Public Cloud Architecture Guide for Amazon Web Services

version 11 Feature Release 20 (11.20)

15th June 2020

Table of Contents

Abstract 4
Commvault with AWS Reference Architectures 5
Amazon Cheat Sheet 11
Shared Responsibility Model 12
Commvault Platform Design Principles for Cloud 13
Architecture Sizing 23
Design Patterns 38
Architecture Considerations 43
Performing Disaster Recovery to the Cloud 51
Protecting and Recovering active workloads in AWS 55
Application Migration 69
Extend to the edge with AWS Outposts 71
How to get started 72
Additional Resources 75
Revision history 76
Solutions, References and Videos 78
Index 79

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.

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

> Commvault utilizes attributes of public cloud to enable cost effective on-demand use cases for both data protection and data management both to and in public cloud platforms.

Cloud resources, bandwidth, and their availability are often localized via massive regional presence to proximity of on-premises corporate assets and human resources, allowing for an easy on-ramp to public cloud. The cost model implications of pay-as-you-go do not just extend to only production workloads, but also to the ever-present challenge of providing a flexible, agile, yet capable, recovery solution for your applications and data. Today, many recovery environments have less compute and storage capacity than their production counterparts, resulting in increased risk of elongated business service outage.

With the public cloud model, the infrastructure availability and refresh aspect are disrupted by removing the need to maintain a hardware fleet that can meet both your recovery requirements and sustain your service level agreements. Public cloud instances can be rapidly provisioned to meet the needs tied to business requirements,

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

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

Abstract

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

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

Currently this guide delivers architecture considerations and sizing recommendations for the Amazon Web Services™ (AWS). Guides for other public cloud environments are available as well.

Terminology

Commvault is a multi-faceted data management solution that contains multiple components that can be consolidated for reduced infrastructure footprint or separated for improved scalability. The following are some common terms that will appear within this document and their definition.

Access Node refers to the component that is responsible for connecting to your primary application and capturing the data to be protected. An Access Node runs a Commvault software package for accessing the source application (i.e. Virtual Server Agent, MySQL Agent, PostgreSQL Agent, Cloud Apps Agent).

Media Agent refers to the component that is responsible for data handling, both transfer and indexing to facilitate optimized backup and recovery operations. A Media Agent runs the Commvault Media Agent software package which communicates directly with secondary storage libraries (disk, cloud, tape).

Any reference to an Access Node (within this document) refers to a system that performs the role of 'Access Node' and 'Media Agent' unless stated otherwise.

Commvault with AWS Reference Architectures

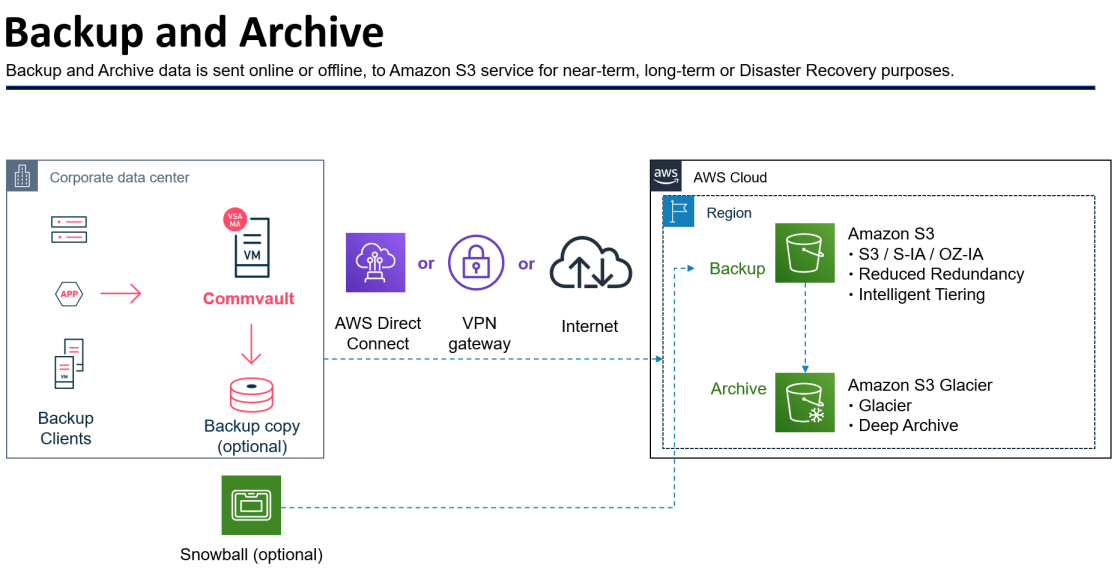
The most common use cases observed at most customer environments by Commvault, related to cloud, fall into three categories depending on the maturity level of initiatives around cloud adoption.

- **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 instances.
- **Manage data in and across clouds** – protecting and life-cycling data and instances 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:

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

Move Data – Backup and Archive to the Cloud

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

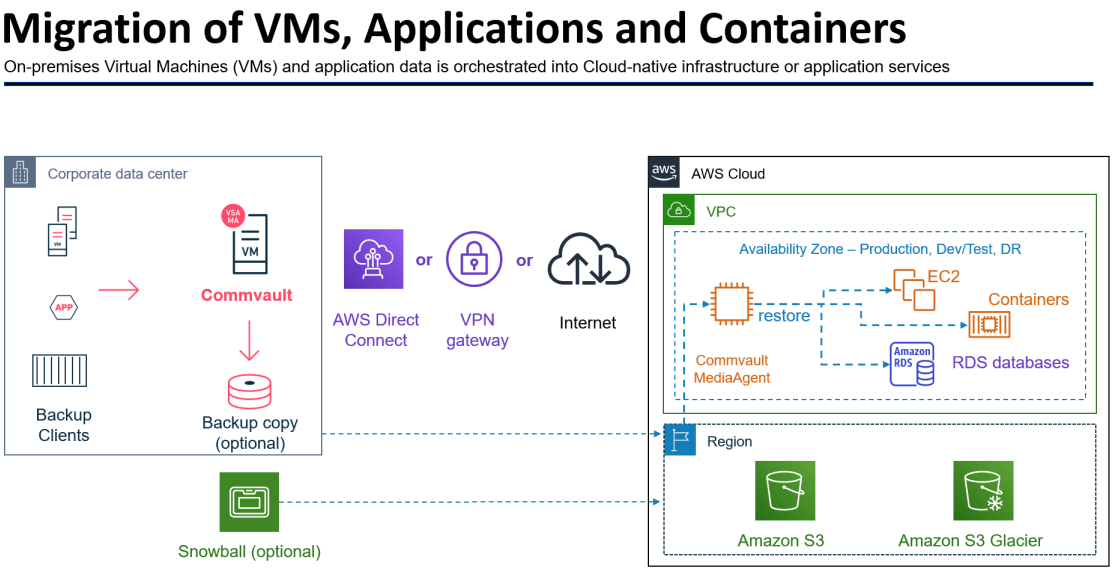


Move Data – Backup and Archive to the Cloud	
Scenario	**Requirements**
• Offsite Storage and "Tape Replacement" Scenario – replace long-term on-site retention with cloud storage. • Native, Direct connectivity to Amazon S3 / Glacier. endpoints – no requirement for translation, gateway, or hardware deduplication 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 a public IaaS provider (such as AWS) or via a Managed Service Provider (MSP). As long a S3 • API endpoint is available. • Local backup copy on-premises is not required. • Remote Office Branch Offices (ROBOs) can send data directly to Cloud.	• Minimum 1x MediaAgent on-premises with no instances in cloud required for backup to the cloud. • Can use direct internet connection or a dedicated network to the cloud provider for optimized data transport performance in a secure manner • Can use offline transfer method by writing backup, archive data of AWS Snowball device(s), then importing into Amazon S3 and Glacier. (lifecycle). • (optional) In-cloud MediaAgent can be created to support a DR solution to cloud using the data that is placed in the cloud storage service (not depicted). • Cloud MediaAgent can be deployed at the time of the DR event or DR test, as required. • Remote offices do not require onsite Commvault infrastructure, they can upload backup data direct to S3 (not depicted)

Move Data – Migration of VMs, Applications and Containers to the Cloud

Business Value: Upon protecting VM and application data at the primary on-premises location, Commvault® software orchestrates the migration of application workloads into the cloud, either at the VM container level or the application

level. While providing the migration lifecycle and workloads are in a transition phase between on- premises and public cloud, data is still protected on-premises.

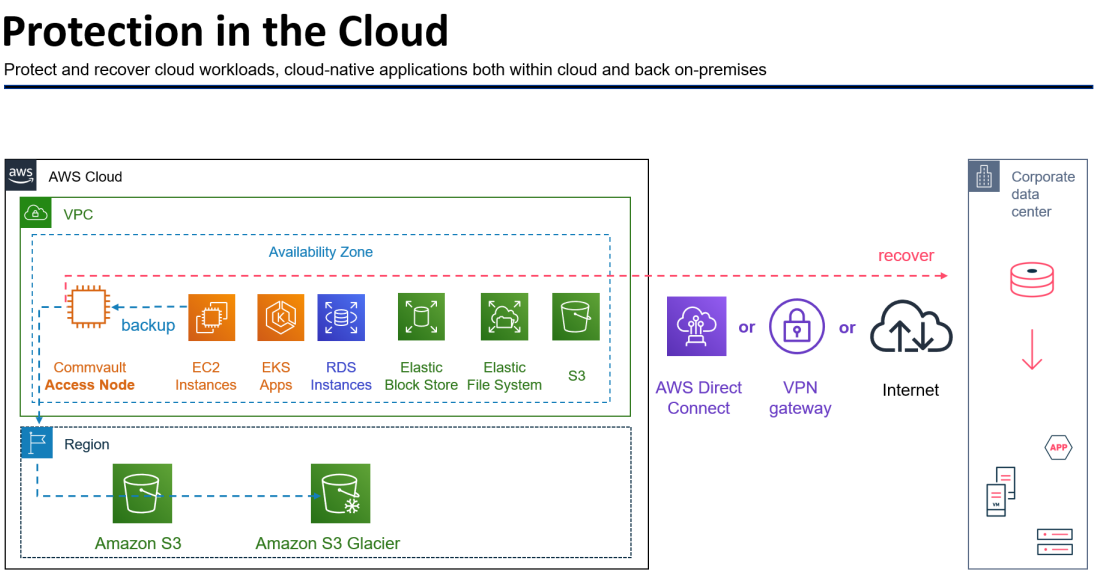


Move Data – Migration of VMs, Applications and Containers to the Cloud	
Scenario	**Requirements**
• **Lift & Shift of virtual machines** – Application-consistent VM backups are used to restore and convert VMware and Hyper-V VMs into Amazon EC2 instances as part of a migration with a phased cut-over strategy reducing on-premises downtime. • **Oracle Migration Feature** (Linux, AIX, Solaris to AWS) – an Oracle Application Migration feature to both synchronize the baseline and any incremental changes as part of a migration lifecycle. • **Application Restore Out-of-Place** –Leverage Commvault® iDataAgents for your supported workload to restore the target application out-of-place to a warm instance 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. • The Oracle migration feature to Amazon EC2 supports Oracle on Linux, AIX and Solaris. For AIX and Solaris source databases, a destination Amazon EC2 instance must exist with Oracle installed. However, for Linux source databases, the destination Amazon EC2 instance can be provisioned as part of the process. • It's highly recommended to use dedicated network links to the cloud provider for best performance (e.g. AWS Direct Connect).

Manage Data – Protection in the Cloud

Business Value: Providing operational recovery for active workloads and data within an external provider's cloud.

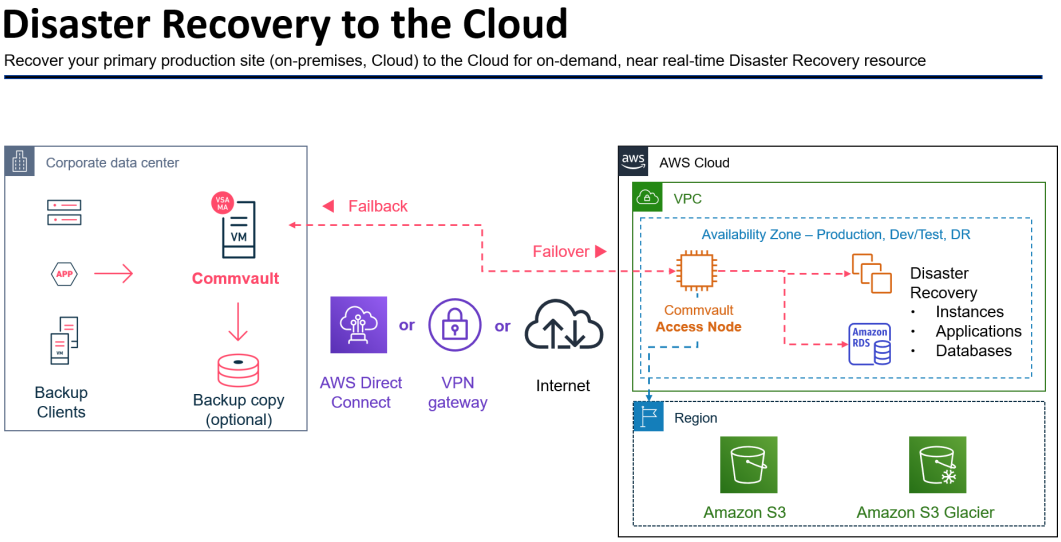
Provide the ability to lifecycle data and cloud instances to meet SLA and cost requirements.



Manage Data – Protection in the Cloud	
Scenario	**Requirements**
• **Data protection for Cloud based workloads** – protecting active workloads within an existing IaaS Cloud (Production, Dev/Test, etc.). • **Agentless Instance Protection** – protect instances with an agentless and script-less protection mechanism through the Virtual Server Agent. • **DASH Copy** data to another region, cloud, or back to on-premises – complete data mobility by replicating to another geographical region within • IaaS provider, a different IaaS provider, or back to on-premises sites. • **Protect Amazon S3** – Backup object storage repositories containing data created by other third-party applications either in cloud, to an alternative provider, or back to on-premises sites.	• An Amazon EC2 **Access Node** provides agentless backup. Applications will require agent-in- guest deployed in-instance. • Applications requiring application-level consistency, and all other cloud providers can be protected via **agents** deployed in each VM/instance within the IaaS provider. • **Minimum 1x MediaAgent in source cloud**, and (optional) minimum 1x MediaAgent at secondary site (whether cloud or on-premises) for receiving replicated copy of data (not depicted). • Recommended to use a dedicated network from cloud provider to on-premises for best performance when replicating back to on-premises (e.g., AWS Direct Connect).

Use Data – Disaster Recovery to the Cloud

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



Use Data – Disaster Recovery to the Cloud	
Scenario	**Requirements**
• **Off-site storage and cold DR site in the Cloud** – only use the Cloud compute infrastructure when a DR event occurs, saving time and money via the elimination of asset allocation with long idle periods between DR operations. • **Live Sync data replication for Warm Recovery** in cloud – Automate the creation of cloud instance and replication of on-premises VMs to Amazon EC2 instances on a periodic cycle basis more frequently than backups. Reduces recovery time to the Cloud. • **VM Restore & Convert** – convert VMware and Hyper-V VMs into Amazon EC2 instances on-demand with data intact. This data transformation automation reduces time and complexity costs. • **Automate Failover and Failback of VMs** - from on-premises VMware to public cloud (AWS). • **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 links to the cloud provider for best performance (AWS Direct Connect).

Amazon Cheat Sheet

Looking to get started quickly? Refer to Cloud Feature Support for Amazon for best practices and instructions.

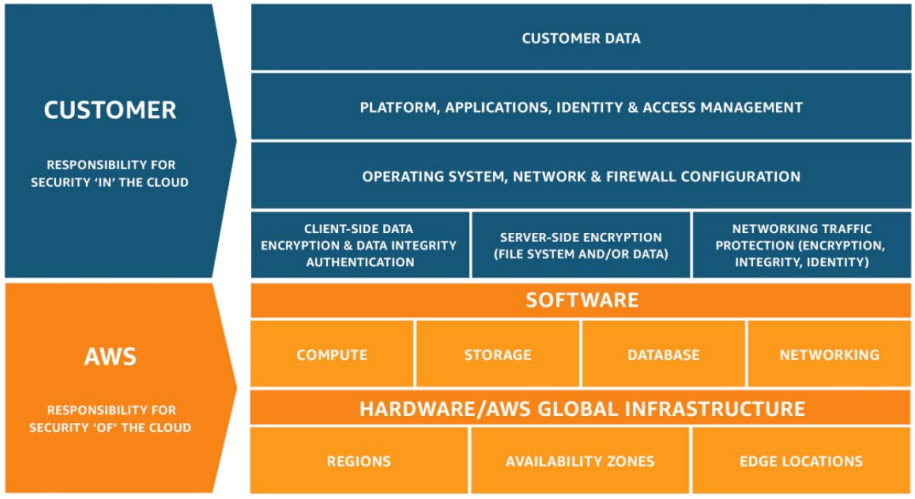
Identity & Access Management 	• JSON Templates for IAM Role Definition and User Permissions
	  • amazon_permission_backup_restore.json
	  • amazon_restricted_role_permissions.json
	  • amazon_s3.json, amazon_permission_conversion.json (vmimport)

Amazon S3 	• Amazon S3 Support, Configuring Cloud Storage, Combined Storage Tiers • Amazon S3 - Access & Secret Access Keys • Amazon S3 - AWS IAM Role Policy • Amazon S3 - AWS STS Assume Role. STS Assume Role with IAM Role Policy
EC2 Protection 	• Protecting Amazon EC2 Instances • Supported agents (for Amazon native snapshot protection) • Application-Aware Backups for the Virtual Server Agent
Container Protection 	• Protecting Elastic Kubernetes Service (containers, persistent volumes, secrets) • Auto-protecting containers by label selector • Application and data migration (cross-cluster, cross-region)
RDS Protection 	• Amazon RDS • Using native snapshots Aurora, MariaDB, SQL Server, MySQL, Oracle, Postgres • Using export/dump MySQL, Oracle, Postgres
EFS Protection 	• Amazon DocumentDB protection • Amazon DynamoDB protection • Amazon Redshift protection
S3 Protection 	• Amazon S3 • Best Practices • Configuring Amazon S3 Bucket Backups Using the IAM Role
FSx EFS Protection 	• Amazon FSx for Windows File Server • AWS EFS (Amazon Elastic File System)
VM Migration 	• Cross-Hypervisor Restores (VM Conversion) • Converting from Azure to Amazon, VMware or Hyper-V to Amazon • Migrating Data to Amazon S3 Using Snowball
Disaster Recovery 	• Live Sync Replication for Amazon • Live Sync Virtual Machine Replication and Application-Aware Backups • Replication for Virtual Machines (Command Center)
App migration 	• Oracle Database Application Migration to an Amazon EC2 Instance • Oracle Database Application Migration to an Amazon RDS Database • Configuring SQL Database Application Migration to an Amazon EC2 Instance • Postgres database migration, and MySQL database migration

Shared Responsibility Model

This document is intended to inform and instruct Commvault customers on how to comply with their **Security 'IN' the Cloud** responsibility as detailed by the AWS Shared Responsibility Model aws.amazon.com/compliance/shared-responsibility-model.

AWS will ensure the security of the region, availability zones (physical sites) and underlying virtualized infrastructure (compute, storage, database, and network). AWS takes their security responsibility seriously with external validation to the leading industry regulations as found at aws.amazon.com/compliance/programs.



This document will help **you** secure your Commvault and AWS infrastructure and provide ***Recovery Readiness*** for your AWS hosted applications and **data**. Applying Commvault to your AWS data landscape you will gain recoverability for the following AWS services both within and across region.

- Compute
  - Amazon Elastic Compute 2 (EC2) instances
  - Amazon Outposts – EC2 instances
  - Amazon Elastic Kubernetes Service (EKS)
  - VMware Cloud on AWS
- Storage
  - Amazon Elastic Block Store (EBS) volumes
  - Amazon Outposts – EBS volumes
  - Amazon Elastic File System (EFS) exports
  - Amazon Elastic FSx for Windows shares
  - Amazon Simple Storage Service (S3) – as source, as target (all storage classes)
  - Amazon S3 Glacier (as target)
  - AWS Storage Gateway (as source)
  - Amazon Snowball Edge, Snowmobile, and Snowcone (as migration device)
- Database
  - Amazon RDS – Oracle, MySQL, PostgreSQL, MS SQL
  - Amazon Aurora
  - Amazon DocumentDB / MongoDB
  - Amazon DocumentDB

- Amazon Redshift

Commvault Platform Design Principles for Cloud

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

Native Cloud Connectivity

The Cloud Library feature of a MediaAgent is the native integration within the Commvault® platform that directly communicates with object storage such as Amazon S3., Amazon S3-IA, One Zone-IA, Intelligent-Tiering, Reduced Redundancy Storage, Glacier. and Glacier Deep Archive. Commvault does not require translation devices, gateways, hardware appliances or Virtual Tape Libraries (VTLs). For more details on AWS storage classes see Amazon S3 Storage Classes.

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

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

For more information on all the supported vendors, please refer to this comprehensive list located at - Supported Cloud Storage - Amazon S3 Vendors.

Scalability

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

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

- Deduplication Building Blocks
- Client-side Deduplication
- Access Node

Deduplication Building Blocks

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

Commvault MediaAgent that will perform the data movement (both ingestion and storage), compression and deduplication.

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

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

More on adaptable scaling may be found in – Architecture Sizing.

Client-side Deduplication

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

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

Access Node

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

Akin to API-based protection for on-premises hypervisors, the Commvault Access Node interfaces directly with Cloud APIs available from the hypervisor layer of AWS to perform protection and management operations of cloud instances. The Access Node not only manages operations such as EBS snapshot creation and orchestration but can also performs automatic instance identification and selective data reads (Change Block Tracking). The Access Node further performs any data format conversions and enables disaster recovery operations for instances to cloud platforms. Working together with the MediaAgent (data mover), the Access Node offers enhanced protection and management of cloud workloads.

*For a complete list of supported Amazon Access Node capabilities please review the online Feature Comparison Matrix - Amazon.

Commvault recommends installation of the Virtual Server Agent package on your MediaAgent initially, to provide a single consolidated **Access Node** for protection and data management. As your cloud protection needs grow, you may look to utilize Access-node auto-scaling to automatically create and destroy Access Nodes on-demand.

Design for Recovery

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

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

Crash Consistency Versus Application Consistency

While crash-consistency within a recovery point may be enough for a file-based dataset or cloud-native applications developed on Amazon EC2 instances, it is not generally appropriate for classic applications such as Microsoft SQL Server or Oracle Database. Database instances need to be quiesced to ensure the database is valid at the time of backup, and recoverable when required. Commvault® software supports both crash- and application-consistent backups, providing flexibility in your design while assuring instance recoverability coupled with application recovery to a specific point in time. Not only are the most common types of applications covered, but a wide variety of classic applications and cloud applications are supported. For a complete list of updated application support please review the online documentation: Backup and Restore Agents.

Storage-level Replication Version Discrete Independent Copies

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

While Commvault® software can support a replicated cloud library model, in which the secondary storage location for backups is replicated using the Cloud vendors storage-based replication tools (see Configuring Replication for Cloud Storage), we recommend that you consider using the Commvault software to create an independent copy of your data, either to another region, or another cloud provider, or back to an on-premises infrastructure to address broader risks. Deduplication is also vital as part of the latter option and this ensures that Commvault software can minimize the cross-region and cross-provider copy time and egress costs by ensuring only the unique changed blocks are transferred outside the cloud provider network.

This recommendation not only ensures recoverability to multiple points in time, it further manages the cost and risk through the assurance that the data is independent of the platform, provider and ensures that different SLAs for protection and retention can be maintained for different classes of data.

Deciding What to Protect

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

Designed for Cloud Efficiency

As already discussed, the ability to provide compression and deduplication for both data to and data in the

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

A common consideration is to utilize multiple tiers of storage for data as the service life of that data reduces. This has been a common practice on-premises and the Commvault platform extends this capability to cloud platforms. By having native integration to primary object storage targets such as Amazon S3 Standard and also having native access to more cost-effective tiers such as Amazon S3 Standard-Infrequent Access (S3-IA), Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA) and Amazon Glacier**.**, data lifecycle management can be performed within the Cloud. For example, it is not uncommon to see Amazon S3 Standard being used as the secondary copy for short-term retention followed by Amazon Glacier being used for long-term retention copies. Having a data management platform that can utilize Service Level Agreement (SLA) policies to orchestrate the data movement and be aware of the location of data for recall and disposition becomes a valuable quality in gaining cloud efficiency.

Storing Data Efficiently

AWS provides a broad selection of Simple Storage Service (S3) **storage classes** to meet your data retention, access frequency and budgetary requirements. It is important to understand the underlying capabilities of these storage classes before placing your data. The first consideration is the **access frequency** or temperature of data:

When storing data within one or more S3 storage classes, consider the following characteristics:

- How often will the data be accessed?
- What portion of the data will be read? (partial, full)
- What is the business expectation for data retrieval time? (**Fast**, **Slow**)
- What is the business expectation for data storage cost, and data retrieval cost?
- What is the business expectation of data durability? *If the data cannot be retrieved – what is the impact?*

Amazon S3 Storage Classes		
Hot	**Warm**	**Cold**
• Standard • Reduced Redundancy	• Infrequent Access (S3 Standard-IA) • One Zone Infrequent Access (S3 One Zone-IA)	• Glacier • Deep Archive

Source: Amazon S3 Storage Classes

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 **expedited 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.

Amazon S3 Storage Classes – for Backup & Archive				
Feature	**Hot**	**Warm**	**Cold**	**Cold**
Storage Class	Standard	Infrequent Access	Glacier	Deep Archive
Annual Store cost (1 TB)	$282.60	$153.60	$49.20	$12.12
Minimum retention	None	30 days	90 days	180 days
First byte latency	milliseconds	milliseconds	mins - hours	12 hours+
Annual Recall cost (1 TB)	$8.64	$122.88	• $122.68 (standard) • $368.64 (expedited) • $34.32 (bulk)	• $245.88 (standard) • n/a • $30.72

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

Amazon S3 Storage Classes to Use-Case Mapping		
Hot	**Warm**	**Cold**
• Cloud native **primary** application (not backup data)	• Backup data (30-90-day retention, infrequent recoveries)	• Long-Term Retention Backup / Archive data (90 day – 1-year retention) • Long-term regulatory archives (1 year+ retention)
• 5% of total org. data	• 10-15% of total org. data	• 70-80% of total org. data

Commvault sees most backup data being **stored** and **infrequently retrieved**, S3.-IA is most cost-effective option for backup data

Architectural best practices for storage class selection

Do's

- **Do** place near-term backup data, used for daily operational recovery in S3.-IA for reduced cost, and low latency access.
- **Do** As data ages beyond operational retention (typically 30-90 days), perform an auxiliary copy for long-term retention into a Commvault Combined storage class that combines warm (S3-IA) and cool (Glacier, Deep Archive) storage classes.
- **Do** utilize a Cloud Media Agent with Cloud MediaAgent Power Management to minimize the running time of Cloud MediaAgent.
- **Do** use Automatic Synthetic Full schedules, to minimize the amount of retrieval activity from Cloud Libraries (relevant when the Primary backup copy is stored within a Cloud library – due to recall cost).
- **Do** consider the cost and latency of recovery when selecting your **cool/cold** storage class, if expedited recovery is anticipated select Glacier.

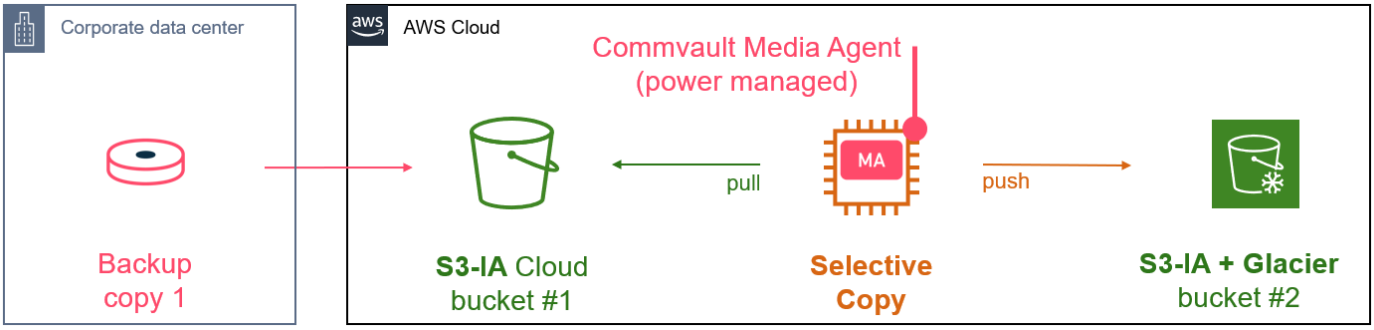
Don'ts

- Do not use S3 Standard for backup and archive data, S3-IA is more cost effective, even with increased minimum retention time.[1]
- Do not utilize extended retention rules on Cloud Copies, instead perform an auxiliary copy to tier data to lower-cost storage class in a separate bucket.
- Do not utilize an on-premise MediaAgent to perform Auxiliary copy as Cloud egress costs (S3 Pricing) will be incurred.
- Do not utilize S3 Standard unless backup data will be frequently retrieved, Commvault does not see a requirement for S3 Standard in most backup use-cases.

[1] In unique instances where frequent, repeated access to primary data copy occurs (Content indexing, dev/test automated recovery) using of **S3 Standard** for primary copy may be more cost effective.

Cloud as your data vault

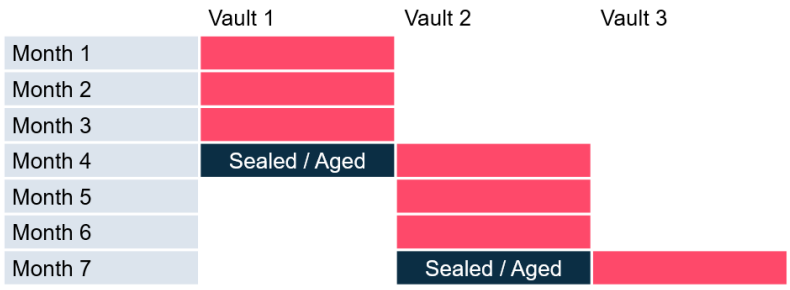
Commvault can be used to implement your air gapped data **vault** for immutable data copies held for recovery from organization wide events that target your Primary data and Secondary (backup) copies.



Important considerations for a **data vault deployment** include:

- Data within the vault is write-once read many (WORM) or immutable
- Vaults should be created to retain half of the total desired retention age (e.g. a 90-day vault requires 2 x 45-day retention buckets)
- Commvault will utilize macro pruning to age out vault data once the retention period has been reached for an entire vault vs. micro-managing storage in the ultra-low-cost storage classes.
- Deduplication is supported and recommended in data vaults.
- Deduplication Databases should be sealed upon creation of a new vault store.
- Retention will always incur at least one additional vault store to handle data aging (see below).

An example of a **180-day retention** data vault with deduplication seals at 3 months. There will always be three (3) copies of the dataset at any time. Each vault represents a full backup of the archival content, however the ultra-low cost of S3.-IA + Glacier./Deep Archive means the storage cost is negligible.

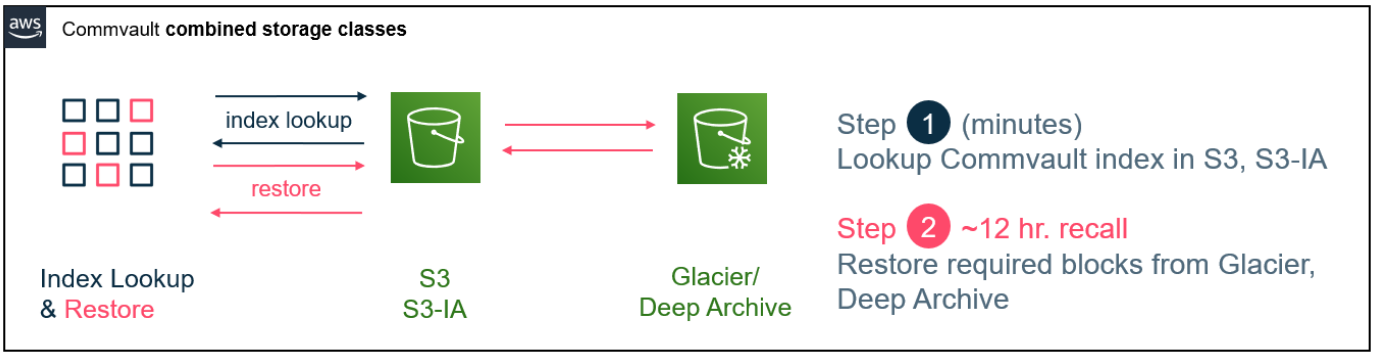


A note on S3 Intelligent-Tiering

Note: You will note that the S3 Intelligent-Tiering storage class is not represented, this is due to the fact that Intelligent Tiering makes data placement decisions based on access frequency. In Commvault, data is split into **warm indexing data** that allows for locating data chunks distributed in large data vaults, and **cool/cold stored data**.

Commvault does not recommend the use of S3 Intelligent-Tiering, but instead advocates the use of **Commvault combined storage classes** (more below) to ensure you can efficiently locate and surgically recall data in minimal timeframes.

The benefit is using **Commvault combined storage tiers**, means that small, warm indexes are kept in low-latency storage classes, available within millisecond first byte latency, meaning a surgical restore for cold/cool data occurs within minimal delay, while leveraging the low-cost of cooler storage classes.



It is important to understand the constraints on each of these storage service classes before placing data. Commvault has excluded the **Reduced Redundancy** and **One Zone-IA** services from analysis as these services are designed for

non-critical or reproducible data. Amazon recommends that customers utilize S3 Standard as a more cost-effective, durable storage location.

▲ For durability, RRS objects have an average annual expected loss of 0.01 percent of objects. If an RRS object is lost, when requests are made to that object, Amazon S3 returns a 405 error.(source)

A note on combined storage tiers

Commvault provides intelligent combined storage classes that place data in the most approach S3 storage class to support timely recovery. Combined storage classes include:

- S3 Standard/Glacier.
- S3 Standard-IA/Glacier.
- S3 One Zone-IA/Glacier.
- S3 Intelligent Tiering/Glacier.
- S3 Standard/Deep Archive
- S3 Standard-IA/Deep Archive
- S3 One Zone-IA/Deep Archive
- S3 Intelligent Tiering/Deep Archive

Commvault recommends that **combined storage classes** be used primarily in archive or long-term retention backup data use-cases. Therefore, Commvault recommends:

- Avoid combined classes using **S3 Standard**, utilize **S3 Standard-IA** for equivalent, lower cost stores.
- Avoid combined classes using **One Zone-IA**, which represents a risk of data loss. Warm tiers will be used to hold Commvault indexes and must be selected for high durability and availability.
- Avoid combined classes using **Intelligent Tiering**, let Commvault place data between warm and cold classes for best recovery outcome.

Commvault combined storage tiers are distinct from the AWS **S3 Intelligent-Tiering >** storage class which automatically migrates content from Standard, Standard-IA, or S3 One Zone-IA to Glacier**.** or Deep Archive based on observed I/O patterns.

Commvault recommends using a combined storage class without Intelligent Tiering to achieve more predictable recovery times.

Architectural best practices for S3 usage

Selecting the appropriate S3 storage class is the key to designing for recovery. Balance your monthly (predictable) storage costs, and retrieval costs (unpredictable) to protect your business. Commvault recommends the following:

Do's

- **Do** use Commvault software-based deduplication with your S3 storage libraries to save stored data cost
- **Do** use a dedicated S3 bucket and storage class for Commvault data, do not share with other workloads
- **Do** use combined storage classes only for cold data (long-term retention or archive)
- **Do** use a dedicated S3 bucket and storage class for Commvault backup data, and another for archive data
- **Do** utilize different AWS credentials for each data type – backup, archive, data vaults (WORM)
- **Do** use Commvault 'combined' storage tiers when placing deduplicated data in Glacier or DeepArchive
- **Do** use Commvault auxiliary copies to migrate data between storage classes when business value changes
- **Do** use Commvault MediaAgent power management to automatically power-down data mover infrastructure
- **Do** seal your **cold copy archives** at (retention time / 2) to compartmentalize archival stores into **data vaults**

Don’ts

- Do **not** use S3 Object Lifecycle management to migrate data between storage classes, as this effects restore predictability
- Do **not** use Glacier./Deep Archive storage directly, instead use Commvault combined storage classes

Cloud Power Management

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

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

See Overview of Cloud MediaAgent Power Management for more information.

Automation

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

- Programmatic Data Management
- Workload Auto-detection And Auto-protection
- Self-service Access and Restore

Programmatic Data Management

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

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

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

Workload Auto-detection And Auto-protection

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

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

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

Self-service access and restore

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

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

Architecture Sizing

Because of the nature of Amazon EC2 and Amazon EBS, Commvault infrastructure components can be elastic to fit your environment's needs. The tables below are a suggestion to handle the upper limits of each component.

Day one sizing

You can simply extend your on-premises Commvault installation into AWS with no requirement for cloud-based infrastructure (no proxies, no long-running EC2 instances). However, if you are simply looking to **Get Started** with Commvault in Cloud, perhaps a Proof of Concept (PoC) or initial set of dev/test workloads – utilize a single server.

Commvault AWS Getting Started Specifications	
Server usage	CommServe, MediaAgent, and Virtual Server Agent components.
Server sizing limits	25 x servers -or- 50 x laptops Up to 32 TB of client data (before deduplication/compression) Up to 50 TB of backup data (post deduplication/compression)
Instance	1 x t3a.2xlarge (EBS optimized) (8 vCPU, 32 GB RAM)
OS or Software Disk	1 x 200 GB (General Purpose SSD – gp2) volume
Deduplication Disk (DDB)	1 x 200 GB (General Purpose SSD – gp2) volume
Index Cache Disk	1 x 400GB (General Purpose SSD – gp2) volume
Operating System	Windows Server 2019, 2016

Source: Hardware Specifications for the CommServe Server, Deduplication Mode, Hardware Specifications for Virtual Server Agent

Note: Commvault has sized an 80% File/VM, 20% Database workload retained as 30 x daily backups, 8 x weekly, 9 x monthlies and 1 x yearly backup. Your workload and retention may allow workloads within this configuration.

The indicative **annual costs** for an on-demand instance of this configuration in US East (N. Virginia) is:

- EC2 instance cost 122.57 x 12 = USD $3,9244.48
- GP2 storage cost 80 x 12 = USD $960.00

TOTAL: 4,884.48

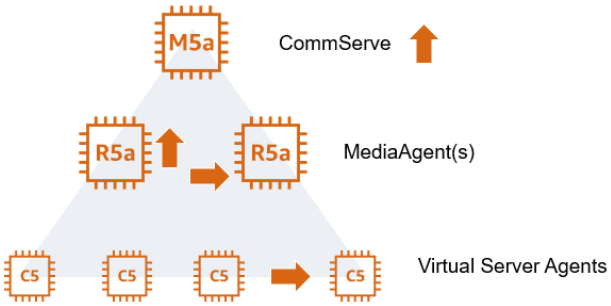
You can download the Commvault software and get started at www.commvault.com/trials with a 30-day trial.

As your CommServe scales, It is likely the MediaAgent and Virtual Server Agent components will be migrated onto dedicated infrastructure that can be powered down when not in use. Power management utilizes a **cloud controller** process to power-up/down MediaAgents when specific actions require them (backup, restore, aux copies, and pruning). Commvault can automatically create and destroy Access Nodes in response to concurrent backup load.

Sizing Guidance

Start with the smallest category (Small or Medium) in the tables below for Amazon EC2 instance size and Amazon EBS volume sizes to fit your data volume. Upgrade them as needed to meet your environment's requirements (e.g. start with an Extra-Small MediaAgent to begin with, then upgrade to a Small MediaAgent as you onboard more client computers).

Consult the AWS Pricing Calculator for the best pricing in your AWS region, Commvault has provided two recommended EC2 instance sizes below for each component. Commvault has selected instance for best baseline performance and cost.



Commvault expects you will scale in three dimensions:

- A single **CommServe** will be scaled vertically to a limit of 20,000 instances protected.
- One or more **Media Agents** will be scaled vertically to a maximum of 1000TB of managed data, then scaled horizontally to a maximum of 4000TB per MediaAgent grid.
- **Virtual Server Agents** will be scaled horizontally with smaller instance types for cost control. Vertical scaling will only be required when the size of a given EC2 instance exceeds desired backup window.

CommServe Sizing

For CommServe servers, adjust the Amazon EC2 instance size upwards when CPU and RAM loads become consistently high; add more space to the Amazon EBS volumes as needed to accommodate the size of the CommServe database as you add more client computers and jobs to your CommCell. CPU and RAM load can be monitored in the AWS Management Console, or in the CommCell Command Console using the Infrastructure Load Report (see Commvault Infrastructure Load Report for details).

At the time of writing (March 2020), the following are the supported Instance types for the CommServe Marketplace image (Ami Id: ami-019dd1467ee3f13f6). Commvault **recommended instances** are highlighted in **raspberry**.

awsmarketplace **CommServe Marketplace Instance Sizes**		
m5a (xlarge-4xlarge, 12xlarge, 24xlarge)	m5 (large-4xlarge, 12xlarge, 24xlarge)	c4 (large-8xlarge)
m5d (large-4xlarge, 12xlarge, 24xlarge)	m4 (large-16xlarge)	p2 (xlarge-16xlarge)

Current recommendation in AWS Marketplace is **c3.2xlarge** which equates to a c4.2xlarge, or **m5a.xlarge** in current generation EC2 instance types.

CommServe Specifications

These specifications are based on Hardware Specifications for the CommServe Server.

Commvault testing has identified Compute Optimized Instances as the most applicable EC2 Instance Type.

Commvault recommends deploying the smallest instance that will service your workload and utilize AWS - EC2 Changing the Instance Type to resize your CommServe when required.

AWS CommServe Specifications			
Small	**Medium**	**Large**	**Extra large**
25 instances or 50 laptops	1000 instances or 5000 laptops	2500 instances or 10,000 laptops	20,000 instances or 50,000 laptops
m5a.xlarge (4 vCPU, 16GB RAM)	m5a.2xlarge (4 vCPU, 16GB RAM)	m5a.4xlarge (16 vCPU, 64GB RAM)	r5a.4xlarge (16 vCPU, 128GB RAM)
100 GB EBS “General Purpose SSD” (gp2) volume for CS Software & CSDB	150 GB EBS “General Purpose SSD” (gp2) volume for CS Software & CSDB	300 GB EBS “General Purpose SSD” (gp2) volume for CS Software & CSDB	300 GB EBS “General Purpose SSD” (gp2) for CS Software & CSDB
Windows Server 2019, 2016, 2012 R2			

Media Agent Sizing

For MediaAgent servers, monitor CPU and RAM utilization and network I/O on the Amazon EC2 instance using AWS native tools such as the AWS Management Console. Upgrade to a larger Amazon EC2 instance when you start encountering bottlenecks in any of these resources. Consult the CommCell Health Report to view the performance of the Deduplication DataBase (DDB). If the DDB is experiencing degraded performance, consider adding more capacity to the Amazon EBS volume holding the DDB or move the DDB to a higher tier of Amazon EBS volume. For more details about the Health Report view in the CommCell Command Console, see Commvault Health Report Overview. The Health Report is downloaded from the Commvault Store.

Avoid using Amazon EC2 instance types with local NVMe SSD volumes (such as the R5d, M5d. C5d, i3, i3en instance types), as these local NVMe volumes will not retain their data if the Amazon EC2 instance is powered off or terminated, which makes them unsuitable choices for the Index Cache or DDB.

If choosing an Amazon EC2 type other than ones listed in the tables below, choose an Amazon EC2 instance type which is “EBS optimized” to ensure good performance with the Index Cache and DDB. Review both the ‘burst’ EBS bandwidth offered for thirty (30) minutes per day, and the baseline EBS bandwidth guaranteed.

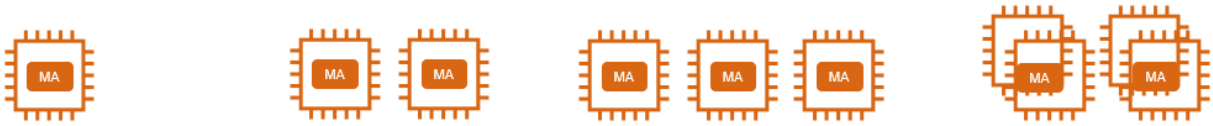
MediaAgent Grids

These specifications are based on Hardware Specifications for Deduplication Mode.

Commvault testing has identified Compute Optimized Instances as the most applicable EC2 Instance Type for MediaAgent workloads, due to compute workload calculating DDB hashes and high-IO DDB activity.

Commvault recommends deploying the smallest instance that will service your workload and utilize AWS - EC2 Changing the Instance Type to resize your MediaAgent when required. A single MediaAgent can service up to 1 PB of

stored backup data before requiring an additional MediaAgent be added. MediaAgents may also be deployed in resilient grids which spread workload across multiple nodes and scale to larger protected data volumes (see below).



1 Node	2 Node	3 Node	4 Node
	AZ outage tolerant	AZ outage tolerant	AZ outage tolerant
50 – 1000 BETB	100 – 2000 BETB	150 – 3000 BETB	1200 – 4000 BETB
32 – 670 FETB	60 – 1396 FETB	102 – 2093 FETB	803 – 2951 FETB

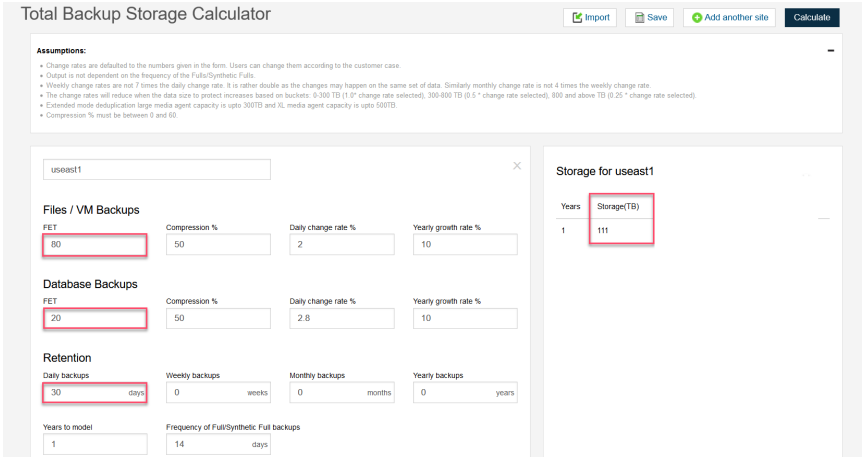
FETB – Front-end TB (data size measured at the client)
BETB – Back-end TB (data size measured in all backup copies)

Source: Hardware Specifications for Deduplication Mode

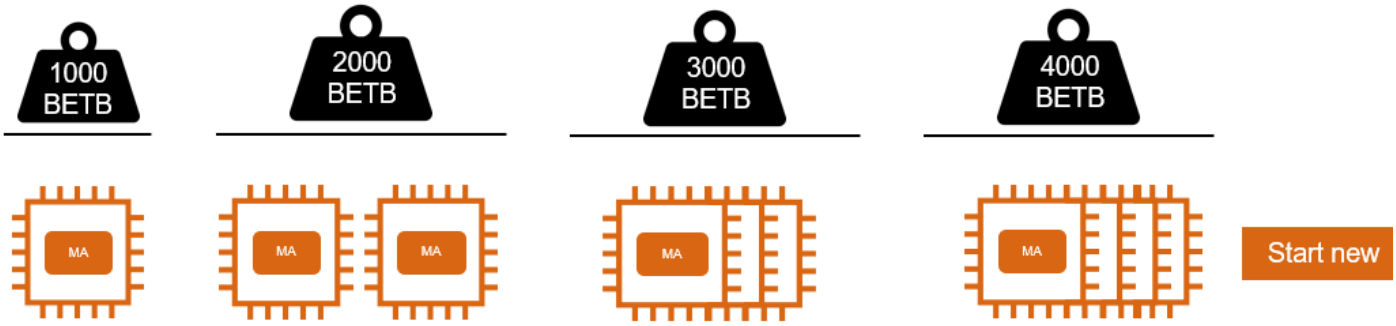
For reference, Commvault has sized the MediaAgent grids based on the following standard backup retention:

- 80% VM/File data
- 20% Database data
- Daily change rate 2%
- Compression rate 50%
- Daily backups are retained for 30 days

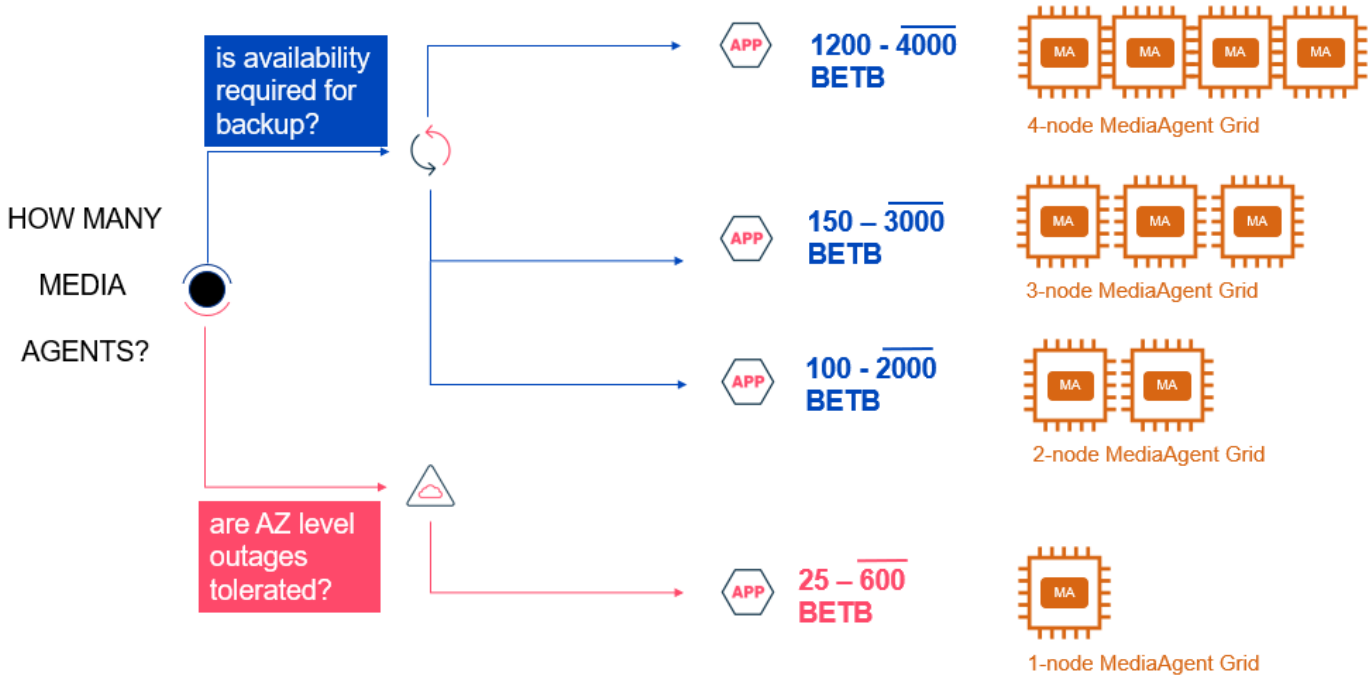
This configuration results in roughly a **1.0-1.2x increase** from front-end to back-end storage, after deduplication and compression is applied (see below for screenshot from Commvault Backup Storage Calculator).



Grid based configurations may be scaled both vertically (increasing CPU, RAM, and Disk) and horizontally (adding additional nodes of same specification). Each node configuration has a maximum amount of backend data that can be managed

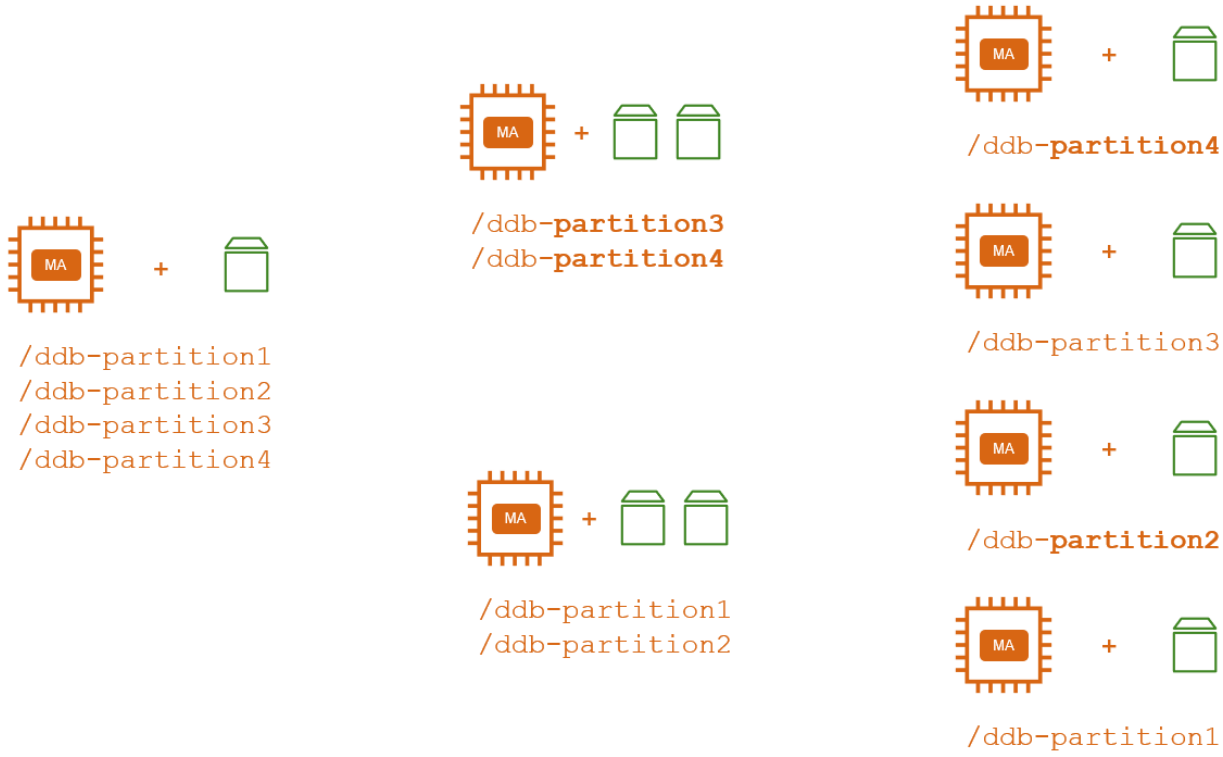


Therefore, when you are sizing a MediaAgent you need to consider the following characteristics



If you plan on scaling your environment to 2-node, 3-node or 4-node MediaAgent configurations, you can plan to ensure minimal cost scaling in future. Employ these architectural best practices to scale efficiently:

- Plan your MediaAgent deployment per Availability Zone
- Deploy a **deduplication partition** for the total number of nodes you want to support at maximum scale (see below0
- Each DDB partition may be deployed in a directory or dedicated volume on a single MediaAgent (day 1)
- When the MediaAgent exceeds its maximum FETB or BETB measurement, move the DDB partition (or volume) to a newly provisioned MediaAgent (see below)



MediaAgent Scalability

The following table outlines the total **back-end storage** that each configuration can support. Items are **green** are the minimum starting footprint for a given grid configuration. It is permitted to upgrade from 1-node, to 2-node, and finally 4-node configurations (as depicted in ddb-partition diagram above).

Total backup data per configuration (in back-end TB)			
Grid Configuration	**Extra Small**	**Small**	**Medium**
1 Node	**50**	100	300
2 Node	**100**	200	600
3 Node	**150**	300	900
4 Node			**1200**

Source: Hardware Specifications for Deduplication Mode

There are several factors to consider when determining your MediaAgent configuration. Ultimately you need to consider:

- The **availability** of backup services if a single availability zone experiences an outage. MediaAgents are required to perform backup and recovery actions.
    - Use of multi-node configuration (2-node, 3-node, 4-node) if you require resiliency from AZ outages.
- The **total volume** of client data, and corresponding backup data you plan to manage within the target availability zone
    - Build for your **immediate needs** and then scale vertically and/or horizontally to gain additional capacity and throughput. **Commvault recommends horizontal scaling (more below)**

If you do not know your **backup data** size, multiple your client data size by 1.6x for an estimate of back-end data for a 90-day retention period (see calculator above)

A significant consideration when building infrastructure in Cloud is cost. Below is a break-down of the on-demand cost of each of the MediaAgent configuration. Costs are shown for on-demand instances, 24x7 operation in U.S East (N. Virginia) as at March 2020. Costs may be reduced utilizing reserved instances, upfront payment, Linux instance and cloud power management.

Annual Cost / max. BETB – Windows MediaAgent Instances			
Grid Configuration	**Extra Small**	**Small**	**Medium**
Instance	r5a.large	r5a.xlarge	m5a.2xlarge
1 Node	$2,756 / 50	**$4,912 / 100**	**$8,637 / 300**
2 Node	$5,512 / 100	$9,823 / 200	$17,274 / 600
3 Node	$8,267 /150	$14,735 / 300	$25,911 / 900
4 Node			$34,548 / 1200

Source: AWS Pricing Calculator

Items shaded in GREY show equivalent volumes of data (100, 600, and 2000BETB each), the least cost alternative is highlighted in **green**, you will note multiple larger nodes is typically lower cost than a vertically scale instance.

Media Agent costs may be reduced further, by utilizing the Linux operating system which is typically offered at a reduced cost to Microsoft windows licensed instances. ***As shown by the table below, using Linux will yield an average annual saving of 34% on MediaAgent cost***.

Annual Cost / max. BETB – Linux MediaAgent Instances			
Grid Configuration	**Extra Small**	**Small**	**Medium**
Instance	r5a.large	r5a.xlarge	m5a.2xlarge
1 Node	$1,950 / 50	**$3,300 / 100**	**$5,413 / 300**
2 Node	$3,900 / 100	$6,600 / 200	$10,827 / 600
3 Node	$5,850 / 150	$9,899 / 300	$16,240 / 900
4 Node			$21,654 / 1200

Source: AWS Pricing Calculator

Architectural best practices for MediaAgent sizing

Do's

- Do perform sizing on a **per availability zone** basis
- Do deploy MediaAgents across multiple Availability Zones (within a region) for resilience from AZ outage. (Note: if a MediaAgent is down, blocks previously deduplicated by that MediaAgent will be re-written during backup by one of the available MediaAgents)
- Do start with the smallest MediaAgent that will meet your immediate needs (3-6 months)
- Do pre-populate your MediaAgent with DDB partitions (or volumes) for the total number of nodes you plan to support within the grid.
- Do distribute DDB partition data (or volumes) across deployed nodes to balance deduplication workload across the grid.
- Do scale horizontally by adding additional MediaAgents when configuration exceeds Commvault MediaAgent sizing specifications
- [Optionally] Do scale your MediaAgent(s) vertically if you simply need additional client data protected, but protection operations are completing within your business accepted backup window.

Don't

- Do not mix MediaAgent Instance Types within a Grid (i.e. a Medium and an Extra-Small)
- Do not distribute MediaAgents across regions. You will require a grid per region, per availability zone.

MediaAgent Specifications

MediaAgent EC2 Specifications		
Extra small	**Small**	**Medium**
50 BETB (max.)	100 BETB	300 BETB
r5a.large (2 vCPU, 16 GB RAM)	r5a.xlarge (4 vCPU, 32 GB RAM)	m5a.2xlarge (8 vCPU, 32 GB RAM)
1 x 200GB EBS “General Purpose SSD” (gp2) volume for Commvault software.	1 x 300GB EBS “General Purpose SSD” (gp2) volume for Commvault software.	1 x 400GB EBS “General Purpose SSD” (gp2) volume for Commvault software.
1 x 200 GB EBS “General Purpose SSD” (gp2) volume for DDB	1 x 400 GB EBS “General Purpose SSD” (gp2) volume for DDB	1 x 600 GB EBS “General Purpose SSD” (gp2) volume for DDB
1 x 400 GB EBS “General Purpose SSD” (gp2) volume for Index Cache	1 x 400 GB EBS “General Purpose SSD” (gp2) volume for Index Cache	1 x 1 TB EBS “General Purpose SSD” (gp2) volume for Index Cache
• Linux / AWS Marketplace (Red Hat Enterprise Linux 7.6) **or** Windows Server 2019, 2016		

Source: Hardware Specification for Virtual Server Agent

BETB refers to a Back-end Terabyte, a TB measured within the Commvault backup system, once processed.

Important: EBS-optimized instances are recommended as they provide dedicated network bandwidth for EBS volumes, improving deduplication and Index Cache performance and freeing up bandwidth to send/receive from clients, other MediaAgents, and Amazon S3 endpoints.

r5a instance types offer higher network bandwidth and lower costs than original r5/m4 instance types.

The BET and FET sizing for maximum capacity are based on a 512 KB deduplication block size, which is the default when protecting Cloud workloads to Cloud libraries.

It should be noted that Commvault requires **SSD disk** (gp2) for the deduplication database. There is no requirement to increase the IOPS specification of this volume to io1 or provisioned IOPS configurations.

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

At the time of writing (June 2020), the supported EC2 instance types for a MediaAgent (Ami id: ami-01bdca37c0d9d4618) are as follows. Items in raspberry are **recommended for Production**.

Note: MediaAgent AMI includes the Virtual Server Agent software, making the component an **Amazon Access Node**, capable of protection and data deduplication.

awsmarketplace MediaAgent Instance Sizes		
• t2 (micro-2xlarge)	• f1 (2xlarge-16xlarge)	• r4 (large-16xlarge)
• t3 (micro-2xlarge)	• g2 (8xlarge)	• x1 (16xlarge, 32xlarge)
• m5a (large-4xlarge, 12xlarge, 24xlarge)	• g3s (xlarge-16xlarge)	• x1e (xlarge-32xlarge)
• m5d (large-4xlarge, 12xlarge, 24xlarge)	• p2 (xlarge-16xlarge)	• z1d (large-12xlarge)
• m5 (large-4xlarge, 12xlarge, 24xlarge)	• p3 (2xlarge-16xlarge)	• d2 (xlarge-8xlarge)
• m4 (large-16xlarge)	• p3dn.24xlarge	• i2 (xlarge-8xlarge)
• c5n (large-18xlarge)	• r5a (large-4xlarge, 12x large, 24xlarge)	• h1 (2xlarge-16xlarge)
• c5d (large-9xlarge, 18xlarge)	• r5d (large-4xlarge, 12xlarge, 24xlarge)	• i3 (large-16xlarge, metal)
• c5 (large-9xlarge, 18xlarge)	• r5 (large-4xlarge, 12xlarge, 24xlarge)	
• c4 (large-8xlarge)	• r4 (large-16xlarge)	

Access Node Sizing

For standalone Access Nodes (without the MediaAgent package), monitor the total **Front-End Terabytes (FETB)** protected by a single Access Node and either scale-up the specification of an existing Access Node, or add an additional Access Node for more backup throughput to your MediaAgents. Remember that each EC2 instance provides addition EBS volume bandwidth and network bandwidth, so scaling horizontally with smaller Access Nodes is recommended over vertical scaling.

Commvault defaults auto-scaled Access Nodes to a **c4.large** instance (default) for optimal cost and performance

Horizontal scaling

Most small Virtual Machines will be protected with a c4.large Access Node. More backup parallelism may be achieved with more concurrent streams and EBS volume mounts, by adding additional Access Nodes.

For **larger VMs** (> 15TB), with multi-TB disks or multi-EBS volumes, scale-up your c4.large to gain more EBS IOPS and associated network bandwidth to complete EBS streaming activity within backup window.

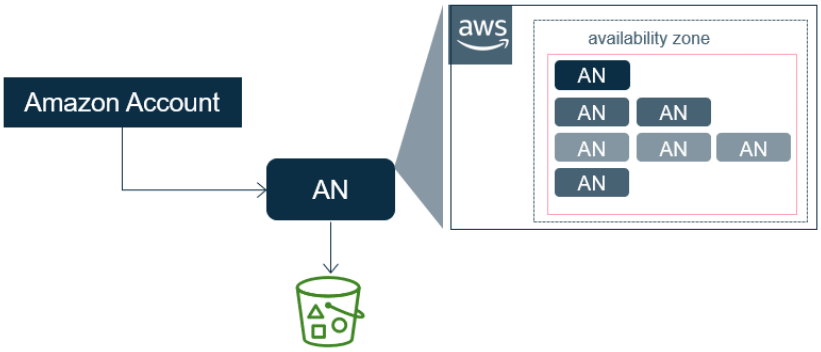
> Amazon has announced released new ARM-based Gravitron2 instances that represent a significant cost reduction from traditional Intel Chipsets (**30% reduction from C4.large**), see Announcing New Amazon EC2 M6g, C6g, and R6g Instances Powered by Next-Generation Arm-based AWS Graviton2 Processors
> **Commvault recommends using C6g Access Nodes as a low-cost auto-scaling Access Node.**

Automatic Access Node Scaling

As of Commvault v11 Feature Release 11.20 Commvault has the ability to perform Automatic Scaling for Amazon Access Nodes during backup activities. Automatic scaling allows further cost optimization of your Amazon EC2 protection infrastructure by only have **Access Nodes** running during protection activities.

Automatic Access Node scaling performs as follows:

- Commvault assesses the total size of the protection activities (number of hosts, TB to protect, estimated backup throughput).
- Commvault automatically spawns the required number of Access Nodes to compete the backup.
- Access Nodes are Linux-based and run same release as the CommServe.
- Access Nodes are automatically registered with Commvault for the period of the backup.
- After protection activities completes, the nodes are shutdown, and then deleted.



- For additional information refer to:
- Getting Started with Automatic Scaling for Amazon Access Nodes
- How to Calculate the Number of Access Nodes for Automatic Scaling

Access Node Specifications

These specifications are based on **Hardware Specifications for Virtual Server Agent >**, and refer to stand-alone Access Nodes that do not contain the MediaAgent software package.

Note: The Virtual Server Agent software package can be installed on the Media Agent, but a limit of twenty-six (26) concurrent volume mounts on a EC2 instance will limit the number of concurrent backups. Commvault recommends utilizing separate Access Nodes, which will send unique data hashes to a centralized MediaAgent and store data in S3 for short/long-term retention. This limit applies to Microsoft Windows and Linux-based MediaAgents.

Commvault recommends **Compute optimized** instances with **EBS optimized**, high IOPS volumes. Testing has shown that many smaller instances vs. fewer larger instances to benefit concurrent protection activities (requirements).

Commvault lab testing has shown a **C5.large instance** protecting 0.5FETB in 1.5 hrs.

The **Virtual Server Agent** is responsible for mounting snapshot volumes of clients and streaming backup data to S3 or equivalent Cloud Libraries. The **EBS bandwidth, throughput, and IOPS** as detailed at Amazon EBS – Optimized Instances is crucial to backup performance.

Note: Amazon calls out specific instance types that can only achieve **maximum performance** for no more than 30 minutes per day.

Amazon Access Node Specifications		
Extra small	**Small**	**Medium**
5 – 10 FETB	10 – 25 FETB	25 – 50 FETB
c5.large (2 vCPU, 4 GB RAM)	c5.xlarge (4 vCPU, 8 GB RAM)	c5.2xlarge (8 vCPU, 16 GB RAM)
1 x 120 GB EBS “General Purpose SSD” (gp2) volume for Live Browse	1 x 120 GB EBS “General Purpose SSD” (gp2) volume for Live Browse	1 x 120 GB EBS “General Purpose SSD” (gp2) volume for Live Browse
Windows Server 2019, 2016 **or** Linux (AWS Marketplace) / CentOS 7.4 / RHEL 7.5, RHEL 7.6 (**recommended**)		

Source: Hardware Specifications for Virtual Server Agent

Access Nodes are only required during protection activities (backup, recovery) and therefore are not considered long-running infrastructure. By co-locating your Access Node with your MediaAgent, and using the Cloud MediaAgent Power Management, the running time of the Access Node can be reduced to predominantly only your nightly backup windows.

Alternatively, if the number of parallel backup activities exceeds the number of MediaAgents available, utilize Amazon Access Node auto-scaling to auto-create and remove Access Nodes on demand.

For reference the annual relative cost of a Commvault recommended Windows-based Access Node (on-demand, us-east-1) is shown below. As evidenced in MediaAgent pricing (above), it is more cost effective to deploy multiple smaller Access Nodes, vs. scaling a single Access Node. Least cost configurations are highlighted in green, with alternatives in grey.

Annual Cost / FETB – Amazon Access Node (Windows)			
Number of Access Nodes	**Extra Small**	**Small**	**Medium**
Instance	c5.large	c5.xlarge	c5.2xlarge
1 Node	$1,695 / 10	$3,245 / 25	**$6,346 / 50**
2 Node	$3,389 / 20	$6,490 / 50	$12,692 / 100
3 Node	$5,084 / 30	$9,735 / 75	$19,038 / 150
4 Nodes	$6,778 / 40	$12,980 / 100	$25,384 / 200

Source: AWS Pricing Calculator

For reference the annual relative cost of Commvault recommended Linux-based Access Nodes (on-demand, us-east-1) is shown below. It should be noted that **Linux Access Nodes do not support all Commvault features**.

Access Node Capabilities – Windows vs. Linux

A summary of the supported features is shown below, if in doubt consult the Feature Comparison Matrix.

Amazon Linux-based Access Node Features		
Feature	**Windows**	**Linux**
Streaming backups	✓	✓
IntelliSnap backups (snapshot)	✓	✓
Backupset and Subclient filtering (Virtual Machine, Disks)	✓	✓
Multi-VM Restores, Restore full VMs, guest files & folders, agentless file recovery, attach disk to VM (existing, new)	✓	✓
Agentless File Recovery	✓	✓
Live Browse (with Collect File Details during backup)	✓	✓
Live Sync (from Amazon to Amazon)	✓	✓
Application aware backups	✓	×
Live File Recovery with UNIX-Based File System Support	✓	×
Access Node Teaming for load distribution	✓	×
Automatic Discovery of Virtual Machines	✓	×
Auto Commit on Kill	✓	×
View All Versions (If backup is done with collect File details enabled)	✓	×
VM Provisioning and Lifecycle Management	✓	×
VM Conversion to Azure	✓	×
VM Conversion to VMware	✓	×
Unicode Support	✓	×

Snap Replication	✓	×
ServiceNow	✓	×
Automated Retry for Failed VM Backups	✓	×
Alerts	✓	×
Cloud MediaAgent Power Management	✓	×
Failover and Failback Orchestration - VMware to Amazon	✓	×
Failover Groups (Application orchestrated failover)	✓	×
File Indexing for Virtual Machines	✓	
Cross account Snapshot Replication	✓	×
Use Resources from admin account (consolidate MediaAgents, Access Nodes)	✓	×

Source: Linux Access Node Support for Amazon Web Services, Feature Comparison Matrix

Annual Cost / FETB – Amazon Access Node (Linux)			
Number of Access Nodes	**Extra Small**	**Small**	**Medium**
Instance	c5.large	c5.xlarge	c5.2xlarge
1 Node	$889 / 10	$1,633 / 25	**$3,122 / 50**
2 Node	$1,777 / 20	$3,266 / 50	$6,245 / 100
3 Node	$2,666 / 30	$4,900 / 75	$9,367 / 150
4 Nodes	$3,554 / 40	$6,533 / 100	$12,490 / 200

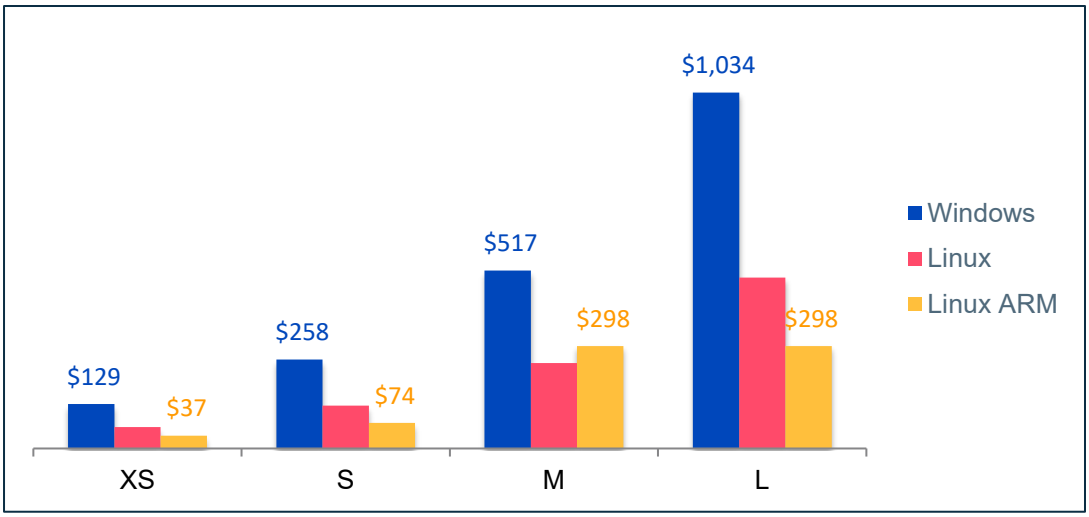
Source: AWS Pricing Calculator

As expected, Linux instances offer a significant price reduction over a Microsoft Windows Access Node, typically reducing per-monthly utility costs by **50%**.

AWS 64-Bit Arm Linux Instances

Additionally, in Feature Release 19 (11.19) Commvault added the capabilities to utilize AWS 64-Bit Arm Linux instances for MediaAgents and Access Nodes (see Feature Release 19 – Virtualization), ARM instances (a1 instance type) represent on average a **70% cost reduction** from equivalent Windows Access Nodes, albeit with a slightly reduced feature set.

Figure 1 - Virtual Server Agent cost by Operating System (per month)



Amazon has announced released new ARM-based Gravitron2 instances that represent a significant cost reduction from traditional Intel Chipsets (**C6g.large is a 30% reduction from C4.large**), see Announcing New Amazon EC2 M6g, C6g, and R6g Instances Powered by Next-Generation Arm-based AWS Graviton2 Processors

Commvault recommends using C6g Access Nodes as a default low-cost auto-scaling unit.

Architecture best practices for Virtual Server Agent sizing

Do's

- Do co-locate the Access Node (Virtual Server Agent) component on your MediaAgent for initial deployments and POCs.
- Do relocate the Access Node (Virtual Server Agent) component to dedicated host(s) when then backup window can no longer be met with existing MediaAgent-based Access Nodes.
- Do scale Access Nodes horizontally when protected client data grows, or backup window can no longer be met.
- Do ensure the latest AWS PV drivers are installed on your Access Nodes to optimize volume mount/unmount activity.

Don'ts

- Do not substitute lower EC2 instances without consulting the Amazon EBS – Optimized Instances throughput.

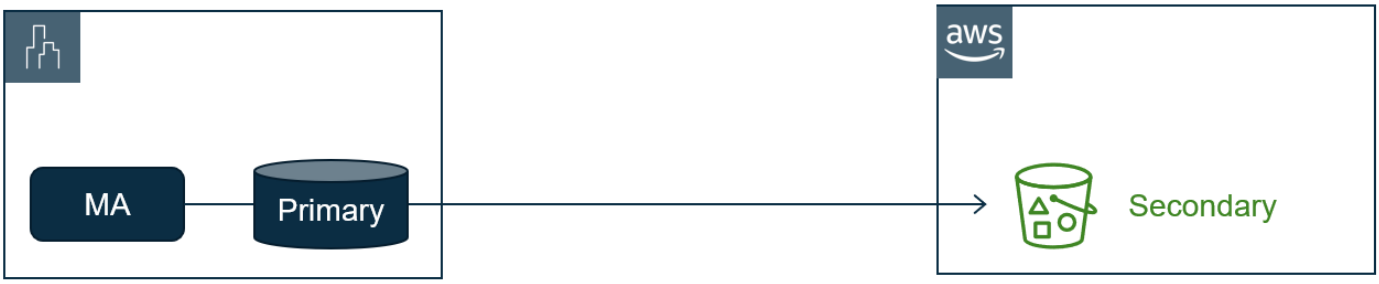
Design Patterns

The following section provides a set of common design patterns used to deploy intelligent data management capability into Amazon. Each pattern solves a single problem and not all patterns must be applied, depending on your needs. Patterns have been grouped by the core problem they address:

- Backup
  - Backup to Cloud – on-premises data copy
  - Backup to Cloud – cloud-only data copy
- Availability
  - High Availability Protection – Single-AZ
  - High Availability Protection – Multi-AZ
- Archive
  - Data Archival – Non-Deduplicated
  - Data Archival – Deduplicated

Backup to Cloud – on-premises data copy

Cloud consumers often begin with public cloud by extending their on-premise data center into the Cloud. In instances where the business has critical workloads still on-premises it is crucial to maintain rapid recoverability on-premises while leveraging elasticity and low-cost of cloud storage.



- Short-term operational recovery copies are held onsite disk/object/tape library (e.g. 7 - 30 days).
- Long-term operational and disaster recovery copies are held in Amazon S3 service.
- Leverages on-premises compute infrastructure (MediaAgent, MA).
- No S3 egress cost unless restoring from Secondary.

Solution minimizes in-cloud infrastructure by leveraging on-premises compute resource.

Backup to Cloud – cloud-only data copy

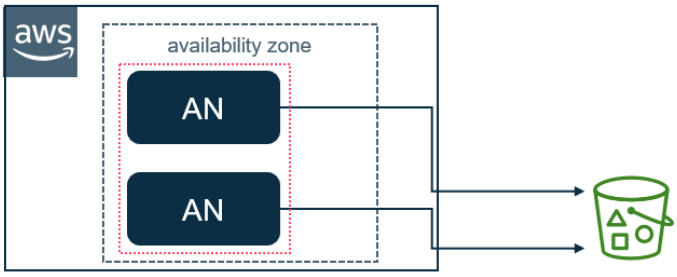
As businesses adopt more public cloud services, there may be a desire to reduce the amount of owned and operated infrastructure. Removal of on-premises backup copies (secondary storage) allows the business to reclaim valuable data center space while leveraging the elasticity and low-cost of cloud storage.

- Short-term operational recovery copies are held offsite in Amazon S3 service.
- Leverages on-premises compute infrastructure (MediaAgent, MA) to optimize (deduplication, compress) data before transfer.
- Allows business to completely remove secondary storage infrastructure from on-premises locations.
- All restores, synthetic full backups, and maintenance activities will incur minor egress charges.

Solution minimizes on-premises infrastructure by storing all backup copies offsite.

High Availability Protection – Single-AZ

Cloud services also require data protection to provide in-region and cross-region data copies for the purposes of operational and disaster recovery. High availability of data protection operations assists in ensuring data protection activities occur as scheduled and are unaffected by temporary planned/unplanned outages. Placement of data management infrastructure must be carefully considered to avoid inter-AZ network transfer fees.



- Access Nodes (**AN**) are grouped within an Availability Zone (AZ) to provide load-balancing and high availability.
- Co-location of Access Nodes avoids inter-AZ network transfer costs (charged in both directions) between ANs.
- Solution protects workloads within the AZ at least cost and may optionally protect other AZs with inter-AZ transfer fees incurred.

Solution provides highly available, least cost, least latency protection for a single availability zone.

High Availability Protection – Multiple-AZ

Cloud services are distributed across availability zones to minimize the impact of planned or unplanned outages. High availability of data protection operations assists in ensuring data protection activities occur as scheduled and are unaffected by temporary planned/unplanned outages. Placement of data management infrastructure must be carefully considered to achieve the highest levels of availability.

- Access Nodes (**AN**) are grouped across Availability Zones (AZs) to provide maximum availability from an AZ outage.
- Distribution of ANs across AZs will incur ongoing inter-AZ transfer fees for all data management activities.
- Solution protects workloads across all AZs within a region, with the acceptance that inter-AZ traffic will occur.
    - Solution provides highly available protection for all availability zones within a region.

Data Archival – Non-Deduplicated

Businesses are retaining more data for historical **business insight** and **analytics**. Historically this meant expensive tape libraries, data preparation and handling activities. Data archival to ultra-low-cost cloud archival services can now provide long-term retention without tape handling.

- Archival data represents a subset of backup data residing on-premises or in cloud storage services.
- Archival data is not suitable for deduplication or compression (lossless, x-rays, CAD, EDF archives)
- Very long-term retention data required for regulatory compliance (i.e. age of patient + 10 years)
- Data is stored in original format, recovery requires **index recall**, followed by **data subset** recall.
- Solution provides long-term data retention in original application format.

Data Archival – Deduplicated

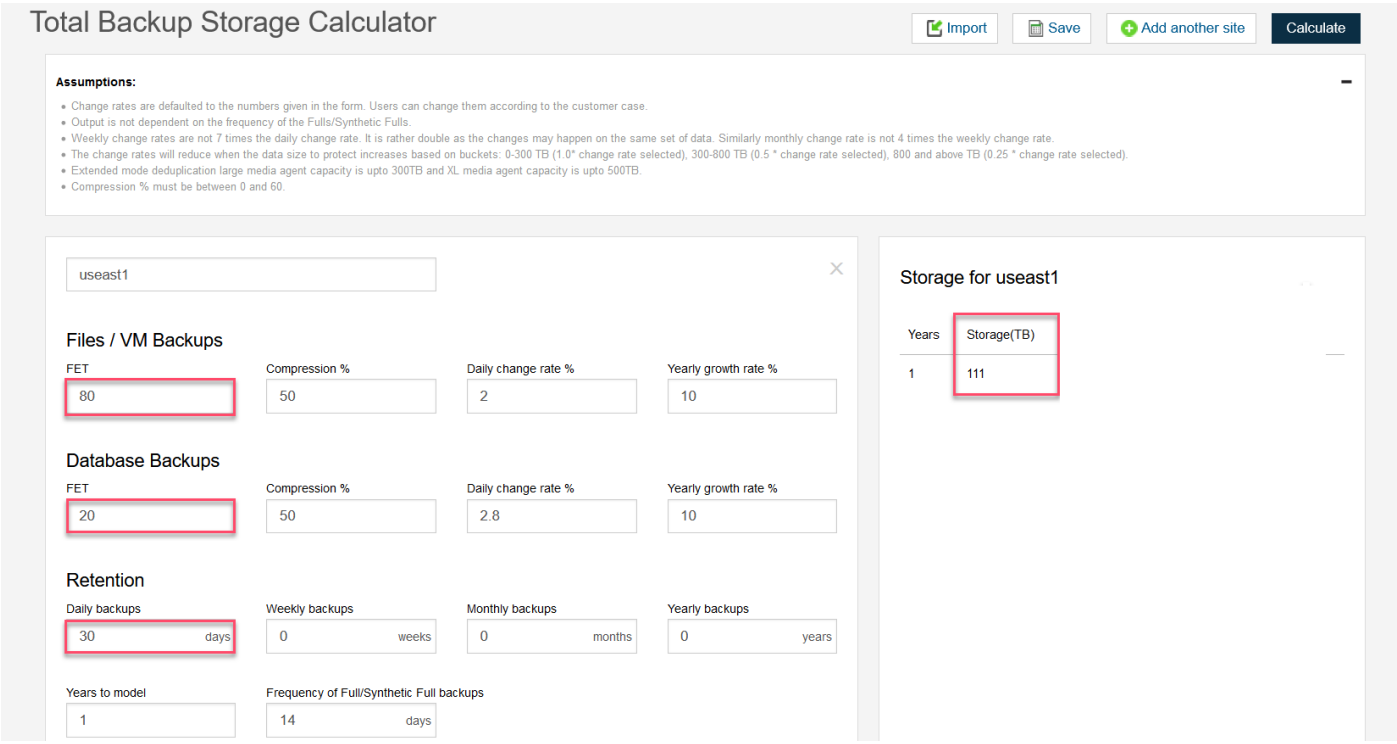
As businesses look to store data for **data analytics**, **visualization** and ultimately **for action** – a cost effective data storage approach is required. Utilizing de-duplication and compression to remove duplicate data before long-term storage can provide cost-optimal storage within minimal additional data handling.

- Archival data represents a subset of backup data residing on-premises or in cloud storage services.
- Archival data is suitable for deduplication and compression (virtual machines, office documents)
- Archive is being kept with intention to restore back into operational state and simplified recall.
- Index data is stored in **warm** storage class (S3, S3-IA), **data** is stored in **cold** storage class (Glacier, DeepArchive)
- Data may be recalled as a simple restore, by leveraging accessible indexes in warm storage.
- Solution provides long-term archive retention with optimized recall process.

An example of sizing

For this example, we want to protect approximately 100 TB of Front-End (FET) capacity. The average size of each instance is assumed to be about 100 GB and each instance has 2 volumes - one for operating system and the other for applications and data. This equates to approximately 1000 instances (100 TB at 100 GB each, using base10 rather than base2 for simplicity in approximation).

It is also assumed that the daily change rate is approximately 2% of net new data that is created per day, or approximately 2 TB worth of new data before it is compressed and deduplicated. The change rate in an environment varies greatly by environment and coupled with retention and deduplication ratio of the data. Both of which are also highly dependent on the specific environment, these three factors affect the back-end storage capacity that is ultimately required. For this example, we shall assume 90 days of retention and approximately 5:1 deduplication ratio. Both are typically observed within most virtual environments running a mix of Windows and Linux operating systems. With retention it is important to note that data that is 90 days old or the first full backup are not deleted until the most recent one is fully committed to the backup target. This accounts for retention+1 storage requirement. This results in approximately 162 TB used for back-end capacity.



Source: Commvault Back end TB Calculator (available to customers at cloud.commvault.com)

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

Using publicly available pricing for AWS resources (as of March 2020) the cost of performing protection in AWS by utilizing any combination of iDataAgents and coupled with Commvault IntelliSnap for AWS snapshot management, the following becomes a rough estimate of the cost of the AWS infrastructure required for a period of 1 year with a Partial upfront payment.

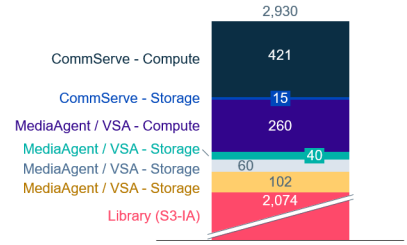


Figure 2 - Monthly Cost Estimate (USD)

The **monthly costs** of each component are show (left), modelled on a full year commitment with 90 days of backup data held (per calculator above). The largest contributor is the **S3.-IA storage** followed by the **CommServe** command and control server.

Commvault can help control the cost of your **MediaAgent** by powering down MediaAgents during periods of inactivity.

	Qty.	Resource type	Unit cost	Cost (mth.)	Cost (annual)
CommServe Medium VM	1	m5a.2xlarge EC2 Standard Reserved instance 1yr partial upfront Windows	$0.288	$210.24	$5,047.00
CommServe OS Disk	1	150 GB gp2 EBS volume	$0.100	$15.00	$180.00
Standard Dedup MA Medium	1	m5a.2xlarge EC2 On-demand instance Windows	$0.712	$519.76	$3,118.56
Access Node OS Disk	1	400GB gp2 EBS volume	$0.100	$40.00	$480.00
Access Node DDB Disk	1	600GB gp2 EBS volume	$0.100	$60.00	$720.00
Access Node Index Disk	1	1 TB gp2 EBS volume	$0.100	$102.40	$1,228.80
Library Capacity (TB) (90 days) **162 TB**	1	S3.-IA bucket capacity	$0.0125	$2,073.60	$24,883.20
Total Annually					**$35,658.44**

Source: Cost (USD) calculator.aws, us-east-1 (Reserved). Excludes ingress, egress, GET/PUT, retrieval costs.

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

Additionally, only the CommServe utilized **reserved instance** pricing, use reserved instances where feasible.

Architecture Considerations

Security, Identity & Compliance

Commvault recommends **security as task #1** for customers in AWS. Specifically, Commvault advises the following best practices for securely deploying and managing your cloud data:

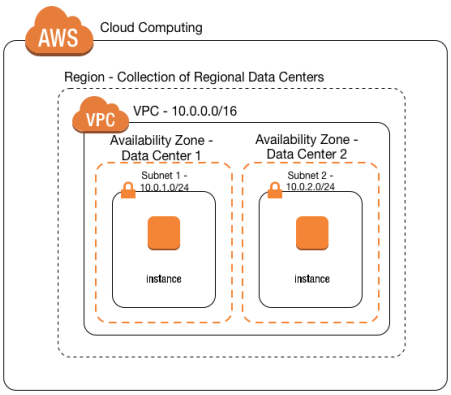
- Ensure **encryption** both on-wire and at-rest is employed for all communications and data handling infrastructure.
- Leverage **Identity & Access Management (IAM)** roles to limit capability and scope of administrative accounts.
- Attach **IAM Roles** to AWS infrastructure to avoid the storing and maintenance of credentials (i.e. secrets).
- Utilize **AWS AssumeRole** capability to provide shared data protection resources that are responsible for protecting per-department AWS accounts.
- Limit the use of **Access Key/Secret Key** deployments (where feasible) to avoid the administrative effort in rotating credentials and updating Commvault stored credentials.
- Optionally, look at simplify your own **Key Management Server (KMS)** located outside of AWS to provide an additional level of control over data access.
- Segregate the **scope** and **level of access** on a per use-case basis. For example:
    - Utilize separate **accounts** and **IAM roles** for Production & Non-Production.
    - Utilize separate accounts for **Primary**, **Secondary** and **Archive** data stores.
- Consider **air gapping** at least one copy of your critical data to ensure an effective response to an unplanned and uncontrolled **ransomware propagation event**.
- Consider utilizing **write-once read many (WORM)** storage mechanisms providing by AWS (S3 Object Lock, Glacier Vault Lock) to provide a protection from **ransomware propagation** into backup stores[1].

[1] Commvault provides native ransomware protections and detection. Protection at the storage layer provides additional controls over propagation of malware.

Networking

Virtual Private Cloud/Networking

AWS has the capability to establish an isolated logical network. This is referred to within AWS as a Virtual Private Cloud (VPC).



Source: Amazon VPC for On-Premises Network Engineers – Part 1

Instances/Virtual Machines deployed within a VPC, by default, have no access to the Public Internet, and utilize a subnet of the customer's choice. Typically, VPCs are used when creating a backbone between Virtual Machines (Amazon EC2 instances), and when establishing a dedicated network route from a customer's existing on- premises network directly into the public cloud provider via AWS Direct Connect.

Bridging on-premises infrastructure – VPN & AWS Direct Connect

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.
- AWS Direct Connect – a dedicated network link is provided at the customer's edge network at an existing on-premises location that provides secure routing into an AWS Virtual Private Cloud Network. Typically, these links are less expensive when compared to a customer's regular internet connection, as pricing is charged on a monthly dual-port fee, with all inbound and outbound data transfers included free of charge, with bandwidth from 10 Mbit/s to 10 Gbit/s.

Infrastructure Access

Hypervisor access in Public Cloud

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

Amazon VPC endpoints

AWS provides VPC Endpoints that enable you to create private connections between a given VPC and another AWS service without having to route via public Internet space. Support for Amazon S3 VPC Endpoints was announced in May 2015, and while it is only supported within the same region as the VPC

Use of VPC Endpoints is highly recommended as it reduces availability risks and bandwidth constraints on the VPC's link through to public Internet.

An Amazon S3 VPC Endpoint must first be defined by creating an Endpoint Policy within the AWS console, but there is no change to the FQDN hostname used to define the Cloud Library within Commvault. Instead, AWS will ensure that DNS queries for the hostname will resolve against the Amazon S3 VPC Endpoint, instead of the public address, and apply appropriate routing (provided the Endpoint Policy is successfully created).

For more information on VPC Endpoints, please refer to this AWS documentation: VPC endpoints.

Data Security

In-flight

By default, all communication with Cloud Libraries utilize HTTPS which ensures that all traffic is encrypted while in-flight between the MediaAgent and the Cloud Library 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. Most Cloud providers require that any seeded data is shipped in an encrypted format. An example of seeding data in AWS is with the use of AWS Snowball or AWS Snowball Edge devices.

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 AWS 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).

Data Seeding

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

Over-the-wire

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

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

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

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

Please refer to Seeding a Cloud Storage Library for more information:

Amazon provides the following offline seeding technologies, all of which are supported by Commvault (per *Seeding a Cloud Storage Library* process above).

- Snowball Edge AWS Snowball Edge
- Snowmobile AWS Snowmobile
- Snowcone AWS Snowcone

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 Amazon S3 Standard instead Amazon S3-IA or Amazon Glacier.

Storage I/O

The input and output operations to storage attached to the virtual machine instance. 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

Amazon S3 storage usually incurs 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 software writes data to a cloud library, the Cloud Library splits the data up into a sub-chunk size of 32 MB. Each 32 MB chunk write or read will incur a GET or PUT request. As of January 2018, AWS charges $0.005 per 1000 PUT requests and $0.004 per 10,000 GET requests for an Amazon S3 Standard bucket for example.

A baseline of 200 GB with a saving of 40% at 32 MB sub-chunk size would result in an approximately 3840 PUT requests. At a charge of $0.005 per 1,000 requests that is a cost of 2 cents.

Note: All cost figures are referenced in USD and based on pricing listed on the AWS website at 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 an agreed time. Storing infrequently accessed data on a low-cost cloud storage solution may be attractive upfront, however Commvault recommends modeling realistic data recall scenarios. In some cases, the data recall charges may be more than the potential cost savings vs. an active cloud storage offering.

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

Performance / Storage

Multi-Streaming with Object storage

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

Cloud Connection Best Practices

There are additional Data Path settings and additional settings used to adjust and fine-tune the performance of the Cloud Library.

The following combined settings are recommended to increase data read/write performance for cloud libraries.

1. Increase **Chunk size** to 4 GB (default) **f**or written backup copies, and 32 GB for archives (Chunk size).
2. Increase **Data Path Block size** to 1024 KB for written backup copies.
3. Increase the number of **Device Streams** above 50 (increases parallel streams during protection activities).
4. Increase **Deduplication Block Size** to 512 KB (for cloud workloads)
5. Increase the **Number of Data Readers** to increase backup read performance.

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

Note: Commvault will automatically tune the Block level deduplication factor for best performance, based on whether the location is a data source (primary copy) or destination (secondary copy).

Compression vs. Deduplication

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

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

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

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

Leveraging multiple mount paths for a cloud library

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

For Amazon S3, using multiple mount paths may help to increase performance due to the nature of how the Amazon S3 subsystem distributes data.

Block storage versus Object storage

While Public IaaS environments allow block-based storage to be provisioned and leveraged as Disk Libraries, the overall cost of those volumes can quickly exceed that of object storage. Based on typical cloud pricing, object storage could store 3x as much data as block-based storage (Amazon EBS “General Purpose SSD” gp2 volumes) for 33% less cost. This will vary based on capacity and performance needed from block-based storage.

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

If you are unsure as to which offering to use, you should consume regular object storage (Amazon S3 Standard).

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 allows for the addition of two extra nodes (up to 4) to an existing deduplication store dynamically, allowing for rapid scale-out configurations. See Commvault online documentation for more information: Partitioned Deduplication.

Micro pruning

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

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

Selecting the right storage class for backup and archive data

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

For example, storing infrequently accessed backups within Amazon S3 Standard-Infrequent Access class (Amazon S3-IA) can significantly lower the cost of your cloud bill, while storing archives in Infrequent Access tier vs. a Deep Storage tier (i.e. Amazon Glacier.) can greatly impact accessibility for end-users to the archived data.

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

- Standard storage – this storage class represents the base offering of any object storage platform – inexpensive, instant access to storage on-demand. Offerings in this category include Amazon S3 Standard, at an average price of $0.024/GB/month (as of April 2018 and depending on geographic region).

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

- Infrequent Access – this is a relatively new offering that addresses what was a gap between Standard offering and Deep Archive storage tiers, in that it is offered at a lower price point than Standard storage ($0.004/ GB/ month for AWS), but is aimed at scenarios where data is infrequently accessed. Offerings in this category include Amazon S3 Standard-Infrequent Access (Amazon S3-IA) and Amazon S3 One Zone-Infrequent Access (Amazon S3 One Zone-IA).

  While the storage is always accessible, like the Standard 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.

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

- Deep Archive – 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. Amazon Glacier. is an example of archive storage which Commvault supports.

  The cost of this storage class is the lowest compared to all three offerings – between $0.002/GB/month to $0.01/ GB/month depending on geographic region – but as with the Infrequent Access class, the Deep Archive class’s cost model is also structured with the expectation that retrievals are infrequent and unusual, and data will be stored for an extended period of time.

  Typically, providers will charge a fee if data is deleted prior to 30-90 days of an object’s creation, and if more than a set percentage of your data set per month is retrieved, then additional costs may apply. 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.

Infrequent access storage class support

Support for the following Infrequent Access storage classes are available in Commvault® Version 11:

- Amazon S3 Standard-Infrequent Access (Amazon S3-IA)
- Amazon S3 One Zone-Infrequent Access (Amazon S3 One Zone-IA), as of V11 SP12
- Amazon S3 Intelligent-Tiering, as of V11 SP14
- Others listed online at **commvault.com**

Storage Accelerator to Amazon S3 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 Amazon S3., completely bypassing the MediaAgent. This is achieved by installing and enabling the Storage Accelerator feature on the client for direct communication to a storage target ad will speed up the backup and reduce costs in these situations.

See Accelerating Backups to Cloud Storage Libraries for more details.

Storage Accelerator can be used to backup data from all Agents.

Deduplication DataBase (DDB) sealing

Commvault provides **software-based deduplication** to reduce the amount of S3 (or equivalent) storage utilized when storing your backup and archival data. The DDB has maximum scale limits and also represents a logical container for your data.

Commvault recommends that you **seal** your DDB every 12 months for secondary backup copies (i.e. long-term retention, archive copies) to segment your long-term data storage more efficiently.

WARNING: Be aware that this will incur a **full rebaseline** or the next data sent to the DDB.

WORM

WORM media means write-once, read-many and incurs different data handling rules.

Commvault recommends sealing WORM stores for half the retention time of the copy – but considering the minimum stay in the storage class.

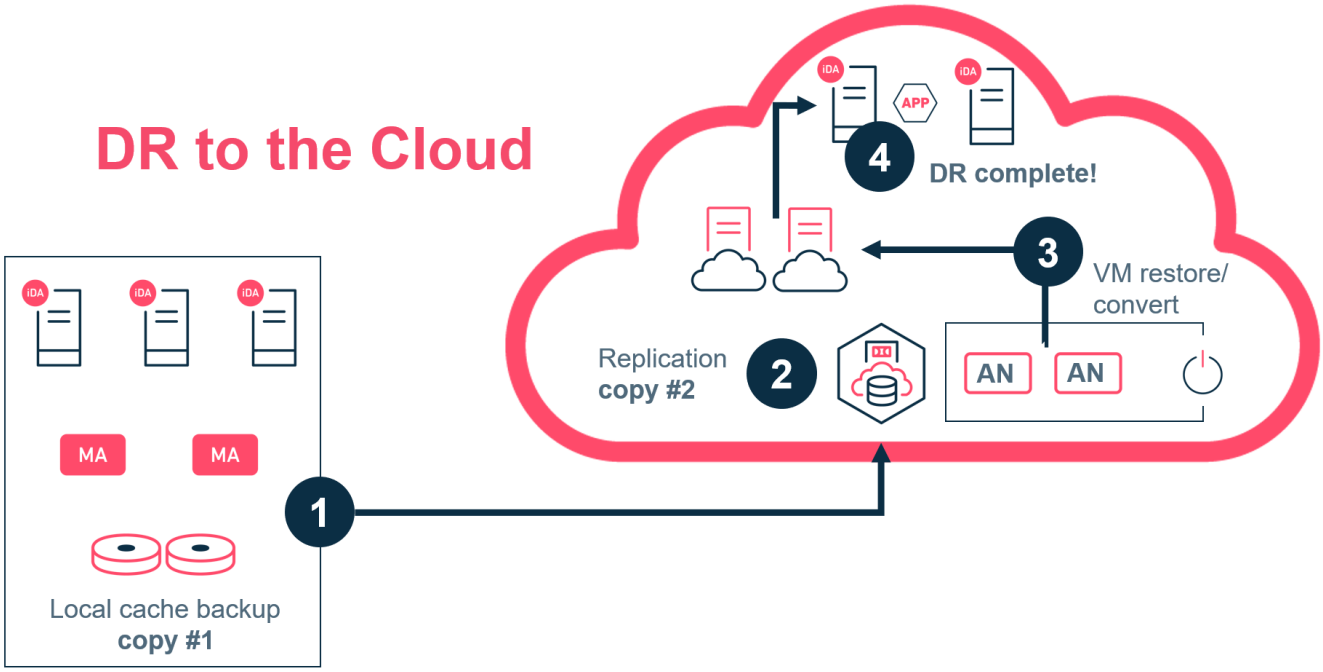
For example:

- A WORM copy with 4 year retention should be sealed every 12 months.
- A WORM copy of 180 day retention, should be sealed every 90 days.
- Never seal more frequently than the minimum stay period.

Performing Disaster Recovery to the Cloud

This section will cover the steps required to perform disaster recovery into the Amazon Web Services 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.

Your Disaster Recovery options are broad when leveraging Amazon Web Services as your on-demand data center. While initial solutions will often utilize a cross-hypervisor migration (i.e. VMware to AWS conversion), subsequent solution(s) may simply replicate/failover workloads across Amazon availability zones and/or Regions.



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 DR scenario or automatically recovered to existing instances for more critical workloads.

Replicating VM workloads with Live Sync

Live Sync also allows you to replicate VMs to public cloud infrastructure. As of December 2016, Amazon EC2 is a supported cloud infrastructure vendor for Commvault®.

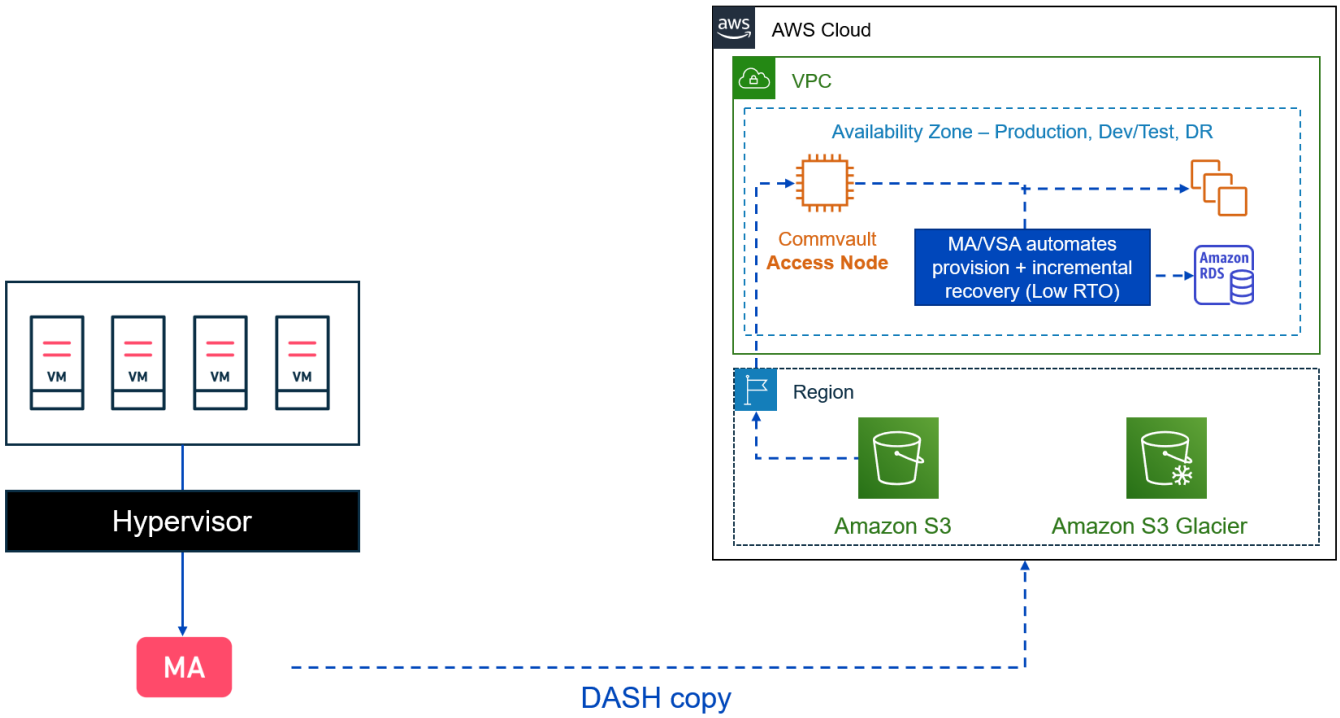
Live Sync combines the VM conversion feature with incremental replication to provide a DR solution utilizing on-demand cloud infrastructure. As Live Sync to cloud integrates with DASH Copy, highly efficient WAN replication is possible by reducing the amount of data being replicated.

As cloud infrastructure is securely shared and protected, only a limited subset of integration points is available for Commvault software to integrate with Amazon EC2. As such, there are extra steps to convert virtual machines and prepare them for recovery. For example, Amazon EC2 stipulates that all imported machines must run through the VM import process – this can take several hours depending on the size of the VMs, increasing the RTO value if a failover

is required while conversion is taking place. This must be considered when designing Live Sync-based DR, as each incremental update will require a full run-through of the import process, delaying the replication process.

As of June 2018 (Commvault V11 SP12), if a Commvault Access Node running inside an Amazon EC2 instance is used as the destination Access Node for the Live Sync replication/conversion, the Commvault software will bypass the AWS VM Import/Export process and will instead natively perform the VM conversion. This results in significant speed increases compared to previous Commvault software releases.

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 the Commvault 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.



Additional information on the Conversion feature is accessible using the link below.

- Live Sync Replication for Amazon Web Services

Replicating EC2 workloads with Live Sync

Live Sync also allows you to replicate VMs within public cloud infrastructure. As of December 2019, Amazon EC2 is a supported cloud infrastructure source and target, providing cross-AZ and cross-region replication (see Replication for Virtual Machines).

It is a good Disaster Recovery strategy to separate your workloads by the scope of the anticipated disaster. Commvault Live Sync allows you to utilize your backup copies to automatically update DR instances in alternate Availability Zones (metro-separation) or Regions (WAN-separation). Consider both the frequency of data replication, but also the types of events your DR solution can service.

Additional information on automating your failover may be found in DR Orchestration and Failover Groups. Failover events are stressful and can be error prone. By utilizing Commvault automation, test booting, and reverse replication, you will gain the confidence that your infrastructure can be recovered efficiently.

Replicating Other Workloads

Commvault Continuous Data Replicator (CDR) allows near time continuous data replication for critical workloads that must be recovered in adherence to Service Level Agreements that exceed the capabilities associated with Live Sync operations. These VMs require a similarly sized Amazon EC2 instance running in AWS to receive any replicated data. For CDR to operate, an Amazon EC2 instance must always be running to receive application changes. Additional information on CDR can be located using the link below.

- ContinuousDataReplicator Overview

Amazon-specific Workloads

Virtual Machine Recovery into EC2 Instances

With the release of Version 11, the Commvault Virtual Server Agent allows the ability to easily perform direct conversion of protected VMware or Hyper-V virtual machines into Amazon EC2 instances, from backups stored either within Amazon S3 (any tier) or another Cloud Library or from an on-premises Disk Library.

This process could be used as part of a disaster recovery strategy using Amazon as a cold DR site, or as a migration strategy (Lift-and-Shift).

Additional information on the Conversion feature can be located using the link below.

- Converting Virtual Machines to AWS (from VMware or Microsoft Hyper-V)

Virtual Machine Recovery from Amazon EC2 into Azure Virtual Machines

With Commvault V11 SP7 or newer, you can now recover Amazon EC2 instances protected with the Virtual Server Agent to an Azure Virtual Machine for disaster recovery or migration purposes. Currently streaming backups are supported for recovery to either Azure Classic or Azure Resource Manager.

Additional information on the Conversion feature can be located using the link below.

- Converting Amazon EC2 instances to Azure VMs

Virtual Machine Recovery from EC2 to VMware

With Commvault V11 SP13 or newer, you can now recover Amazon EC2 instances protected with the Virtual Server Agent to VMware virtual machines for disaster recovery or migration purposes. The Amazon EC2 instance to be converted into a VMware virtual machine must be one that was previously converted by the Commvault software from VMware into Amazon EC2.

Additional information on the Conversion feature can be located using the link below.

- Converting Amazon EC2 instances to VMware

Using Commvault Workflows to Automate DR

The Commvault Workflow engine provides a framework in which the DR runbook process, covering the deployment of new instances, 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 the Commvault Personalization Services team, please contact Commvault or your Commvault Partner Account team.

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

Protecting and Recovering active workloads in AWS

This section outlines the basics on protecting active workloads running in AWS. 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-in-guest (streaming)

An agent-in-guest approach is used to protect a wide variety of operating systems and applications. Agent-in- guest is used on the production workload and protected to the MediaAgent residing in AWS, leveraging client-side deduplication to reduce the network consumption within the cloud. This 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 instances for the more critical workloads.

When to Use Agent-in-Guest Approach

- When you require application-consistent backups – use an agent-in-guest either standalone or in conjunction with an Amazon Access Node backup. Deployment of agents can either be deployed via automation by Commvault® software, incorporated into AMI templates 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.

Architecture Requirements for Agent-In-Guest

- Minimum 1x iDataAgent per Amazon EC2instance for the intended dataset (i.e. SQL database, File). Multiple iDataAgents can be deployed on the same Amazon EC2 instance.
- Minimum 1x MediaAgent per region. MediaAgents connect to the target object storage and can either be deployed on the same Amazon EC2 instance as the dataset iDataAgent, or on a dedicated host for a fan-in configuration. The Amazon EC2 instance specifications of the MediaAgent should match the MediaAgent specifications within this Architecture Guide.
- Check the Systems Requirements section in documentation to determine if the iDataAgent supports your application (see Backup Agents).

Architecture Recommendations

- Use of multiple readers to increase concurrency to the object storage target is recommended.
- Use of the Amazon S3 VPC Endpoint is highly recommended to improve throughput to/from Amazon S3 buckets.

Snapshot Based Agent In-Guest (EBS IntelliSnap®)

In addition to the standard agent-in-guest streaming approach, for supported configurations the agent can integrate the Amazon EBS snapshot facility into a seamless framework that combines fast storage-based snapshots with application consistency, without the need for a collection of scripts and disparate snapshot, backup and recovery tools. Native API integration with Amazon EBS APIs allows creation, protection and replication of encrypted volume snapshots using AWS KMS integration using Key/Secret or IAM authentication.

While the Snapshot-based Agent-in-Guest approach is currently only supported for Oracle on Linux, SAP HANA on Linux, and the Linux File System iDataAgent, there is no difference with the actual agents themselves – agents are

capable of both Streaming and Commvault IntelliSnap® backup approaches, and it only requires that Commvault IntelliSnap be configured.

The agent is deployed on the production Amazon EC2 instance, orchestrating work between the target workload and the Amazon EBS API, with additional support to extracting data from those snapshots to be stored in a deduplicated, compressed format, within Amazon S3 Standard or Amazon S3-IA buckets for long-term retention at a lower cost point. This process of extracting from the snapshot is known as a Backup Copy job. This ensures that consumption costs are minimized while also protecting against failure of single Availability Zones and Regions by combining both Amazon EBS and Amazon S3 storage options and redundancy offerings.

When to Use Snapshot-Based Agent-in-Guest Approach Instance of The Streaming-Based Approach

- When fast, snapshot-based recovery is required – use the snapshot-based agent-in-guest to leverage Amazon EBS snapshots while maintaining application consistency and granular-level recovery options for workloads that require fast recovery.
- When application-consistent backups with a shorter backup window are required – the snapshot-based agent-in-guest approach allows you to execute fast backups, reducing the amount of time normally reserved for a backup window in order to extract the blocks.
- When minimal load against the production server is required– the snapshot-based agent-in-guest approach operates with minimal load against the production host and can be instructed to perform the Backup Copy (extracting from snapshot to Amazon S3 Standard or Amazon S3-IA buckets) using a separate Amazon EC2 instance as a Access Node.

Architecture Requirements for Snapshot-Based Agent-in-Guest

- Minimum 1x iDataAgent per Amazon EC2 instance for the intended dataset (i.e. Oracle, File). Multiple iDataAgents can be deployed on the same Amazon EC2 instance.
- Amazon must be configured as an “array” in Commvault under Array Management. For more information, see Setting Up the Amazon Array Using Array Management using the Commvault online documentation.
- Any selected Access Node must be in the same availability zone and region in order to both access any volumes created from snapshots and for best performance.
- (Backup Copy) Minimum 1x MediaAgent per region. MediaAgents connect to the target object storage and can either be deployed on the same Amazon EC2 instance as the client, or on a dedicated host for a fan- in configuration. The Amazon EC2 instance specifications of the MediaAgent should match the MediaAgent specifications within this Architecture Guide.
- Check the Systems Requirements section in documentation for your specific application (see Backup Agents)

Architecture Recommendations

- For AWS environments with multiple accounts you can configure a virtualization client for Amazon Web Services to use a separate Admin account for data protection operations. This approach reduces the impact of backup operations and restore operations on production accounts. For more information on Cross-Account operations, see Processing for Cross-Account Operations.
- While Amazon EBS snapshots are independent of the original volume, they are still only redundant within a single Availability Zone. Extracting the data from the snapshot into object storage is highly recommended to gain higher redundancy, whether just single region (store in Amazon S3 Standard or Amazon S3-IA buckets) or multiple regions (store in an Amazon S3 Standard or Amazon S3-IA bucket, or DASH Copy/replicate to another region).
- (Backup Copy) Use of multiple readers to increase concurrency to the object storage target is recommended.

- (Backup Copy) Use of the Amazon S3 VPC Endpoint is highly recommended to improve throughput to/from Amazon S3 buckets.

Architecture Recommendations for Oracle Databases

- Take the sizing of the production Oracle instance into consideration when selecting a backup frequency.
- As part of the backup, the target Oracle instance is placed into backup mode until the Amazon EBS snapshot has been successfully created, then Commvault software will instruct Oracle to exit backup mode.
- While this process is generally quite efficient, if this is the first snapshot taken for the Amazon EBS volumes, the Oracle instance may be placed into backup mode for longer than expected (minutes instead of seconds) and the Oracle instance and Amazon EC2 instance should be sized appropriately for increased redo log activity until the conclusion of the snapshot.
- If the Amazon EC2 instance running Oracle is provisioned as part of a Marketplace AMI, the volumes packaged with that AMI cannot be targeted for Amazon EBS snapshots as per AWS design. Ensure that any desired instance data is stored on volumes created post-launch to enable snapshot operations.
- Use separate Amazon EBS volumes and mount points for both Oracle data and archive log files.
- Do not use nested file system mounts for Oracle IntelliSnap.
- The data files and archive log files for each database should be isolated on their own volume groups. If volumes are shared for multiple databases, use the 'Multi Instance Snap Optimization' feature, which snaps and reverts all the databases together. For more information, see Multi-Instance Snap Optimization.
- If you enable the block change tracking file, for faster incremental backups, specify its location on a device that also contains the data files for the database. If the Amazon EC2 instance is a Marketplace sourced AMI, you should not place this on a volume that was packaged with the AMI, or the snapshot creation operation will fail.

NOTES

- The Oracle iDataAgent, when in a Commvault IntelliSnap configuration, will only snapshot Amazon EBS volumes that contain the target database. At time of backup the agent automatically interrogates the Oracle instance to determine the source path for the data and log destinations back to the OS mount path, before then resolving this to an Amazon EBS volume ID that is used during the snapshot API call.
- For more information on Oracle IntelliSnap, please refer to Protecting Oracle Databases that are Hosted by an Amazon EC2 Account.

Agent-Less Amazon EC2 Instance Protection (Amazon Access Nodes)

Introduced in Version 11 Service Pack 3, the Virtual Server Agent for AWS (VSA for AWS) delivers an agent-less, block-level capture of Amazon EC2 instances and their attached Amazon EBS volumes. Amazon Access Nodes include Changed Block Tracking (CBT) support, included with v11 Feature Release 20, which helps accelerate incremental backup performance. Amazon Access Nodes utilize Amazon EBS direct APIs (documentation) to request changed blocks from the Amazon EBS service, dramatically accelerating backup time. CBT backups are enabled by default for new configurations from 11.20. Restoration options include both Full Virtual Machine recovery and granular-level file recovery.

When to use the Amazon Access Nodes for AWS

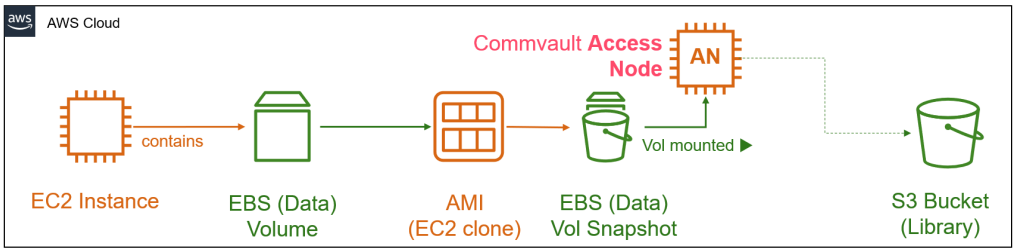
- Agent-less protection approach for Amazon EC2 instances and file-level data – no agents are required in-guest to perform a block-level backup to provide Instance and File-level recovery.

When not to use the Amazon Access Nodes for AWS

- Protecting worker/stateless instances – 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 using Amazon Access Nodes or agent-in-guest, depending on the required level of backup (crash vs. application consistent).
- When aggressive RPO and RTO's are required – Due to the way Amazon processes Amazon EC2/EBS snapshots, backup and recovery operations may perform slower than expected using this data protection method. This can be partially mitigated by leveraging Commvault® IntelliSnap data snapshot functionality and ensuring at least one snapshot is always present on the VM. It is also possible to use a Commvault® agent-in-guest method for faster data protection.

How Instances Are Qualified for Protection

- Each Amazon Access Node can be configured with one or more subclients. Each subclient defines a rule set on which to auto- detect and protect Amazon EC2 instances, based on a user-defined criterion of Instance Name, Region, Availability Zone or Tags.
- During the discovery phase of the backup job, the Access Node will use the subclient rule to qualify instances for protection within that job.



- Snapshot is mounted to Access Node, indexed and scraped
- Granular file-level data can then be restored, independent of original snapshot
- AMI is removed at completion of job

Commvault software does not require access to the AWS hypervisor-level, instead using the Amazon EC2/EBS REST APIs to create a Full VM Clone (AMI creation) of each Amazon EC2 instance, attaching the Amazon EBS volumes to a nominated Access Node to read and deduplicate the blocks before writing out to an Amazon S3 Bucket.

Commvault IntelliSnap® functionality introduced in Version 11 June 2016 for the VSA for AWS modifies this behavior by simply creating an Amazon EBS snapshot and retaining the snapshot based on the Storage Policy's Snap Primary retention setting.

This has the effect of reducing the amount of time required to protect data, providing fast, snapshot-based restoration capability and offloading the task of extracting blocks from the snapshot into Amazon S3 for longer-term retention through the Backup Copy operation.

On snapshot cost: Use of the Commvault IntelliSnap method does mean that snapshots remain available for longer than the backup window, however Commvault recommends that snapshots be 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.

Architecture Requirements For Amazon Access Node

- Minimum 1x Access Node per Region, Recommended 1x Access Node per Availability Zone
- Each Access Node represents a single Amazon EC2 instance with the Virtual Server Agent and MediaAgent packages deployed. The Amazon EC2 instance specifications should match the MediaAgent specifications within this Architecture Guide.
- In order to perform a recovery, an Access Node is required to be present in the target destination Availability Zone.
- Deploy deduplication MediaAgents in the same availability zone as the workloads being protected.
- Deploy MediaAgents in a common GridStor® configuration within a single zone.
- Deploy Access Nodes in a common pool within a single zone for availability and load-balancing (see Auto-scaling notes below)
- [optional] Deploy MediaAgents across multiple availability zones in same region where high-availability is sought, with the understanding that this will incur Data Transfer – Inter AZ fees (refer to Data Transfer within the same AWS Region)
- [optional] Deploy Access Nodes across multiple availability zones in same region where high-availability is sought, with the understanding that this will incur Data Transfer – Inter AZ fees (refer to Data Transfer within the same AWS Region)

Cautions:

- Commvault does not support nor recommend distributing MediaAgents in a common GridStor™ configuration across regions, as latency will affect backup performance when the source host and MediaAgent are separated by WAN distances.
- Commvault does not support nor recommend distributing Access Nodes in a pool that spans regions, as latency will affect backup performance when the source host and Access Node are separated by WAN distances.

Considerations for Access Node Placement

The following table details where an Access Node can be placed relative to the workload it is protecting.

Placement of Access Node (relative to protected data)	Backup	Restore
Same Region	✓	✓
Different Region	✗	✗
Same Availability Zone (AZ)	✓	✓
Different Availability Zone (same region)	▲[1]	✗
Different Availability Zone (different region)	✗	✗

[1] Possible, will incur inter-AZ transfer fees during backup activity.

Architecture Recommendations For Amazon Access Nodes

For AWS environments with multiple accounts you can configure a virtualization client for Amazon Web Services to use a separate Admin account for data protection operations. This approach reduces the impact of backup operations and restore operations on production accounts. For more information on Cross-Account operations, see Processing for Cross-Account Operations.

Use of the Commvault IntelliSnap® configuration is highly recommended (requires Version 11 Service Pack 4 available as of June 2016 or forward) to improve backup and restore times. 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, Access Node backups of Amazon EC2 instances are crash consistent. To receive application consistent backups, leverage the AppAware feature available as of Commvault Version 11 Service Pack 7 released in March 2017. Additional information regarding AppAware can be found at Application-Aware Backups for the Virtual Server Agent.

If the AMI was sourced from the AWS Marketplace, any volumes that were deployed as part of that AMI cannot be targeted for snapshots. Only new volumes created and attached post-instance launch can be snapshotted. This is by Amazon Web Services design.

(Backup Copy / Streaming) Use of the Amazon S3 VPC Endpoint is highly recommended to improve throughput to/from Amazon S3 buckets.

(Backup Copy / Streaming) Configuring more than 10 readers for a single Access Node may cause snapshot mount operations to fail. Consider scaling out with multiple Access Nodes proxies if higher concurrency is required.

(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 Amazon EBS volumes.

(Backup Copy / Streaming) To restore from advanced Linux file systems such as EXT4, XFS and others, you can deploy a file recovery enabler by deploying a Linux-based Virtual Server Agent. When browsing data on advanced file systems, Commvault software will leverage the FREL to access the file system. Introduced in December 2018 with Commvault V11 SP14, the FREL can now be deployed from an AWS Marketplace AMI.

Current Best Practices for Amazon EC2 Instance Protection

Below is some standard benchmark comparison of backup methods (5 instances, 1 TB used data, incremental 20% change rate, and run over a 4-job schedule) to help illustrate some of the operation outcomes with more typical enterprise workloads running in the IaaS environment. The AWS example chart is included below based on Commvault V11 SP14 testing using Amazon EBS “Provisioned IOPS SSD” (io1) volumes.

BACKUP STATISTICS

Backup Method Used	Time Taken for Backup			
	Full	Incr 1	Incr 2	Dash Full
VSA for AWS	4 Hours	4 Hours, 15 min	4 hours, 10 min	18 min
File level via in guest	3 Hours, 20 min	25 min	28 min	19 min
Block Level via agent in guest	2 Hours, 30 min	21 min	20 min	4 min
Storage Accelerator	2 Hours	2 min	2 min	5 min

BACKEND STORAGE CONSUMED

Backup method used	Backend storage consumed				
	Full	Incr 1	Incr 2	DASH Full	Total Stoage
VSA for AWS	995 GB	201.15 GB	201.16 GB	3.39 GB	1.4 TB
File level via in guest	1008.15 GB	200.25 GB	200.25 GB	5.9 GB	1.4 TB
Block Level via agent in guest	1003.15 GB	200.25 GB	200.10 GB	0.8 GB	1.4 TB
Storage Accelerator	1008.15 GB	200.25 GB	200.25 GB	5.9 GB	1.4 TB

RESTORE STATISTICS

Backup Method used	Time Taken for Restore			
	Granular restore of 100 files	Granular restore of 4098 files	Full Volume, File level	Full Volume Block Level
VSA for AWS	1 hr, 20min	19 hours	x	3 hrs, 29 min
File level via in guest	2 min, 33 sec	31 min, 49 sec	3 hrs, 45 min	x
Block Level via agent in guest	8 min, 48 sec	56 min, 42 sec	x	2 hrs, 43 min

ACCESS TIME (LIVE MOUNT)

Backup method used	Time to run DIR /S on a "Live Mounted Volume
VSA for AWS	26 min
File level via agent in guest	x
Block level agent in guest	10.5 min

These tests reveal that because of the movement of data during agent-less snapshots from Amazon EBS volumes attached to Amazon EC2 instances to a non-selectable Amazon S3 bucket, and the rehydration of data back from this non-selectable shared Amazon S3 bucket to the selected Amazon EBS io1 volumes attached to the Commvault Access Node, the initial full backup can take an elongated time.

Subsequently backups via the Amazon EBS volume snapshot method, despite having incremental changes, are subject to the same data movement, therefore incremental agent-less backups can consume a considerable time. Once this movement of data from Amazon EBS to Amazon S3 back to Amazon EBS is completed then the Commvault platform can retrieve, deduplicate, and backup the data to a target.

Change Block Tracking Improvements

As of Commvault v11 Feature Release 11.20, Commvault supports the Amazon ES direct read APIs which provide Change Block Tracking (CBT) capability to EC2 and EBS protection. Commvault has enabled CBT by default for all new Amazon EC2 subclients created on or after release 11.20. Additionally, you may enable CBT on your existing Amazon subclients. Commvault expects to enable CBT on all subclients in future by Feature Release update.

Change Block Tracking reduces backup time by 90% commvault lab testing

The following test results from Commvault labs demonstrates the **significant** improvement of backup time using CBT.

	Total time taken	Transfer time taken
Full (CBT **off**)	9 hrs, 22 mins	7 hrs 15 mins
Full (CBT **on**)	1hr 3 mins	58 mins
Incremental 1 (CBT **off**)	5 hrs, 21 mins	4hrs, 30 mins
Incremental 1 (CBT **on**)	1 hr, 27 mins	1 hr, 22 mins
Incremental 2 (CBT **off**)	4hrs, 27 mins	4hrs, 23 mins
Incremental 2 (CBT **on**)	1hr, 44mins	12 mins

Due to improved **Change Block Tracking (CBT)** capabilities, the best approach to protect data is to utilize a hybrid protection strategy that comprises:

- Utilize Commvault IntelliSnap for AWS to create low RPO and RTO point-in-time cloud-native snapshots of Amazon EC2 instances and Amazon EBS volumes. These can be used to quickly recover entire Amazon EC2 instances and entire Amazon EBS volumes. These provide an RPO of minutes to hours, depending on frequency of snapshots. To keep costs low, these snapshots can be retained for a short period of time ranging from hours to days purely dependent on business requirements and cost.
- Optionally utilize Commvault IntelliSnap snapshots to perform Point in Time (PiT) backup, but automatically copy the snapshot to Commvault deduplicated S3 storage. Snapshots are mounted on a Access Node and unique or new data is streamed off in a completely agentless approach. Utilizing the enhanced EBS Direct Read APIs, this approach can be utilized on any EC2 instance size and data volume. This approach allows a tiered model with a small number of native snapshots for rapid recovery, followed by long-term retention in Commvault deduplicated S3 storage.
- Optionally utilize block-level backups using agents-in-guest to protect the larger file systems with a typical RPO of 12 - 24 hours. These streaming backups will move only incremental block changes at the file system level and coupled with source-side deduplication will minimize the data movement. The data can be stored in a deduplicated format at a destination such as Amazon S3.-IA storage for a longer retention period ranging from days to months. These backups can be further migrated to lower cost storage such as Amazon S3 Glacier. for elongated retention.
- For remote office location, 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 Amazon S3, completely bypassing the MediaAgent. This is achieved by installing and enabling the Storage Accelerator feature on the client for direct communication to a storage target and will speed up the backup and reduce costs in these situations.
- Consider the scope of your recovery. Utilize single file, folder or data volume recover and attach to limit the scope of recovery to only the affected application data. Operating system data and drives can remain untouched to speed recovery.
- Attaching restored data volumes to newly created application hosts is also recommended for application upgrade testing, by provisioning new application infrastructure and then attaching restored data volume(s).

This type of hybrid architecture will provide protection from multiple types of scenarios – entire instances and volumes can be recovered near instantaneously while granular recovery of individual files, folders and application data can be performed quickly from the Commvault platform, utilizing browsing of indexed data without having to stage and recover the entire volume first. This hybrid approach also maintains a balance of costs since snapshot retention and life-cycling can be managed via Commvault IntelliSnap and the data in the longer-term retention storage can be stored in a deduplicated format. Some Amazon EC2 instances may simply utilize one approach in its entirety depending on the requirements for speed for backup, RPO, RTO and retention. To further investigate and discuss this architecture, please work with your Commvault Pre-sales Systems Engineer.

Amazon RDS Protection

The Amazon Relational Database Service (RDS) offers customers access to databases without having to manage the infrastructure supporting the databases. However, eliminating the need to administer the underlying operating system also comes with a drawback. AWS restricts users from accessing the operating system and restricts privileged connections to the databases. This limits the options that are available for database protection.

Using Amazon RDS Snapshots

AWS does not allow for privileged connections to the Amazon RDS instances. AWS controls the backup and recovery processes for these databases via snapshots, not via the individual vendor's proprietary tools such as RMAN (Oracle) or Microsoft SQL Management Studio (SQL Server). Using the Commvault Amazon Access Node, orchestration of the Amazon RDS snapshots is performed. An Amazon RDS virtualization pseudo-client is created. Here the Amazon RDS instances can be automatically discovered. Commvault will invoke the Amazon RDS snapshots and retain the snapshots based on configured storage policy. The Commvault software can also orchestrate the sharing of these Amazon RDS snapshots to another AWS region for disaster-recovery purposes. It's important to note that Amazon RDS snapshots are always full backups. Another limitation with the Amazon RDS snapshot backup method is that there is no control over the transaction logs or archive logs, so there is no ability to restore to a specific point-in-time, only to the time of the snapshot.

Additionally, the Commvault Amazon Access Node is encryption aware and allows for an encrypted RDS instance to be protected, replicated, and accessed in its new destination without the need to decrypt/re-encrypt the data. This provides the confidence that data in-flight or at rest is always handled in an encrypted state via Key/Secret or IAM-based authentication.

Please check the online documentation for supported Amazon RDS instances: Overview - Protecting Amazon RDS Instances.

SQL Consistent Streaming Backup of Amazon RDS SQL Instances

While most of the Amazon RDS offerings can be protected via cloud-native snapshots, and consequently the Amazon Access Node, an Amazon RDS SQL instance can also be protected via the Commvault SQL Server iDataAgent in addition to snapshots (as of Commvault V11 SP12, released in June 2018). This method does not utilize Amazon RDS snapshots, but instead performs an export of the SQL Server database. This export can be used to port the data to another Amazon RDS instance as well as between SQL Server databases that may reside either on-premises or in the cloud. While this method does provide the flexibility of a portable backup, it will generally be slower than the snapshot orchestration outlined earlier. Please check Commvault online documentation for details regarding this Amazon RDS SQL instance protection method: Backup and Restore Amazon RDS SQL Databases.

Alternative Protection and Migration Methods for Amazon RDS Oracle Instances

The Commvault software includes workflows to allow the migration of Amazon RDS Oracle instances to either a Microsoft Azure Database or to an Oracle instance running in a Microsoft Azure VM. This provides flexibility for your Oracle workloads and enables a multi-cloud strategy. Please see Commvault online documentation for details: Amazon RDS Database Migration to an Azure Database.

In addition to the Amazon RDS snapshot protection method described in the previous section, as of Commvault V11 SP13 (September 2018) an Amazon RDS Oracle instance can also be protected using the Oracle Data Pump utility installed on a Windows Access Node. The protected Oracle database can subsequently be restored to an Amazon RDS Oracle instance, a traditional Oracle database instance running in an Amazon EC2 instance, or to a traditional Oracle database instance running on-premises or in another Cloud provider. For more details, see Commvault online documentation: Amazon RDS Oracle Protection Using Oracle Native Export Utility.

Storing Amazon RDS Data Pump exports allows long-term retention of database copies in S3, Glacier, and Deep Archive. This represents a significant cost reduction over RDS snapshots

---

Database-Consistent Streaming Backup of Amazon RDS MySQL and PostgreSQL Instances

As of Commvault V11 SP13 (September 2018), the Commvault software platform can now perform streaming backups of Amazon RDS MySQL and Amazon RDS PostgreSQL instances using the Commvault MySQL iDA or Commvault PostgreSQL iDA installed on a Linux Access Node. This Access Node must be an Amazon EC2 instance running in the same AWS region as the Amazon RDS instance to be protected. The protected Amazon RDS MySQL and Amazon RDS PostgreSQL instances can be subsequently restored to an Amazon RDS instance, to a traditional database instance running in an Amazon EC2 instance, or to a traditional database instance running on-premises or in another Cloud provider. For more details, see Commvault online documentation:

Amazon RDS MySQL Protection Using Native Database Export Utility Amazon

RDS PostgreSQL Protection Using Native Database Export Utility

Maintaining Database Performance and Resilience With DB Parameter Groups

Amazon provide DB Parameter groups to ensure that consistent data configuration, security and tuning is applied across like database types. Commvault is DB Parameter group aware, and as of Commvault V11 SP17 (September 2019), will restore DB Parameter groups on RDS restores.

Refer to Restoring an Amazon RDS Backup for restore processes.

Amazon S3 Object Storage Backup

As of December 2016, Commvault Version 11 includes the ability to perform backups of Amazon S3 object storage created by third party applications. This capability allows Commvault® software to capture the data contained inside an Amazon S3 bucket, allowing for restore back to an Amazon S3 bucket or file system client.

Additional information on Amazon S3 backups can be located using the link below.

- Amazon S3 Backup

When to Use Amazon S3 Backup

- Backing up object storage – Protect objects in Amazon S3 which have been created by third-party applications.

When Not to Use Amazon S3 Backup

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

Architecture Recommendations

- For large datasets, consider using multiple subclients to increase scan performance and lower the amount of time taken to traverse 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.

Amazon FSx for Windows File Server Protection

As of March June 2019, Commvault Version 11 Feature Release 11.16 includes the ability to perform backups of Amazon FSx for Windows File Server (Amazon FSx) SMB snares. Commvault performs granular protection and recovery of files/folders across accounts and regions. Commvault utilizes a windows-based Access Node to mount and protect selected content within the FSx share .

Additional information may be found at:

- Amazon FSx for Windows File Server
- Getting Started with Amazon FSX for Windows File Server
- Adding a NAS Server for Network Share Backups

When to Use Amazon FSx Backup

- Protection of fully managed Amazon FSx SMB shares.

When Not to Use Amazon FSx Backup

- Protection of self-hosted SMB shares running on Amazon EC2 Instances (use VM Group protection[)

Architecture Recommendations

- Leverage your existing Amazon Access Nodes to protect Amazon FSx shares.
- Utilize Access Nodes within the same availability zone as your FSx shares to avoid inter-AZ data transfer fees.

Amazon EFS (Amazon Elastic File System) Protection

As of March June 2019, Commvault Version 11 Feature Release 11.16 includes the ability to perform backups of Amazon Elastic File System (Amazon EFS) NFS exports. Commvault performs granular protection and recovery of files/folders across accounts and regions. Commvault utilizes a linux-based Access Node to mount and protect selected content within the EFS export.

Additional information may be found at:

- Cloud Apps – AWS EFS (Amazon Elastic File System)
- Getting Started with EFS
- File Servers – Network Share

When to Use Amazon EFS Backup

- Protection of Amazon EFS NFS-based exports.

When Not to Use Amazon EFS Backup

- Protection of self-hosted NFS exports running on Amazon EC2 Instances.

Architecture Recommendations

- Leverage your existing Amazon Access Nodes to protect Amazon EFS exports.
- Utilize Access Nodes within the same availability zone as your EFS export to avoid inter-AZ data transfer fees.

Amazon DocumentDB Protection

As of March June 2020, Commvault® Version 11 Feature Release 11.19 includes the ability to perform backups of Amazon DocumentDB MongoDB NoSQL databases. Commvault protects DocumentDB instances across multiple accounts and regions. Commvault utilizes native integration with Amazon to create snapshots and perform full cluster recovery from a single self-service interface.

Additional information may be found at:

- Command Center – Amazon DocumentDB
- Getting Started With Amazon DocumentDB Backups
- Restoring a DocumentDB Cluster

When to Use Amazon DocumentDB Backup

- Protection of Amazon DocumentDB clusters (data structures, MongoDB workloads)

When Not to Use Amazon DocumentDB Backup

- Protection of self-built MongoDB instances running on Amazon EC2 instances (or equivalent).
- Use the Commvault MongoDB agent for infrastructure built and managed Mongo instances.

Architecture Recommendations

- Leverage your existing Amazon Access Nodes to protect Amazon DocumentDB clusters.
- Utilize Access Nodes within the same availability zone as your cluster to avoid inter-AZ data transfer fees.

Amazon DynamoDB Protection

As of March June 2020, Commvault Version 11 Feature Release 11.19 includes the ability to perform backups of Amazon DynamoDB NoSQL key-value / document stores. Commvault protects DynamoDB instances across multiple accounts and regions. Commvault performs full and incremental protection of complete clusters or individual DynamoDB tables. Commvault leverages native integration with Amazon to create snapshots and perform full cluster and partial (table-level) recovery from a single self-service interface.

Additional information may be found at:

- Command Center – Amazon DynamoDB
- Getting Started With Amazon DynamoDB Backups
- DynamoDB Restores (full table, cross account)

When to Use Amazon DynamoDB Backup

- Protection of fully managed Amazon DynamoDB clusters

When Not to Use Amazon DynamoDB Backup

- Protection of self-built key-value or document stores.
- Use the Commvault MongoDB agent for infrastructure built and managed Mongo instances.

Architecture Recommendations

- Leverage your existing Amazon Access Nodes to protect Amazon DynamoDB clusters.
- Utilize Access Nodes within the same availability zone as your cluster to avoid inter-AZ data transfer fees.

Amazon Redshift Protection

As of March June 2020, Commvault Version 11 Feature Release 11.19 includes the ability to perform backups of Amazon Redshift datawarehouses. Commvault protects Redshift instances across multiple accounts and regions. Commvault utilizes native integration with Amazon to create snapshots and perform full cluster recovery from a single self-service interface.

Additional information may be found at:

- Command Center – Amazon Redshift
- Getting Started With Amazon Redshift Backups
- Restoring a Redshift Cluster

Architecture Recommendations

- Leverage your existing Amazon Access Nodes to protect Amazon Redshift clusters.
- Utilize Access Nodes within the same availability zone as your cluster to avoid inter-AZ data transfer fees.

Amazon EKS Protection

As of December June 2020, Commvault Version 11 Feature Release 11.20 includes the ability to perform backups of Amazon Elastic Kubernetes Service (Amazon EKS) workloads. Commvault will protect all containers and api-resources associated with stateful EKS applications. Commvault leverages the Container Storage Interface (CSI) and ebs.csi.aws.com driver to provide crash-consistent snapshots of container storage.

Additional information may be found at:

- Command Center - Kubernetes
- Auto-protecting new applications with label selectors
- Migrating applications and data cross-cluster or cross-region

When to Use Amazon EKS Backup

- Protecting stateful applications with persistent volumes and data that exists beyond the container invocation.
- Protecting Kubernetes orchestrated containers and applications.

When Not to Use Amazon EKS Backup

- Protecting **stateful** data written by containers to other Amazon services (i.e. Amazon S3, Amazon EFS).
- Using customized container-based services (Elastic Container Services, Amazon Fargate)..

Architecture Recommendations

- Leverage your existing Amazon Access Nodes to protect EKS clusters.
- Utilize Access Nodes within the same availability zone as you EKS cluster to avoid inter-AZ data transfer fees.

Application Migration

Commvault can assist in application migration efforts when shifting from on-premises facilities to Public Cloud providers such as AWS. By leveraging the power of the data management platform, workloads can be migrated through several methods.

Virtual Machine Restore & Convert (Lift and Shift to AWS)

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

With this image, the Virtual Server Agent can then restore and convert (VMware and Hyper-V) the Virtual Machine into Amazon EC2 instances directly, and the process can handle single or multiple Virtual Machines.

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

Oracle Application Migration

Introduced in Version 11 Service Pack 5, the Oracle iDataAgent can now perform a migration of an Oracle Instance from a physical or virtual machine on-premises into an Amazon EC2 instance.

The process requires the Oracle iDataAgent for Linux to be deployed on the source machine. At which point the migration process will automate:

- Capturing the application server physical configuration
- Protecting the application data
- Provisioning the compute instance and storage
- Perform RMAN conversion if source is big endian platform
- Restoring the data
- Validating the restore
- Continuously applying changes from the production database backups to the migrated database by using the Live Sync feature.

For more information regarding the migration of an Oracle instance into an Amazon EC2 instance, see Commvault online documentation: Oracle Database Application Migration to an Amazon EC2 Instance.

The Commvault software also can migrate a traditional Oracle database instance into an Amazon RDS Oracle instance. See Commvault online documentation for more details: Oracle Database Application Migration to an Amazon RDS Database.

Microsoft SQL Server Migration

Introduced in Version 11 Service Pack 7, the Microsoft SQL Server iDataAgent can now perform a migration of a SQL Server database from a physical or virtual machine on-premises into an Amazon EC2 instance. Microsoft SQL Database Migration.

Kubernetes Application Migration

Introduced in Version 11 Feature Release 20 (11.20), the Kubernetes Virtual Server Agent can now perform migration of Kubernetes orchestrated containers and persistent data between Kubernetes clusters. Protected applications, configuration (manifests, API resources, secrets) and data (persistent volumes, persistent volume claims) may be restored to any supported Kubernetes Cluster.

See Migration Use Cases for Kubernetes for additional details regarding:

- Migration from on-premises to managed Amazon Elastic Kubernetes Service (Amazon EKS).
- Migration between Amazon EKS clusters, regions or availability zones.
- Migration between Amazon EKS in region and Amazon EKS on Outposts (or vice versa)

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.

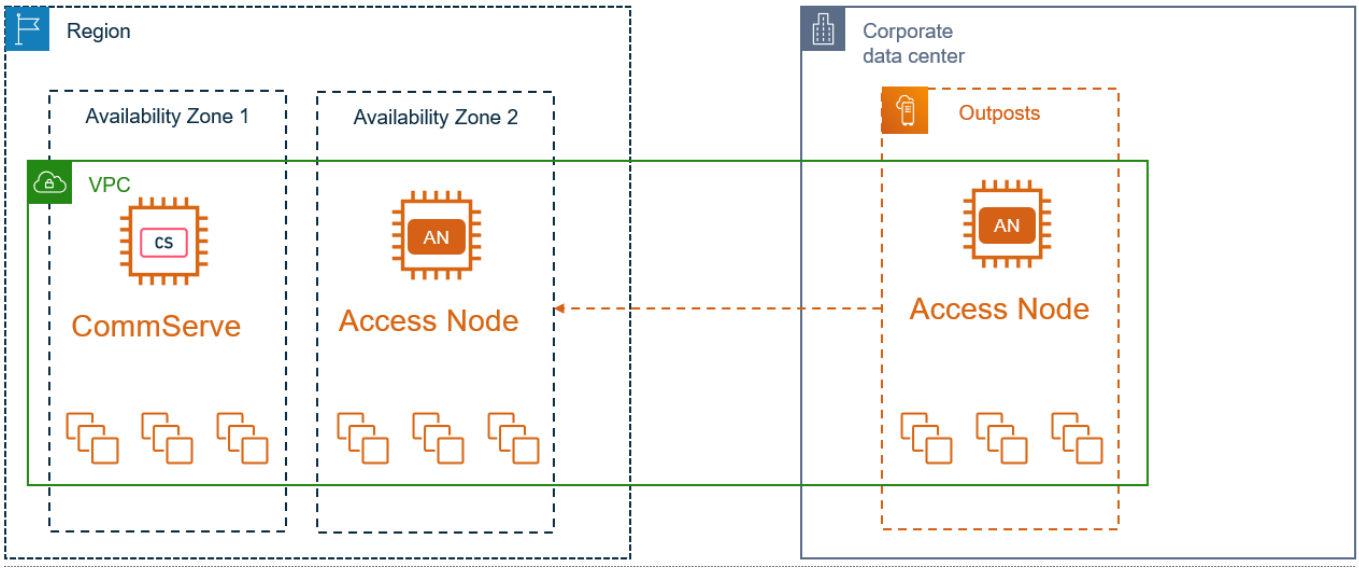
The process requires the supported iDataAgent to be deployed on both the source instance, and the destination Amazon EC2 instance.

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

Extend to the edge with AWS Outposts

In December 2019, Amazon announced general availability of AWS Outposts (refer Announcing General Availability of AWS Outposts). AWS Outposts provides local compute, storage and networking resource from Amazon, within remote owned/operated data centers.

AWS Outposts is designed for workloads that require low-latency or must remain on-premises due to regulatory or data residency requirements. Commvault announced support for Amazon Outposts in December 2019 (refer Commvault Data Protection Software Fully Tested and Validated to Support AWS Outposts) and continues to extend this capability in line with Amazon advancements.



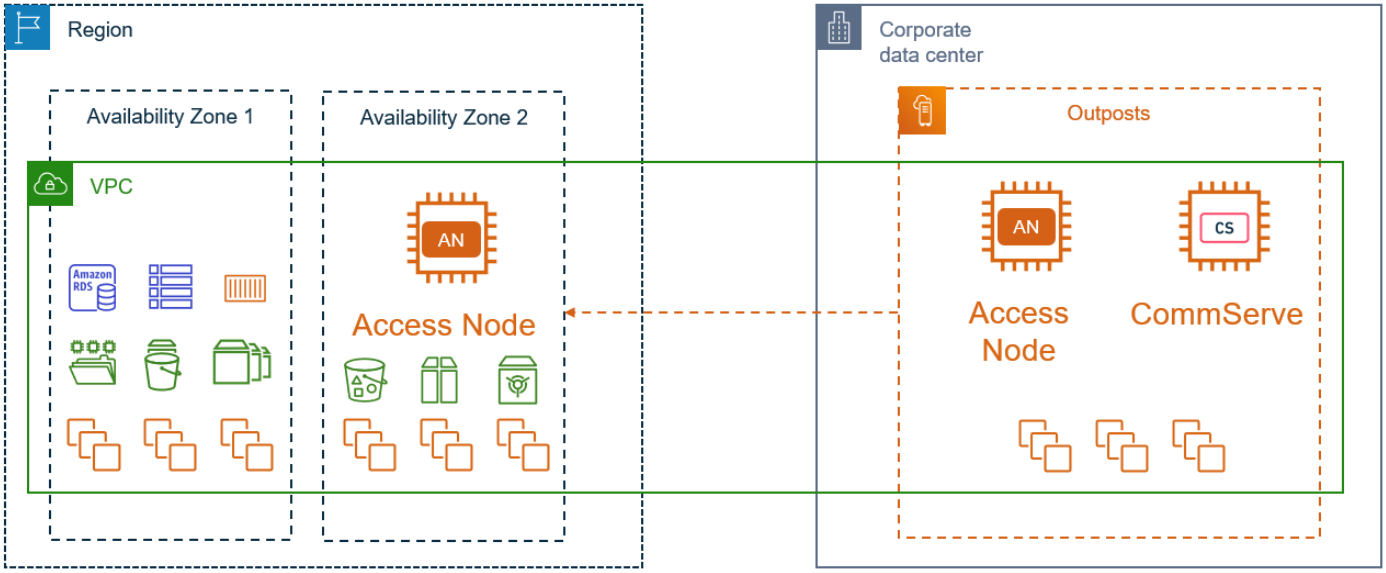
Commvault supports Amazon Outposts for two primary use-cases:

- Commvault Data Management components are fully supported when running on Amazon Outposts
- Commvault can protect Amazon Outposts data, both within the Outposts and back to the Amazon region.
- Commvault protects the following Amazon Outposts workloads – EC2 instances, EBS volumes.
- Refer to AWS Outposts Support for additional details.

AWS Outposts supports General purpose (M5), Compute optimized (C5), and Memory optimization (R5) instances which are all supported and recommended instance types for Commvault CommServe, MediaAgents, and Access Nodes (refer to AWS Outposts – Features).

Note: At the time of writing (June 2020), AWS Outposts does not support the Simple Storage Service (S3). If data is required to remain on-premises (within the Outpost), the an EBS-based Disk Library may be utilized. All S3-based Cloud Libraries will reside within the Amazon region remote to the Amazon Outposts workload. Additionally, an EBS snapshots created on the Outpost will be stored remotely within the Amazon region (see AWS Outposts – FAQs)

When utilizing AWS Outposts to host Commvault workloads, a hybrid protection approach is recommended. Commvault Access Nodes may be deployed locally within the Availability Zone they protect (see below)



How to get started

Getting start with Amazon Marketplace

Commvault publishes AMI images for both a CommServe and Access Node (MediaAgent + Virtual Server Agent) within the Amazon Marketplace.

Navigate to the Commvault Marketplace Page to obtain the image for your environment.

Need some help getting started, view the following quickstart - Getting Started with Commvault in Amazon Marketplace.

Remote Access / Bring Your Own Software

As with all IaaS offerings, remote access to Virtual Machine instances can be achieved with your favorite protocol/ software (RDP for Windows, SSH for Linux instances) and Commvault® module deployment can be achieved with the current procedures listed in documentation.

Installation Basics

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

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

CommServe Disaster Recovery Solution Comparison

The following link covers CommServe® DR Solution comparisons for building a standby DR CommServe in the Cloud, or simply restoring on-demand (DR Backup restore): CommServe Disaster Recovery.

Pre-Packing Commvault Software within a VM Template

For environments where deployment time is reduced by preparing software and configuration within VM templates, such as Amazon Machine Images (AMIs), the Commvault iDataAgents can also be deployed in Decoupled mode. This means that the iDataAgent is deployed within the instance but will only be activated upon registration with the CommServe.

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

- Custom Packages

Automating Deployment with Continuous Delivery

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

Refer to the Commvault ansible library to automate your Commvault operations
github.com/Commvault/ansible

Automate your RESTful development and testing with the Commvault POSTMAN collection
github.com/Commvault/Rest-API-Postman-Collection

Use the Commvault Python SDK to automate repeatable Commvault operational tasks
github.com/Commvault/cvpysdk

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

- Unattended Installation

For more information on using Commvault software's XML / REST API interface to control configuration post-deployment, please refer to the online documentation links below to review options available for each iDataAgent:

- REST API – Overview
- Command Line – Overview

Cloud Library Configuration

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

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

The link below lists all the supported direct cloud storage targets.

- Supported Cloud Storage

The link below covers cloud storage target setup and management.

- Cloud Storage - Overview

Details on performance tuning are covered below.

- Cloud Library Performance Tuning

Unsupported Cloud Storage Configurations

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

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

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

For devices that are not publicly accessible, please contact your account manager.

Additional Resources

Documentation

Cloud Storage

The Cloud Storage documentation covers technical procedures and information on Supported Cloud Targets, Advanced procedures, Troubleshooting and FAQ sections for Commvault customers.

AWS IAM Permissions

All required Amazon user permissions can be accessed from documentation here:

- JSON Templates for IAM Role Definition and User Permissions
    - Performing Backups to an S3 Library
    - Performing Backups with Restricted Access
    - Amazon Web Services User Permissions for Backups and Restores
    - Amazon Web Services User Permissions for RDS Backup and Restores
    - Amazon Web Services User Permissions for VM Conversion
    - Amazon Web Services User Permissions for DocumentDB Backup and Restores
    - Amazon Web Services User Permissions for Redshift Backup and Restores
    - Amazon Web Services User Permissions for RDS Snapshot-based Backup and Restores
    - Amazon Web Services User Permissions for RDS Dump-based Backup and Restores
    - Amazon Web Services VM Import Role
- AWS Permissions – Change Block Tracking

A full breakdown of all permissions required by Commvault for each task may be found at Amazon Web Services Permission Usage.

For quick and easy **IAM Role** and **Policy** definition, you may download these permission definitions from the Commvault github repo – here github.com/Commvault/aws-permissions.

Revision history

Version	Data	Changes
1.0	March 2015	• Added Commvault IntelliSnap® functionality into VSA for AWS
1.1	May 2015	• Added Migration to the Cloud use case and Application Migration section
1.2	June 2015	• Added Live Sync DR for Amazon EC2 and revised DR structure • Added Performance test results for AWS backup methods: VSA & agent-in-guest
1.3	July 2015	• Minor updates
1.4	August 2015	• Added new links to video content
1.5	September 2015	• Added Selecting the right Storage Class section
1.6	November 2015	• Updated requirements for Disaster Recovery to the Cloud • Added Unsupported Cloud Configurations section
2.0	March 2016	• New Virtual Server Agent methodologies, deployment and changes to use cases • Backup to the Cloud, DR to the Cloud and Protection in the Cloud scenarios • Added Automating Deployment with Puppet/Chef
2.1	June 2016	• Added Commvault IntelliSnap functionality into VSA for AWS
2.2	September 2016	• Added Migration to the Cloud and Application Migration
2.3	March 2017	• Added Live Sync DR for Amazon EC2. • Added Amazon S3./Blob storage backup feature
2.4	April 2017	• Updated Cloud pricing considerations and modelling
2.5	August 2017	• Added Windows 2016 (v11 SP7) to CS and MA
2.6	September 2017	• Added Oracle E-Business Suite migration functionality
2.7	February 2018	• Updated sizing for Media Agents running in AWS
2.8	May 2018	• Updated MediaAgent Instance sizing and updated S3 classes, EC2 instances.
2.9	October 2018	• Updated document with SP12 and SP13 functionality
2.10	January 2019	• Updated document with SP14 functionality • Revised Amazon EC2 sizing recommendations for Media Agents

2.11	March 2019	• Updated document with SP15 functionality • Additional Best Practices for EC2 instance protection • Architecture recommendations for Cross-Account Operations
2.12	March 2020	• Updated document with SP16-SP19 functionality • Added AWS encrypted snapshot copy support for RDS and EBS volume snap • Added AWS to AWS Live Sync from Command Center functionality • Revised AWS EC2 sizing requirements for MediaAgents, Virtual Server Agents • Added support for ARM Linux MediaAgents and Access Nodes
2.12	June 2020	• Updated document with Feature Release 20 (FR20) functionality • EBS Direct Read API support (Change Block Tracking). • Auto-scaling AWS Access Nodes for backup support. • Added protection of Elastic Kubernetes Service (EKS) support. • Simplified MediaAgent and Access Node sizing and placement. • Added AWS Snowcone support.

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

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

Solutions, References and Videos

Datasheets

Whitepaper: Commvault for Amazon Web Services (AWS) (link) **new**

Datasheet: Commvault for VMware® Cloud on Amazon Web Services – Datasheet (link) **new**

Commvault.com/aws – Amazon Web Services (AWS) Solutions (link)

Solution Briefs

Backup Done Differently: One Solution to Solve & Simplify All Your Backup Needs (link) **new**

Disaster Recovery On Demand: Keep Your Enterprise Up and Running (link) **new**

Commvault Data Protection Software Fully Tested and Validated to Support AWS Outposts. (link)

Government Cloud Solutions from Commvault and AWS (link)

Backup to AWS (link)

References

Webinar: Why Parsons Considers Its Data Protection Strategy a Business Advantage (AWScloud) new

Webinar: Ransomware: Staying ahead of the global threat (commvault.com) new

University of Canberra Integrates Commvault Complete Backup & Recovery With Amazon S3. (YouTube, link)

College of the Holy Cross: Long-Term Retention with Commvault and AWS Cloud (YouTube)

Moving Forward Faster: How Monash University Automated Data on AWS with Commvault (YouTube)

Videos

Getting Started with Commvault in Amazon Marketplace (Vidyard) **new**

AWS re:Invent 2019: Ensuring data protection readiness across hybrid environments (ENT323-S)

Backups to AWS Made Simple With Commvault (YouTube)

Multi-Cloud Data Protection Made Simple With the Commvault Command Center (YouTube)

5 Real Life Pieces of Advice to Make Your Cloud Strategy a Success (YouTube)

2 Clicks to the Cloud with AWS and Commvault (YouTube)
Focuses on creating an Amazon S3 container and configuring as a Cloud Library within Commvault® v11.

Index

Best Practices 14, 16, 29, 35
Commvault
  Cloud Library 9, 40, 42, 43, 48, 62, 66
  CommServe 19, 20, 21, 23, 37, 38, 61, 62
  MediaAgent 4, 5, 6, 7, 8, 9, 10, 11, 14, 17, 19, 20, 21, 22, 24, 25, 26, 27, 29, 30, 31, 32, 33, 35, 36, 37, 40, 45, 46, 50, 51, 52, 54, 55, 56, 57, 61, 64
  Storage Accelerator 45, 57
  Virtual Server Agent 7, 10, 18, 19, 30, 31, 32, 35, 48, 53, 55, 58, 60, 61, 64
Compute
  EC2 ..... 6, 7, 8, 11, 18, 19, 20, 21, 22, 23, 24, 30, 31, 32, 36, 38, 39, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 59, 60, 61, 64, 65
Database
  Oracle 6, 11, 50, 51, 52, 53, 58, 59, 60, 61, 64
  RDS 57, 58, 59, 61, 65
  SQL Server 11, 50, 58, 61
Networking
  Availability Zones 29, 47, 51
  Direct Connect 6, 7, 8, 39, 41
  Regions 46, 47, 51
Replication
  Live Sync 8, 33, 46, 47, 48, 60, 64, 65
Security, Identity & Compliance
  IAM 51, 58, 63
  Key 51, 58
  KMS 51
  Secret 51, 58
Sizing
  CommServe 23
  example 36
  MediaAgent 30
  Virtual Server Agent 32
Storage
  Deep Archive 9, 44
  EBS 11, 18, 19, 21, 22, 23, 30, 31, 32, 36, 38, 43, 51, 52, 53, 54, 55, 56, 57, 65
  Glacier 4, 9, 12, 13, 14, 15, 16, 17, 42, 44, 45, 57, 66
  Intelligent-Tiering 9, 15, 16, 45
  Reduced Redundancy Storage 9
  S3 ..... 4, 5, 7, 9, 12, 13, 14, 15, 16, 17, 30, 31, 37, 38, 40, 42, 43, 44, 45, 48, 50, 51, 52, 54, 55, 56, 57, 59, 63, 64, 66
  S3 One Zone-IA 12, 13
  S3-IA 9, 12
  Snowball 4, 40, 41
  VPC Endpoint 40, 50, 52, 55
Use-cases
  Disaster Recovery 8
  Migration 5
  Move Data 4
  Protection 6