**Version 1.0**

**October 2024**

# Archive Data to Spectra Logic S3-Compatible Object-based Tape Storage Using Cohesity DataProtect

**ABSTRACT**

*Tape storage remains cost-effective for long-term data retention, crucial for security, compliance, and legal requirements. Cohesity Platform™ and DataProtect™ offer robust on-premises solutions for enterprise data protection and storage. Cohesity Platform™ extends its capabilities and integrates seamlessly with SpectraLogic S3-Compatible Object-based Tape Storage solution, enhancing archival capabilities and simplifying the solution efficiently.*

# Table of Contents

# Figures

# Tables

# The Need for Archival to Tape

Organizations must adopt robust long-term data retention strategies to manage new data influx effectively. According to studies, the majority of the data is not frequently used but needs to be preserved for business or compliance reasons. Software-defined object storage offering an S3-Compatible REST API using standardized tape technology as the storage medium, like LTO, provides a solution to this challenge.

Archival to tape is crucial in enterprise backup solutions due to its cost efficiency, data security, and long-term reliability. Tape storage is highly durable meeting mandatory compliance, regulatory, long-term storage requirements, and legal requirements, with minimal incremental costs of backing up to tapes in most environments. You can safely move the sensitive data that you want to retain for future reference or regulatory compliance to an archive environment. It also serves as an "air-gap" strategy to secure your data from threats and reduce primary storage consumption and ownership costs.

When choosing a tape archival solution, consider the following:

- Data Security
- Backup and Disaster Recovery
- Performance
- Reliability and Durability
- Capacity Requirement and Interoperability
- Scalability

Meeting all the key requirements, Cohesity has partnered with Spectra Logic to deliver a comprehensive, data and tape-agnostic archival solution that simplifies long-term data retention and archival goals. This collaboration combines Spectra Logic's enterprise-class tape archival systems with Cohesity's cutting-edge, hyperconverged platform for managing data backups and archives.

The integration ensures seamless setup, scheduling, and management of archives on a Spectra Logic tape library through the Cohesity Platform, offering enterprises a scalable, secure, and cost-effective approach to long-term data storage and reliable data recovery. This guide offers the solution to integrate Spectra Logic S3-Compatible Object-based tape solution with Cohesity, to ensure streamlined operations and enhanced data protection.

# Long-Term Retention to Object-based Tape Features and Benefits

Cohesity's Archival to Object-based Tape provides many key features, each of which delivers several benefits to organizations and their IT administration staff. Specifically:

Table 1: Long-Term Retention to Object-based Tape Features and Benefits

| Features | Benefits |
|---|---|
| Policy-based archival | • Easy to use.<br>• Archive unique data differently by mapping Protection Policies to the required SLA.<br>• Reduce bandwidth and storage costs. |
| Cost Efficiency | Significantly lower cost ($/TB) |
| Off-site copies | • Geo-redundancy<br>• Disaster recovery<br>• "Air-Gap" enhancing data protection |
| Compression | Efficient data transfer and storage. |
| Recovery | • Instantly locate objects to Recover<br>• Recover just what you need. |
| Encryption | Data is secure both in flight and at rest. |

**NOTE**: This document covers only Cohesity DataProtect operations for archiving to tape using S3-Compatible Object-based Tape storage. For archiving to public cloud vendors, see guides for AWS, Azure, and Google Cloud Platform, and for archiving to NAS and S3-Compatible Storage, please refer to the above guides.

# Use Spectra Logic S3-Compatible Object-based Tape with Cohesity DataProtect Platform

Long-term data and application retention is critical to organizations that seek to prevent data loss and meet security, legal, and compliance requirements. Cohesity DataProtect offers robust on-premises solutions for enterprise data protection and storage. Cohesity's CloudArchive and Cloud Recover bring data protection and recovery together in a single coherent solution, both on-premises and in the cloud, with S3-Compatible Object-based Tape Storage.

With Cohesity, IT organizations save time by quickly archiving data to multiple targets—public clouds, private clouds, any S3-Compatible device, as well as NAS-NFSv3 from storage vendors, and S3-Compatible managed tape libraries, eliminating the need for cloud gateways and point solutions to connect to the cloud, while increasing operational efficiency and lowering the Total Cost of Ownership (TCO).

The S3-Compatible Object-based Tape Storage Class uses S3 Glacier API interface to communicate between Spectra Logic BlackPearl and Cohesity Cluster for Archival and Recovery.

## Archival Process for S3-Compatible Object-Tape Based Storage

Object-based Tape Storage Archival has two phases to perform Data archival from Cohesity backup.

Figure 1: Archiving Cohesity backup data to Spectra Logic S3-Compatible Object-based Tape Storage
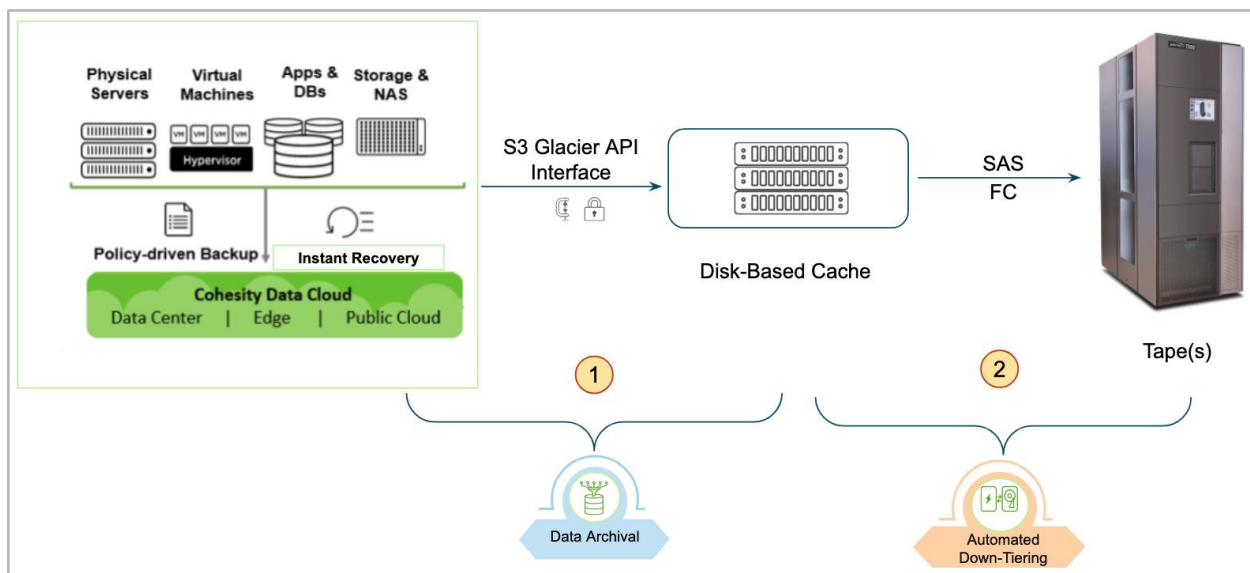


1. **Data Archival**: Once the backup to Cohesity Cluster is complete, the Archival data is ingested using S3-Compatible APIs (S3-Glacier) and lands in the Cache of Spectra Logic BlackPearl.
2. **Automated Down-Tiering**: Data is moved in parallel from the BlackPearl Cache to Tape drives.

# Restore Process for S3-Compatible Object-Tape Based Storage

When performing a restore, we perform the following actions to read the data.

Figure 2: Restoring data from Spectra Logic S3-Compatible Object-based Tape to the Restore Target
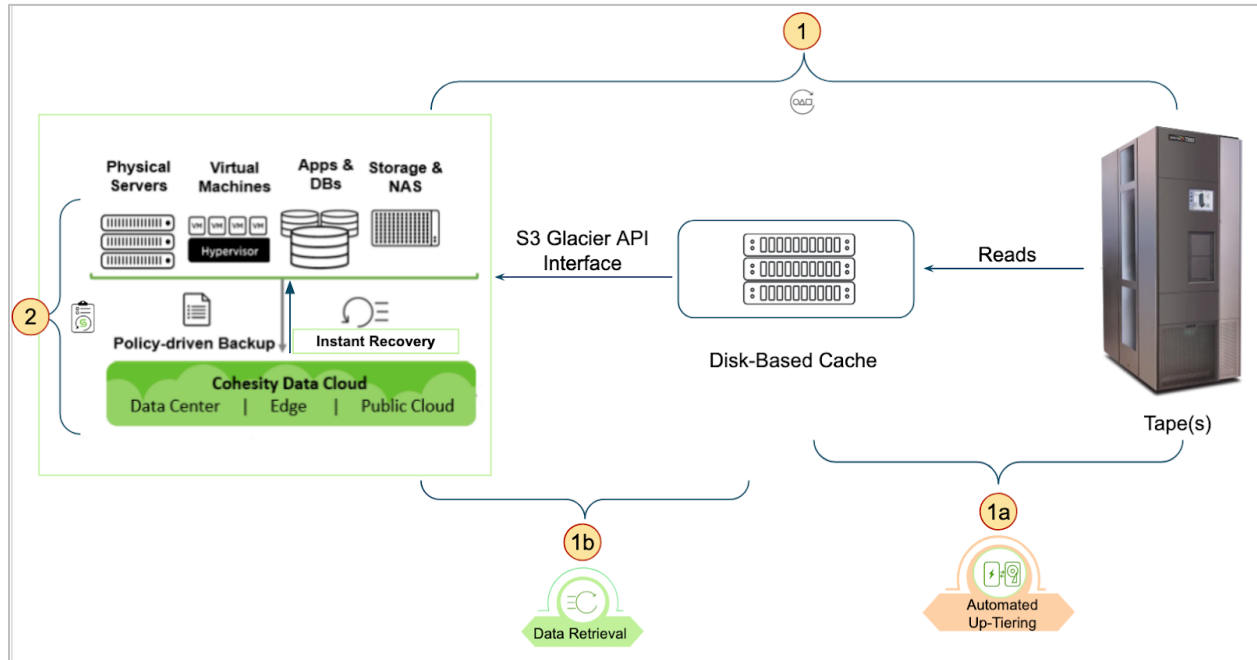


1. Data is Read from the Tape and Written to the Cohesity Cluster using S3-Glacier APIs

   a. **Automated Up-Tiering**: Data is Up-tiered from Tape to BlackPearl Cache.
   b. **Data Retrieval**: Data is Read from the BlackPearl Cache and Written to the Cohesity Cluster.

2. **Data Restore**: Data is Read from the Cohesity Cluster and Written to the Restore Target.

# Prerequisites

Archiving to Spectra Logic S3-Compatible Object-based Tape Storage is supported from the below versions:

Table 2: Minimum Cohesity and Spectra Logic BlackPearl software version needed

| Software | Cohesity DataPlatform Version | Spectra Logic BlackPearl Version |
|---|---|---|
| Minimum Version Required | 7.1.1 | 5.7.2 |

# Supported Storage Class for Archive to S3-Compatible Object-based Tape External Target

Cohesity Archive supports Archival to S3-Compatible Object-based Tape External Target from Cohesity version 7.1.1 onwards.

- Any S3-Compatible Object-based Tape Storage targets that use tape storage in the backend must be registered as an **Object-based Tape** Storage Class.

- These targets have capabilities similar to those of AWS S3 Glacier Storage Class.

- The Object-based Tape Storage Class uses S3 Glacier APIs for archival and recovery.

## Considerations

The following are the critical considerations with S3-Compatible Object-based Tape Storage Solution.

1. The archival format is *Always Full*, with Encryption and Compression enabled by default.

2. The recovery workflow retains the restored data from the tape on the S3 cache for a minimum of nine days. Please contact Cohesity Support to reduce the retention period of restored data on the Spectra Logic BlackPearl Cache,

3. The number of parallel streams is set to 10 by default. Contact Cohesity Support to change the values.

4. Does not support Deduplication and Incremental archives.

5. Does not support CloudRetrieve.

6. Does not support file-level recovery.

# CloudArchive High-Level Workflow

At the high level, leveraging Archival to S3-Compatible Object-based Tape Storage involves below 4 tasks:

1. Create an S3 bucket on the Spectra Logic BlackPearl.

    a. Capture your S3 bucket's Access Key ID and Secret Access Key.

2. Register the S3 bucket created above with the Cohesity Cluster as an S3-Compatible Object-based Tape External Target.

3. Archive your data to the External Target.

    a. Create a Cohesity Protection Policy.
    b. Create a Cohesity Protection Group.

4. Recover your data from the External Target.

Figure 3: Leverage Spectra Logic S3-Compatible Object-based Tape Storage with Cohesity
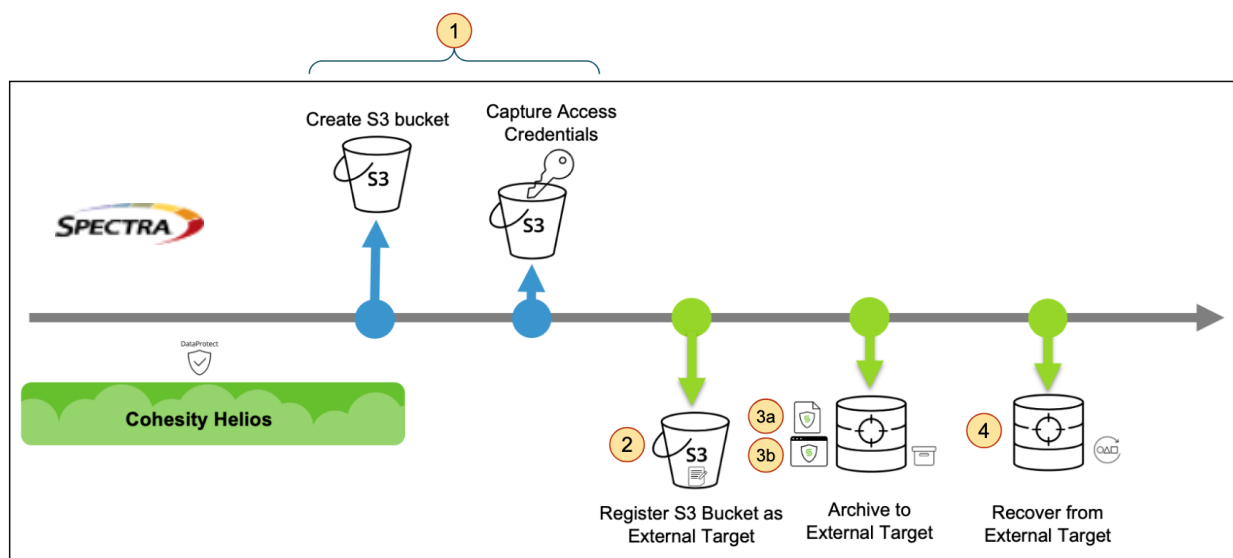


## Create Your S3 Bucket on Spectra Logic BlackPearl

Creating an S3 Bucket on Spectra Logic BlackPearl involves creating an IAM user, the S3 bucket, and capturing the below required external target fields to register the S3 bucket as an External Target on the Cohesity Cluster.

## Required External Target Fields to Register Your S3 Bucket

To register your S3 bucket as an External Target, Cohesity Cluster requires the following fields:

- Access Key ID
- Secret Access Key
- Bucket Name
- Endpoint
- Port on which the S3 bucket is exposed
- AWS Signature Version

To get started, we need to create an IAM user on the Spectra Logic BlackPearl management console.

## Create IAM Users on Spectra Logic BlackPearl Management Console

1. Log in to the Spectra Logic BlackPearl management console.

2. In the upper right corner of the BlackPearl management console, click the **gear** icon and select **IAM Users & Groups.**
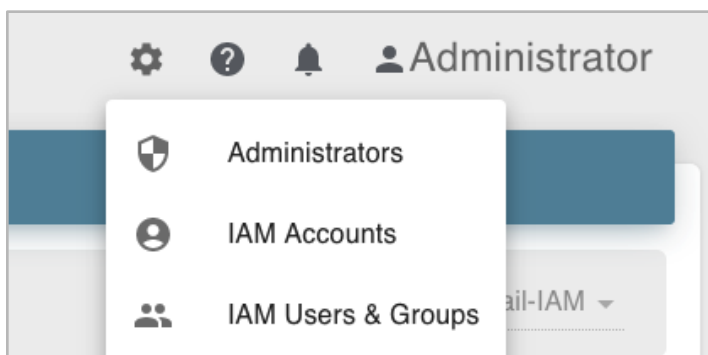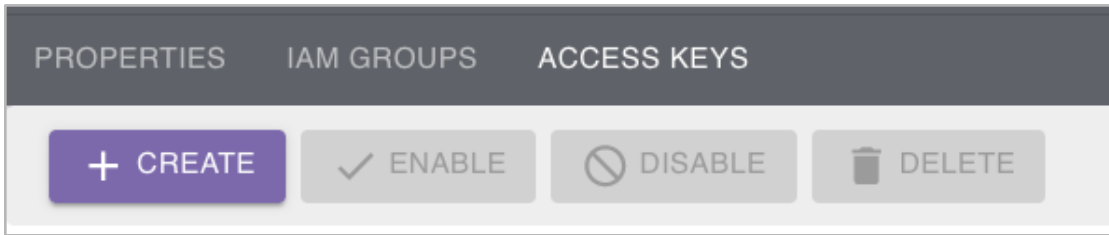


3. Under the IAM Users banner, click **Create**.
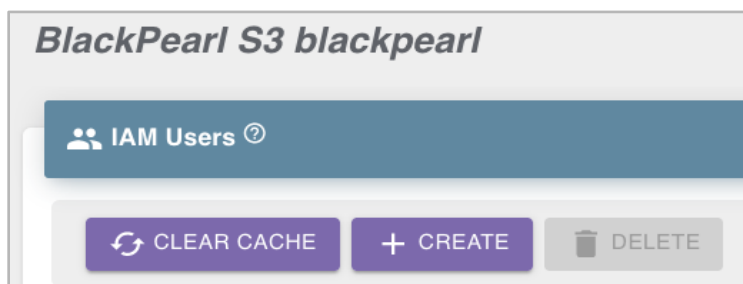


4. Enter the **Username** for the new IAM user and click **Submit.**

5. Select the user created and click on **View details.**

6. Add the user to IAM Groups.

7. Click on **Access Keys** and **Create** to generate the Access Key ID and Secret Key**.**



8. Copy the **AWS Access Key ID** and **AWS Access Secret Key** that is generated.
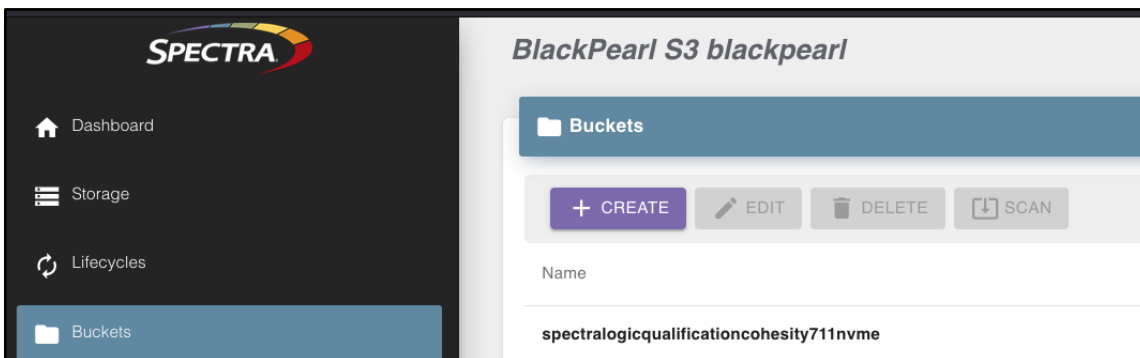
**NOTE**: This is the **only** time this secret access key can be viewed. You cannot recover it. Store the Secret Access Key safely.

This will be used to register this bucket as an external target archival.
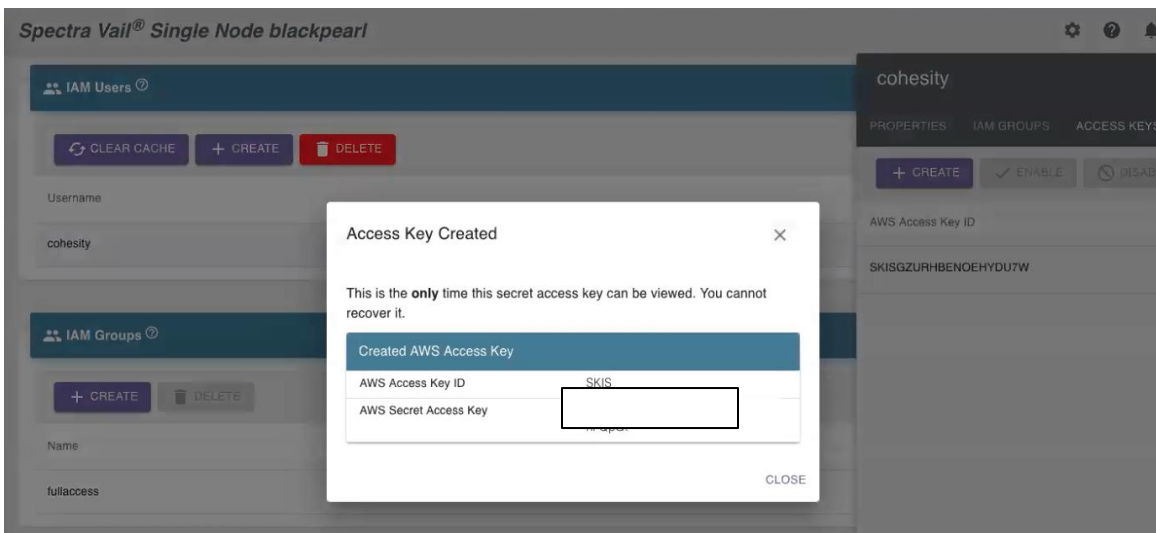


## Create Your S3 Bucket

1. In the Spectra Logic BlackPearl management console taskbar, click **Buckets**.

2. Under the Buckets banner, click **Create.**

3. Enter the desired **Bucket Name**.

> **NOTE**: Spectra Logic recommends following Amazon AWS documentation naming conventions and restrictions to maintain consistency and avoid potential issues. See AWS Bucket Naming Rules for more information.

4. Use the **Bucket Owner** drop-down menu to select a user to own the bucket. The bucket owner sets permissions for the bucket.

5. Encryption and Compression will be enabled in External Target Registration on Cohesity Cluster UI, so leave it unchecked.

6. Edit the example Policy code as required. Policy permissions are used if you want to exclude IAM user(s) under the main AWS account from accessing the BlackPearl bucket.

> **NOTE**: For additional information on configuring a policy, see the Amazon S3 Actions documentation.

7. Click **Next.**

8. Select Lifecycle - None.

> **NOTE**: The Object-based Tape Storage Class uses S3-Glacier APIs for archival and recovery. So, no lifecycle needs to be configured to transfer the data.



9. Review the configuration, then click **Submit** to create the bucket.

# Protect Your Data

Now that you have the bucket created, and the IAM user created, you're ready to connect the S3 Bucket to the Cohesity Cluster as an External Target.

## Register Your S3 Bucket as an External Target

To Register an External Target with your cluster, follow these steps.

1. Log in to Cohesity Cluster UI.

2. Click on **Infrastructure** > **External Targets**.
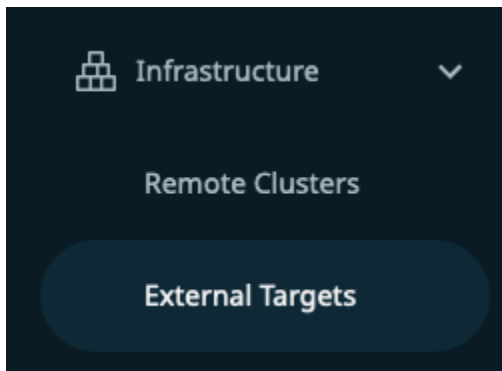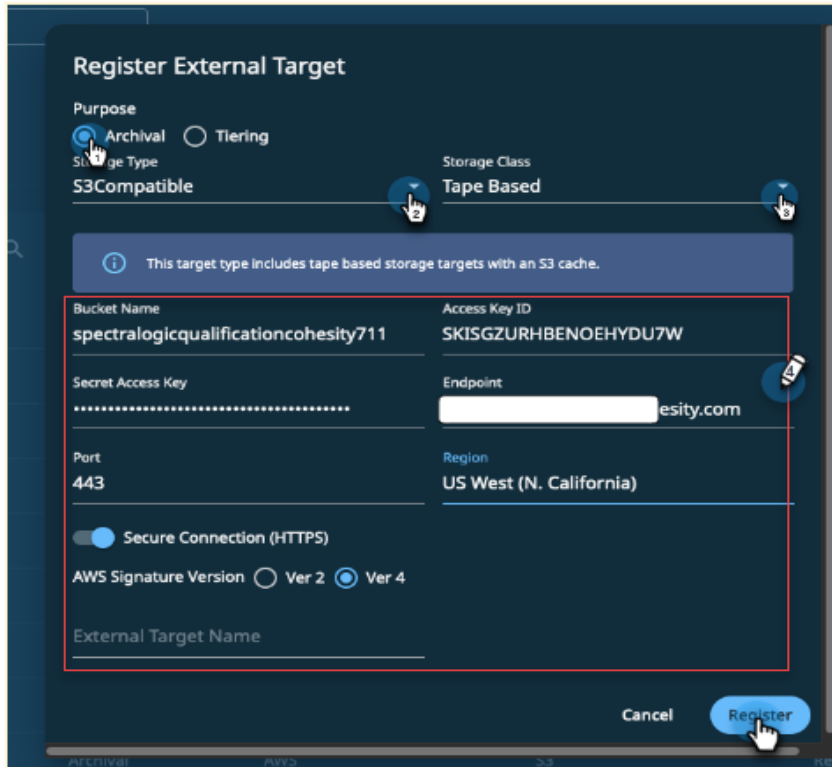


3. Click on "**Add External Target**".



4. Choose Purpose as "**Archival**", Storage Type as "**S3Compatible**", and Storage Class as "**Tape Based**".
   In the form that opens:

   o Enter the **Bucket Name**.

   o Enter the **Access Key ID** and **Secret Access Key** that you captured when you created the users in SpectraBlackPearl UI.

   o Enter **Endpoint, Port**, and **Region**.

   o **Secure Connection (HTTPS)**: Enabled by default.

   o **AWS Signature Version 4**: Spectra Logic exposes the S3 bucket on Port 443 and supports AWS Signature Version 4.

   o Enter a unique **External Target** Name.

   o Region field can be left blank, this is *ignored* in the backend.

   > **NOTE**: Archival Format will be Always Full as it is S3-compatible object-based Tape Storage.

5. Enable **Encryption** and Choose between the options provided.
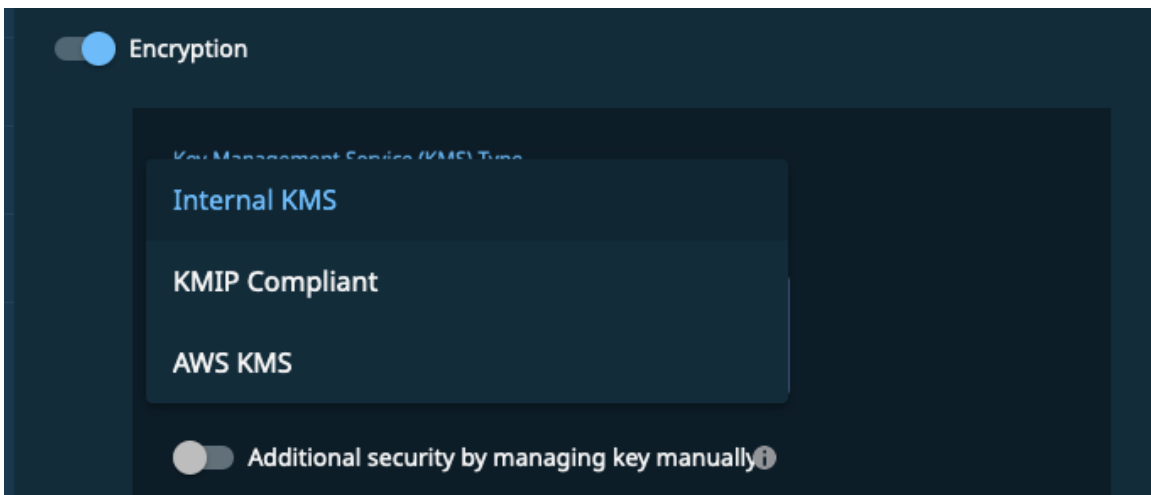
6. **Compression** is enabled by default**.**



7. Click **Save.**

Your registered External Target is now available to select when you create a Cohesity Protection Policy for use in Protection Groups.

## Create a Protection Policy

Once Cohesity Cluster registers your S3 bucket as an External Target, you will Create a Protection Policy according to your business needs. A protection policy is a reusable set of settings that define how and when objects are protected, replicated, and archived. It allows you to incorporate the External Target that you created above as an archive target with a specific retention period.

In the Policy, you configure how virtual and physical servers, databases, and unstructured data are protected:

- Backup frequency and retention period.

- Whether to have your backups archived, how often, and how long to retain.

   **NOTE**: You can add more than one archival schedule to the same Policy, and use the same or a different External Target, with the same or different frequency and retention.

- Which External Target to use (in this case, your newly registered S3 bucket)?

## Create a Protection Group

Once an external target is registered on Cohesity Cluster UI, we need to create a Protection Group to protect an object, which will be used to archive. Protection Groups combine operational requirements with the business requirements that are defined in a Protection Policy.

In the Protection Group, you select the source, which data objects from that source to store, the Protection Policy and the Storage domain (the named storage location) to use, and operational details such as Start Time, End Date, QoS Policy, Pre & Post Scripts, and more. Once you save a Protection Group, it will run on the schedule you define.

**NOTE**: Multiple Protection Groups can use the same Protection Policy, but each Protection Group can have only one policy.

## Performance: Optimize Archival Throughput

The archival process involves several steps, including reading data from the source, compressing it, storing the primary copy in the Cohesity Cluster, and then writing it to archival storage. If not managed effectively, these tasks can increase disk I/O and CPU utilization, potentially affecting the cluster's overall performance.

NOTE: We by default allow 10 streams to perform Archival to Spectra Logic S3-Compatible Object-based Tape External Target.

To achieve optimal archival performance, collaborate with Cohesity support to take advantage of Cohesity and Spectra Logic Archival Performance.

## Recover Your Data from The Tape Archive

When recovering your archived data, Cohesity Cluster allows you to Restore Entire Data Objects (VMs, databases, NAS, etc.).

Please review the considerations mentioned here.

NOTE: Monitor alerts and emails from Spectra Logic when any restore is done to ensure the required tapes are available.

## How long is the data kept in the BlackPearl cache?

IMPORTANT: When a Restore is initiated, the object restored from the S3-Glacier / Tape is kept in the Spectra Logic Cache for 7 days by Cohesity (this is to avoid the need to rehydrate the data multiple times if the restore fails/cancels for any reason).

The cache is managed by Spectra Logic BlackPearl and keeps the data in the Cache until midnight UTC + 1 day (i.e., if you restore at 5 pm UTC on the 20th, the data will stay in the cache until 12:00 am UTC on the 29th).

In total, the restored objects are kept in Spectra Logic Cache for 7 + 2 = 9 days.

Please contact Cohesity Support to change the configuration setting to reduce the number of days the restored data is kept in the BlackPearl Cache.

Now that we understand the cache behavior when a Restore is initiated, let us understand the Restore workflow for the below common scenarios.

## Restore Timeline and Workflow for Common Scenarios

When a user initiates the Restore on the Cohesity Cluster UI, below are the actions performed to retrieve the data. Let's take a look at the Restore Timeline.

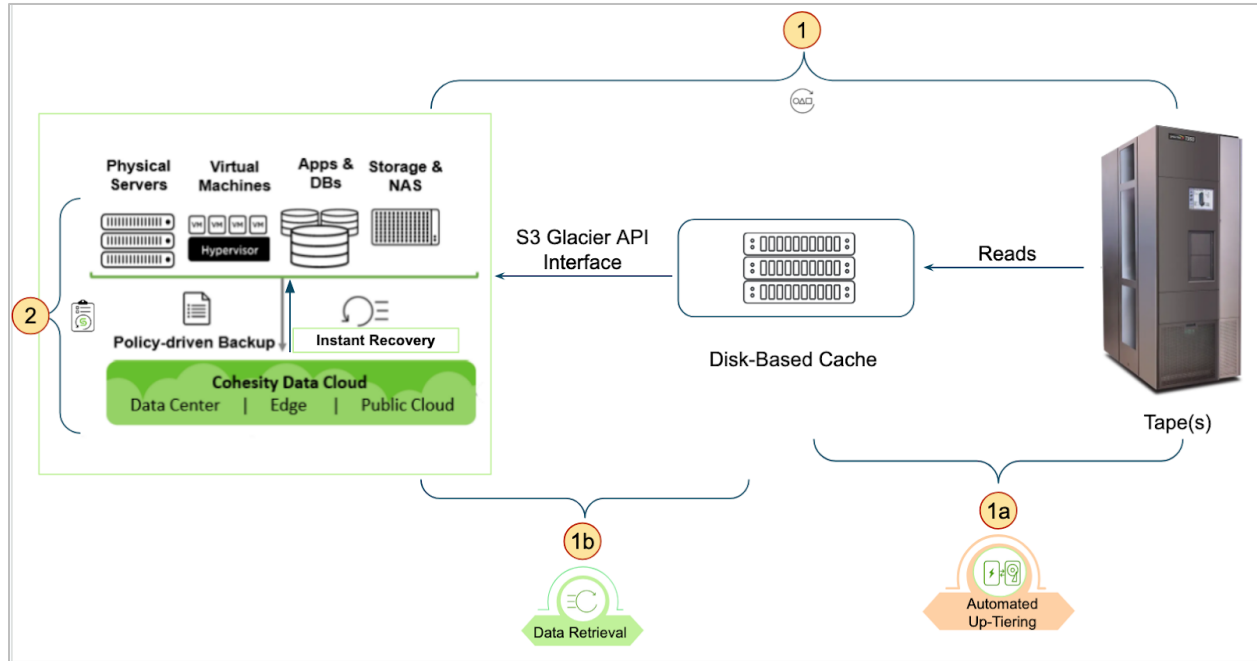Figure 4: Restoring data from Spectra Logic S3-Compatible Object-based Tape to the Restore Target



Table 3: Restore Workflow for Common Scenarios

| | Tape Data in Cache | Tape is Online | Tape is Offline |
|---|---|---|---|
| User Intervention | ❌ | ❌ | Monitor for Alerts and load required Tape. |
| Restore Time Taken | Quick | Take additional time (due to Up-Tiering Data from Tape to Cache) | Restore will make progress only when the required tape is online. (Takes similar time as Tape is Online scenario, once required Tape is online.) |
| Steps | 1b, 2 | 1a, 1b, 2 | User Intervention, 1a, 1b, 2 |

Below are the illustrations for 3 different common scenarios:

1. **Tape Data in Cache**: When a user initiates a Restore, the restore starts immediately as the data is present in Cache, therefore there is no need to read data from Tape.

There are 2 Steps involved:

    a.   **Data Retrieval**: Data is Read from the BlackPearl Cache and Written to the Cohesity Cluster.

    b.   **Data Restore**: Data is Read from the Cohesity Cluster and Written to the Restore Target.
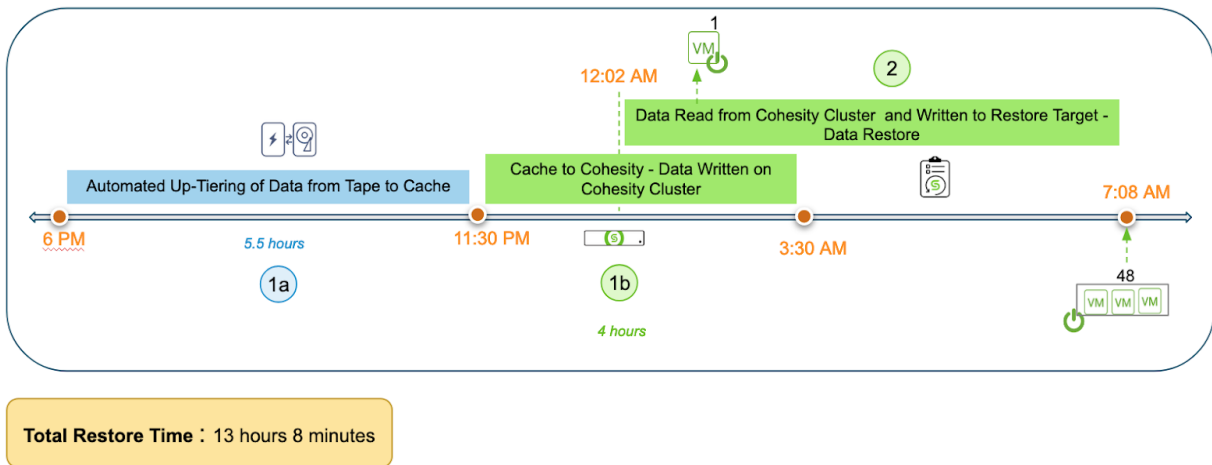
Figure 5: Tape Data in Cache
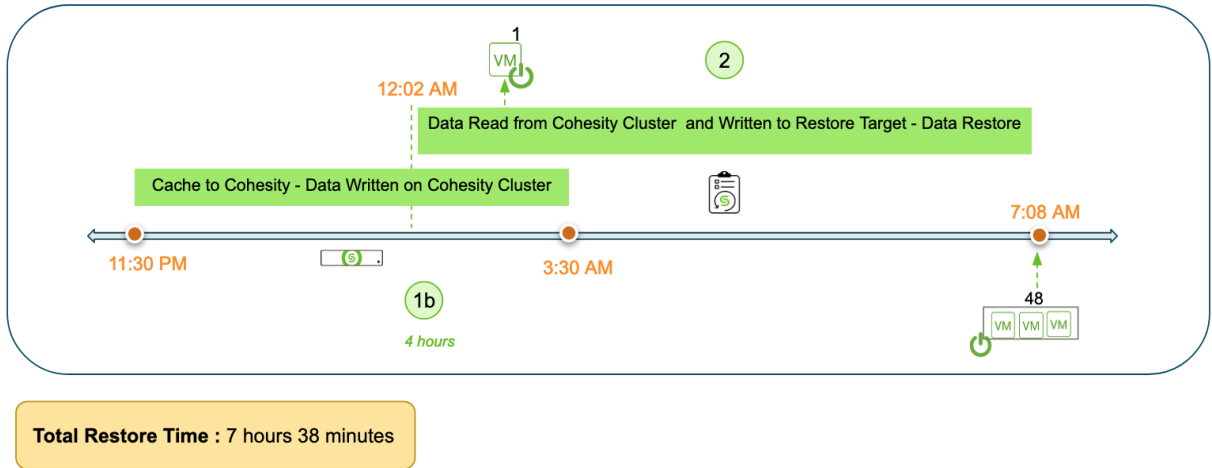


2. **Tape is Online**: When a user initiates a Restore, the requested data is not present in the Cache, requiring Up-tiering of data from Tape.

There are 3 Steps involved:

    a.   **Automated Up-Tiering**: Data is Up-tiered from Tape to BlackPearl Cache.

    b.   **Data Retrieval**: Data is Read from the BlackPearl Cache and Written to the Cohesity Cluster.

    c.   **Data Restore**: Data is Read from the Cohesity Cluster and Written to the Restore Target.
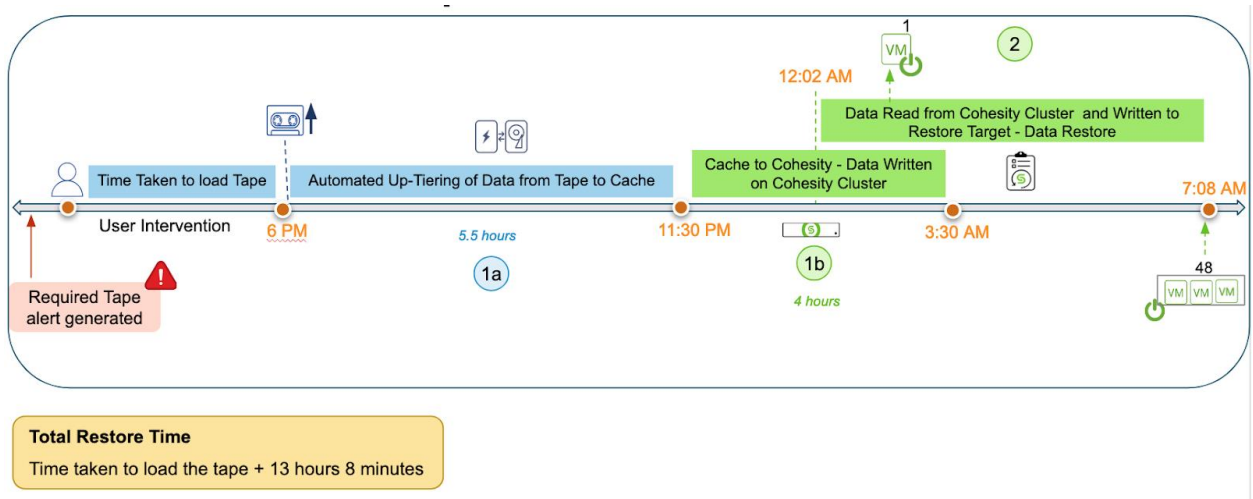
Figure 6: Tape is Online



3. **Tape is Offline:** When a user initiates a Restore, the requested data is not present in the Cache, requiring up-tiering of data from tape, but the required tape is Offline.

There are 4 Steps involved in this process:

a. **User Intervention:** Required Tape alert is generated in SpectraLogic BlackPearl and an email is also sent. The user is required to get/load the required tape. Once the tape is online, we run the below steps.

b. **Automated Up-Tiering**: Data is Up-tiered from Tape to BlackPearl Cache.

c. **Data Retrieval**: Data is Read from the BlackPearl Cache and Written to the Cohesity Cluster.

d. **Data Restore**: Data is Read from the Cohesity Cluster and Written to the Restore Target.

Figure 7: Tape is Offline



Figure 8: Snapshot of Spectra Logic Alerts requesting missing Tapes in the Tape Library



When a restore is in progress, looking at Spectra Logic BlackPearl Performance chart helps in analyzing:

1. When the data is brought from the Tape(s) to the BlackPearl Cache.

2. When the Data is transferred from the BlackPearl Cache to the Cohesity Cluster.

# Appendix

## CloudArchive Terminology

There are several important terms to understand as you learn how CloudArchive works.

Table 4: CloudArchive Terminology

| Term | Definition | Notes |
|---|---|---|
| Cohesity Platform | Cohesity Platform consolidates secondary data and applications, including backups, files, objects, test/dev, and analytics on a single, software-defined platform. Inspired by web-scale architecture, Cohesity Platform is a scale-out solution based on a unique distributed file system, SpanFS®. | |
| Cluster | An instance of Cohesity Platform. | |
| External Target | Any storage to which data is sent outside the source Cohesity Cluster | Archive to Cloud, Tape, NAS, and replication targets are all External Targets in Cohesity Platform. |
| Protection Group | Defines operational requirements, such as which source objects to protect, the Protection Policy to use, and operational considerations like indexing, exclusions, inclusions, alerts, app consistency, and more | Each Protection Group has a schedule of Group Runs, and each archive is a collection of those Group Runs. |
| Protection Policy | Reflects business needs of Recovery Point Objectives (RPOs) and Recovery Time Objectives (RTOs) by defining the frequency and retention requirements of backup, archival, and replication. | |

| Term | Definition | Notes |
|---|---|---|
| Recover | Retrieve an entire data object, such as a VM or database, or granularly recover files and folders from an External Target onto the original cluster | |

# Your Feedback

Was this document helpful? <u>Send us your feedback</u>!

# About the Authors

**Jedidiah Sonavane** is a Product Solutions Architect and product Solutions WWFO at Cohesity. In his role, he focuses on enterprise data protection, solution validation, solution testing, solution qualification, and software usability.

Other essential contributors included:

- Adaikkappan Arumugam, Sr Director, Product Solutions
- David Feller, CTO, Spectra Logic
- Jeff Biley, Application Engineer, Spectra Logic
- Tarey Treasure, Network Product Engineer, Spectra Logic
- Evan Lamson, Application Engineer, Spectra Logic
- Mary Juliya, Technical Editor, Technical Solution Engineering

# Document Version History

| VERSION | DATE | DOCUMENT HISTORY |
|---------|------|------------------|
| 1.0 | October 2024 | First release |

# ABOUT COHESITY

Cohesity is a leader in AI-powered data security and management. Aided by an extensive ecosystem of partners, Cohesity makes it easier to protect, manage, and get value from data – across the data center, edge, and cloud. Cohesity helps organizations defend against cybersecurity threats with comprehensive data security and management capabilities, including immutable backup snapshots, AI-based threat detection, monitoring for malicious behavior, and rapid recovery at scale. Cohesity solutions are delivered as a service, self-managed, or provided by a Cohesity-powered partner. Cohesity is headquartered in San Jose, CA, and is trusted by the world's largest enterprises, including six of the Fortune 10 and 44 of the Fortune 100.

Visit our website and blog, follow us on Twitter and LinkedIn and like us on Facebook.

2000055-001-EN 11-2024