COHESITY

Using Cohesity with Amazon Web Services (AWS)

Achieve your long-term retention and archival objectives for secondary data

Cohesity DataPlatform is a hyperconverged secondary data and application solution that simplifies IT and business operations by consolidating your backup, files, objects, test/dev, and analytic datasets onto a web-scale platform. The combination of Cohesity and native-cloud integration with Amazon Web Services (AWS) offers your enterprise significant cost and agility advantages over tape backups for long-term retention and archival of critical data and applications.

With a flexible, pay as you grow cost model, Cohesity automates workflows based on defined business policies while taking advantage of built-in public cloud vendor protections, reducing IT burdens even as the pace of your business accelerates.

Achieving your long-term data retention and archival objectives for secondary data is simpler with Cohesity. Cohesity integration with AWS enables your organization to

- **Save time and lower TCO** – Leverage cloud scalability for the long-term retention and archival of secondary data without cloud gateways and disparate point solutions connecting to the cloud.
- **Improve efficiency –** Use advanced Cohesity algorithms for true global deduplication—across clusters, workloads, and protocols and compression in the cloud to optimize capacity efficiency and lower the cost of AWS for archival.
- **Derive greater value from your data** – Gain fast access and retrieval of data from AWS to make data more useful to business teams seeking to uncover meaningful insights from previously untapped data.

This technical brief describes how to run Cohesity DataPlatform with Amazon S3 and Amazon Glacier for long-term data and application retention and archival.

Cohesity and AWS for Long-Term Retention and Archiving

CloudArchive is the feature built into Cohesity DataPlatform giving you the flexibility to extend from your core data center to AWS (see Figure 1). Cohesity CloudArchive, which holds data and applications infrequently accessed, is regulated by data policies IT has set based on business or industry security and compliance requirements.

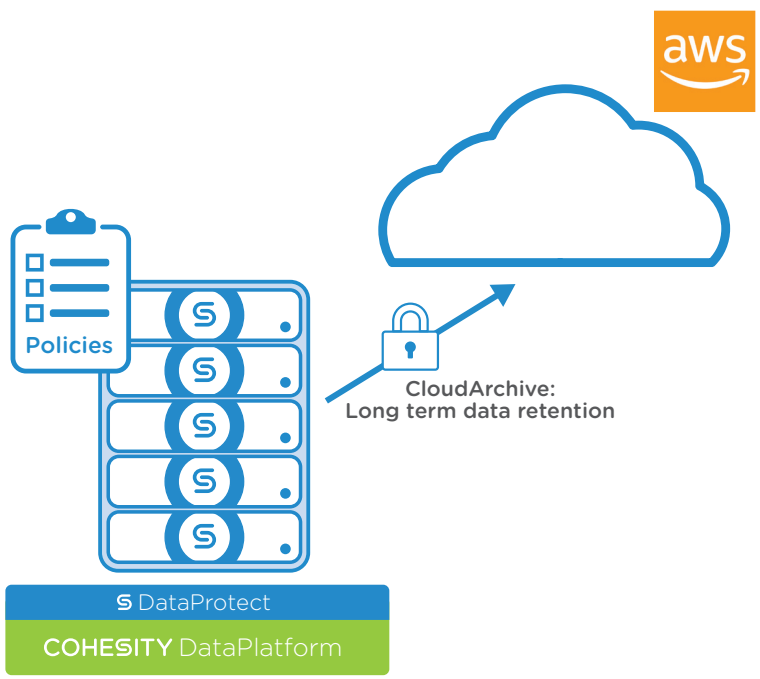

Figure 1. Cohesity CloudArchive works seamlessly with AWS

CloudArchive reduces reliance on tape and lowers TCO while providing an easy way to move data and applications to cloud storage, then retrieve them data back on-premises from the cloud and recover to a different site.

How It Works

Cohesity protects a wide variety of workloads—from virtual and physical to databases and NAS. All unstructured workloads are first backed up to on-premises Cohesity clusters before data is pushed to cloud storage based on the archive schedule (see Figure 2).

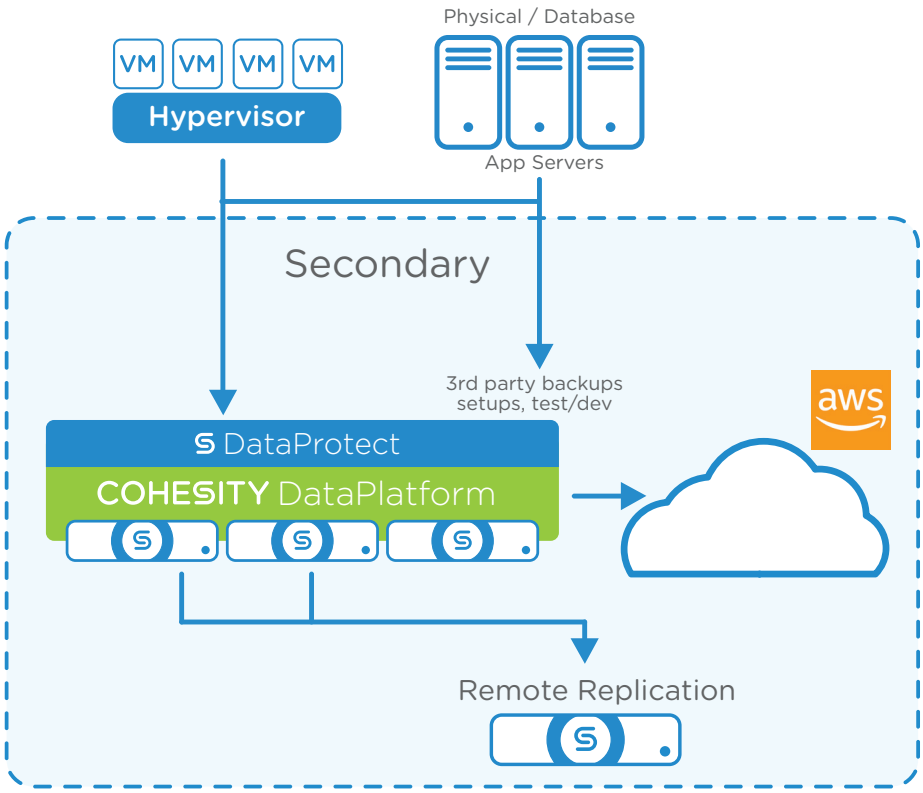

Figure 2. Cohesity CloudArchive supports the same workloads on-premises and in the cloud.

Cohesity DataPlatform divides archived objects into smaller segments, then compresses and encrypts them before transferring all of them to the cloud. Each segment goes through a lookup in a fingerprint database to compare its metadata hash. If the fingerprint is a match, only the metadata is archived. If no fingerprint matches, both the data and metadata are archived. This helps to optimize how much data is transferred to the cloud.

Native APIs for Amazon S3 and Amazon Glacier are used by Cohesity to upload data to the cloud. The secondary data platform encrypts all incoming data and decrypts all outgoing data in real time. It uses the industry-standard 256-bit Advanced Encryption Standard (AES) algorithm, accelerated by the Intel CPU's AES instruction set, and standard Cipher Block Chaining (CBC) mode-based encryption. The encryption/decryption process is automated within Cohesity, yet transparent to all inbound/outbound protocols and applications. When data is archived to Amazon S3 or Amazon Glacier, Cohesity encrypts the data using a cloud encryption key and transfers data over secure channels (HTTPS) to the cloud target.

Preparing to Use CloudArchive

There are three steps to enabling CloudArchive for archival and data recovery.

Step 1: Register Cloud Storage with Cohesity DataPlatform

To use Amazon S3 or Amazon Glacier with Cohesity, you must first register an external target with an on-premises Cohesity cluster (see Figure 3), following these instructions:

1. Access the user interface (UI) for Cohesity DataPlatform by pointing your browser to the IP address of any one of the nodes in an on-premises Cohesity cluster.
2. Choose **External Targets** under the **Platform** tab to access the external target registration page.
3. Register the cloud target by providing your credentials

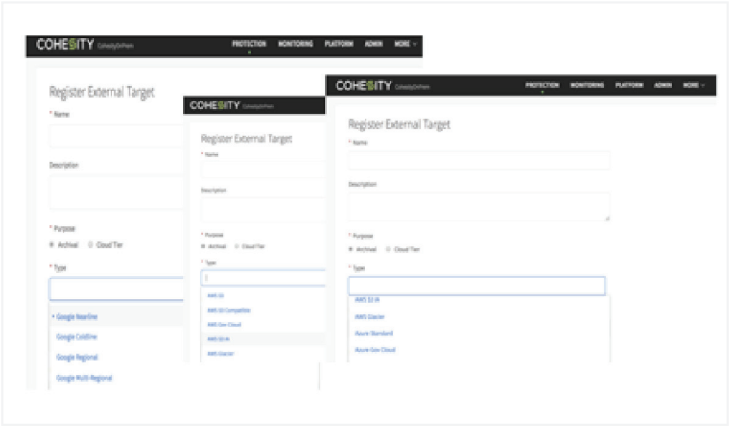

Figure 3. Registering cloud storage with Cohesity.

Step 2: Set a Job Protection Policy to Archive to Cloud Storage

A Cohesity protection policy defines how virtual and physical servers, databases, and other unstructured data and applications will be protected, how frequently they will be backed up, and how long backups will be retained. Each job can have a unique protection policy that allows the user to incorporate the cloud storage external target created in step one as an archive target within the desired retention period. You can set a new protection policy (see Figure 4) by following these instructions:

1. Select **Policy Manager** under the **Protection** tab of the UI.
2. Name the new policy, then set your backup, replication, and archival requirements.

In the example, a user is creating a policy to archive data to a public cloud once a month with the data retained in the cloud for 730 days (2 years). A protection job will then leverage this policy to select the set of objects that need to be archived to the cloud based on this policy.

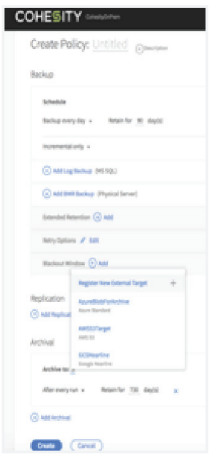

Figure 4. Setting a new protection policy in Cohesity.

Step 3: Recovery of Data and Applications Stored in the Cloud
Cohesity DataPlatform includes an indexing engine that opens underlying files and indexes the metadata as virtual machines (VMs) and physical servers are backed up. This indexing engine powers the rapid search and recovery of files and VMs from backups stored both on-premises and in the cloud, delivering extremely fast, wild-card search results that are then used for nearly instantaneous granular restores. You can search and recover a file from Cohesity (see Figure 5) by following these instructions:

1. Choose **Recovery** under the **Protection** tab to access the search screen.
2. Enter the search term to reveal a list of VMs and jobs that match.
3. Choose the **Continue** button, and select the relevant snapshot to recover from Amazon S3 or Amazon Glacier.

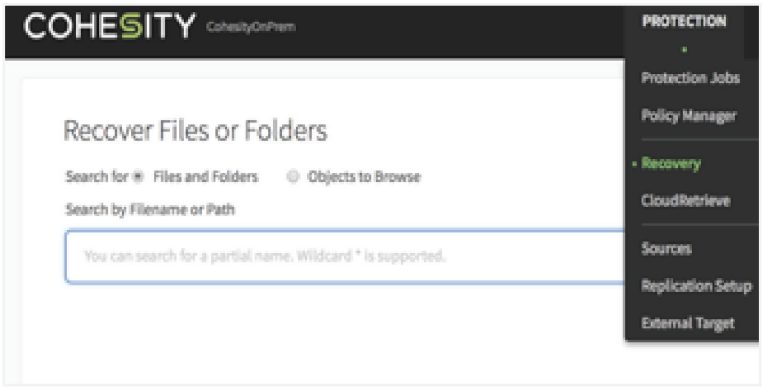

Figure 5. Data recovery from the cloud using Cohesity.

Cohesity DataPlatform also gives you the option to recover data from an archive to an alternate Cohesity platform. CloudRetrieve is the feature built into Cohesity DataPlatform that enables the downloading of job data archived to an external target by another Cohesity cluster. You can begin the alternate recovery process (see Figure 6), as steps one and two may each take a few hours, by following these instructions:

1. Choose **Search** to look for clusters and protection jobs archived to the target.
2. Download job metadata and snapshots of selected protection jobs.
3. Recover the protection jobs using the recovery flow.

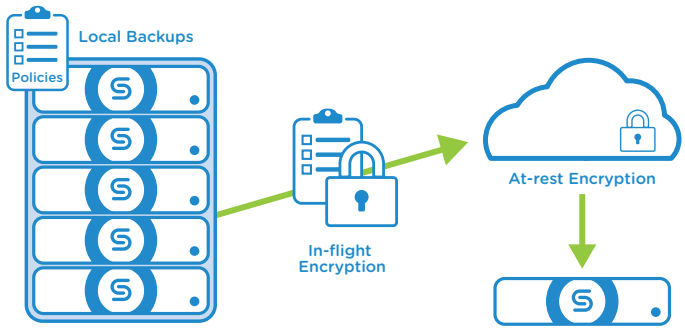

Figure 6. Data recovery from the cloud to another Cohesity cluster.

Learn More about Cohesity for Long-Term Retrieval and Archival

Together, Cohesity and AWS help you achieve your long-term data retention and archival objectives for secondary data. Native-cloud integration and Cohesity DataPlatform simplify deployment, improve protection, speed disaster recovery, and reduce your capital costs.

Learn more about how to solve your long-term data and application retention and archival challenges with Cohesity:

- Cohesity Data Protection White Paper
- Cohesity Cloud Integration Web Page

Cohesity, Inc.
Address 300 Park Ave., Suite 300, San Jose, CA 95110
Email contact@cohesity.com www.cohesity.com
@cohesity