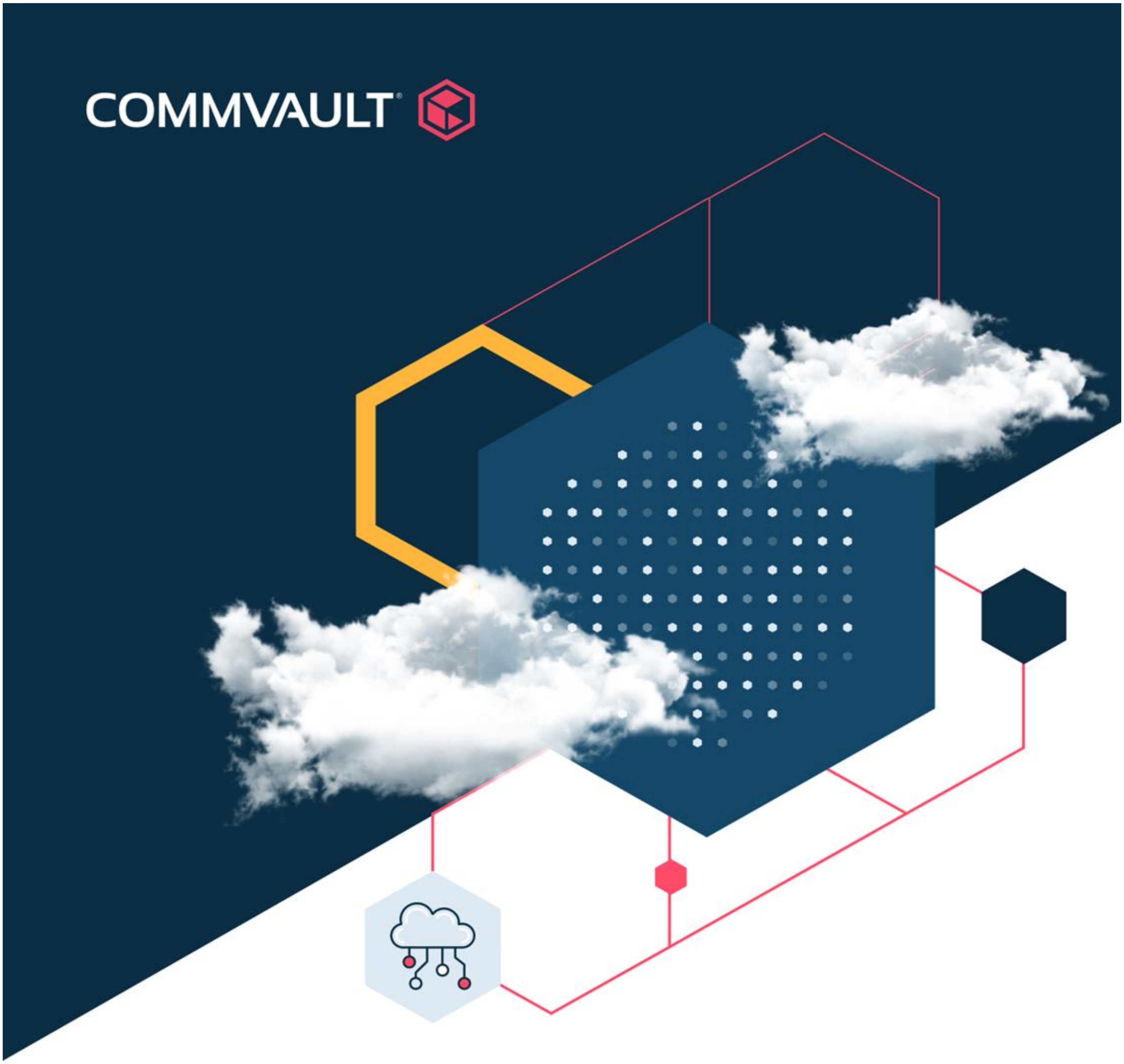


Public Cloud Architecture Guide for Microsoft Azure

CPR 2023E

Aug 2023

Table of Contents

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.

New in this version

Version	Data	Changes
11.19	April 2020	• Updated document with SP17-SP19 functionality
11.25	Sept 2021	• Updated document with 11.20 to 11.25 functionality • Update reference diagram • Azure blob storage metadata backup • Azure immutable storage • Azure Generation 2 VM • VSA Access Node Auto-Scale • Availability Zones support • VM Centric operations • Owner detection • Data verification recommendation
11.28	July 2022	• Updated document with CPR 2022E functionality • Formatting update to standardize Architecture reference • Adapt a Well-Architected format • Include Service catalog for Azure services and protection
11.30	Jan 2023	• Updated document with CPR 2023 functionality • Revision of Service catalog without exhaustive list
11.30	Feb 2023	• Removed skeleton well architected guide.
11.32	June 2023	• Updated for 2023E

Commvault remains committed to ensuring the **Cloud Architecture Guide** remains relevant to currently available public cloud capabilities.

The latest copy of this document is available at **Virtualization White Papers >**

Preface

Microsoft Azure is a comprehensive cloud platform that offers a wide range of services, including compute, storage, networking, and more. Commvault is a leading provider of data protection and management solutions. By combining Commvault with Azure, organizations can achieve a high level of data protection and availability while also taking advantage of the scalability, flexibility, and cost-effectiveness of the cloud.

This guide provides an overview of the architecture for integrating Commvault with Azure. It covers the following topics:

- The benefits of using Commvault with Azure
- Quick links to guides for installing and operating Commvault Solutions for Azure
- The best practices for integrating Commvault with Azure

This guide is intended for system administrators and IT professionals who are responsible for data protection and management. It is assumed that the reader has a basic understanding of Commvault and Azure.

Why Commvault?

Consider the following capabilities when assessing your data management needs in Azure.

	Broad protection for cloud-native, SaaS, and traditional workloads Commvault has the broadest industry support for cloud-native, SaaS, and traditional applications, hypervisors, and storage arrays. Backup isn't often the first capability productized by new service providers, Commvault is there to perform protection for your current and future applications.
	Protection for data, regardless of location Commvault protects all Azure regions and availability zones. Commvault automatically discovers your cloud workloads by tag, then manages the Azure protection copy lifecycle to meet your business rules and cost objectives.
	Cloud-native protection – by default Commvault orchestrates the creation of cloud-native snapshots with encryption. Commvault automates snapshot copies within and across regions and accounts.
	Cloud mobility without compromise Workloads may be in Azure today, back on-premises tomorrow running in Azure Stack HCI and perhaps in another cloud provider for new dev/test initiatives (AWS, Google, Oracle). Commvault provides mobility for Containers, VMs, Databases, and Application data across clouds – meaning flexibility for your business.

	Self-service backup, recovery, and disaster recovery Enable authorized end users to perform the recoveries using the Commvault Command Center™ to self-service simple and complex recovery needs, without specialist Azure skills or knowledge.
	Recovery readiness and insight Commvault Command Center™ provides visibility into SLA compliance, backup/recovery history, and data access requests (eDiscovery). Your business will know what data it has, whether it is protected and whether it represents a risk (PII data, insecure data, orphaned data).
	Information Management across your entire data estate Commvault provides insight into your data through File Storage Optimization (FSO), Data Governance, and eDiscovery and Compliance capabilities. Visualize and identify data risks, and inefficiencies across protected and live data sources

Azure Quick Links

Ready to get started? Refer to the following links to get help and best practice recommendations from our Documentation.

Cloud Feature Support for Azure

Quick Links	
Configure Azure Blob for Cloud Storage	• How to configure cloud storage as a backup target • **Spoiler**: You should probably be using Cool tier storage ○ **Configuring Cloud Storage** ○ **Recommended: Azure IAM VM Role** • Improving performance in the Cloud for large VMs ○ **Accelerating Backups to Cloud Storage Libraries**
Backups and Restores of Azure VMs	• Installation of an Access Node on a virtual machine. • An Access Node manages backup and restore operations for guest VMs. ○ **Creating an Access Node for Azure** • Installation and configuration of an Azure Hypervisor ○ **Microsoft Azure Overview** • How the backup process works • Considerations and limitations • Different types of Backups supported. ○ **Backup Overview** ○ **Recommended**: **Using managed Identities** • How CBT with Azure works ○ **Changed Block Tracking for Microsoft Azure** • How the restore process works • Considerations and limitations ○ **Restore Overview** • Enabling snapshots on Azure with IntelliSnap ○ **IntelliSnap for Azure Backups**
Azure Databases	• The following databases are supported in Azure ○ **Cosmos DB Core** ○ **Cosmos DB Table** ○ **MariaDB** ○ **MySQL** ○ **PostgreSQL** ○ **Azure SQL** ○ **Azure SQL Managed Instance (MI)** • **Azure Table Storage**
Azure Blob Storage	• Protecting Azure Blob Storage ○ **Azure Blob Storage Overview** ○ **Cross-Cloud Restores of Metadata Between AWS S3 and Azure Blob**
Azure Data Lake Gen2	• Protecting Azure Data Lake Gen2 (ADL) ○ **Azure Data Lake Storage Gen2** • Backup support for ADL using the Hadoop (HDFS) Agent ○ **Azure Data Lake Storage Using the HDFS Agent**
Azure File Storage	• Protecting Azure Files ○ **Azure File Storage** ○ **Azure Files**

Azure NetApp Files	• Protecting Azure NetApp Files • Azure NetApp Files
Importing Backup Data to Azure	• Moving a large amount of backup data without large network bandwidth is possible using Azure Data Box ○ Migrating Data to Microsoft Azure Using Azure Data Box
Commvault Enhanced VM Conversion	• How to configure and get started ○ Converting to Azure • Compatibility table for all VM conversions ○ Cross-Hypervisor Migration • Common conversions to Azure ○ AWS to Azure ○ VMware to Azure ○ Hyper-V to Azure ○ Physical to Azure
App Migration	• Several applications support being migrated into Azure ○ Migrating Amazon Oracle RDS to Azure ○ SQL Application to an Azure VM ○ On-prem SQL to Azure SQL ○ On-prem SQL to Azure SQL MI
Cost Management	• Enabling cost management features on the platform can have significant impacts ○ Media Agent/Access Node Power Management ○ Auto-Scaling Azure Access Nodes ○ Appropriate Storage Tier use and Combined Tier use when necessary

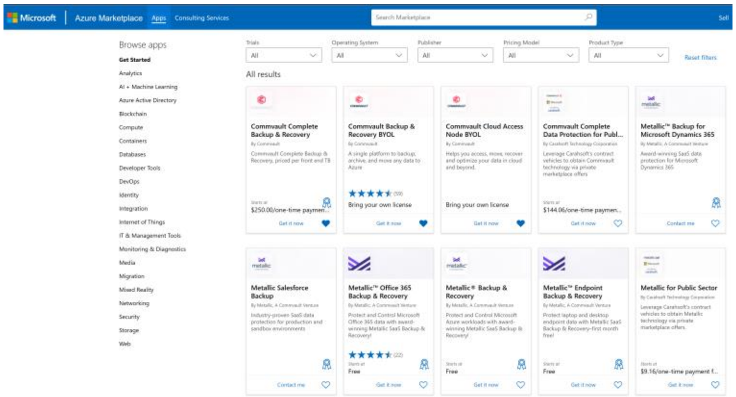
Getting started in the Marketplace

Commvault has provided several options in the Azure Marketplace so that our customers have a convenient and intuitive way to discover, purchase and use Commvault software making it easier for them to find and start using our solutions.

To get started, navigate to the **Azure Marketplace and search for Commvault**.

You will find both paid and bring your own license (BYOL) options available for the CommServe. Commvault Complete Backup and Recovery is billed directly through Azure Marketplace. Commvault Backup and Recovery BYOL does not need a metered purchase through Azure.

Access Node options are also available for Commvault Complete and Metallic in the Marketplace. An Access Node is a virtual machine deployed in Azure to manage communication and move data between Azure and Commvault data protection infrastructure.



Commvault Complete Backup & Recovery
By Commvault
Commvault Complete Backup & Recovery, priced per front end TB
Starts at
$250.00/one-time paymen...

Commvault Backup & Recovery BYOL
By Commvault
A single platform to backup, archive, and move any data to Azure
★★★★☆ (59)
Bring your own license

Commvault Cloud Access Node BYOL
By Commvault
Helps you access, move, recover and optimize your data in cloud and beyond.
Bring your own license

Commvault Complete Data Protection for Publ...
By Carahsoft Technology Corporation
Leverage Carahsoft's contract vehicles to obtain Commvault technology via private marketplace offers
Starts at
$144.06/one-time paymen...

Metallic™ Backup for Microsoft Dynamics 365
By Metallic, A Commvault Venture
Award-winning SaaS data protection for Microsoft Dynamics 365

Metallic Salesforce Backup
By Metallic, A Commvault Venture
Industry-proven SaaS data protection for production and sandbox environments

Metallic™ Office 365 Backup & Recovery
By Metallic, A Commvault Venture
Protect and Control Microsoft Office 365 data with award-winning Metallic SaaS Backup & Recovery!
★★★★☆ (22)
Starts at
Free

Metallic® Backup & Recovery
By Metallic, A Commvault Venture
Protect and Control Microsoft Azure workloads with award-winning Metallic SaaS Backup & Recovery!
Starts at
Free

Metallic™ Endpoint Backup & Recovery
By Metallic, A Commvault Venture
Protect laptop and desktop endpoint data with Metallic SaaS Backup & Recovery–first month free!
Starts at
Free

Metallic for Public Sector
By Carahsoft Technology Corporation
Leverage Carahsoft's contract vehicles to obtain Metallic technology via private marketplace offers.
Starts at
$9.16/one-time payment f...

- **Commvault Complete Backup & Recovery**, priced per front end TB. It is deployed on a single Azure Windows VM with a pre-defined configuration.

- **Commvault Backup & Recovery BYOL**. Deployed on a single Azure Windows VM, it is a fast way to setup and start protecting your Azure environment. It comes with a 150-day trial, so testing and prototyping with this instance type is also popular.

- **Commvault Cloud Access Node BYOL**. An Azure virtual machine that acts as an API coordinator and data mover for Azure infrastructure. It can scale both vertically, by increasing its instance type "size" or horizontally, by distributing workloads across low-cost virtual machines. In this way Commvault can ensure that the appropriate architecture is paired to the workload required for the least cost needed.

Commvault Protection of Azure Cloud Products

Azure Product	Commvault Link	Summary
Azure Virtual Machines	Azure Virtual Machine	You can use the Commvault software to back up and recover Azure VMs that are part of an Azure application. You can configure a hypervisor to represent the Azure application. **Data You Can Back Up** • Azure Generation 1 and Azure Generation 2 virtual machines. • Unmanaged and managed disks with Changed Block Tracking (CBT) enabled. For more information, see **Changed Block Tracking for Azure (CBT)**. • VMs that have encrypted blobs. • Virtual machines that are encrypted by Azure Key Vault. Backups of these VMs can include the operating system information, the data disks information, secrets (for example, token and password information), and encryption keys (for example, algorithm information). The encryption keys can be managed by Microsoft or by customers. • For Azure-managed disks, information about the configured Availability Zones, which are specific (physical) locations within an Azure region. • Azure-managed disks that are enabled with encryption at the host, on Windows or Linux VMs. • Proximity placement groups. • If the source VM is associated with an application security group, then the application security group is included in the backup.
Azure Active Directory	Azure Active Directory	You can use the Commvault software to back up and restore Azure Active Directory (Azure AD) and Azure Active Directory B2C account. **Data You Can Back Up** • Individual Azure Active Directory objects • Object attributes, such as User, Group, Enterprise application, and App registration
Azure Blob Storage	Azure Blob Storage	You can use the Commvault software to back up and restore Azure Blob Storage. **Data You Can Back Up** • Metadata of the object • Tags of the object

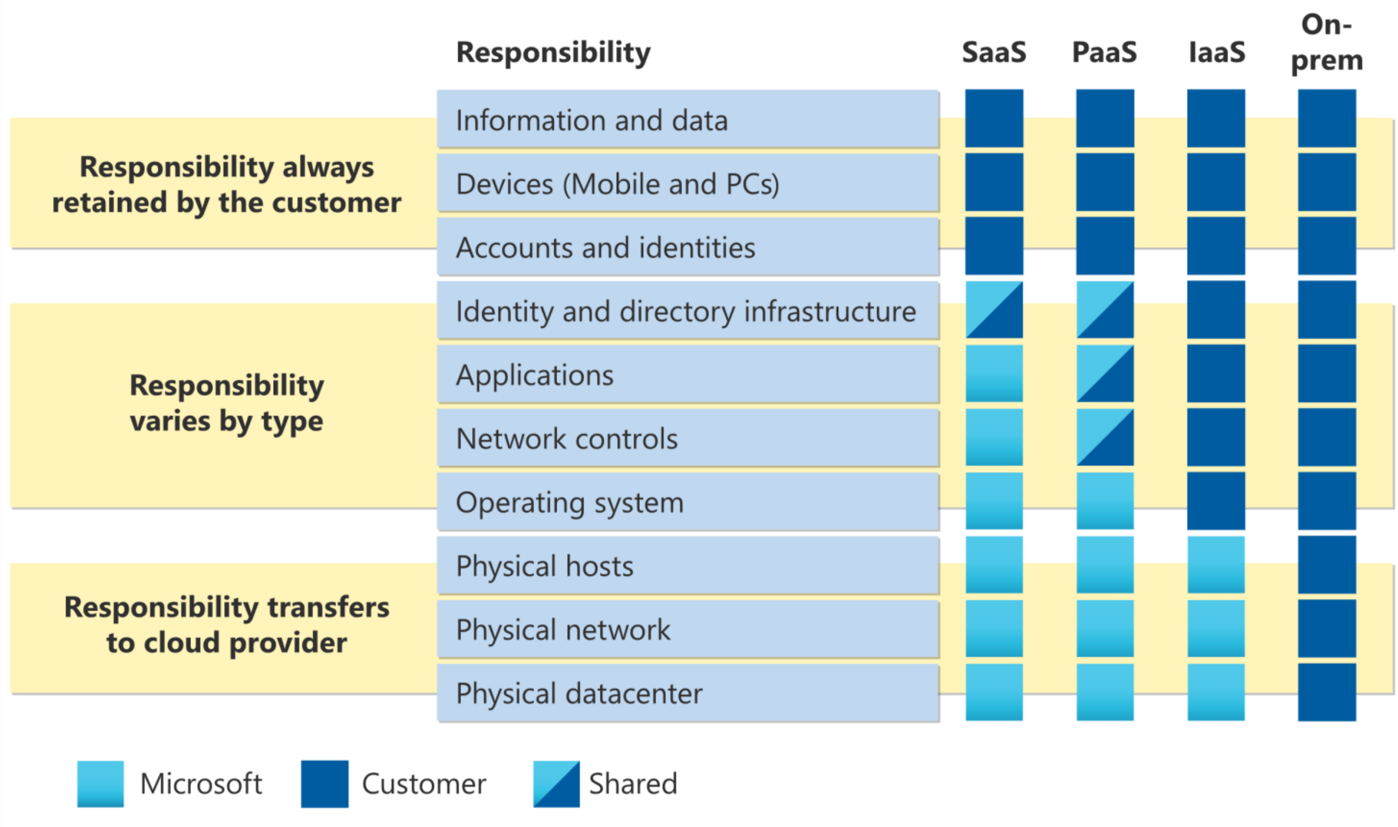
Azure Cosmos DB SQL API	Azure Cosmos DB SQL API	You can use the Commvault software to back up and restore Azure Cosmos DB database containers across multiple Azure accounts and regions. You can schedule full and incremental backups of database containers to meet recovery point objectives (RPO), long-term retention, and compliance needs. **Key Features** • You can manage containers from multiple Cosmos DB accounts, all from the Command Center. • You can back up and restore individual containers or databases, or the entire account.
Azure Cosmos DB Table API	Azure Cosmos DB Table API	You can use the Commvault software to back up and restore Azure Cosmos DB tables across multiple Azure accounts and regions. You can schedule table backups to meet recovery point objectives (RPO), long-term retention, and compliance needs. **Key Features** • You can manage tables from multiple Cosmos DB accounts, all from the Command Center. • You can back up and restore individual tables, or all the tables using the table group in an account or tables that belong to single or multiple accounts. • During backups and restores, if the primary access node is not accessible, you can automatically fail over the access node. • For all backups and restores, you can change the access node.
Azure Data Lake Storage Gen2	Azure Data Lake Storage Gen2	**Data You Can Back Up** • Metadata of the object • ACLs of the object
Azure DevOps	Azure DevOps	You can use the Commvault software to back up and restore Azure repositories. **Data You Can Back Up** • Azure repositories • Metadata (repositories inherit project settings): ○ Project visibility (public repository, private repository) ○ Project description • Pull request: ○ Title ○ Description ○ State ○ Comments, reviewers, and tags

Azure Files	Azure Files	You can use the Command Center to protect and recover data on the Azure cloud files share. Azure files is an Azure native service from Microsoft that provides fully managed file shares in the cloud that are accessible via the industry-standard SMB and NFS protocols.
Azure NetApp Files	Azure NetApp Files	You can use the Command Center to protect and recover data on the Azure NetApp Files service. Azure NetApp Files is an Azure native service from Microsoft and NetApp that provides high performance and low latency file services or file storage targets within the Azure environment.
Azure Table Storage	Azure Table Storage	You can use the Commvault software to back up and restore Azure Table storage data from storage accounts across multiple Azure subscriptions and regions. You can schedule full and incremental backups of Azure tables to meet recovery point objectives (RPO), long-term retention, and compliance needs. **Key Features** • You can back up and restore individual tables or all the tables in storage account. • You can back up and restore all the tables using the table group or tables that belong to single or multiple storage accounts. • During backups and restores, if the primary access node is not accessible, you can automatically fail over the access node. • For all backups and restores, you can change the access node.
GitHub	GitHub	You can use the Commvault software to back up and restore Git repositories. **Data You Can Back Up** • Git repositories • Metadata: ○ Repository visibility (public repository, private repository) ○ Repository description • Pull request ○ Title ○ Description ○ State ○ Pull request issue comments ○ Pull request review comments
Azure Database for MariaDB	Azure Database for MariaDB	You can use the Commvault software to back up and restore Microsoft Azure Database for MariaDB (PaaS).

		Refer to **Commvault's Books Online** for complete listing of data that is protected.
Azure Database for MySQL	Azure Database for MySQL	You can use the Commvault software to back up and restore Microsoft Azure Database for MySQL. You can back up Azure stack databases using an Azure VM that has MySQL installed. **Data You Can Back Up** - MySQL user databases (data)
Microsoft 365	Microsoft 365 Exchange Online OneDrive for Business SharePoint Online Teams	You can back up and restore Office 365. Refer to **Commvault's Books Online** for complete listing of data that is protected.
Azure Database for PostgreSQL	Azure Database for PostgreSQL	You can use the Commvault software to back up and restore PostgreSQL data on the Azure database for PostgreSQL (PaaS). **Data You Can Back Up** - PostgreSQL system databases - PostgreSQL user databases
SharePoint Server	SharePoint Server	You can use the Commvault software to back up and restore SharePoint Server. **Components You Can Back Up** - SharePoint databases - Farm components, such as Global Search settings, InfoPath Forms services, and State services. - Web application - Content databases - Configuration database - Service applications and service application proxies - SharePoint documents - Subsites - Lists and list items. - Document libraries and library items - Transaction logs
SQL	SQL	You can use the Commvault software to back up and restore Microsoft SQL Server databases. **Data You Can Back Up** - Database - Log files

		Backups You Can Perform • Block-level backups • Differential backups • Full backups • Full backups, using IntelliSnap • IntelliSnap backup copy • Transaction log backups
SQL Managed Instance	SQL Managed Instance	You can use the Commvault software to back up and restore Azure SQL Managed Instance. **Data You Can Back Up** • Managed instances **Backups You Can Perform** • Full backups
Azure SQL Database	Azure SQL Database	You can use the Commvault software to back up and restore for Azure SQL instance. **Data You Can Back Up** • Azure SQL Instance **Backups You Can Perform** • Full backups

Cloud Shared Responsibility Model



	Responsibility	SaaS	PaaS	IaaS	On-prem
Responsibility always retained by the customer	Information and data	■	■	■	■
	Devices (Mobile and PCs)	■	■	■	■
	Accounts and identities	■	■	■	■
Responsibility varies by type	Identity and directory infrastructure	◩	◩	■	■
	Applications	🟦	◩	■	■
	Network controls	🟦	◩	■	■
	Operating system	🟦	🟦	■	■
Responsibility transfers to cloud provider	Physical hosts	🟦	🟦	🟦	■
	Physical network	🟦	🟦	🟦	■
	Physical datacenter	🟦	🟦	🟦	■

🟦 Microsoft ■ Customer ◩ Shared

The Shared Responsibility Model is a security concept used by Microsoft in its Azure cloud computing platform to clearly define the roles and responsibilities of both Microsoft and the customer when it comes to securing their cloud environment.

In the Shared Responsibility Model, Microsoft is responsible for securing the infrastructure of Azure, including the physical facilities, servers, and network infrastructure, as well as ensuring the secure operation of the underlying cloud services. This includes maintaining the security of the Azure platform and keeping it up to date with the latest security patches and updates.

On the other hand, customers are responsible for securing their own workloads that are deployed in the Azure environment. This includes securing their virtual machines, storage accounts, network security, and application security. Customers are also responsible for securing their own data and ensuring its confidentiality, integrity, and availability.

Commvault can help customers achieve their responsibility requirements in the following ways.

Data Protection: Commvault can protect and manage customer data stored in Azure, ensuring its confidentiality, integrity, and availability. This includes features such as data encryption, deduplication, and compression to optimize storage utilization and reduce costs.

Compliance: Commvault can help customers meet regulatory and compliance requirements by providing auditing, reporting, and data retention capabilities. This helps customers ensure that their data is being managed and protected in accordance with industry standards and regulations.

Disaster Recovery: Commvault can help customers implement disaster recovery strategies, such as replicating virtual machines to a secondary site for failover in the event of an outage. This helps customers ensure that their workloads remain available and accessible even in the face of a disaster.

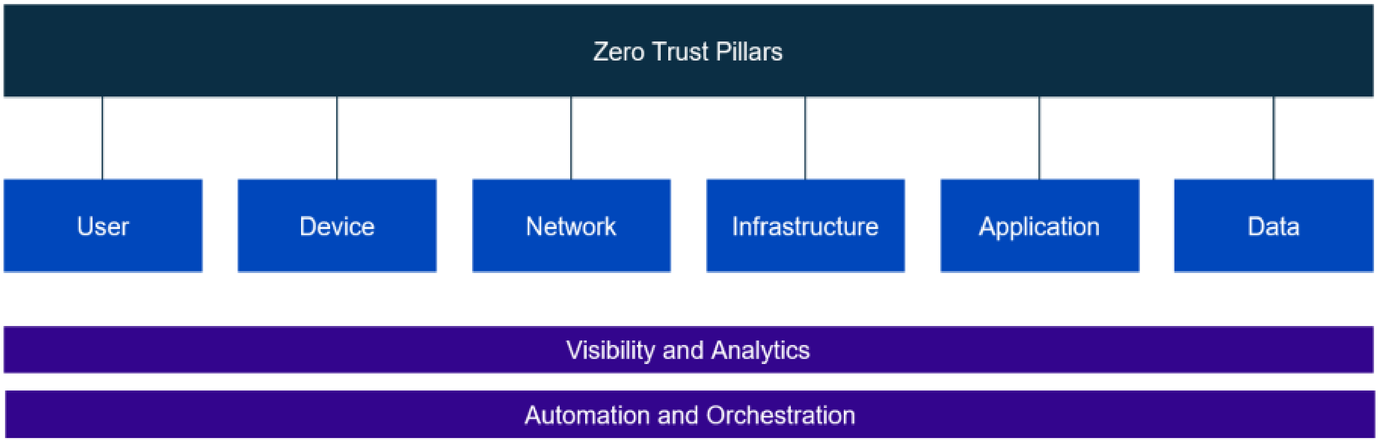
Zero trust architecture

The adoption of the cloud has meant that traditional perimeter defense and defense-in-depth approaches to securing an organization's applications and data are no longer acceptable. Zero trust (ZT) cybersecurity models are built on the following base assumption:

Assume that an attacker is always in the environment and that enterprise-owned or operated environments are no different or trustworthy than any non-enterprise-owned environment.

NIST Special Publication 800-207 – Zero Trust Architecture
https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf

This approach drives a fundamental change in how applications are architected, designed, and operated and is built on five (5) to eight (8) pillars – Commvault has features and functions that allow the implementation of a Zero trust architecture which are detailed below.



User Trust

Commvault software provides a centralized identity store that identifies individual users and groups permitted to interact and operate the multi-tenanted Commvault data management platform.

Authentication of users may occur utilizing Single Sign-On (SSO) centralized identity stores such as Active Directory (AD), secure LDAP, and externalized identity and access management (IAM) systems accessed using SAML 2.0 and OAuth (i.e., Okta, Ping, SecureAuth). The use of externalized IAM systems allows dynamic authentication rules that can incorporate location, time of day, and other metadata to determine if a user authentication request should be granted.

Additionally, access to Commvault administrative interfaces supports Multi-Factor Authentication (MFA) / Two-Factor Authentication (2FA) with support for Commvault and industry-leading web authentication applications (Google and Microsoft Authenticator). Hardware-based devices supporting Fast Identity Online (FIDO2) like Yubico Yubikey can be used as a second-factor device.

Common Access Cards (CAC) may also be used to perform password-less authentication.

Once authenticated, Commvault has a privileged access management system that combines the user/group, a role, and one or many entities (application, virtual machine, containers) that the user is permitted to act upon. All sessions are authenticated with a configurable default timeout of 30 minutes.

Events within the Commvault system that indicate a penetration or threat is present can programmatically modify a user or user group access dynamically by utilizing the Commvault REST API. Commvault has multiple anomaly detection and machine-learning algorithms that track and automatically respond to events and user behavior that is indicative of a threat (i.e., multiple failed login attempts).

Commvault software implements a secure multi-tenancy model where tenants are created as securely separated 'companies'. Commvault considers tenant isolation and granular role-based access control and a specialized form of macro-segmentation between companies. Individual tenant user rights and access controls may be considered micro-segmentation within the tenant.

One of the key features of Azure is its Zero Trust architecture, which is a security model that assumes breach and verifies each request as though it originated from an uncontrolled network. Zero Trust architecture is based on three guiding principles: verify explicitly, use least privileged access, and assume breach.

To verify explicitly, Azure can use Multi Factor authentication with conditional access that considers user account risk, device status, and other criteria and policies that you set. Azure also uses identity and access management solutions such as **Azure Active Directory** to manage user identities and enforce granular access policies.

To use least privileged access, Azure limits user access with **Just-In-Time** and **Just-Enough-Access** (JIT/JEA), risk-based adaptive policies, and data protection. Azure also uses role-based access control (RBAC) to assign permissions to users, groups, and applications based on their roles and responsibilities.

To assume breach, Azure minimizes blast radius and segments access by using **Azure Virtual Networks (VNets)** to isolate workloads and resources. Azure also uses end-to-end encryption and analytics to get visibility, drive threat detection, and improve defenses. Azure also leverages **Microsoft Defender for Cloud** to provide threat protection and intelligence across identities, devices, applications, data, infrastructure, and networks.

By applying Zero Trust principles to Azure, you can ensure security with user trust and protect your applications and data from cyberattacks.

Device Trust

Commvault performs active vulnerability management and reporting, for issues that impact Commvault products at **documentation.commvault.com/2022e/essential/146231_security_vulnerability_and_reporting.html.**

Vulnerabilities may require an update to Commvault software and/or Operating system and third-party software on client devices.

Commvault endpoint protection software provides device security and enterprise mobility management for the protection of high-risk mobile data (i.e., laptops, tablets, phones). Commvault endpoint and edge protection include mechanisms to perform data-loss prevention periodic encryption, identify device location, and initiate automated remote wipes in response to lost devices.

Commvault software maintains a unique device identity by enrolling all protected devices with a device-specific cryptographic certificate that is managed and rotated periodically by Commvault. All communications (control and data plane traffic) implement device authentication using the certificate to establish and validate identity. All communications between the device and Commvault data management infrastructure are encrypted using an AES256 cipher.

Commvault software assists in device management by providing centralized device configuration reporting and insights. Additionally, Commvault software automatically downloads and deploys software patches and updates per the defined policy on all client and core data management devices/appliances. Commvault can also manage the deployment of Windows Operating System (OS) updates if required.

Commvault centralized reporting provides a device inventory of all protected hardware and details all installed protection modules and configuration.

With Azure’s implementation of Zero Trust Architecture, Device Trust means that only devices that meet certain criteria, such as being healthy, compliant, and up to date, can access sensitive resources and data. **Azure Active Directory** (Azure AD) is the policy decision point that enforces device trust policies based on insights on the user, device, target resource, and environment. Azure AD also integrates with other Azure services, such as **Microsoft Defender for Cloud** and **Azure Web Application Firewall**, to provide threat protection and end-to-end encryption for devices accessing web applications hosted in Azure.

Network Trust

Network trust is the degree of confidence that a network is secure and reliable. Traditionally, network trust assumes that the perimeter of the network is well-defined and protected, and that anything inside the network is trustworthy. However, this assumption is no longer valid in the era of cloud computing, mobile devices, and cyberattacks.

Commvault software allows deployment into any communications network topology and enforces the network rules or network access control policy for the organization. Commvault software supports direct connections, port-forwarding network gateways, DMZ-based network gateways, authenticated HTTP proxies, and advanced network topologies that enforce one-way, two-way, and bi-directional tunneling on discrete ports.

Commvault software effectively provides a software-defined network topology and/or software-defined perimeter that mimics or mirrors the network used for data protection. Network topologies may be modified programmatically using CLI, SDK, and REST APIs to provide dynamic network controls and configuration in response to detected threats.

Commvault software provides micro-segmentation at the workload level. Each device running the Commvault software core package includes an application-level firewall that allows discrete access control on defined ports and protocols. Additionally, macro-segmentation is provided via network topologies and network gateways/firewalls that dictate how data can flow between Commvault data management components.

Commvault encrypts all data in-transit (control plane, data plane) using per-device symmetric cryptography where the same key is used for encryption and decryption, and AES-256 cipher suite is used by default. Communication is session-based with re-authentication required periodically to ensure devices and users reestablish their privileges.

Commvault software crypto library implementation has been certified as **FIPS-140-2 cryptographic module validation program compliant**.

Zero trust architecture in Azure is a set of principles and best practices that help you implement zero trust security in your cloud environment. It involves applying granular policies and controls to every layer of your network, from the identity and device level to the application and data level. Use of **Network Security Groups** and **Application Security Groups** can provide capabilities for securing workloads at a use case level without needing to define or manage network addressing for security increasing security and flexibility. By adopting zero trust architecture in Azure, you can reduce your attack surface, improve your visibility and compliance, and enhance your user experience.

Infrastructure Trust

Commvault keeps enterprise workloads secure while migrating between cloud environments by encrypting control and data-plane traffic in-transit. Additionally, VM conversion activities that utilize temporary cloud storage locations may write to encrypted-only object storage buckets, with cloud-provider or customer-managed keys (CMKs) specified.

Commvault in essence is a cloud access security broker (CASB), with multi-cloud, multi-account permissions to protect (read) and restore or migrate (write) data from and to cloud environments. Commvault uses privileged cloud access credentials to perform cloud data management and provides an access control layer that authenticates and authorizes users before allowing access to cloud resources. All actions attempted and executed against a cloud are logged in the Commvault immutable audit log for traceability and forensic analysis.

Azure provides several services and features that can help you implement zero trust infrastructure in your cloud environment. Some of these are:

- **Azure Active Directory** (AAD): A cloud-based identity and access management service that enables single sign-on, multi-factor authentication, conditional access policies, and identity protection.
- **Azure Sentinel**: A cloud-native security information and event management (SIEM) platform that collects and analyzes data from various sources, such as Azure services, devices, applications, and users. It also provides threat detection, response, and hunting capabilities.
- **Microsoft Defender for Cloud** (formerly Azure Security Center): A unified security management service that provides visibility, assessment, and protection for your Azure resources. It also integrates with Azure Defender, a cloud workload protection platform that offers advanced threat protection for different types of Azure resources.
- **Azure Firewall**: A stateful firewall service that provides network and application-level protection for your Azure virtual network. It also supports features such as network address translation (NAT), traffic filtering, threat intelligence, and service tags.
- **Azure Key Vault**: A secure storage service that allows you to manage and encrypt your secrets, such as passwords, keys, certificates, and tokens. It also supports features such as access policies, logging, backup and restore, and hardware security modules (HSMs).

Application Trust

Commvault software utilizes multiple web application components to service HyperText Transfer Protocol (HTTP) requests to Commvault Command Center™, WebConsole, and REST API endpoint(s). Commvault software supports the use of an Operating System (OS) firewall and attack surface reduction rules to automatically block known malicious behaviors.

Commvault recommends implementing cloud provider web application in front of Commvault web-services to provide an additional level of threat detection and mitigation.

Commvault development practices employ multiple methods to develop and maintain secure data handling at all stages of data management. Commvault development practices require peer review from multiple parties including security domain specialists. Commvault performs quarterly static vulnerability scanning and remediation on Commvault software and third-party libraries and performs penetration testing both internally and via third-party engagements. Commvault is committed to detecting and resolving security issues rapidly and provides methods for individuals and organizations to report security defects for prioritized resolution.

Commvault utilizes the Quay/Clair vulnerability scanner on container images utilized by the Commvault software and ensures zero (0) issues are reported. Commvault software uses the :latest tag for actively maintained official docker images to ensure Commvault software is always using the most current and patched OS image.

Commvault software and by extension Metallic Data Management as a Service (DMaaS) both provide a secure access cloud layer that provides a privileged access control layer to authenticated and authorized organizational users and user groups. Granular access to perform read-only reporting restores to new locations or restores to the original location for service recovery can all be granted.

Commvault software employs a least privilege approach to cloud data management requesting only the minimum permissions required to protect an Azure service. Commvault role-based access controls (RBAC) are then overlayed to further restrict the individual user and/or user group.

Commvault software is accessible via the HTTPS portal and provides an ‘any device access’ approach to enterprise backup and archival data. Data may be accessed from any device (PC, tablet, mobile phone) and downloaded or restored to any location by authorized users.

Data Trust

Commvault software provides software and hardware encryption for data-in-transit and data-at-rest in the cloud and on-premises data storage locations. Commvault has a built-in FIPS-140-2 compliant cryptographic library for generating and rotating encryption keys stored securely within the Commvault Database (CSDB). Alternatively, customers may choose to utilize Cloud-provider Key Management Services like **Azure Key Vault**. Commvault integrates natively with cloud-provider key management services to transparently access encrypted data in-use in unencrypted form. Commvault accesses encrypted application data in unencrypted format and transfers securely via an encrypted tunnel to Commvault encrypted storage. When transferring cloud-native snapshots between encryption boundaries, Commvault decrypts and then re-encrypts snapshots with customer-selected encryption keys.

Commvault software provides data security by allowing data classification and tagging by detecting sensitive content, data protection based on data classification, and data masking to prevent data leakage/spillage in cross-environment data restore (i.e., production to development seeding). Commvault prevents data leaking by performing redaction for sensitive email and indexed data fields during export to PDF for sharing with external parties.

Commvault software helps businesses comply with relevant industry, geography, and organizational regulations and policies. Commvault has been successfully deployed and certified to many industry standards, a published list of relevant standards where Commvault implementations have been certified is here:

Certifications and Compliance

https://documentation.commvault.com/2022e/expert/110316_certifications_and_compliance.html

Commvault is responsible for the integrity of data persistent within Commvault-controlled cloud, HyperScale™, disk, and tape storage locations. Commvault performs periodic data verification jobs to validate stored data has not been modified since the initial backup. Data verification in cloud storage locations is not recommended due to the recall or retrieval cost. The durability of cloud storage and storage of multiple independent copies prevents the need to perform periodic costly data verification.

Commvault File Storage Optimization, Data Governance, and eDiscovery and Compliance index and classify backup, archival and live data to organize and optimize data by risk and value. Data may be analyzed based on access and modification frequency, access control settings, and the presence of personally identifiable information. Legal and compliance requests can perform granular keyword searches and exports to respond to external and internal forensic investigations.

Visibility and Analytics

Commvault software provides threat intelligence via a granular audit log of all activities occurring within the data management platform. Events may be forwarded to external Security Incident and Event Management (SIEM) or Security Orchestration Automation and Response (SOAR) systems via Syslog, SDK, or custom action.

Commvault recommends that audit events are forwarded to **Azure Monitor** to visualize, report, and automate responses to critical events and threats. Additionally, Azure Monitor integrates with **Microsoft Sentinel** and **Microsoft Defender for Cloud** so that visibility of threat progression can be observed.

Commvault provides continuous diagnostics and mitigation capability via centralized reporting and alerts for infrastructure (servers, virtual machines, desktop/laptops) that are lagging behind the most current software patches and updates and require reboots or configuration upgrades.

Automation and Orchestration

Commvault software represents a centralized policy engine (PE) that is responsible for evaluating a user or user group request for access to a resource. The policy is implemented as a three-way relationship between a user/user group, role, and a Commvault entity (server, VM, database, application, etc.).

Commvault provides the Commvault Firewall Daemon (CVFWD) on all data management infrastructure, which performs the role of a policy administrator (PA) establishing and terminating encrypted communication tunnels for authorized data management activities.

Commvault software also performs policy enforcement (PE) from the centralized CommServe® instance which monitors, initiates, and terminates communication as required to complete data management activities. The CommServe Job Manager communicates with individual policy administrator instances to monitor and terminate connections on-demand.

Azure Automation is a cloud-based service that helps you automate and manage your workloads securely across your Azure and non-Azure environments. It offers features such as process automation, configuration management, update management, change tracking, and inventory.

With Azure Automation, you can:

- Create and run graphical, PowerShell, or Python runbooks to automate frequent, time-consuming, and error-prone tasks.
- Monitor and update Windows and Linux systems within a defined maintenance window.
- Author and manage PowerShell configurations to ensure consistent management for Windows and Linux machines.
- Collect an inventory of operating system resources and track changes across services, daemons, software, registry, and files.
- Integrate with other Azure services and third-party systems using serverless runbooks and webhooks.

Azure Automation helps you simplify cloud management, reduce errors and overhead costs, and deliver more reliable services faster.

Additional resources

NIST Special Publication 800-207 Zero Trust Architecture

https://csrc.nist.gov/publications/detail/sp/800-207/final

- NIST Special Publication 800-63-3 Digital Identity Guidelines

https://pages.nist.gov/800-63-3/

- Whitehouse.gov – Memorandum For The Heads of Executive Departments And Agencies

https://www.whitehouse.gov/wp-content/uploads/2022/01/M-22-09.pdf

- U.S Cybersecurity & Infrastructure Security Agency (CISA) Cloud Security Technical Reference Architecture

https://www.cisa.gov/cloud-security-technical-reference-architecture

- U.S General Services Administration – Zero Trust Architecture Buyers Guide

https://www.gsa.gov/cdnstatic/Zero Trust Architecture Buyers Guide v11 20210610.pdf

- Zero Trust security in Azure

https://learn.microsoft.com/en-us/azure/security/fundamentals/zero-trust

- How to build a Zero Trust Recovery Solution with Commvault and Metallic

https://www.commvault.com/blogs/build-a-zero-trust-recovery-solution-with-commvault-and-metallic

- Zero Trust Networks, Evan Gilman, Doug Barth, Oreilly Publishing

https://oreilly.com/library/view/zero-trust-networks/9781491962183/ch01.html

- Microsoft Azure Well-Architected Framework

https://learn.microsoft.com/en-us/azure/well-architected/

Ransomware protection

Ransomware and malware protection are key considerations in any cloud-based architecture. Commvault has the following ransomware protection capabilities that should be layered into your layered data protection plan.

Hardening your CommServe

Consider **hardening your CommServe**® instance by restricted access to authorized users, authorized hosts, and authorized ports.

Access to the centralized configuration MS SQL Server database should also be limited, see **Securing the CommServe Database**.

Account segregation

Utilize multiple Azure subscriptions to segregate users, departments, and backup copies. Commvault supports the protection and storage of data across multiple accounts, regions, and availability zones.

Air-gapped backup copies

An air gap back copy is a specialized type of backup copy that provides additional protection over the traditional day to-day copy. An air gap is defined by the NIST Computer Security Resource Center (CSRC) as:

“An interface between two systems at which (a) they are not connected physically and (b) any logical connection is not automated (i.e., data is transferred through the interface only manually, under human control).”

Source(s): CNSSI 4009-2015 from IETF RFC 4949 Ver 2

- Consider air-gapping at least one copy of your critical data by powering down the MediaAgent that provides access to the Data. This approach ensures an effective response to an unplanned and uncontrolled ransomware propagation event.
  - Commvault Cloud MediaAgent Power Management can be used to automate the power-down/power-up of the MediaAgent.
  - (optional) Consider utilizing Azure Scheduler to schedule the power-up/power-down of air-gapped MediaAgents.
  - A powered-down MediaAgent is considered an ‘air-gapped’ copy from your primary operational recovery infrastructure.
- Always utilize Azure Blob cloud storage with dedicated credentials and authorized MediaAgents
  - The use of object-based storage without end-user access is considered another airgap for your data.
  - Traditional storage mediums (disk-based storage) can be directly accessed by malware and infected if security credentials are discovered/breached.

See Offline Backup Copies for additional information.

WORM backup copies

Consider utilizing write-once-read-many (WORM) storage mechanisms. Commvault provides the Enable Worm on Cloud Storage workflow to configure cloud and Commvault best practices in one easy step.

See the following information for more details:

- Configuring the WORM Storage Mode on Cloud Storage
- **Setting Up immutable storage in Azure blob**
- Workflow for Configuring WORM Storage Mode on Cloud Storage

Note - WORM should ideally be enabled on the backup library before writing any backup data, to ensure all data is immutable from initial creation.

Azure Blob Best Practices

When the DDB is sealed, the sealing process closes the DDB and starts a new DDB. When the new database is started, the next instance of each data block processed creates a new signature tracking entry and the data block is written to the disk again as a new initial baseline.

This will result in a full copy of the backup content being resent to the Cloud Library. This is intentional and provides multiple, segregated data copies to protect from corruption or other unforeseen data access issues.

This will require access to the Java GUI, which will have limited availability in later versions. See **Sealing the Deduplication Database**.

Design Patterns

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.

Commvault recommends the following design principals when making architecture decisions:

Right-sizing over Forecasting

Always select the smallest recommended size applicable to its use. Cloud resources are elastic, able to expand and contract as needed. Regular assessment and adjustment will provide significant cost savings.

The following three **Azure VM Types** are recommended as of this publication. The leading letter defines the category of VM type and corresponds to its ratio of CPU and RAM. The number in the second position determines its size, usually indicating how many vCPU's it has allocated. If the number is followed by an "a" then it is an AMD CPU, if not then it is an Intel. Following the "a", and "s" indicates that it can use Standard SSD, Standard HDD, Premium SSD and Ultra Disk.

Easv5	E Series has a high RAM to vCPU ratio of 8:1	AMD's 3rd Generation EPYCTM 7763v
Dasv5	D Series has a RAM to vCPU ration of 4:1	AMD's 3rd Generation EPYCTM 7763v
Fsv2	F Series has a RAM to vCPU ration of 2:1	3rd Generation Intel® Xeon® Platinum 8370C

Lastly, the v5 indicates the generation of CPU that the VM will be able to allocate. The difference in performance between generations can be as much as 100% improvement with only an incremental difference in cost, if any at all.

The recommended streams per Access Node is based on its CPU and available RAM. In Azure this translates into a machine type. For a D Class machine, each VM size should process ~25 streams, started at a machine type 2. So, an D2as_v5 would support 25 streams and an D8as_v5 would support 100 streams.

Match protection to business impact

Understanding your data classification model and matching that with retention and compliance will reduce cost and help manage SLAs. Rarely can a single data retention be applicable for all data and RPO/RTO requirements. Aligning business critical data into rapid recovery layers, like snapshots, and long-term data into archive blob storage ensures appropriate cost.

Central Management of Azure Snapshots

Managing Azure snapshots and Azure's New Recovery Points using Commvault provides a single point of management and help eliminate dark data and orphaned assets that can drive cost up.

Optimize at Rest

Ensure that all copies of data are stored at the least cost optimized format and storage-tier applicable for their recovery. If a copy of data no longer requires rapid recovery but continues to require retention, consider moving it to archive tiers for lower cost long term retention.

Azure supports several tiers for **Blob** data storage:

- Hot tier - An online tier optimized for storing data that is accessed or modified frequently. The hot tier has the highest storage costs, but the lowest access costs.
- Cool tier - An online tier optimized for storing data that is infrequently accessed or modified. Data in the cool tier should be stored for a minimum of 30 days. The cool tier has lower storage costs and higher access costs compared to the hot tier.
- Cold tier - An online tier optimized for storing data that is infrequently accessed or modified. Data in the cold tier should be stored for a minimum of 90 days. The cold tier has lower storage costs and higher access costs compared to the cool tier.
- Archive tier - An offline tier optimized for storing data that is rarely accessed, and that has flexible latency requirements, on the order of hours. Data in the archive tier should be stored for a minimum of 180 days.

Commvault recommends using LRS (Locally Redundant) blob storage in a Cool tier for the most cost-efficient data retention.

Optimize on Wire

Cross Region, cross Availability Zone (AZ) and Cloud egress can all incur cost. Ensuring the data in flight is compressed and deduplicated will decrease this cost and reduce transit times. Additionally, Commvault supports selective replication of data. This can provide the means to protect applications with tiers of availability and reduce wasteful replication if the data isn't needed.

Azure has a bandwidth **cost** for ingress and egress from Availability zone and Regions. If the amount of data crossing data zones has a significant cost, then positioning Access Nodes and Blob Storage within AZ and Regions of data can reduce the transmission cost. This always needs to be balanced against the cost of the VM.

Separation of Duty

Commvault recommends leveraging fine-grained **role-based access control** (RBAC) to provide users, admins, and business analysts the least privilege rights to protect, recover and report on business-wide data management.

Encrypt Everything

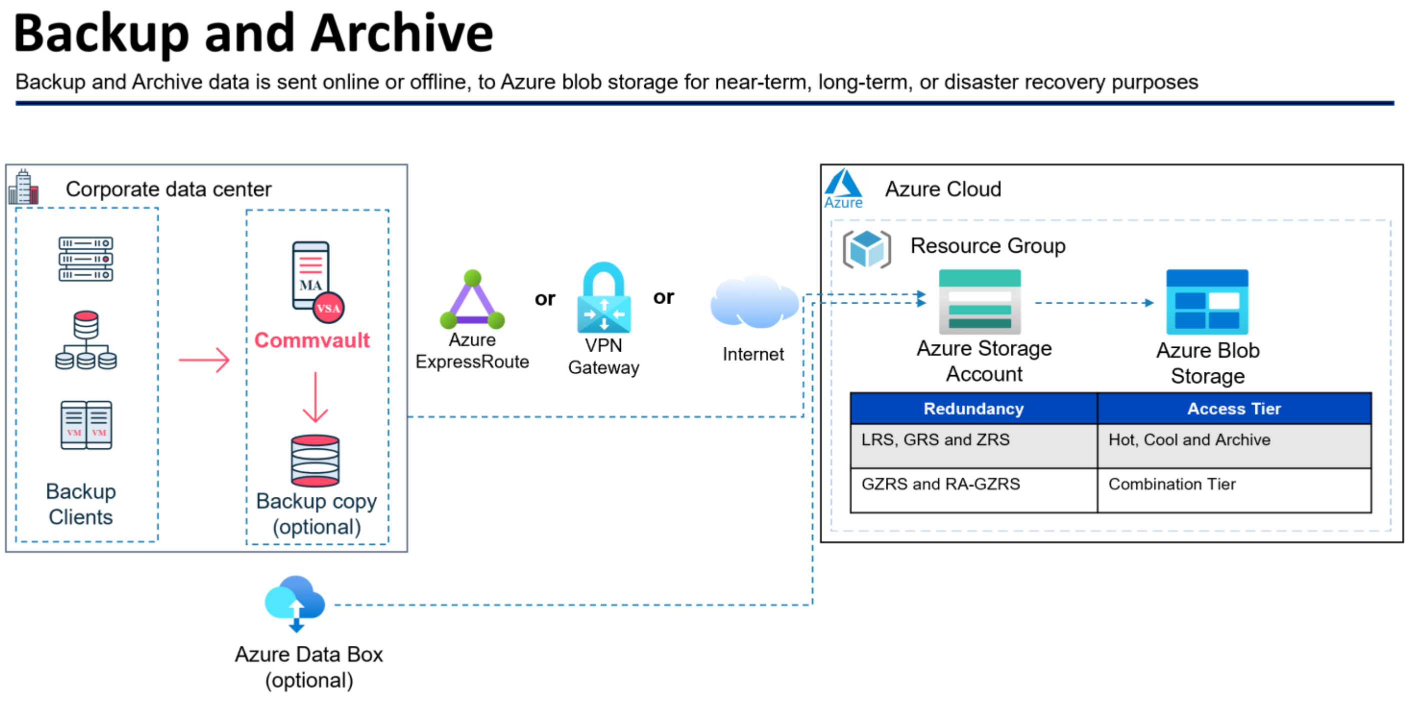
Encrypting everything with backups is a good practice to protect your data from unauthorized access or theft. Commvault backups support encryption at rest and in transit, meaning that your data is encrypted both when it is stored in blob storage and when it is transferred over the network. Encryption does not add noticeable performance overhead but does add an extra layer of protection.

Automate Over Runbooks

Automate operations to scale with speed and remove human error from daily operations. Commvault API, SDK, and CLI allow integration with Azure Resource Manager, Terraform, and **Azure Policy** , and **Azure Automation** to automate operations where required.

Backup and Archive to Azure

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.



Backup and Archive

Backup and Archive data is sent online or offline, to Azure blob storage for near-term, long-term, or disaster recovery purposes

Redundancy	Access Tier
LRS, GRS and ZRS	Hot, Cool and Archive
GZRS and RA-GZRS	Combination Tier

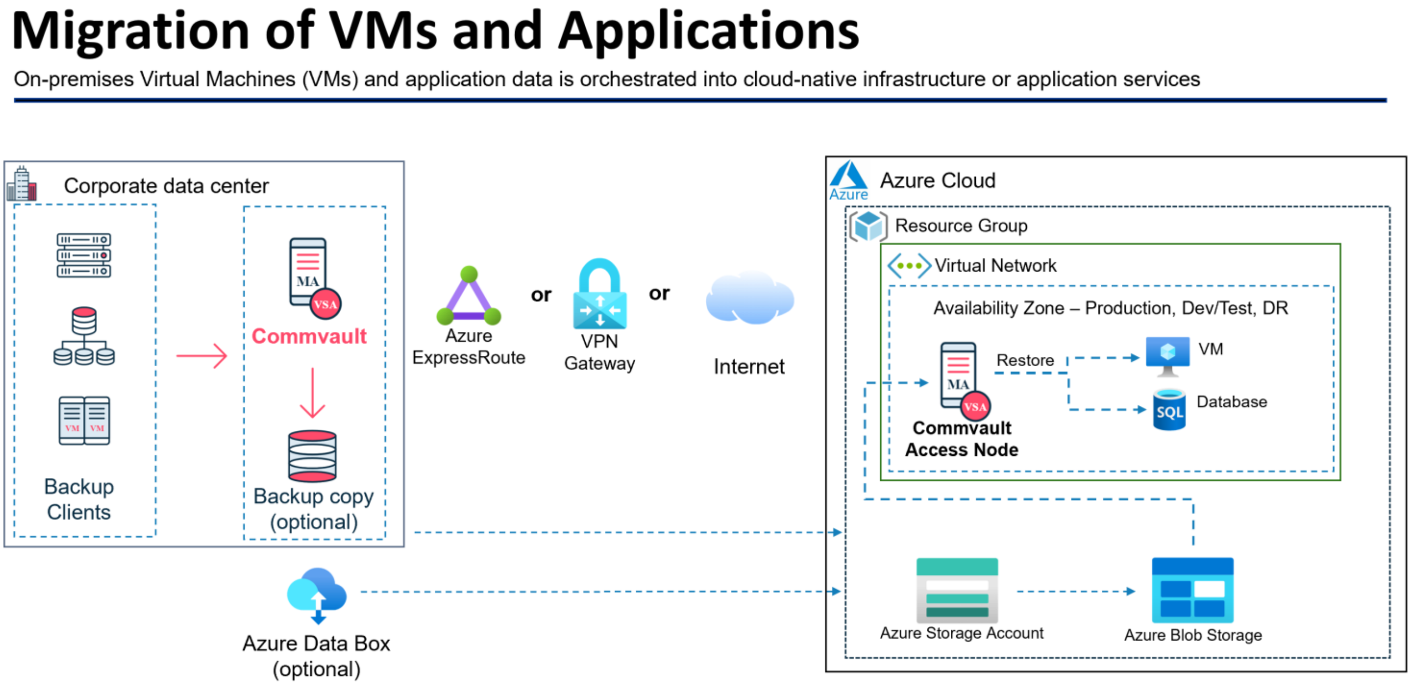
Move Data – Backup and Archive to the Cloud	
Scenario/Suitability	**Requirements**
Offsite storage and “tape replacement” scenario – replace long-term on-site retention with cloud storage. Native, direct connectivity to 40+ object storage endpoints, including Azure Blob Storage – no requirement for translation, gateway, or 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 Azure or Metallic® Recovery Reserve™. Local backup copy on-premises is not required.	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**). Can use offline transfer method by writing backup, archive data to **Azure Data Box** device(s), then importing Azure Blob storage. (optional) In-cloud MediaAgent can be created to support tiering, DR solution in the cloud using the data that is placed in the cloud. This can be done at time of DR or DR Test, as required.

Remote Office Branch Offices (ROBOs) can send data directly to Cloud.	Cloud MediaAgent can be deployed at the time of the Disaster Recovery event or test, as required. Remote offices do not require onsite Commvault infrastructure, they can upload backup data direct to Azure Blob storage.

Design Considerations
Selecting the correct storage in this scenario will ensure cost effective retention. The Storage Account will determine the redundancy of the Blob storage as well as the default Region. The Blob container can belong to any Region, but it will determine the accessibility of the data. Commvault recommends the use of Locally Redundant Storage (LRS) and a Cool tier. There is no performance difference between Cool and Hot. The Archive tier does have an access time penalty.
The use of Key Value tags for all resources will make billing audits easier and management of resources easier to manage.
Commvault provides an alternative to native Azure Blob storage with Metallic® Recovery Reserve™. It utilizes Azure Storage on the backend but is managed by Commvault with a flat rate and no egress charges. This can help eliminate budget surprises due to recovery efforts.
This design is commonly referred to a Tape elimination scenario where a customer simply needs to eliminate the offsite tape or secondary copy location for their data protection strategy. Azure makes an ideal solution for this solution.

Migration of VMs and applications to Azure

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.



Migration of VMs and Applications

On-premises Virtual Machines (VMs) and application data is orchestrated into cloud-native infrastructure or application services

Move Data – Migration of VMs and application to the cloud	
Scenario/Suitability	**Requirements**
Lift & shift of virtual machines – Application-consistent VM backups are used to restore and convert VMware, Hyper-V, Acropolis, Oracle VM, and AWS 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**).

Design Considerations
Selecting the correct storage in this scenario will ensure cost effective retention. The Storage Account will determine the redundancy of the Blob storage as well as the default Region. The Blob container can belong to any

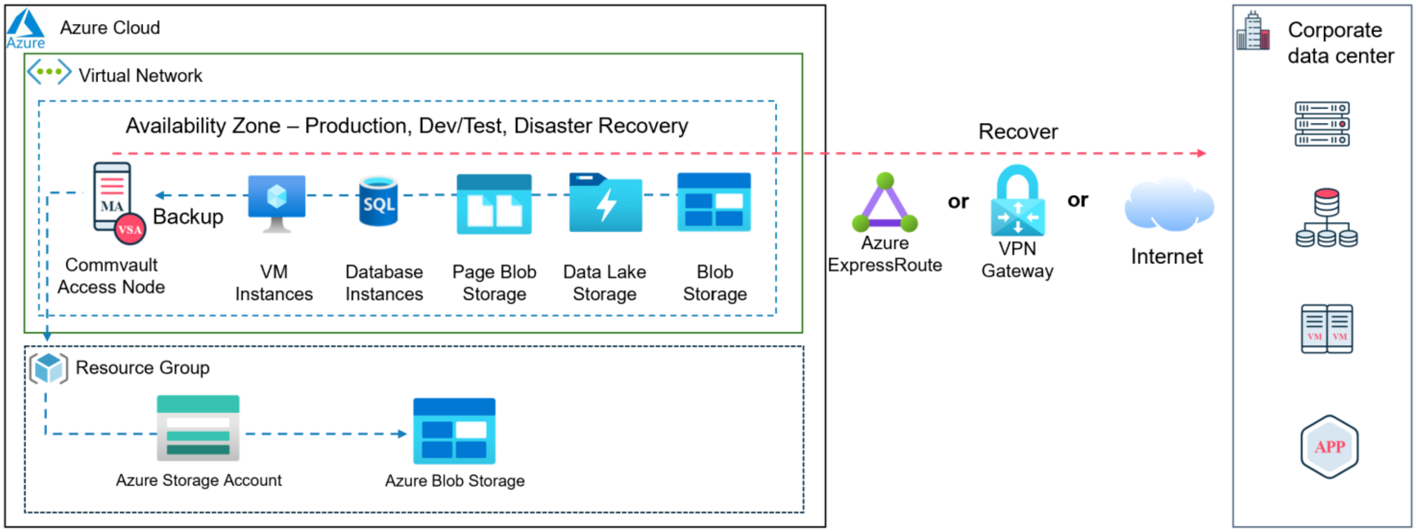
Region, but it will determine the accessibility of the data. If the data is only landing in Azure Blob Storage for a short period of time as it is being imported into a VM, the Commvault would recommend using LRS Hot storage to avoid the 30 day storage fee and the transaction fees for reading the data immediately.
The use of Key Value tags for all resources will make billing audits easier and management of resources easier to manage.
A clear migration strategy and roll out plan will ensure that resources are not over allocated. The performance of migrating to the cloud is dependent on bandwidth available between the customer site and Azure and the resources allocated to perform the work. More MediaAgents can move more data and build more streams, but if the bandwidth is the bottleneck, those resources are wasted.
This is a common “Journey to the Cloud” milestone. As customers move VM’s in mass to Azure to take advantage of account credits and longer support periods for older machine types.

Protection in Azure

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.

Protection in the Cloud

Protect and recover cloud workloads, cloud-native applications both within cloud and back on-premises



Manage data – Protection in the cloud	
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 with application awareness. 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 Microsoft 365 – mail, OneDrive for Business and SharePoint online either in cloud or back to on-premises	Azure Virtual Server Agent and MediaAgent deployed on an Access Node within IaaS provider for agentless backup. Applications not supported by VSA application awareness and requiring application-level consistency can be protected via agents-in-guest. 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).

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.	

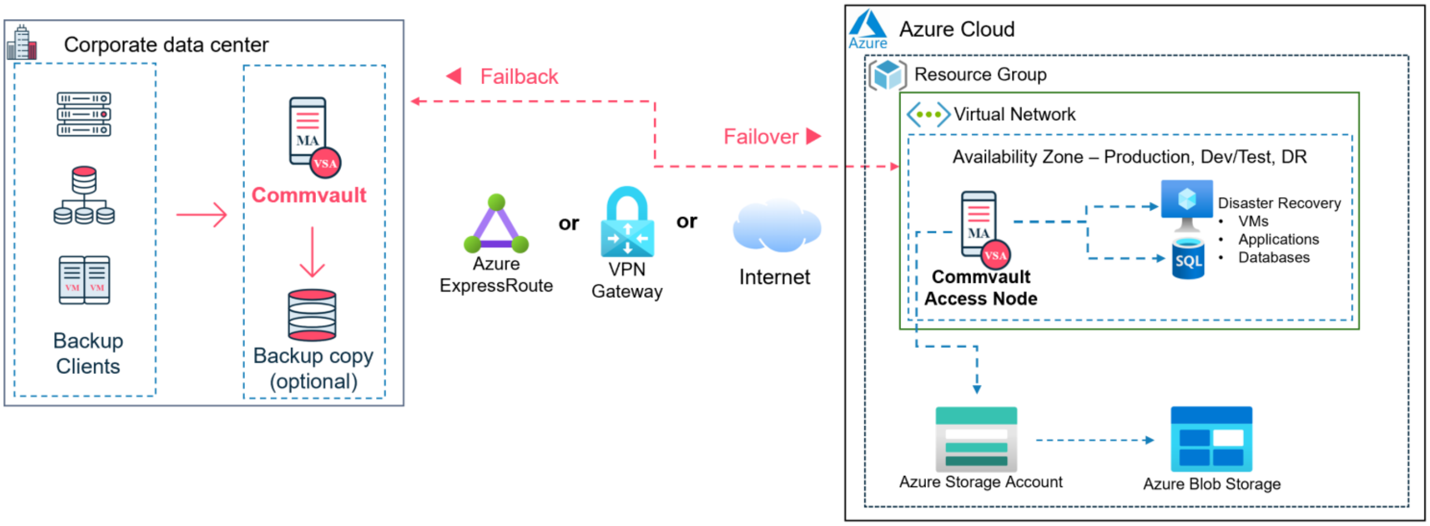
Design Considerations
Selecting the correct storage in this scenario will ensure cost effective retention. The Storage Account will determine the redundancy of the Blob storage as well as the default Region. The Blob container can belong to any Region, but it will determine the accessibility of the data. Commvault usually recommends LRS Cool Blob Storage for data protection purposes. It is the most common to this solution. Longer term retention data can be sent to a combined tier using Cool and Archive for additional cost savings.
The use of Key Value tags for all resources will make billing audits easier and management of resources easier to manage. If possible (and it should be) data protection should be defined by the use of tags so that those resource designations can be dynamically recognized by Commvault VM Groups. This eliminates management overhead from the end user and ensures that resources that follow proper name tag policy will be protected.
Access nodes can auto-scale. Using a select number of MediaAgents within Regions and allowing auto-scale to create ephemeral resources when needed helps control infrastructure costs. Another is to make sure all infrastructure in Azure is using Power Management if available. These two design principals will help control idle CPU cost.
It's not as common to see customers making a local copy of their data back to an on-prem resource, but it does happen. In that scenario make sure that Dash Copies are used, and the data is selectively copied to incur the lowest egress costs.

Disaster recovery to Azure

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

Disaster Recovery to the Cloud

Recover your primary production site (on-premises, cloud) to the cloud for on-demand, near real-time disaster recovery resource



Use Data – Disaster Recovery to the Cloud	
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. Commvault Periodic 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, Hyper-V, Acropolis, Oracle VM, and AWS 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 and Hyper-V 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**).

Design Considerations
Selecting the correct storage in this scenario will ensure cost effective retention. The Storage Account will determine the redundancy of the Blob storage as well as the default Region. The Blob container can belong to any Region, but it will determine the accessibility of the data. Commvault usually recommends LRS Cool Blob Storage for data protection purposes. It is the most common to this solution. Longer term retention data can be sent to a combined tier using Cool and Archive for additional cost savings.
The use of Key Value tags for all resources will make billing audits easier and management of resources easier to manage. If possible (and it should be) data protection should be defined by the use of tags so that those resource designations can be dynamically recognized by Commvault VM Groups. This eliminates management overhead from the end user and ensures that resources that follow proper name tag policy will be protected.
For the latest list of support for cross-platform replication: Cross-Platform Feature Support for Replication (commvault.com) For the latest list of support for cross-platform restore: Cross-Hypervisor Restores (VM Conversion) (commvault.com)
This solution has the most cost saving potential as a customer can leverage a warm cloud repository over a complete secondary data center location. Design principals to ensure this solution is viable are to optimize over the wire and power management. Those will be critical to controlling cost.

Architecture Sizing

The nature of cloud scale and right sizing lies in the ability to be flexible. This is a notable change in architecture from an on-prem design. It needs to be re-evaluated and resized regularly to ensure that new demands are met.

Start with the smallest category of an Azure instance size that fits your current data protection size. Then scale them as needed to meet your environment's requirements. Determine a regular interval that the performance of the instance will be evaluated. Increasing an instance in vCPU and RAM can increase the performance in many cases, but there are significant cost benefits from reducing the size of an instance that is underutilized. Scaling horizontally by increasing the number of instances is also a cost-effective possibility in the cloud that wouldn't be an on-prem consideration.

For the CommServe® server, adjust the instance size upwards when CPU and RAM loads become consistently high (over 80%); add more space to the second volume as needed to accommodate the size of the CommServe® server database as you add more clients and jobs to the CommCell. CPU and RAM load can be monitored in the Azure Portal using **Azure Monitor** or other performance tools or in the CommCell Command Center using the **Infrastructure Load Report.**

The Azure **Pricing Calculator** is available to predict cost for any given combination of infrastructure. It's recommended to estimate the overall solution's financial impact. Additional insight is available in the new **cost analysis preview**. This will provide granular detail on cost down to the resource type.

The Azure Instance Size Reference website provides insight into all the available instances that Microsoft currently offers. As the pricing and availability of these instances can and will vary over time, please refer to this link to validate your sizing selection. Market place images are available in Azure Marketplace for most of the following server recommendations.

Azure Server Specifications

Note: For more detailed information, please refer to the following link: Hardware Specifications for the CommServe Server

Commvault Recommended Azure VM Instance Types	
	Recommended Instance Type
Seed All-In-One CommServe®	Standard_E4as_v5 E Series has a high RAM to vCPU ratio of 8:1
Seed MediaAgent with Deduplication and Streaming Workloads	Standard_D4as_v5 D Series has a RAM to vCPU ration of 4:1
Access Node – Snapshot only	Standard_F2s_v2 - Standard_F8s_v2 F Series has a RAM to vCPU ration of 2:1
Access Nodes with Streaming (VSA)	Standard_D2as_v5 - Standard_D16as_v5
Scale-out CommServe® Base Configuration	Standard_D8as_v5 (snapshot) Standard_E8as_v5 (streamed and snapshot)
Scale-out MediaAgent Grids Base Configuration	Standard_D8as_v5 (snapshot) Standard_E8as_v5 (streamed and snapshot)

All-In-One CommServe®

The following is the recommendation for a day-one minimum configuration for getting started with Commvault Software.

Azure Quick Start Specifications – Seed All-In-One CommServe®		
Configuration	An all-in-one deployment with Commvault CommServe®, MediaAgent, Access Nodes, and Cloud Controller in one Azure Virtual Machine. It hosts deduplication cloud libraries for backups.	
Instance Type	Standard_E4as_v5 – AMD Standard_E4s_v5 - Intel	
Operating System	Red Hat Enterprise Linux 8 **Full List**	Microsoft Windows 2022 **Full List**
In Azure Marketplace	No	**Yes** (Microsoft Windows 2019)
Required Azure Storage	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 128GB – P20: Commvault Software 64GB – P20: Temp DB Drive 256GB – P20: Index Cache Disk 256GB – P20: Deduplication DB Disk	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 128GB – P20: Commvault Software 64GB – P20: Temp, Jobs, Cache Drive 256GB – P20: Index Cache Disk 256GB – P20: Deduplication DB Disk (32K blocks)
Primary Use Case	Snapshot and streaming protection of Azure services within the same region, Azure Stack Hub, Azure Stack HCI, trail use cases and small scale POCs.	
Protection Pattern	Azure Active Directory, Azure Blob Storage, Azure Cosmos DB, Azure Data Lake Storage Gen2, Azure DevOps, Azure Files, Azure NetApp Files, Azure Table Storage, NetApp Files, Office 365, SharePoint Server, SQL, SQL Managed Instance, SQL Server	
Configuration Notes	This is intended as a starter, test/dev or POC CommServe. It shouldn't exceed more than 100 Azure VMs. The amount of DDB space generally dictates the amount of data that can be stored. This is approximately 200:1. An example of this would be a 1TB DDB drive should be able to support 200TB of protected data. For an All-in-one starter CommServe with 256GB of DDB space, it should be limited to 50TB of protection. Protection within Region is recommended to reduce egress cost.	

MediaAgent

The following is an example of a minimum configuration for a single MediaAgent. The VM type can vary greatly according to workload necessity. In many cases a combination of scaling the Media Agents both horizontally (scale out) and vertically (monolithic) can be effective.

Azure Quick Start Specifications – Seed MediaAgent with Deduplication and Streaming Workloads		
Configuration	A single node MediaAgent, Access Node (Virtual Server Agent), CloudApps, and IntelliSnap® instance that hosts the index and deduplication database for disk or cloud libraries.	
Instance Type	Standard_D2as_v5 – Standard_D32as_v5	
Operating System **Full List**	Red Hat Enterprise Linux 8 Oracle Linux 8	Microsoft Windows 2022
In Azure Marketplace	**Yes** (Oracle Linux 8)	**Yes** (Microsoft Windows 2019)
Required Azure Storage	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 128GB – P20: Commvault Software 128GB – P20: Temp DB Drive 256GB – P20: Index Cache Disk 256GB – P20: Deduplication DB Disk	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 128GB – P20: Commvault Software 128GB – P20: Jobs Drive 256GB – P20: Index Cache Disk 256GB – P20: Deduplication DB Disk (32K)
Primary Use Case	Snapshot and streaming protection of Azure services within the same region, Azure Stack Hub, Azure Stack HCI. The MediaAgent acts as the data mover, writer, and manager for all data reading agents within Commvault software.	
Protection Pattern	This MediaAgent configuration is appropriate for streaming backups and IntelliSnap (snapshot management). Each Region should host its own MediaAgent to reduce inter-region traffic charges.	
Vertical Scaling	The recommended streams per node is based on its CPU and available RAM. In Azure this translates into a machine type. For an E Class machine, each VM size should process ~25 streams, started at a machine type 2. So, an E2as_v5 would support 25 streams and an E8as_v5 would support 100 streams. Refer to a **Grid Configuration** for scaling this configuration horizontally.	
Drive Scaling	Azure uses drive size categories for standardized size and performance. These values can be overridden but usually this is only done for performance requirements. An example would be a managed disk that is 128GB, which only has the drive performance of P10 for 500 IOPS but uses the performance characteristics of P20 to achieve 2300 IOPS. DDB and Index will both need to scale at approximately 2% and 2.5% respectively of the amount of deduplicated data per node. Large VM types will support 2 DDB drives to allow for scaling a single MediaAgent even more without needing to scale out horizontally.	

Access Nodes for IntelliSnap

The following is the recommended day-one minimum configuration for initial MediaAgent deployment (single node) responsible for performing a snapshot-only backup architecture. Snapshot only MediaAgents can be added to large pools of management groups for resiliency.

Azure Quick Start Specifications – Access Node – Snapshot Only		
Configuration	A single node all-in-one MediaAgent, Access Node (Virtual Server Agent), CloudApps, and IntelliSnap® instance that hosts non-deduplicated cloud libraries for snapshot backup indexes.	
Instance Type	Standard_F2s_v2	
Operating System **Full List**	Red Hat Enterprise Linux 8 Oracle Linux 8	Microsoft Windows 2022
In Azure Marketplace	**Yes** (Oracle Linux 8)	**Yes** (Microsoft Windows 2019)
Required Azure Storage	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 128GB – P20: Commvault Software 64GB – P20: Index Cache Disk	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 128GB – P20: Commvault Software 64GB – P20: Index Cache Disk
Primary Use Case	Performing Azure snapshot creation, sharing, and copying between regions. Writing backup activity for snapshot-only jobs to a hosted non-deduplicated cloud library. (Optional) Sending backup activity for snapshot-only jobs to a remote cloud library.	
Protection Pattern	This Access Node will be used to coordinate the snapshot and retention of virtual machines in Azure. As it is not needed to move data, network bandwidth and capacity are not factors.	
Vertical Scaling	Performance shouldn't be a factor with the VM size, but as the number of VMs with snapshots increase it may become necessary to scale these snapshot Access Nodes horizontally to add parallel job management and additional API handlers.	
Drive Scaling	Azure uses drive size categories for standardized size and performance. These values can be overridden but usually this is only done for performance requirements. An example would be a managed disk that is 128GB, which only has the drive performance of P10 for 500 IOPS but uses the performance characteristics of P20 to achieve 2300 IOPS. Index will need to scale at approximately 2.5% of the amount of data protected per node.	

Access Nodes for Streaming

The following is an Access Node suitable to act as an API gateway, network proxy, snapshot coordinator and CloudApps agent. It acts as the data reader for Cloud Applications and Virtual Machines.

Azure Quick Start Specifications – Access Node – Streaming		
Configuration	A single node all-in-one Access Node (Virtual Server Agent), CloudApps, and IntelliSnap® instance.	
Instance Type	Standard_E2as_v5	
Operating System **Full List**	Red Hat Enterprise Linux 8 Oracle Linux 8	Microsoft Windows 2022
In Azure Marketplace	**Yes** (Oracle Linux 8)	**Yes** (Microsoft Windows 2019)
Required Azure Storage	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 128GB – P20: Commvault Software	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 128GB – P20: Commvault Software
Primary Use Case	Performing Azure snapshot creation, sharing, and copying between regions. Streaming backups to deduplication MediaAgent grids. Acting as an API gateway within a region. (Optional) Sending backup activity for snapshot-only jobs to a remote cloud library.	
Protection Pattern	This Access Node will be used to protect Azure Virtual Machines and applications. Network acceleration is always recommended. Commvault will refer to an Access Node as a proxy, VSA or FREL throughout documentation. They can be used interchangeably.	
Vertical Scaling	Performance can be a factor depending on the workload. It is recommended to try **Auto-Scaling** agents as an option instead of vertically scaling this VM. If needed it can be increased as high as E16, but customers might receive diminishing returns for monolithic scale.	
Drive Scaling	Scaling the drive will not be needed unless the VM is used to cache restores. In that scenario, a drive can be added for the jobs cache and be scaled up as needed. A reasonable starting point would be 128GB E10 Standard SSD.	

Scale-out CommServe® Base Configuration

The following is a base configuration of a CommServe that will act as the central management component of a CommCell as it scales up. It is important to understand that although an All-in-one configuration is available and it is recommended for deployment, many of its services will be functionally disabled so that the CommServe takes on the singular role of management. The VSA, Access Node, IntelliSnap and MediaAgent functionality will not be needed and will be offloaded to other components within the CommCell so that they can scale independently of one another.

Azure Quick Start Specifications – Scale-out CommServe – Base Configuration		
Configuration	A scale-out deployment with Commvault CommServe®, MediaAgent, Access Nodes, and Cloud Controller in one Azure Virtual Machine. This base can be used for all Commvault Reference Architecture.	
Instance Type	Standard_D4as_v5 - Standard_D32as_v5 – slightly lower cost, but less RAM Standard_E4as_v5 - Standard_E32as_v5 – Increased RAM may provide scaling options if RAM becomes limited as opposed to increasing the machine size which usually doubles the cost.	
Operating System	Red Hat Enterprise Linux 8 **Full List**	Microsoft Windows 2022 **Full List**
In Azure Marketplace	No	**Yes** (Microsoft Windows 2019)
Required Azure Storage	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 256GB – P20: Commvault Database 128GB – P20: Temp DB 128GB – P20: Logs and Cache	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 256GB – P20: Commvault Database 128GB – P20: Temp DB 128GB – P20: Logs and Cache
Primary Use Case	Snapshot and streaming protection of Azure services within any region, Azure Stack Hub, Azure Stack HCI, on-prem and any remote location with network connectivity.	
Protection Pattern	Azure Active Directory, Azure Blob Storage, Azure Cosmos DB, Azure Data Lake Storage Gen2, Azure DevOps, Azure Files, Azure NetApp Files, Azure Table Storage, NetApp Files, Office 365, SharePoint Server, SQL, SQL Managed Instance, SQL Server	
Vertical Scaling	Performance can be a factor depending on the workload. The base configuration should be able to scale to hundreds of virtual machines. It should be evaluated on a regular basis and if performance indicators are consistently 90% or higher, the next machine size should be used.	
Drive Scaling	The drive sizes may never need to be increased, but as the machine size increases and workloads increase, the performance of the drives may need to increase. Isolating the CommServe database and the temp DB drives will become critical. Once you have reached a Machine Size of E8 and higher and disk speed of P40+ is recommended.	
Scale Factors	Workloads under 1000 VMs can generally be handled with 4-vCPU class servers. At 1000, an 8-vCPU server will be more appropriate. Consider stepping each order of magnitude with a Machine Size. A midrange CommServe of an E8 size should be able to accommodate thousands of VMs. The performance of the CommServe should indicate your size increase. Many streaming jobs have a much higher impact than many snapshot jobs. The resource management is lower so the CommServe can accommodate more.	

	Your drive sizes are also just guidelines. With a dedicated CommServe just managing jobs, your database will never grow to a significant size. A CommServe managing a couple hundred VMs for a year will likely never have a database more than a few GBs. Dedicating drive space to the database so that it can have appropriate block size and better/tunable performance is more important than capacity. Dedicating a 64K formatted 128GB database disk and 64K formatted 64GB temp disk is a good start. Adjusting the IOPS to P20 level (~2400 IOPS) will keep them from bottlenecking. As the performance needs increase, you can just increase the P-level and size of the drive as needed. The OS will need to recognize the change, but this is usually not an issue.

Scale-out MediaAgent Base Configuration

Azure Quick Start Specifications – Scale-out MediaAgent – Base Configuration		
Configuration	A base configuration of a MediaAgent, Access Node (Virtual Server Agent), CloudApps, and IntelliSnap® instance that hosts the index and deduplication database for disk or cloud libraries. This would be suitable for a grid configuration or a cross regional deduplication engine.	
Instance Type	Standard_D4as_v5 – Standard_D32as_v5 Standard_D4s_v5 – Standard_D32s_v5	
Operating System **Full List**	Red Hat Enterprise Linux 8 Oracle Linux 8	Microsoft Windows 2022
In Azure Marketplace	**Yes** (Oracle Linux 8)	**Yes** (Microsoft Windows 2019)
Required Azure Storage	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 128GB – P20: Commvault Software 128GB – P20: Temp DB Drive 256GB – P20: Index Cache Disk 256GB – P20: Deduplication DB Disk	Premium SSD – LRS for all volumes Addition Commvault Disks Requires: 128GB – P20: Commvault Software 128GB – P20: Jobs Drive 256GB – P20: Index Cache Disk 256GB – P20: Deduplication DB Disk (32K)
Primary Use Case	A typical grid configuration needs to start with at least 2 and up to 4 MediaAgents per region for redundancy and load balancing. Resources specific to Availability Zone (AZ) would reduce transmission cost even more. These cost factors need to be evaluated to determine need.	
Protection Pattern	This MediaAgent configuration is appropriate for streaming backups and IntelliSnap (snapshot management). Each Region should host its own MediaAgent to reduce inter-region traffic charges. If there is significant inter-Availability Zone traffic, then resources can be created within those AZs to reduce cost.	
Vertical Scaling	The recommended streams per node is based on its CPU and available RAM. In Azure this translates into a machine type. For an E Class machine, each VM size should process ~25 streams, started at a machine type 2. So, an E2as_v5 would support 25 streams and an E8as_v5 would support 100 streams.	
Drive Scaling	Azure uses drive size categories for standardized size and performance. These values can be overridden but usually this is only done for performance requirements. An example would be a managed disk that is 128GB, which only has the drive performance of P10 for 500 IOPS but uses the performance characteristics of P20 to achieve 2300 IOPS. DDB and Index will both need to scale at approximately 2% and 2.5% respectively of the amount of deduplicated data per node. Large VM types will support 2 DDB drives to allow for scaling a single MediaAgent even more without needing to scale out horizontally.	

Disk Drive Considerations

Drive Mount Details and Sizing Guideline Options – Windows – CommServe All-in-one					
Drive Letter	Size	Block Size	Performance Type	Performance Tier	Purpose
C:\	64GB Default	4K	Premium SSD	P15	OS
E:\	256GB	4K	Premium SSD	P20	Binaries, Software Cache, logs, MS SQL Binaries
F:\	128GB	64K	Premium SSD	P20	SQL DB
G:\	64GB	64K	Premium SSD	P20	SQL Transaction Logs
H:\ *	256GB	32K	Premium SSD	P20	Dedupe Database
I:\ *	256GB	32K	Premium SSD	P20	Index Cache
J:\ *	128GB	32K	Premium SSD	P20	Jobs, 3DFS Cache

Drive Mount Details and Sizing Guidelines – Linux – Media Agent / Access Node					
Mount	Size	Block Size	Performance Type	Performance Tier	Purpose
/	64GB Default	4K	Premium SSD	P15	OS
/mnt/cvsoftware	128GB	4K	Premium SSD	P20	Commvault Binaries, SQL Binaries, SQL database
/mnt/commvault_jobs	128GB	4K	Premium SSD	P20	Jobs Results, 3DFS Cache, Temp space *
/mnt/commvault_indexcache	256GB	4K	Premium SSD	P20	Index *
/mnt/commvault_ddb	256GB	4K	Premium SSD	P20	DDB *

*** Optional for All-in-one configurations**

Architecture Workshop – Protecting Azure in Azure

Taking the principals and applying them to a functional design can prove useful for extrapolating larger scenarios. In this example I will walk through the design and implementation of a midsize CommCell.

Architecture Outline		
Size	2000 VMs	200TB (average 100GBs, use Azure Resource Graph Explorer for details)
	SQL MI	40GB
	Azure Data Lake	50TB
	Blob Storage	2TB
	Azure Files	100TB
	AKS	250GB
Parameters	This solution needs to exist completely in Azure. The core of the data will reside in Central US, but there is redundancy to the East US region for DR and testing. Roughly 70/30. Customer requires minimal cost with a selective copy of critical systems to the DR Region	

CommCell Recommendations

The size of the CommCell does not need to be large enough to manage the growth for the next 3 years, planning on 3 months of growth is adequate. A regular cadence of inspection and adjustment needs to be followed. The difference between deploying an E4 versus an E8 is literally twice the cost. If you are sizing for what you will use in a year, you are losing the cost of these resources for a year. Right sizing the solution is key.

This solution will be protecting at least 2000 VMs with ~200TB of capacity and ~150TB of application data. The CommServe can reside in a single Region and Availability Zone. If Region redundancy is needed, remember that you can now replicate between CommServes creating a fault tolerant pair for a multi-regional control plane.

First, this design wouldn't be appropriate for an All-in-one CommServe. The image can be used for the CommServe, and the extra components can be left in place, but the day-to-day protection and indexing will be handled through scale-out MediaAgents. This is usually still preferable to installing a CommServe from a software kit to a generic VM.

For this solution, the following configuration should be adequate for the CommServe.

CommServe Size		
VM Size	Standard_E8as_v5	This is an 8-vCPU instance with 64GB of RAM. Arguably you could also use a D8as_v5. The difference being the amount of RAM. If the resource seems underutilized after the inspection period, it can be reduced.
Disk Size and Configuration	C:\ - 128GB – 4K	OS Disk
	D:\ - 512GB – 4K	Commvault software and cache
	E:\ - 256GB – 32K	CommServe Database. P20 Performance tier is recommended for midrange capacity.

	F:\ - 64GB – 32K	CommServe Database T-Logs. P20 Performance tier is recommended. Some machine types also have ephemeral disks provided at high speeds. This is supported for a CommServe installation.
Notes	Additional scaling information is available from the **Commvault scaling** guide.	

MediaAgent Recommendations

The size and location of your MediaAgents is dictated completely by the load and cost of transit. In this scenario, we are cost conscious so having an MA in each Availability Zone within our central Region prevents cross AZ cost and allows for lower cost machines. Some cost will exist between AZs soon, refer to **Microsoft's Bandwidth Cost Guide** for more information. In the Central US Region, we will employ a horizontal scale and in the East US we'll use a vertical scale.

MediaAgents (MA) – Multi-AZ Scale-out for Central US		
VM Size	Standard_E8as_v5	This is an 8-vCPU instance with 32GB of RAM. We are going to use a dedicated MediaAgent in each Availability Zone within the Central US Region. It provides resiliency across all AZs. If an AZ becomes unavailable, any of the MediaAgents can act as the Access Node for another.
Disk Size and Configuration	/ - 64GB – 4K	OS Disk
	/mnt/cvsoftware 512GB – 4K	Commvault Software and SQL Binaries
	/mnt/commvault_jobs 128GB – 4K	Jobs
	/mnt/commvault_indexcache 2TB – 4K – P40	Index Cache
	/mnt/commvault_ddb1 1TB – 4K – P30	Deduplication Database 1
	/mnt/commvault_ddb2 1TB – 4K – P30	Deduplication Database 2 – At this size range you can add an additional DDB to the MediaAgents and double their management capabilities.
Notes	The initial design looks to use 3 of these VMs for data protection in the Central Region. We'll do a cost analysis at the end to see if this is justified or if there is room to make a change without sacrificing performance. The metric used here is that protecting ~350TB of VMs/Data. ~250TB of it will be in Central US and 100TB will be in East US. 2% of 250TB is 5TB. Divided between 3 MAs gives about 1.6TB per MA. With 2 – 1TB DDB drives, you will have appropriate capacity and performance. Azure Managed Disks have set capacities so you will need to choose the best, closest match. 1TB drives	

	are classified as **P30**. This gives them 5,000 IOPS, burstable to 30K, and 200 MB/second, burstable to 1,000 MB/second. The single Index drive should be configured for P40 as this will give it enough capacity at 2TB (250TB * 2.5% /3) and performance. Obviously, these VMs can begin to incur a significant cost. As the CommServe and MediaAgents are infrastructure components, using reserved pricing is recommended.

MediaAgents (MA) – Single Media Agent for Auto-scale in the EAST US		
VM Size	Standard_E8as_v5	This is an 8-vCPU instance with 32GB of RAM. We are going to use a single MA for this Region and rely on auto-scaling features to reduce cost but still provide the same level of performance.
Disk Size and Configuration	/ - 64GB – 4K	OS Disk
	/mnt/cvsoftware 512GB – 4K	Commvault Software and SQL Binaries
	/mnt/commvault_jobs 128GB – 4K	Jobs
	/mnt/commvault_indexcache 2TB – 4K – P40	Index Cache
	/mnt/commvault_ddb1 1TB – 4K – P30	Deduplication Database 1
	/mnt/commvault_ddb2 1TB – 4K – P30	Deduplication Database 2 – At this size range you can add an additional DDB to the MediaAgents and double their management capabilities.
Notes	The initial design looks to use a single VM for data protection in the East US Region. The metric used here is that protecting ~350TB of VMs/Data. ~250TB of it will be in Central US and 100TB will be in East US. 2% of 100TB is 2TB. With 2 – 1TB DDB drives, you will have appropriate capacity and performance. Azure Managed Disks have set capacities so you will need to choose the best, closest match. 1TB drives are classified as **P30**. This gives them 5,000 IOPS, burstable to 30K, and 200 MB/second, burstable to 1,000 MB/second. The single Index drive should be configured for P40 as this will give it enough capacity at 2TB (100TB * 2.5%) and performance. This index drive is a little under sized, but it can be resized at a later time if the capacity gets too high. All capacities and performance markers need to be alert items for management so that just-in-time adjustments are made if necessary. Obviously, these VMs can begin to incur a significant cost. As the CommServe and MediaAgents are infrastructure components, using reserved pricing is recommended.	

Agent Coverage

Virtual Machines

The size of the MediaAgents will be adequate for the capacity of the workload within this environment, but there may be bottlenecks that are created naturally by the diverse workload. The Commvault's Azure Virtual Server Agent (VSA) supports **auto-scale**. This means that should the protection of Azure VMs become resource constrained, the CommServe can deploy access nodes that are ephemeral. They will be created, used, powered off and even decommissioned when not needed. This is supported for VM backups and VM IntelliSnap for creating deduplicated backups of an Azure Snapshot.

SQL MI

There bulk of these backups can be performed directly from the MediaAgents. Additional Agents, Proxies or Access Nodes should not be required. These are full backups so monitoring these streams to ensure they do not create a job bottleneck is important. As this resource grows, it may become necessary to allocate resources to it.

Azure Data Lake

ADL can grow to a very large size, by design. Commvault supports an incremental forever style protection for ADL that should reduce the amount of transit time daily. A 50TB ADL with only a 1% change rate will take a little time on the first full backup, but then will only be moving ~500GB a day. As this resource grows it may become necessary to allocate dedicated resources. Additionally, there may be requirements for performance that necessitate additional Access Nodes just to read data. These can be light weight data movers and will likely need to scale out horizontally as ADL as no issue scaling out in parallel streams.

Blob Storage

2TB of Blob storage is not a lot, but it is probably tied to an application or website, so it could be very critical. The protection of this storage can be handled directly by the Media Agents.

Azure Files

This data type is protected as a **proxy agent method**. You can use the existing MediaAgents as the proxy for this data. Be away of file system limitations and "crawling" through massive file collections. It is often that a customer would fill the Azure File Service up just like an on-prem file server. That is why it was invented. The same complication can come with that design. If needed and the file system is taking too long to read, additional light weight Access Nodes can be created as data readers to establish parallel streams. This type of agent may need fine tuning based on the type of data stored. It is much more common for customers to replace a system like this with OneDrive.

AKS

Commvault supports a wide array of **CNCF-certified Kubernetes** distributions, including Azure's managed version Azure Kubernetes Service (**AKS**). 250 GB of AKS is not a lot of space, so there may not be any need for dedicated resources on Day-1, but these types of environments can often grow exponentially. Adding a lightweight Access Node to increase data reading throughput can be significant.

General Environment Scaling Notes

Several bullet points in the design ask for monitoring and adjustments. The most common would be as a resource grows, you should be able to grow accommodating infrastructure with it. The static resources like the CommServe and the MediaAgents can be reserved, and this drops their cost significantly. Customers can also negotiate custom pricing for resources they know will be in their environment for multiple years. The discounts for the compute range from 40-60% based on 1-3 year commitments and ~20% for storage based on 1 year commitments.

The lightweight Access Nodes referenced could be something in the D2 or F2S family of VMs. They do not index or deduplicate, so their storage requirements are negligible. Point in fact, Commvault uses the F2S instances as our auto-scale access nodes. They are ideal data readers. The only concern you might have for a Size 2 VM is that there are limitations from Azure on the amount of data that can move through their network interface. There are applications that you might have to use a larger machine size just to accommodate the amount of data you are trying to move.

Another Commvault advantage is that these Access Nodes can be lightweight because they are reading data and deduplicating it and moving that data to the MediaAgent to storage in its index, deduplication data base and into Azure Blob storage. There is another architectural advantage available as well. Each Access Node can use **Storage Accelerator** to write the backups directly to Blob Storage. Deduplication and indexing still occurs on the MediaAgents (MA), but the data is read and instead of passing through the MA it is written directly to Azure Blob. This effectively offloads the data streams from the MA's and allows the entire landscape to scale up more effectively.

Lastly, lets review the principles to make sure we have checked all the boxes.

✓	Right-sizing over Forecasting	This is an appropriate Day-1 configuration. Do not build for 3 years of growth.
✓	Match protection to business impact	There may need to be additional resources deployed for business need. There is a smaller site in East US that is used for Dash Copies. There may be a need for rapid recovery there. It will increase the cost of infrastructure and storage, but that could be appropriate for a select group of AAA VMs needed for Business Continuity.
✓	Central Management of Azure Snapshots	One-off solutions for snaps and backups will pollute the landscape. Commvault can manage the snapshot used for data protection for the entire environment. With our tagging and management in place the customer has better billing and visibility.
✓	Optimize at Rest	Encryption, compression, and deduplication should be used in the lowest cost tier appropriate for data recall. Snapshot only is fast (and common), but a costly solution for long term storage. Commvault can use snapshot for a couple days creating a rapid recovery layer and still send long term retention off to low cost object storage. No compromise, lowest cost.
✓	Optimize on Wire	Encryption, compression and deduplication at the data reader ensures the least amount of data sent over the wire. Especially important for inter-AZ and inter-Region ingress and egress costs.
✓	Separation of Duty	Managed Identities should be used when possible. It eliminates a significant security issue in any environment when you must manage credentials. This assigns the VM that is doing the work with the capabilities for that work. If a bad actor on another VM or another time zone tries to access those resources, it is simply denied no matter the credentials. This also means that someone with Commvault Admin credentials would not be able to access those resources directly. Keeping cloud management and backup management separate.
✓	Encrypt Everything	Everything. From the moment data is read, over the wire, down to the storage and into the cloud library.
✓	Automate Over Runbooks	The use of Power Management to reduce the cost of infrastructure online and the ability to auto-scale are two examples of Commvault Automation. The customer is encouraged to automate the roll-out and configuration of their environment as well.

		Infrastructure as Code (**IaC**) has had a lot of traction recently. Much of it has come from the Kubernetes influence as it is a platform reliant on automation. The days of "If you have to do it twice, build a script" are gone. "**Automate Everything**" is becoming the new normal.

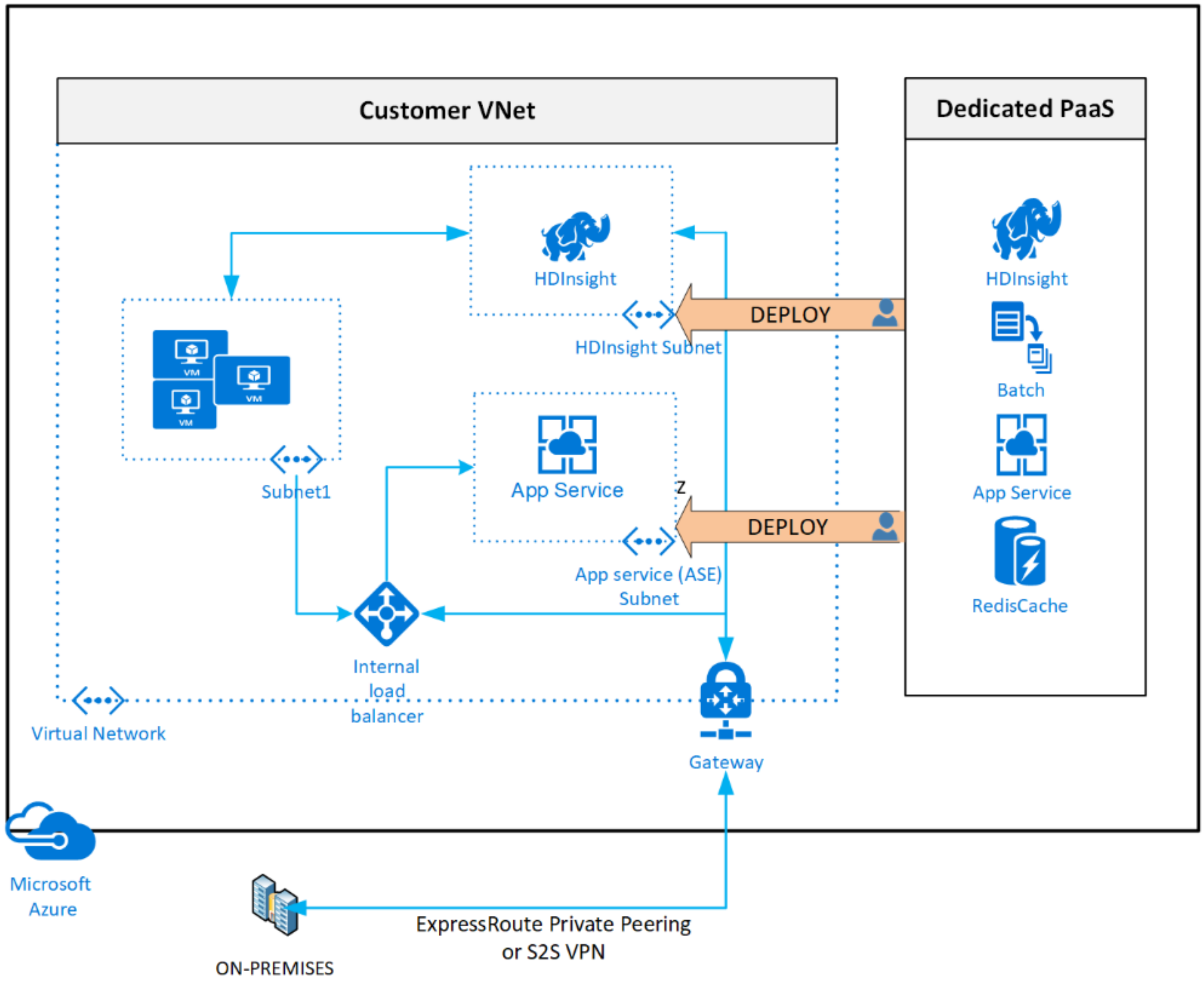
Architecture considerations

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).

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 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 and 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.

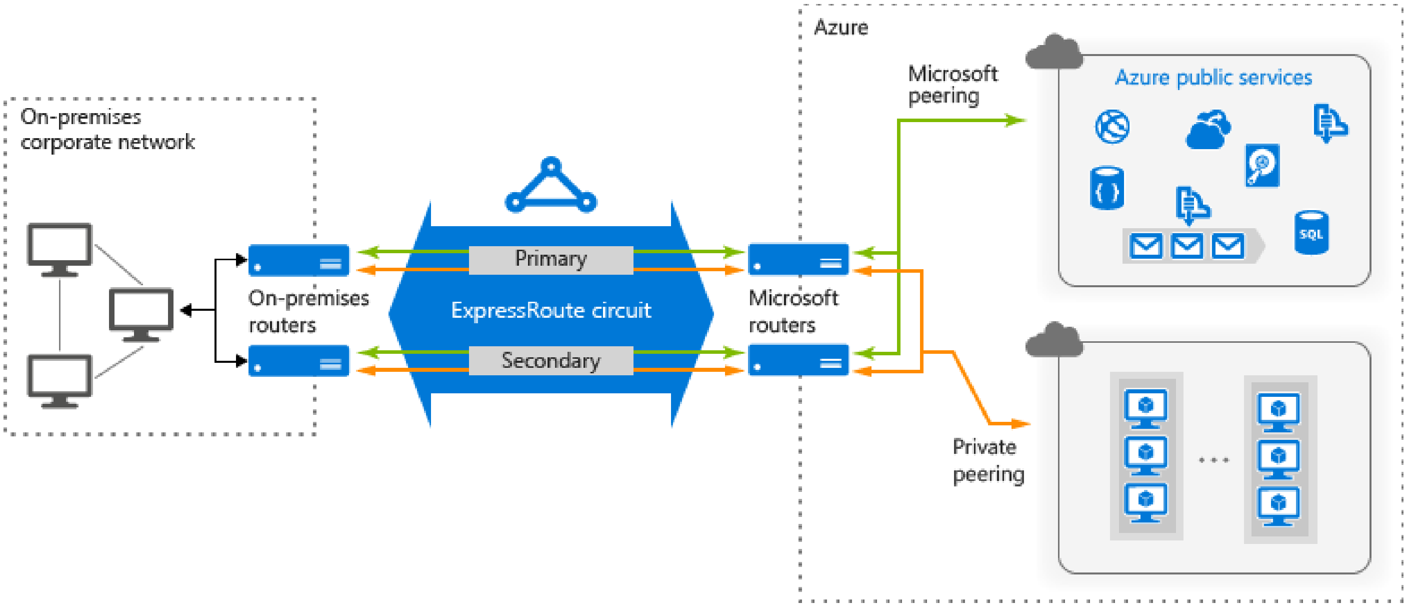
This is particularly relevant to solutions where you wish to Backup/Archive directly to the cloud or create deduplicated secondary data copies (DASH Copy) of existing backup/archive data to object storage within a cloud provider.

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. Available at various bandwidth speeds for both metered/unmetered charges for outbound data transfers. Details can be found at **Azure ExpressRoute.**

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 (Unlimited Data plan), with bandwidth from 50Mbps to 100Gbps.



Firewall ports

When communicating with cloud services and resources, security is a high priority, and the use of firewalls are recommended. Commvault has registered specific default ports for its services that can be changed if desired. Several Commvault services listen for incoming network traffic on these ports.

Static ports

The CVD service port (8400) is used for communication of the CommServe server, MediaAgents, and clients.
The EvMgrS service port (8401) is used by the CommServe server for receiving events from MediaAgents and/or clients.
The FWD service port (8403) is used for tunneling connections across firewalls.

A single custom port can also be designated for firewall configurations that require port tunneling.

Commvault service	Port number	Protocol
Commvault Communications Service (GxCVD, found in all client computers)	8400	TCP
Commvault Server Event Manager (GxEvMgrS, available in CommServe server)	8401	TCP
Commvault Firewall (GxFWD, tunnel port for HTTP/HTTPS)	8403	TCP

Dynamic ports

Dynamic ports are opened and closed by the running Commvault software to permit certain types of transient traffic. The CVD service dynamically uses free ports between 49152 - 65535 to communicate during data protection and data recovery jobs. The system dynamically assigns a number of free ports to be used by each job to allow parallel data movement. After the job is finished, if no other job is pending, the dynamic ports are released.

Additional information can be found by referring to **TCP Ports Used for Services.**

Azure Virtual Network

Azure Virtual Network (VNet) 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

Cost reduction is a top priority for cloud users and there are many ways to reduce these costs. Here are a few key points to keep in mind when designing the networking in Azure:

- Azure networking costs are limited to egress or outbound bandwidth. All incoming data to Azure data centers is free while the cost of outgoing data is tiered based on usage.
- Outbound data transfer charges apply when the data transfer is to another region.

- Region to region networking is performed through the use of **Global VNet Peering**.
- **Virtual Network peering** allows you to route traffic between virtual networks using private IP addresses and does incur both inbound and outbound bandwidth charges at the rate of $0.01 per GB.
- VMs deployed in availability zones incur charges for inbound and outbound data transfers between availability zones at a rate of $0.01 per GB.
- Data transfers within the same availability zone incur no charges.
- Data transfers from a VNet resource in an availability zone and a public address in the same Azure region incur no charges.

When configuring VNet peering, the address spaces must be non-overlapping. This means you cannot have both VNet1 and VNet2 with an IP subnet of 10.1.0.1. VNet2 would have to have an IP address subnet of 10.2.0.1 as an example.

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

Data security

In-flight

By default, all communication with Cloud Libraries utilizes HTTPS which ensures that all traffic is encrypted while in-flight between the MediaAgent and the Cloud Library endpoint, 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 and 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. To meet this requirement, Azure provides the capability to encrypt IaaS VMs with Azure Disk Encryption. Azure Disk Encryption supports both Windows and Linux VMs and it is integrated with Azure Key Vault to manage and control the disk-encryption keys and secrets. Along with IaaS VM encryption, 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

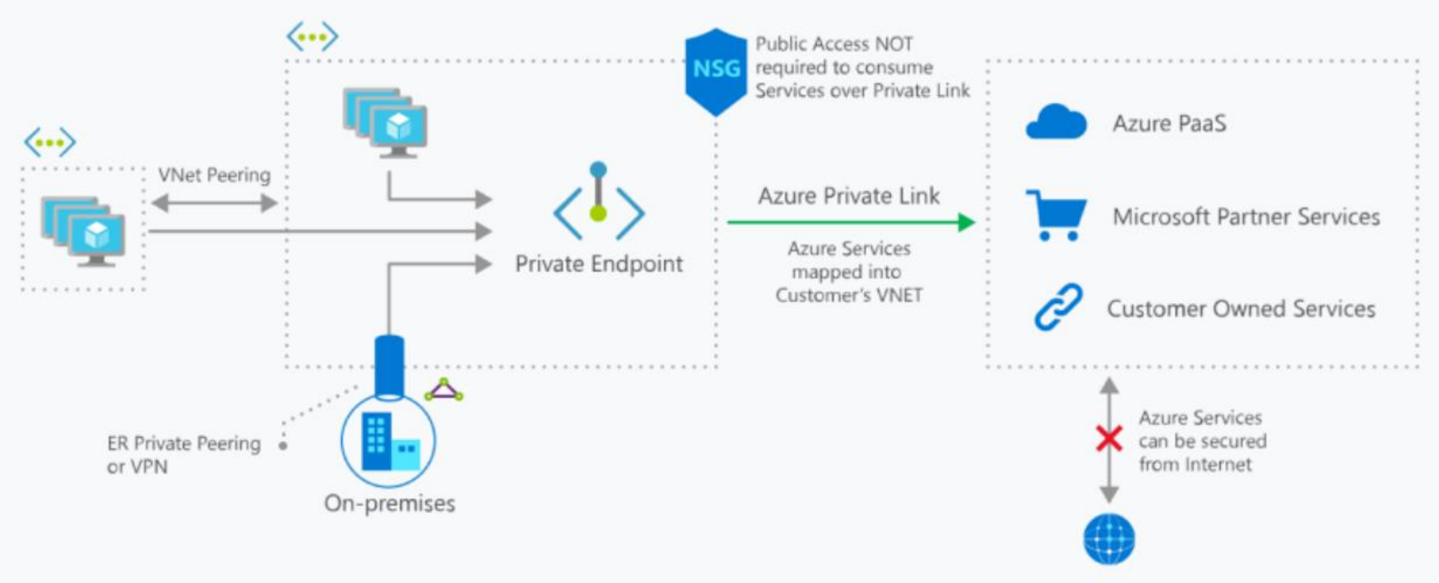
Please take note of any HTTP(S) proxies between MediaAgents and Endpoints, whether via public Internet or private space, as this may have a performance impact upon any backup/restore operations to/from an object storage endpoint. Where possible, Commvault software should be configured to have direct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

Account separation

Consider utilizing separate Azure accounts for production and non-production data protection activities. Commvault supports cross-account restores, allowing segregation of data access but also authenticated and authorized data mobility where required (e.g. Dev/test re-seeding restores).

Azure Private Link

Private Link is a service from Microsoft that provides private connectivity from a virtual network to Azure platform as a service (PaaS), customer-owned, or Microsoft partner services. It allows you to create a private endpoint in your virtual network and simplifies the network architecture. This endpoint provides you an internal IP address for the resource and all traffic from your virtual network to the resource goes over the Microsoft backbone instead out over the public internet. All traffic to the service can be routed through the private endpoint, so no gateways, NAT devices, ExpressRoute or VPN connections, or public IP addresses are needed. There are several benefits to this service but most importantly data security since only the mapped resource would be accessible.


NSG
Public Access NOT required to consume Services over Private Link
VNet Peering
Private Endpoint
Azure Private Link
Azure Services mapped into Customer's VNET
Azure PaaS
Microsoft Partner Services
Customer Owned Services
ER Private Peering or VPN
On-premises
Azure Services can be secured from Internet

How does Private Link differ from Service Endpoints? Service Endpoints provide a way to lock down access to resources to a virtual network, but you are still accessing the public endpoint. You are only locking down access to a service and not to a specific resource. With Private Link you are only allowing access to the defined resource, which greatly enhances data security and control of your resources. Private Link creates an endpoint with a private IP address and data flows solely inside your virtual network, which does not require Network Security Group (NSG) rules allowing outbound traffic beyond your virtual network.

There is no cost for the Private Link service, however there is a cost for the Private Endpoint ($0.01 per hour) as well as all data inbound ($0.01 per GB) and outbound ($0.01 per GB).

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. There are two primary methods for seeding data to an external cloud provider:

“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.

If you require an understanding of the possible throughput between two Azure VMs to assist with transfer times, you can use the Azure NTTTCP tool.

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
10 Mbit	14 min	2.2 hrs	22.2 hrs	9.2 days	92.6 days	-	-	-
100 Mbit	1 min 20 s	13 m 20 s	2.2 hrs	22.2 hrs	9.2 days	92.6 days	-	-
1 Gbit	8 s	1 m 20 s	13 m 20 s	2.2 hrs	22.2 hrs	9.2 days	92.6 days	-
10 Gbit	0.8 s	8 s	1 m 20 s	13 m 20 s	2.2 hrs	22.2 hrs	9.2 days	92.6 days

Drive seeding

If the dataset 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. The Azure Data Box family of products is supported and can be used to migrate data to Azure.

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:

Import Services

Azure Databox

Migrating Data to Microsoft Azure Using Azure Data Box

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. **Azure egress fees** are charged per GB per month, so ensure you are aware of the costs and design your data movement in the most efficient way possible.

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 April 2020, Azure charges $0.05 per 10,000 PUT requests and $0.004 per 10,000 GET requests for its LRS Hot storage tier for example.

A baseline of 2,000 GB with a saving of 40% at 32 MB sub-chunk size would result in an approximately 37,500 PUT requests. At a charge of $0.05 per 10,000 requests, the cost would be 20 cents.

Note: All cost figures are referenced in USD and based on pricing listed at the *Azure website* at the time of this document's publication.

Data recall

Low-cost cloud storage solutions may have a cost associated with accessing data or deleting data earlier than the minimum days requirements of the tiers. Hot tier has no requirements, Cool tier is 30 days, Archive tier is 180 days. 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 by leveraging the **Azure cost calculator.**

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 other 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.

Device Streams > 50

Device Streams are logical channels that connect client data to the media, where data that is secured by backup operations are stored. For a Plan, the number of device streams that is configured must be equal to the number of drives or writers of all libraries that are defined in the storage policy copy. Increasing the number of Device Streams above 50 increases parallel streams during protection activities.

Deduplication block size = 512 KB

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 512 KB.

Data Readers

Data Readers determine the number of parallel read operations while the data is backed up. Configuring multiple data readers per subclient on disk arrays can improve the backup performance of clients. Increase the Number of Data Readers to increase the backup read performance.

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, long-term retention designs in the Archive tier (tape replacements use case), 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, using multiple mount paths may help to increase performance but typically it's not required.

Block blob versus Page blob object storage

With Azure, be aware that both block object blobs as well as page blobs, are potential options when creating a storage account. Public IaaS environments allow premium (page blob) based storage to be provisioned and leveraged as disk libraries, by attaching directly to the MediaAgent. The overall cost of those volumes can quickly exceed that of standard (block blob) object storage. 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.

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

If you are unsure as to which offering to use, you should consume standard **block object storage blobs**

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. Commvault version 11 allows for the addition of two extra nodes (up to 4) to an existing deduplication store dynamically, allowing for rapid scale-up configurations.
Configuring Additional Partitions for a Deduplication Database

Micro pruning

Micro pruning support for object storage is effective for any new data written into the active deduplication database. 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 cloud storage library until the deduplication database has been sealed. Once the active library is sealed, there will no longer be a need for continued periodic sealing against that library. Micro pruning is supported in the Azure Hot and Cool storage tiers. Micro pruning is **not** supported in the Azure Archive storage tier.

Choosing the correct Azure storage tier

Azure storage redundancy

Azure provides redundancy of data stored in a storage account to ensure durability and high availability. The redundancy options available are:

Locally Redundant Storage (LRS)

Zone-Redundant Storage (ZRS)

Geo-Redundant Storage (GRS)

Geo-Zone-Redundant Storage (GZRS)

Read-Access Geo-Redundant Storage (RA-GRS).

These **redundancy levels** are transparent to Commvault and can be used if required.

Commvault recommends the use of LRS with Cool storage as a good utility tier. It has a good price point while offering performance that is equal to more expensive tiers. Architecture requirements will vary.

Selecting the right storage class for backup and archive data

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

For example, storing infrequently accessed backups within an intermediate storage tier (Cool blob) can significantly lower the cost of your cloud bill, while storing data in an archive storage tier (Archive blob) may greatly impact accessibility for end-users to the archived data but have the added benefit of reducing storage costs further. To delve into further detail, these storage classes can be broken into three primary categories:

Standard storage (Hot) – 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 with a very short-term retention configuration or where restores are frequent.

Intermediate storage (Cool) – this is a storage tier that addresses a gap between the standard Hot storage tier offering and 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. This tier functions well as a secondary copy for 30+ days.

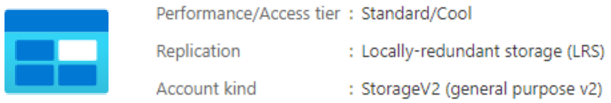
Archive storage (Cold) – sometimes referred to as “cold storage”, this tier is intended for data that will probably not be accessed again, but must be retained in the event of compliance, legal action, or another business reason, Azure Archive Blob storage tier is an example of archive storage which Commvault software supports. The cost of this storage class is the lowest compared to all three offerings – $0.001/GB/month, but as with the intermediate class, the 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 early deletion period of 180 days. The charge is prorated based on the amount of time the data 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 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.

Additional information: **Supported Cloud Storage Products >**

Choosing the correct Azure storage tier for backup data is crucial when considering performance, cost, and accessibility. If, after reading through the information below, you are still unsure where to store backup data, we recommend:

Azure Cool Locally redundant storage (LRS)
General Purpose v2 storage



Performance/Access tier : Standard/Cool
Replication : Locally-redundant storage (LRS)
Account kind : StorageV2 (general purpose v2)

It has the best cost/performance of all tiers and should be used for most primary copies. Keep in mind this is just a recommendation and there are always exceptions.

Each tier of storage has specific characteristics which need to be fully understood. However, before choosing the storage tier, you first need to determine what the business requirements demand for accessibility and retention.

For example:

Data storage requirements	Hot	Cool	Archive
Data retention (Days, Months, Years)	Days	Months	Years
Accessibility (Frequent, Occasional, Rare)	Frequent	Occasional	Rare
Time to recover (Minutes, Hours, Days)	Minutes & hours	Minutes & hours	Hours & days

Additional factors:

Location of data to be protected (On-premises or in the cloud)

Location of MediaAgent (On-premises or in the cloud)

Location of backup storage (On-premises or in the cloud)

Internet bandwidth (Public internet or Azure ExpressRoute)

The answers to these questions may be the same for all data being protected, or they may be different depending on the type of data (databases, healthcare, financial, filesystems, etc). We can then determine the storage tier best suited for the data.

Azure Hot storage is designed for fast and easy access to data, but at the highest cost ($0.0184 USD) per GB. Data stored in Hot tier is recommended for short term retention (primary backup copy).

Azure Cool storage is technically the same as Hot storage and is designed for fast and easy access to data. However, it's cheaper at ~45% less, ($0.01 USD) per GB. The reason for the price break is because Azure is expecting the data to remain in Cool storage for a minimum of 30 days. If the data is removed prior to 30 days, the storage cost is prorated. Data stored in Cool tier is recommended for short-term and long-term retention that requires access for restores on frequent or occasional use (primary backup copy and long term retention).

Azure Archive storage is the lowest price storage tier, ($0.002 USD) per GB. The purpose for this tier is for a very specific use case. Data placed in Archive storage is expected to remain for a minimum of 180 days. Any data removed prior to this is subject to an early deletion cost and is prorated. For example, if data is moved to the Archive tier and then deleted or moved to the Hot/Cool tier after 75 days, the customer is charged an early deletion fee for 105 (180 minus 75) days of storage in the Archive tier.

Azure offers reserved capacity for blob storage, which provides reduced cost of storage through discounts. This is done by committing to either a one-year or three-year term for a fixed amount of storage capacity. The cost savings achieved depend on the duration of your reservation, the total capacity you choose to reserve, and the access tier and type of redundancy that you've chosen for your storage account. Reserved capacity provides a billing discount and doesn't affect the state of your Azure Storage resources. Reserved capacity is purchased in units of 100 TB and 1 PB per month for a one-year or three-year term.

Azure Blob Storage pricing

Optimize costs for Blob storage with reserved capacity

Commvault Combined Storage tier

In addition to Azure Hot, Cool, and Archive storage, Commvault also provides two additional storage options and are unique to Commvault. We have combined Azure Archive storage with Hot and Cool storage to assist when writing deduplicated data. **Hot/Archive** and **Cool/Archive** are beneficial to those who want to store data (SFILE) in Archive storage, but still have fast access to the chunk metadata (CHUNK_META_DATA) for fast deduplication lookups.

Note: As of Feature Release 11.23 the job index data is stored in the Hot\Cool tier for fast browsing of files in the job for restores. There is additional cloud storage costs for the storage of the index, but the efficiency and smoother user experience far outweighs the small additional costs. This does not improve the restore time of the actual data from Archive tier drastically which could still be upwards of 15 hours. Therefore, Commvault recommends using combined storage tiers for backup data with long term retention using deduplication.

Combined storage tiers are not recommended for primary backup copies. The reason for this is because of the time required to recall the data as well as the cloud costs incurred for the recalls. If you need fast access to the data to recover the business operationally – use the Hot or Cool storage tier and not a combined storage tier.

For customers that are looking to use combined storage tiers and currently have their data in Hot or Cool, you will need to configure and add a combined storage tier to the Commvault Plan. Set the retention policy for the combined storage and define the primary copy as the source. This will aux-copy the data from the primary copy to the combined storage tier. We do not 'move' the data, it's a copy. Once the data has been copied to the combined storage tier, you can change the retention setting of the primary copy, which will prune the data from the primary copy. The result will be reduced cloud storage costs in the primary storage copy. However, you should make sure the primary copy retention is sufficient for typical data recall requirements, so as not to incur Archive recall costs.

Combined Storage Tiers

Commvault metadata stored in Azure Hot\Cool tier

Backup Job Indexes

Backup Job Indexes store the details of data retained in each individual backup job or job copy (auxiliary copy).

Backup job indexes are mandatory to present a list of valid backups for recovery.

Pruning Indexes

Pruning indexes hold details on which individual blocks in the backup store are used or expired.

Commvault uses pruning indexes to determine when data can be 'deleted' from the backup store.

Deduplication metadata

When **deduplication** is utilized, deduplication metadata is stored in the Hot\Cool tier. This metadata is used for restoration and DDB reconstruction activities.

Commvault stores **all** indexing data within the 'Hot or Cool' storage class when using **Combined Storage Tiers,** for seamless recovery

Combined Storage Tiers

Restore Types

Automated

- Commvault **alert** triggers Cloud Archive Recall Workflow
- Workflow orchestrates recall of selected data
- Limited to **specific agents**
- Granular restore of files, folder, tables, etc.

Manual / On-Demand

- **Cloud Archive Recall Workflow** is manually executed
- **Job ID** of required job is provided
- **Cloud Archive Recall** recalls the data for selected Job ID.
- Supported by **all commvault agents**

Best Practices

Secondary copies only

Commvault combined storage tiers are designed for infrequently used data.

They are not designed to be used as the **Primary copy** location.

Use for Storage copies

Commvault combined storage tiers represent an additional **storage copy** configured in your Plan/Storage Policy.

Commvault does not move data between cloud libraries

Just add a storage copy…

If combined storage tiers are adopted after a cloud library is already populated…

Individual **jobs** must be copied, then expired from primary location.

Assess data needs day one

To avoid lengthy data migration processes, consider your **near-term**, **mid-term**, and **long-term** data accessibility when creating your plans.

	Hot/Archive	Cool/Archive	Archive
Deduplicated backup	Metadata and job indexes will be written in the Hot tier and data will be written in the Archive tier	Metadata and job indexes will be written in the Cool tier and data will be written in the Archive tier	Metadata, job indexes, and data will be written in the Archive tier
Non deduplicated backup	All data will be written in the Hot tier	All data will be written in the Cool tier	All data will be written in the Archive tier
Restore	**Workflow** will recall the data from the Archive tier to the recall destination. (Recall destination can be set in the Workflow. Can be Hot or Cool. Default value is Cool in the workflow.)	**Workflow** will recall the data from the Archive tier to the Cool tier. (Recall destination can be set in the Workflow. Can be Hot or Cool. Default value is Cool in the workflow.)	For deduplicated data, **Workflow** will recall both the metadata and data from the Archive tier, directly to the restore location. For non-deduplicated data, all data will be recalled from the Archive tier, directly to the restore location.
After restore	Data will be moved back to the Archive tier by the **Workflow**	Data will be moved back to the Archive tier by the **Workflow**	For deduplicated data, **Workflow** will move the metadata and data back to the Archive storage, once the recall period provided in the Workflow is met. For non-deduplicated data, all data will be moved back to the archive storage

Note: Micro-pruning is not supported in the Azure Archive storage tier.

Consider the following characteristics when selecting your cloud storage – there is cost in storing, cost in retrieving, and cost associated with the **latency** of your data retrieval. While **High Priority retrieval** options are possible, these should be considered exceptional events and choosing a warmer storage class is recommended if a likelihood for rapid recovery is expected.

Azure Storage Classes – for Backup & Archive			
Feature	**Hot**	**Warm**	**Cold**
Storage Class	Azure Hot	Azure Cool	Azure Archive
Annual Storage cost (1 TB)	$220.80	$120.00	$24.00
Minimum retention	None	30 days Note: Recalls under 30 days will be pro-rated	180 days Note: Recalls under 180 days will be pro-rated
First byte latency	milliseconds	milliseconds	Up to 15hrs Typically around ~4hrs
Recall cost (1 TB)	$0.016	$10.056	• Standard: $40.00 High-Priority: $300

Note: Pricing is for LRS storage from region 'East US 2' and is in $USD. Minimum retention is the time required for data to remain in storage or the data will be pro-rated and charged early deletion fees. Recall costs are based on a single recall of data from storage and includes the cost of 'Read Operations' and 'Data Retrieval'.

https://azure.microsoft.com/en-us/pricing/details/storage/blobs/

Commvault recommends a **tiered approach** to consuming Cloud storage from your Primary application data, Secondary (backup) data, and finally your Tertiary (archive) data vaults.

Warning: Make sure the Azure Life-Cycle policy is disabled for any Storage Account being used for Archive data. There is a conflict where Azure re-archives any data we recall before we can use it for the restore.

More information can be found here:
https://docs.microsoft.com/en-us/azure/storage/blobs/storage-lifecycle-management-concepts?tabs=azure-portal

Azure Immutable Storage (WORM)

Immutable storage, also known as WORM (Write Once, Read Many) storage, enables users to store business-critical data with the requirement that it cannot be modified or deleted based on a user defined retention period. Designing a solution with offsite copies to protect against ransomware and cyber threats is imperative. Commvault provides the ability to utilize cloud immutable storage with Azure Blob storage for enhanced data security.

Immutable storage for Azure Blob storage supports two types of immutability policies:

Time-based retention policies: a time-based retention policy which users can define for a specified time. Data can be created and read but cannot be modified or deleted until the retention period has expired.

Legal hold policies: a user defined retention policy which explicitly holds data until the policy is cleared. A user can define for a set of data as legal hold. This subset of data can be created and read but cannot be modified or deleted until the legal hold is cleared.

When enabling **Azure immutable storage**, Commvault makes the following changes:

Configures a default **retention period >** to 2x the data retention period set on the Server Plan (or Storage Policy)

Enables **default retention mode >** (legal hold) on the container, for newly written objects.

Configures **periodic DDB sealing** to match the Server Plan (or Storage policy).

Micro-pruning is disabled on the Cloud Library

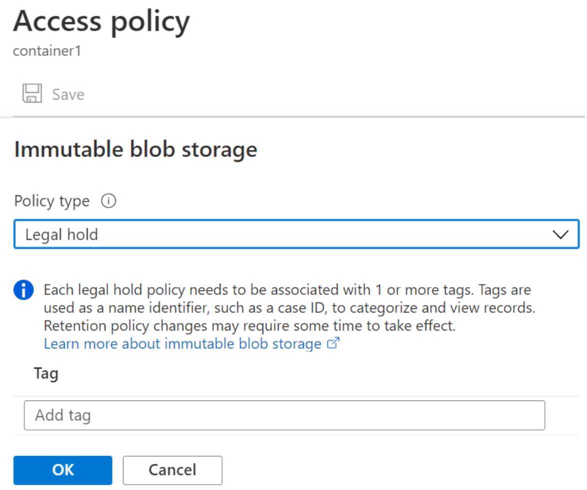
Configuration Example

For associated deduplication-enabled storage pools, retention at the container level is the sum of the value at the storage policy copy level and the DDB seal frequency value on the cloud. The seal frequency is set to either a maximum of 180 days or half the retention days of the storage policy copy, whichever is greater. For example, if the retention is set for 60 days at the storage policy copy level, the seal frequency is set to 180 days, and then at the cloud storage level, the retention is set for 60+180=240 days. If the retention at the storage policy copy level is 2 years, then the seal frequency is set to 1 year, and the cloud storage level retention is set to 2 +1=3 years.

A cloud library server plan with 90 days retention

Default retention period will be 180 days

The DDB will be sealed every 90 days



Policy type: Legal hold

Each legal hold policy needs to be associated with 1 or more tags. Tags are used as a name identifier, such as a case ID, to categorize and view records. Retention policy changes may require some time to take effect.
Learn more about immutable blob storage

Tag: Add tag

Note: Immutability should ideally be enabled on the backup library before writing any backup data to ensure data is immutable from initial creation.

Effects of DDB sealing

When the DDB is sealed, the sealing process closes the DDB and starts a new DDB. When the new database is started, the next instance of each data block processed creates a new signature tracking entry and the data block is written to the disk again as a new initial baseline.
This will result in a full copy of the backup content being re-sent to the Cloud Library. This is intentional and provides multiple, segregated data copies to protect from corruption or other unforeseen data access issues.

Additional information can be found here:

Configuring the WORM Storage Mode on Cloud Storage

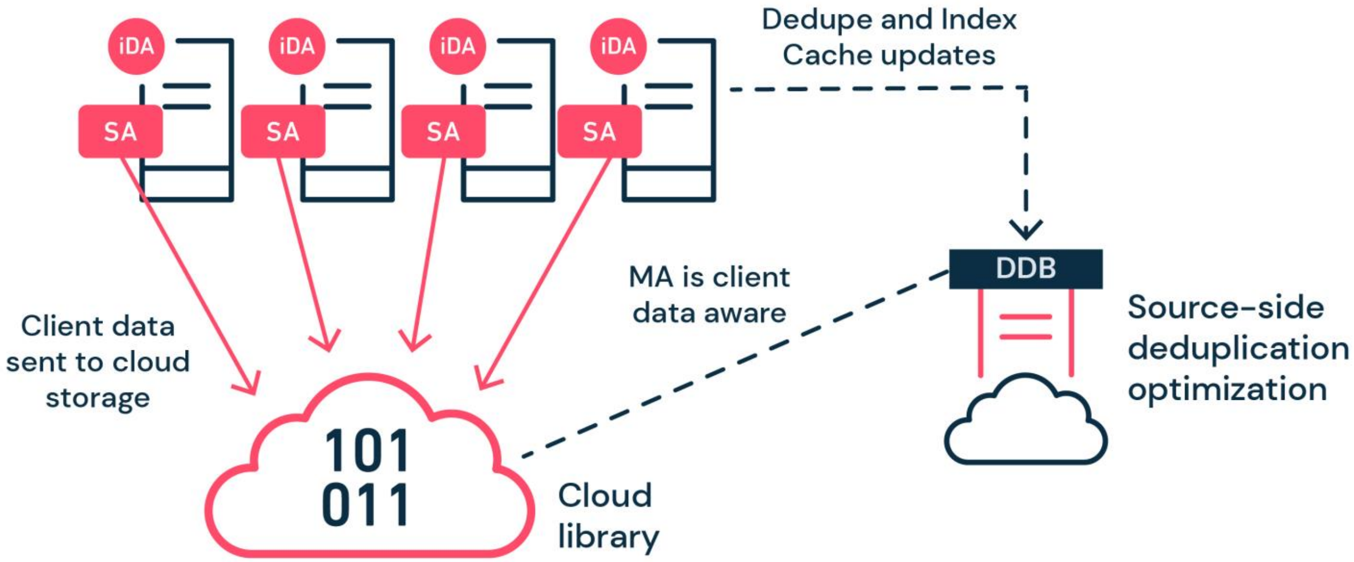
Workflow for Configuring WORM Storage Mode on Cloud Storage

Sealing the Deduplication Database

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. Workloads with large datasets may benefit highly by this use case, as an example VMs that are greater than 1TB. 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 MediaAgent 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**


iDA
iDA
iDA
iDA
SA
SA
SA
SA
Dedupe and Index Cache updates
DDB
MA is client data aware
Client data sent to cloud storage
Source-side deduplication optimization
101
011
Cloud library

The example above shows the storage accelerator installed on the clients. During a backup job, the clients communicate with the MediaAgent deduplication database (DDB) and only send the unique blocks directly to the cloud storage target.

Restores are handled through direct transfer of data from the cloud storage to the client, bypassing the MA, which greatly reduce restore times. This is especially powerful when many restores are happening at the same time to different clients. The throughput of the MediaAgent is not a limiting factor. The throughput will depend on the cloud storage and the client. Since the MediaAgent is only handling dedupe tasks and not the transfer of data for these clients, the load is reduced and can potentially be sized smaller.

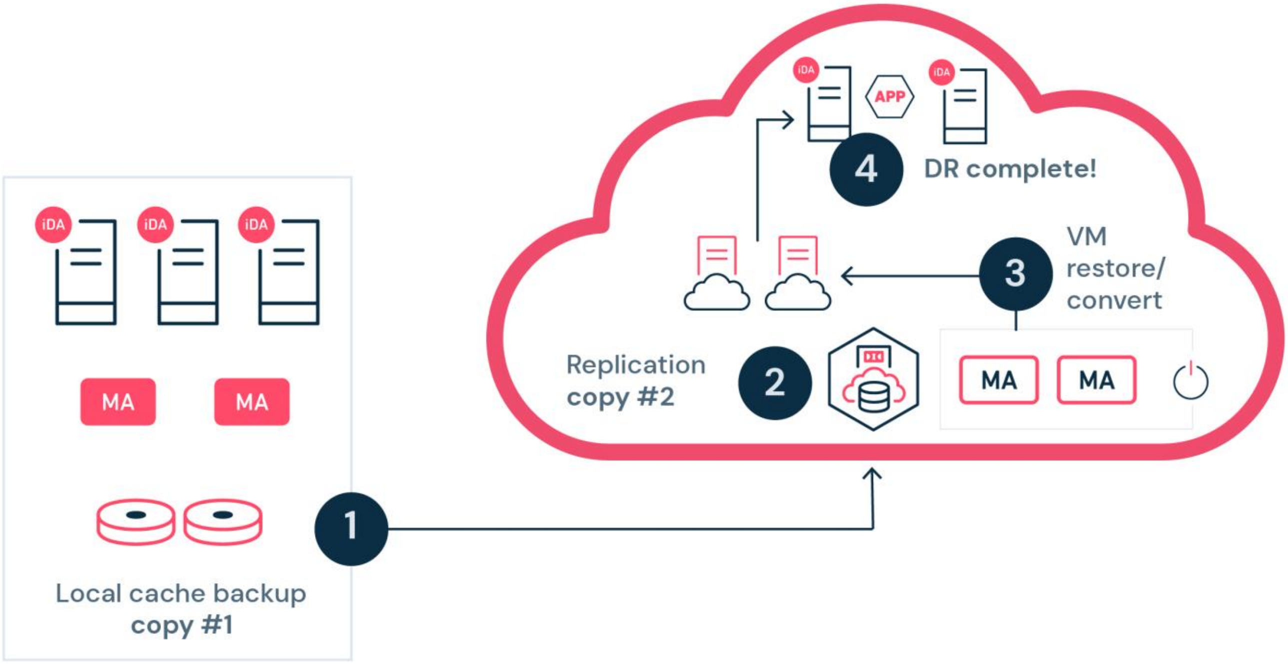
During testing the performance was obvious and provided faster backups, 3 to 5 times faster.

Note: The storage accelerator role consumes additional compute and network resources while performing data protection operations and should be managed so as not to interfere with production use.

Use-Cases / Deployment

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)

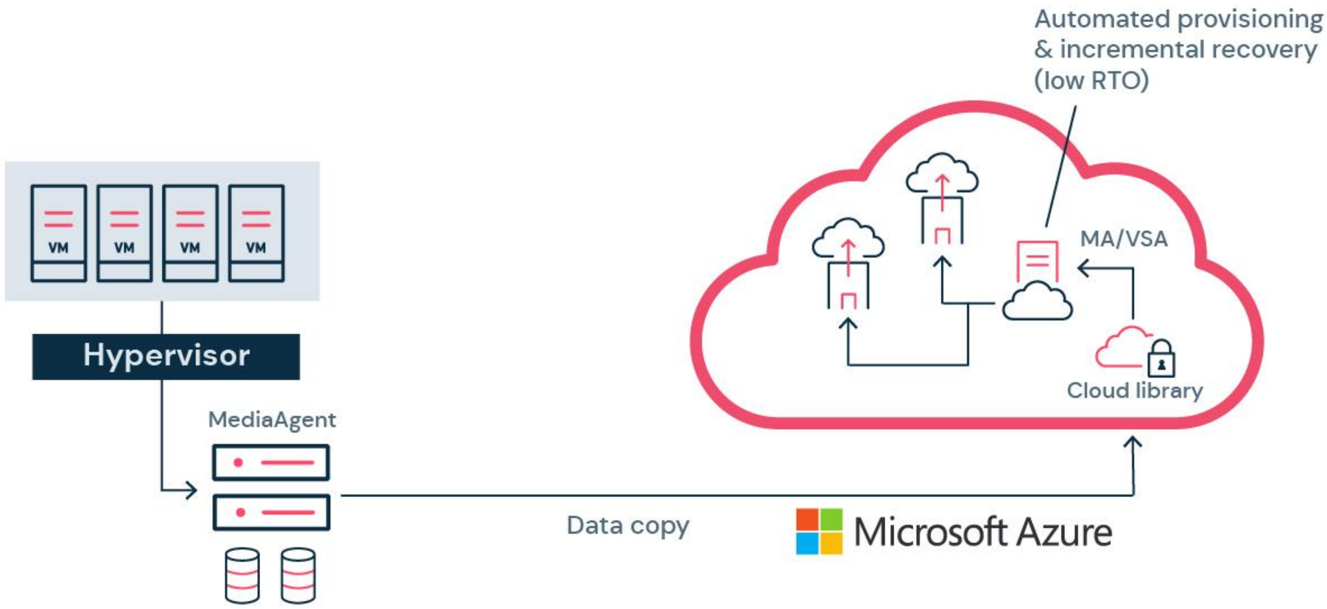
An agent-in-guest approach allows for the recovery of a wide variety of operating systems and applications. These can be captured at the primary site and replicated to the cloud based MediaAgent in a deduplicated efficient manner. Once replicated, the data can be held and restored in the event of a disaster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

VM Replication (Live Sync)

Live Sync allows you to replicate VMs to the same or a different hypervisor, including public cloud infrastructure. As of December 2016, Azure is a supported cloud infrastructure vendor for Commvault v11. Live Sync combines the VM conversion feature with incremental replication to provide a disaster recovery 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, Azure Stack HCI, 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. Live Sync supports two configuration options:



Create and maintain a VM in Azure in a stopped state – This provides a low RTO for a customer, but the customer incurs Azure costs for resources (storage, IP, etc.) utilized by the VM.

Create a VM at the time of failover – This saves the customers cost because no resources are used until the VM is created. However, this can elongate the customer's RTO (usually only takes a few minutes) because the VM needs to be created after the failover occurs.

Note: There are limitations when writing to an Azure Managed disk which affects our replication methods. The API provides the ability to create a Managed disk but doesn't allow us to reopen the disk for incremental writes in the same way we can with an Unmanaged disk. In order to overcome this, we stage the VM disk being replicated in blob storage so that we can write the incremental data. Once the VM is required in a Failover state we create the Managed disk from the blob storage.

Get additional information on the **Conversion feature.**

Virtual machine recovery from Amazon EC2 to Azure VM

Since the release of Commvault v11 SP7, you can recover Amazon EC2 instances protected with the Virtual Server Agent (VSA) 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 VMs to Azure VMs from the following hypervisors:

- Amazon
- Azure Classic
- Azure Resource Manager
- Azure Stack Hub
- Hyper-V
- Nutanix AHV
- Oracle VM
- VMware

Backups can be 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 disaster recovery site, or as a migration strategy (Lift-and-Shift).

Additional details can be found **Cross-Hypervisor Restores.**

Using Commvault Workflows to automate Disaster Recovery

The Commvault Workflow engine provides a framework in which the disaster recovery 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 the Commvault Partner Account team.

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

Protecting and recovering active workloads in Azure

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.

Agent-less VM protection (Virtual Server Agent for Azure)

Introduced in Commvault V11 SP4, the Virtual Server Agent (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 (CBT), included with v11 SP5, which helps accelerate incremental backup performance. With v11 SP15, Commvault added support for 8 TB VM disks and backup and restoration of VMs that are using Azure Disk Encryption with Azure Key Vault was added with SP15. These features provide customers with additional functionality when protecting and recovering IaaS VMs in Azure. The VSA Proxy is also referred to as the **Access Node**.

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. 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 Unmanaged Disks since v11 SP4. As of v11 SP10, protection of Azure Managed Disks is also supported.

Azure Generation 2 VM

Azure added support for Generation 2 VMs in February 2020. Gen 2 VMs use the new UEFI-based boot architecture, as opposed to the BIOS-based architecture used by Generation 1 VMs.

Some of the benefits of Gen 2 VMs:

- Build larger VMs (up to 12 TBs)
- Provision OS disks sizes that exceed 2 TBs
- Supports Windows and Linux Managed VMs

As of FR 11.19 Commvault is able to protect Azure Generation 2 VMs. During protection operations, the “VM Generation” is identified so that a Gen 2 VM is protected with the ability to restore, replicate, and convert.

The following features are supported:

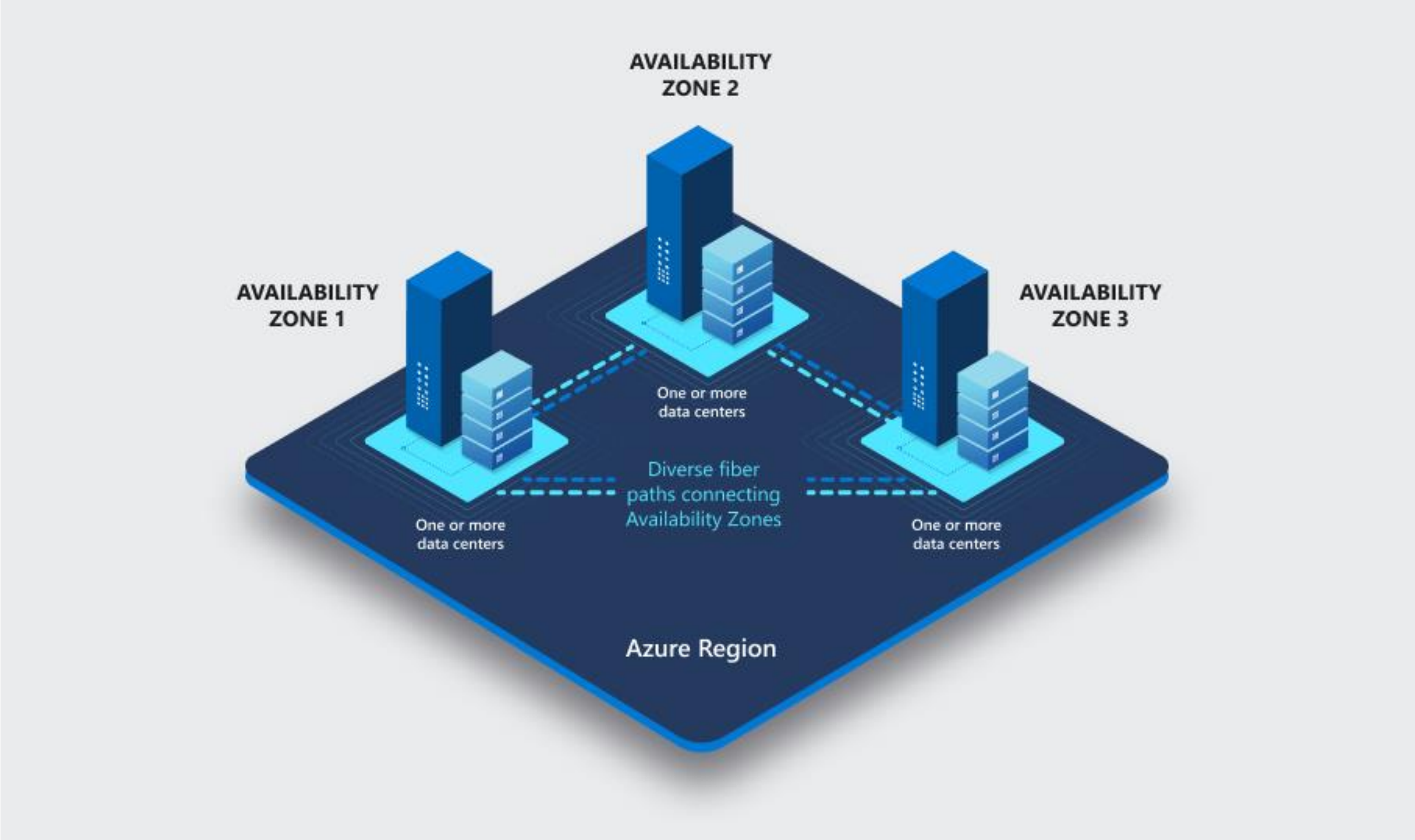
- Streaming backup
- Snapshot
- Backup Copy
- Azure to Hyper-V conversion
- Hyper-V to Azure conversion
- Azure to Azure replication
- VMware to Azure conversion
- Changed Block Tracking
- Managed VMs only supported by Microsoft

Additional information:

https://docs.microsoft.com/en-us/azure/virtual-machines/windows/generation-2

Azure Availability Zones

An Azure Availability Zone is a high availability offering that protects applications and data from datacenter failures. They are unique physical locations within an Azure region. Each zone is made up of one or more datacenters equipped with independent power, cooling, and networking. To ensure resiliency, there's a minimum of three separate zones in all enabled regions. During VM creation the availability zone can be configured based on the users requirements. Commvault will capture this information during VSA backup. When a restore is initiated, the VM will be restored back to the availability zone originally defined. This feature is enabled by default in FR 11.21 and does not require any configuration. Microsoft currently only supports Managed VMs with availability zones.



Replicate and migrate VMs to a specific Azure Availability Zone

As of Feature Release 11.23 VMs can be replicated and migrated to an Azure Availability Zone from the Commvault Command Center. This reduces the complexity and time required to ensure important workloads are configured with redundancy based on their requirements. When migrating non-Azure VMs to Azure you can specify the Availability Zone (1, 2, 3, None, or Auto). A job will fail if the VM is not able to be migrated or replicated to a specifically defined Availability Zone (1, 2, 3, or None). Leaving it set to 'Auto' will choose the Availability Zone setting of the source VM. If the source VM is not an Azure VM or does not have an Availability Zone defined, the VM will not be placed in an Availability Zone. The 'Auto' setting is a best-effort only and no job failures will be associated with it.

Additional information:

- Replication with Availability Zones

- Azure Availability Zones

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 block-level backup to provide application aware VM and File-level recovery. The following matrix shows which applications are supported for application aware protection for Microsoft Azure. Application aware protection for Azure was added with V11 SP16. Application Consistency was added in CPR 2022E (11.28) with Restore Point Collections. It is recommended to use the agentless approach unless Azure does not support application consistency with the VM and protection beyond crash consistency is needed.

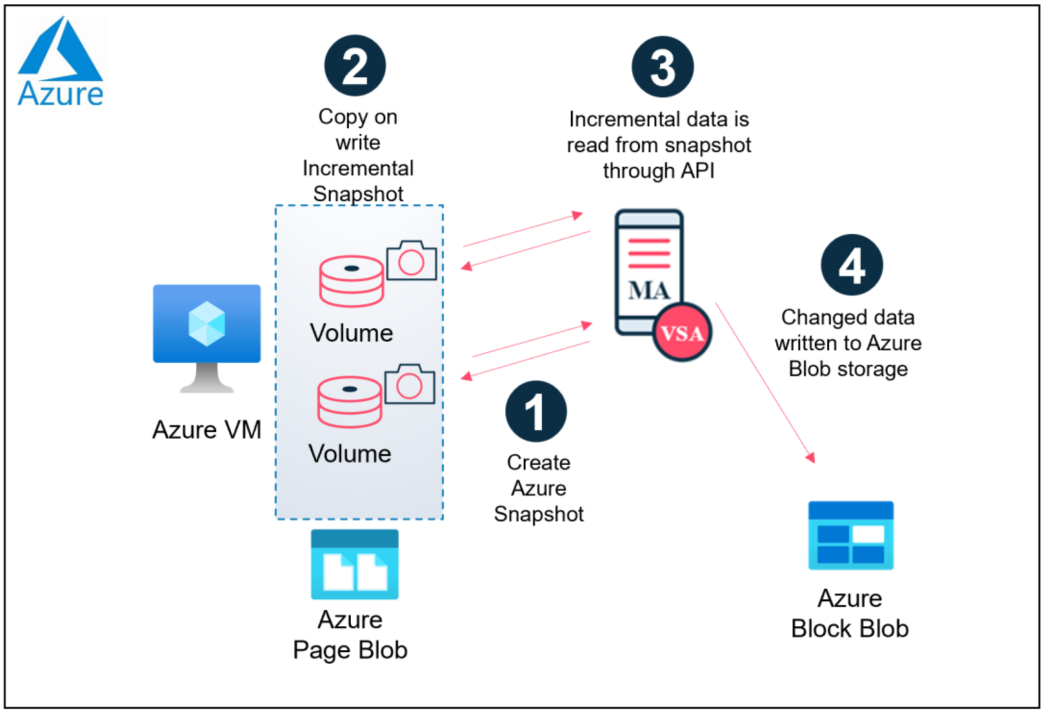
When not to use the VSA for Azure

- 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).
- An example of this architecture would be an application that uses ephemeral worker VMs in Azure Scale Sets. The database may exist in Azure Data Lake or Azure Managed SQL or even a farm of VM database servers that are reserved instances. Those data locations should be protected using VSA or database agents, but the ephemeral VM's in the scale sets should not be.

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, Resource Group, Region, Storage Account, Tags, or Power state. 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 (through APIs) the snapshot to a nominated Access Node (Azure VM-based VSA / MediaAgent) to read and de-duplicate the blocks before writing out to an Azure storage.


Azure
2
Copy on write Incremental Snapshot
3
Incremental data is read from snapshot through API
MA
VSA
4
Changed data written to Azure Blob storage
Volume
Volume
Azure VM
1
Create Azure Snapshot
Azure Page Blob
Azure Block Blob

Commvault IntelliSnap® functionality introduced in V11 SP7 for the VSA for Azure. Commvault IntelliSnap® allows snapshots to be retained on the VM configured via Storage Policy's Snap Primary retention setting.

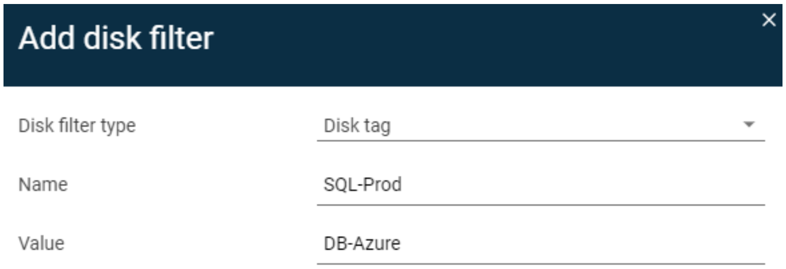
Commvault IntelliSnap® backups enable reducing backup windows considerably, providing fast, snapshot-based restoration capability. Azure also allows for the use of CBT (Changed Block Tracking) which can accelerate incremental backup performance considerably.

Azure changed block tracking (CBT) is available for Azure unmanaged and managed disks and are both supported by Commvault.

Get additional details **IntelliSnap for Microsoft Azure**

Filtering Disks using Azure Tags

- Tagging is a feature used by cloud providers to simplify the tracking of cloud resources. Disk Tags are a user-defined name and value pair which can be assigned to a cloud disk. For example, you can apply a disk tag name of "SQL-Prod" and give it a value of "DB-Azure". Tagging all disks with this pair will help organize and identify all SQL database disks that are in Azure.
- Commvault supports Azure Disk Tags and can be used during data protection operations to group/filter VM disks. The Tag Name or Tag Value that are added as filters will be used to search through the Azure subscription associated to the VSA client.


Add disk filter

Disk filter type	Disk tag
Name	SQL-Prod
Value	DB-Azure

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 “VSA/MA” node represents a single Azure VM with the Virtual Server Agent VSA and MediaAgent MA modules deployed. The Azure VM specifications should match the MediaAgent specifications within this Architecture Guide.
- When considering the VM instance type for the Access Node, the Commvault platform does not mount the guest VM snap disks on the Access Node for backup. This means that if you are protecting VMs with Premium Disks, you don’t necessarily need to select an Access Node 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 Access Node VM type is not able to keep up with the required compute and IOPS demanded of it.

Architecture recommendations

- Use of the Commvault IntelliSnap® configuration is highly recommended to improve backup and restore times by enabling native Azure snapshot retention. Use of this method does mean that snapshots remain available for longer than the backup window, however Commvault recommends that snapshots are retained only for as long as required. The Snap Primary can be configured to retain at least one snapshot, keeping snapshot costs at a minimum while providing fast backup and restoration capabilities.
- By default, VSA backups of Azure VMs are crash consistent. To receive application consistent backups, leverage the AppAware feature.
- While the readers count can be increased to improve concurrency per VSA/MA node, consider scaling out with multiple VSA proxies. Azure recommendations mention that for optimal performance, you will want to limit the number of highly utilized disks attached per worker node to avoid possible throttling.
- Use of Premium Storage SSD 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 MediaAgent 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.

- (Backup Copy / Streaming) Disable Granular Recovery of files if granular recovery is not required, or agents-in-guest are used to collect large file system datasets. This will improve the backup window by removing the need to 'walk' the file system structure within the Azure VM volumes.

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

Best practice: The SQL server on the proxy client should run the latest SQL server release to ensure the proxy server is in sync with the Azure instance. Add the SQL server instance to the CommCell Console or Command Center. 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 as a reference.
- Check the Systems Requirements section in Commvault documentation to determine if the iDataAgent supports your application (Backup Agents).

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® supports Azure managed disks and unmanaged disks.

With Commvault IntelliSnap® support for Microsoft Azure, you can:

- Perform full and incremental snapshot backups of virtual machines.
- Perform backup copy jobs from snapshot backups.
- Perform backup copy jobs with Changed Block Tracking (CBT).
- Perform full VM restores from snapshot backups, and full VM and file level restores (using Live Browse) from backup copy jobs.

Get additional details **here**

Azure blob storage backup

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

For more information refer to **Azure Blob Storage Overview**

When to use Azure Blob storage backup:

- Backing up object storage – Protect data/objects created and stored in Azure Blob storage.

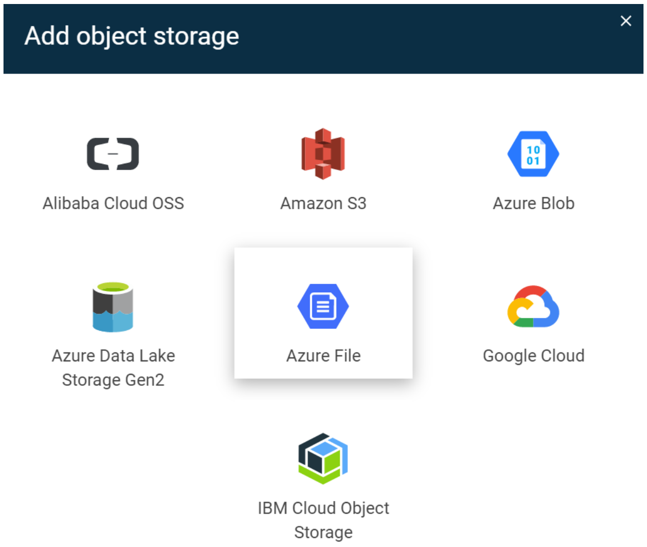
Architecture Recommendations

- For large datasets, consider using multiple subclients to increase scan performance and lower the amount of time taken to traverse and scan the bucket contents.
- Configure data operations for multi-streaming using multiple readers for best performance.
- To improve backup performance, we recommend you disable logging or to redirect the logging to another bucket in a user-defined subclient.
- Metadata is protected during backup

When not to use this approach:

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

Azure File Share



Azure File Share is a feature derived from Azure Blob storage. Commvault can protect Azure blob storage and can also protect Azure Files. 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 File share folder of your choice, including the metadata at a folder level. You can back up an Azure Files Share by providing the Azure File Share URL (Host URL= **file.core.windows.net**) when you create an Azure File virtual client from the Command Center or the CommCell® Console. Once configured, you'll have Full and Granular restore capabilities of the data, in-place or out-of-place.

This feature is currently available for streaming backups and snapshots. Azure File Share Snapshot backup was added in FR 11.23 and offers protection of locked files by default. You can also choose to backup all files using Azure File Share snapshots.

For more information refer to Azure File Storage

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.

Deployment

Installation basics

The following links cover the steps when installing the CommServe® server in the cloud. This is only needed when the primary CommServe® server 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)

Multi-mobility (flexibility)

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 several methods.

Virtual Machine restore & convert (lift and shift to Azure)

The Virtual Server Agent can capture virtual machines from VMware, Hyper-V, Nutanix AHV, Oracle VM and AWS based platforms in an application-consistent method 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 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® Command Center or automated via Commvault Workflow and 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® Command Center or automated via Commvault Workflow and API calls.

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

Datasheets

- Commvault for Microsoft® Azure Cloud
- Go ahead, move your most important applications and databases to Azure
- Accelerating data migration to Azure
- We manage and protect mission-critical apps running on Azure Ultra SSD
- The right choice for enterprise class recovery of SAP workloads on Azure
- Commvault and Azure NetApp files
- Thinking of managing your open source databases on Microsoft Azure PaaS?
- Microsoft Datasheet: AS/400 Backup to Azure
- A checklist for cloud data management
- Microsoft trusts Commvault. Shouldn't you?

Solution briefs

- Rapid data migration for Azure
- Azure Stack Backup and Recovery

Videos

- VMware to Azure Migrations With Commvault
- Virtual Machine Backup Demo Using the Commvault Command Center
- Keys to Fast, Secure Azure Migration With Data Protection
- Shikun and Binui: Cloud Data Protection Across Azure, Office 365 and SaaS environments
- Commvault Customer Champions: UConn Health

Case studies

- University of Central Florida
- HarperCollins
- Yura Corporation reduces costs and boosts business continuity
- Chart Industries
- UConn Health
- Laing O'Rourke

commvault.com | 888.746.3849