

- Backing up object storage – Created by 3rd party applications.

WHEN NOT TO USE THIS APPROACH:

- Protecting Commvault® cloud libraries (Backup Data) – To protect data contained in cloud libraries, use secondary copies (DASH copies or aux copies) in the storage policy instead.

AZURE FILES SHARE

Azure Files is a feature based off of Azure Blob storage. Since we can protect Azure blob storage, we can also protect Azure Files. We use the same Azure Cloud Apps connector to protect Azure Files with a slight change. You can protect it from the Blob perspective and backup everything in the file/folder structure, or you can direct the connector to protect the Files share folder of your choice. You can back up an Azure Files Share by providing the Azure File Share URL (file.core.windows.net) when you create an Azure Blob Storage virtual client from the CommCell® Console. Once configured, you'll have Full and Granular restore capabilities of the data, in-place or out-of-place. This is currently available for streaming backups only.

For more information on how to create an Azure File Share virtual client, see Creating an Azure Blob Storage Virtual Client.

ARCHITECTURE RECOMMENDATIONS:

- For large datasets, consider using multiple subclients to increase scan performance and lower the amount of time taken to traverse the container contents.
- Multi-stream using multiple readers for best performance.

1 "Virtual Machine disk limits", from Azure Documentation.

APPLICATION MIGRATION

Commvault can assist in application migration efforts when shifting from on-premises facilities to public cloud providers such as Microsoft Azure. By leveraging the power of the Data Management platform, workloads can be migrated through a number of methods.

VIRTUAL MACHINE RESTORE & CONVERT (LIFT AND SHIFT TO AZURE)

The Virtual Server Agent can capture virtual machines from VMware and Hyper-V based platforms in an application-consistent method (VSS call / VMware Guest Tools hooks / Hyper-V Integration Tools) to ensure that a consistent image of the guest, and the applications residing within, are captured correctly.

With this image, the Virtual Server Agent can then restore and convert (VMware & Hyper-V) the virtual machine into Azure VMs directly, and the process can handle single or multiple virtual machines.

This process is performed interactively through the CommCell® Console, via Commvault® Workflow or API calls.

APPLICATION OUT-OF-PLACE RESTORE (ALL SUPPORTED PLATFORMS)

All application iDataAgents support the capability to restore a given source dataset out-of-place to an alternate location. In this method, the data is captured from the source system (physical, or virtual), and then either directly from the source copy or replicated to cloud (DASH Copy), a restore to the destination is submitted.

The process requires the supported iDataAgent to be deployed on both the source VM, and the destination Azure VM.

This process is performed interactively through the CommCell® Console, via Commvault® Workflow or API calls.

DEPLOYMENT

REMOTE ACCESS / BRING YOUR OWN SOFTWARE

As with all IaaS offerings, remote access to virtual machine can be achieved with your favorite protocol / software (RDP for Windows, SSH for Linux VMs) and Commvault® module deployment can be achieved with the current procedures listed in Online Documentation.

INSTALLATION BASICS

The following links cover the steps when installing the CommServe® in the cloud. This is only needed when the primary CommServe® will be running on the hosted cloud VM or used for DR recovery. Multiple modules can be deployed in a single installation pass to streamline deployment.

- Installation Overview
- Installing the CommServe
- Installing the MediaAgent
- Installing the Virtual Server Agent (Azure)

COMMSERVE® DISASTER RECOVERY SOLUTION COMPARISON

Learn more about CommServe DR Solution comparisons for building a standby DR CommServe in the cloud, or simply restoring on-demand (DR backup restore), here.

PRE-PACKAGING COMMVAULT® SOFTWARE WITHIN A VM TEMPLATE

For environments where deployment time is reduced by preparing software and configuration within VM templates, the Commvault® iDataAgents can also be deployed in Decoupled mode. This means that the iDataAgent is deployed within the VM, but will only be activated upon registration with the CommServe®.

For more information, please refer to the Installing the Custom Package instructions within Online Documentation:

- Installing the Custom Package on Windows
- Installing the Custom Package on Linux

AUTOMATING DEPLOYMENT WITH CONTINUOUS DELIVERY

For environments using Continuous Delivery toolsets such as Puppet, Chef or Ansible, Commvault supports deployment methods that allow administrators to both control agent deployment and configuration to provide an automated deploy & protect outcome for applications and servers.

For more information on creating an unattended installation package for inclusion in a recipe, please refer to the Unattended Installation guide within Commvault® Online Documentation:

For more information on using Commvault® software's XML / REST API interface to control configuration post-deployment, please refer to the Command Line – Overview section to review options available for each iDataAgent.

CLOUD LIBRARY CONFIGURATION

This section covers the steps needed to configure cloud storage as a primary, secondary, tertiary, etc. storage target. Please keep in mind that use cases outside of archive will require Commvault® infrastructure in the cloud to recover any protected data.

For most on-premises backup use cases (except for very small environments limited to 100 GB in payload size), cloud as a direct storage target for the primary copy is not recommended. For performance and responsiveness, a primary copy should be stored on an on-site disk library and a secondary copy should be hosted on the cloud storage. The secondary copy should be setup as an encrypted network optimized DASH copy to the cloud.

- Supported Cloud Storage
- Cloud Storage - Overview
- Cloud Library Performance Tuning

UNSUPPORTED CLOUD STORAGE CONFIGURATIONS

If a cloud Storage target is not listed in the Cloud Storage – Support table, but the cloud storage endpoints are publicly accessible, and provide either an S3-compatible or OpenStack-compatible REST API, you can verify the compatibility of the storage offering with Commvault.

Depending upon your cloud device type you may choose to verify the compatibility between:

- Amazon S3 supported vendors and Commvault®
- OpenStack object storage supported vendors and Commvault®

For devices that are not publicly accessible, please contact your account manager or Commvault support for more information on the Cloud Storage certification process.

ARCHITECTURE SIZING

AZURE

AZURE COMMSERVE® SPECIFICATIONS

SMALL (25 VMS OR 50 LAPTOPS)	MEDIUM (1000 VMS OR 5000 LAPTOPS)	LARGE (2500 VMS OR 10,000 LAPTOPS)	EXTRA LARGE (20,000 VMS OR 50,000 LAPTOPS)
• Standard_ D4s_v3 VM size (4 vCPU, 16 GB RAM) • 100 GB volume for CS Software & CSDB • Windows Server 2012 R2 or Windows Server 2016 (Commvault® v11 SP7+)	• Standard_D8s_v3 VM size (8 vCPU, 32 GB RAM) • 150 GB volume for CS Software & CSDB • Windows Server 2012 R2 or Windows Server 2016 (Commvault® v11 SP7+)	• Standard_D8s_v3 VM size (8 vCPU, 32 GB RAM) • 300 GB Premium Storage volume for CS Software & CSDB (P20 type) 2300 IOPS • Windows Server 2012 R2 or Windows Server 2016 (Commvault® v11 SP7+)	• Standard_D16s_v3 VM size (16 vCPU, 64 GB RAM) • 300 GB Premium Storage volume for CS Software & CSDB (P20 type) 2300 IOPS • Windows Server 2012 R2 or Windows Server 2016 (Commvault® v11 SP7+)

Note: For more detailed information, please refer to the following link:

https://documentation.commvault.com/commvault/v11/article?p=1645.htm

AZURE MEDIAAGENT

EXTRA SMALL 60 TB BET	SMALL 120 TB BET	MEDIUM 240 TB BET	LARGE 600 TB BET	EXTRA LARGE 800 TB BET
• Up to 60 TB estimated back end data • Standard_D2s_ v3 VM size (2 vCPU, 8 GB RAM) • 200 GB Premium Storage volume for DDB (P15 type) 1100 IOPS • 400 GB Storage volume for Index Cache (non- premium) • Linux or Windows Server 2012 R2 or Windows Server 2016 (Commvault® v11 SP7+)	• Up to 120 TB estimated back end data • Standard_D4s_v 3 VM size (4 vCPU, 16 GB RAM) • 400 GB Premium Storage volume for DDB (P20 type) 2300 IOPS • 400 GB Storage volume for Index Cache (non- premium) • Linux or Windows Server 2012 R2 or Windows Server 2016 (Commvault® v11 SP7+)	• Up to 240 TB estimated back end data • Standard_D8s_v 3 VM size (8 vCPU, 32 GB RAM) • 600 GB Premium Storage volume for DDB (P30 type) 5000 IOPS • 1 TB Storage volume for Index Cache (non- premium) • Linux or Windows Server 2012 R2 or Windows Server 2016 (Commvault® v11 SP7+)	• Up to 600 TB estimated back end data • Standard_D16s_ v3 VM size (16 vCPU, 64 GB RAM) • 1.2 TB Premium Storage volume for DDB (40 type) 7500 IOPS • 1 TB Storage Premium volume for Index Cache • Linux or Windows Server 2012 R2 or Windows Server 2016 (Commvault® v11 SP7+)	• Up to 800 TB estimated back end data • Standard_D32s_ v3 VM size (32 vCPU, 128 GB RAM) • 2 TB Premium Storage volume for DDB (P40 type) 7500 IOPS • 2 TB Storage Premium volume for Index Cache • Linux or Windows Server 2012 R2 or Windows Server 2016 (Commvault® v11 SP7+)

Note: The above recommendations serve as a baseline. As environments continue to grow, inspect DDB performance health check to ensure it remains in a healthy state. If DDB health becomes non-optimal provide an additional disk in that tier. This method reduces upfront cost by only scaling to IOP demands when necessary.

Get more detailed information here.

Important: As of March 2017 Azure VMs provide a single, fast SSD free of charge, however this storage is temporary and all data is lost if the VM is moved to another host or is rebooted. For performance reasons, you can move the CommServe® TempDB SQL Database to this volume, and additional automation is required to ensure that on-reboot any required directory structures are re-created prior to SQL Server startup; otherwise, the SQL Instance (and the CommServe® services) will not successfully start. Commvault® Professional Services can provide this bespoke automation as part of its solution delivery process.

See Using SSDs in Azure VMs to store SQL Server TempDB for more information

The BET and FET sizing for maximum capacity are based on a 512K deduplication block size, which is the default for writing to cloud libraries.

EXAMPLE OF SIZING IN CLOUD

ASSUMPTIONS & COSTING

For this example configuration, we want to protect approximately 100TB of Front-End (FET) capacity. The average size of each virtual machine instance is assumed to be about 100GB and each instance has 2 volumes - one for operating system and the other for applications and data. This equates to approximately 1000 VM (100TB at 100GB each, using base10 rather than base 2 for simplicity in approximation).

It is also assumed that the daily change rate is ~2% of net new data that is created per day, or ~2TB worth of new data before it is compressed and deduplicated. The change rate in an environment varies greatly by environment and coupled with retention and deduplication ratio of the data. Both of which are also highly dependent on the specific environment, these three factors affect the back-end storage capacity that is ultimately required. For this example, we shall assume 90 days of retention and ~5:1 deduplication ratio. Both are typically observed within most virtual environments running a mix of Windows and Linux operating systems. With retention it is important to note that data that is 90 days old or the first full backup are not deleted until the most recent one if fully committed to the backup target. This accounts for retention+1 storage requirement. This results in approximately 117TB for the back-end.

It will also be assumed that absolutely no infrastructure to manage the cloud protection environment is present outside the cloud and that Commvault cloud MediaAgent power management feature is enabled, enabling shutdown of resources when backups and recoveries are not occurring. While most backup windows are usually 8 hrs., the assumption is that with restore account for another 4 hrs. per day allowing for power management to operate for only half of a given day.

Using publicly available pricing for Azure resources the cost of performing protection in Azure by utilizing any combination of iDataAgents, agentless VSA based backups, and coupled with Commvault IntelliSnap for Azure snapshots, the following becomes a rough estimate of the cost of the Azure infrastructure required for a period of 1 year:

COMMVAULT COMPONENT	QTY.	AZURE TYPE	AZURE COST/ HR OR /GB	AZURE COST/YEAR	AZURE COST/YEAR W/POWEROFF
CommServe® Medium VM	1	D8s_v3	$0.597	$5,229.72	$5,229.72
CommServe OS Disk	1	S15_256GB	$0.016	$137.53	$137.53
Standard Dedup MA Medium VM	1	D8s_v3	$0.597	$5,229.72	$2,614.86
MA + VSA OS Disk	1	P20_512GB	$0.102	$890.89	$890.89
MA DDB Disk	1	P30_600GB	$0.102	$897.02	$897.02
MA Index Disk	1	P30_1TB	$0.171	$1,497.96	$1,497.96
Disk Library (TB) (Full 90 Days)	230	Cool_Blob	$0.01	$14,568.65	$14,568.65
		TOTALS:		$28,959.58*	$26,090.68*

*It must be noted that this is a sample configuration utilizing estimated sizing data and that actual costs will vary depending on data type, retention and numerous other factors. This assumes scaled up to 100TB FET, starting with a much smaller footprint and growing as the source grows is perfectly acceptable.

ADDITIONAL RESOURCES

DOCUMENTATION

ONLINE DOCUMENTATION – CLOUD STORAGE

The Cloud Storage section from Commvault's Online Documentation covers technical procedures and information on Supported Cloud Targets, Advanced procedures, Troubleshooting and FAQ sections for Commvault customers. Learn more.

VIDEOS

2 CLICKS TO THE CLOUD WITH AZURE AND COMMVAULT®

Focuses on creating an Azure Storage Library within Commvault® v11.

BACKUP IN AZURE (TECHNICAL FEATURE, VSA FOR AZURE)

Technical showcase for the Virtual Server Agent for Azure (Version 11 SP4).