COHESITY

Version 2.1

July 2024

CloudArchive & CloudRetrieve Deployment & Recovery Guide for GCP

Store Your Protected Data in the Cloud for Long-Term Retention and Disaster Recovery

ABSTRACT

Long-term data and application retention is critical to organizations that seek to prevent data loss and meet security, legal, and compliance requirements. Cohesity Data Cloud offers robust on-premises solutions for enterprise data protection and storage. Cohesity's CloudArchive and CloudRetrieve bring data protection and recovery together with cloud storage.

Table of Contents

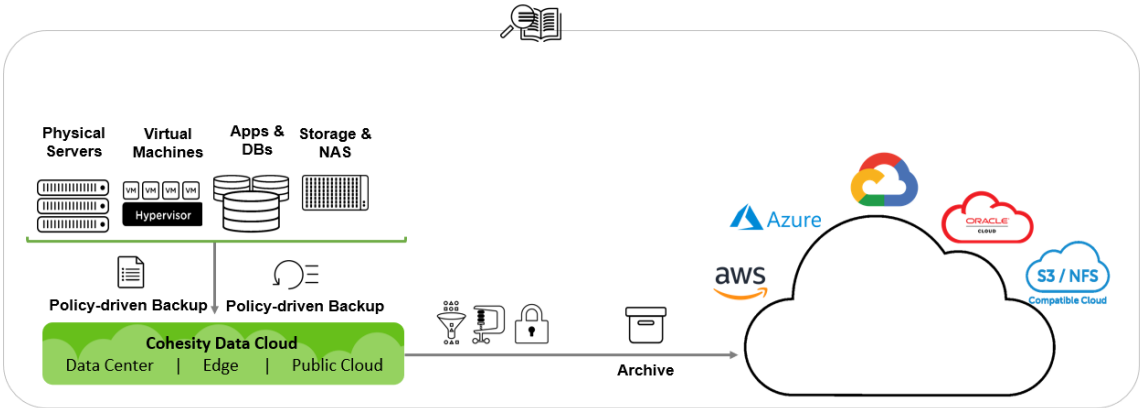
Figures

Tables

CloudArchive Connects Cloud Storage to Cohesity Data Cloud

Long-term data and application retention is critical to organizations that seek to prevent data loss and meet security, legal, and compliance requirements. Cohesity Data Cloud (previously known as “Cohesity Platform” and hereafter referred to as “Data Cloud”) offers robust on-premises solutions for enterprise data protection and storage. Cohesity’s CloudArchive and CloudRetrieve bring data protection and recovery together in a single coherent solution, both on-premises and in the cloud:

Figure 1: CloudArchive Connects Cloud Storage to Data Cloud



With Cohesity Data Cloud, 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 QStar managed tape libraries. Cohesity CloudArchive’s cloud-native integrations with AWS, Azure, and GCP eliminate the need for cloud gateways and point solutions to connect to the cloud, while increasing operational efficiency and lowering total cost of ownership (TCO).

NOTE: This document covers only Cohesity Data Cloud operations for archiving to the cloud and *not* tape or NFS targets. For archiving to tape, see Long Term Retention to Tape with Cohesity DataProtect solution guide.

CloudArchive Versions

Cohesity CloudArchive has two versions:

- CloudArchive Incremental with periodic full
- CloudArchive Incremental forever — available from 6.6.0a onwards (currently supports only AWS and Azure)

Before you configure CloudArchive; review the differences between the versions and the supported sources.

CloudArchive Features and Benefits

CloudArchive supports all of the leading object storage from cloud providers, any S3-compatible device, as well as NAS from storage vendors. Specifically:

Table 1: CloudArchive Features and Benefits

FEATURES	BENEFITS
Policy-based cloud archival	• Easy to use • Archive unique data differently by mapping Protection Policies to the required SLA • Reduce bandwidth & storage costs.
Off-site copies	• Geo-redundancy • Disaster recovery
Deduplication and compression	Efficient data transfer and storage.
Granular recovery	• Instantly locate VMs, files, and folders. • Recover just what you need.
Encryption	Data is secure both in flight and at rest.
WORM/Object Lock	End-to-end WORM capability through DataLock at Cohesity end, and WORM support at storage target end.

Classes of Supported Storage for CloudArchive

CloudArchive supports all of the leading object storage from cloud providers, any S3-compatible device, as well as NAS from storage vendors. Review the support matrix to understand the supported storage classes for CloudArchive: External target support matrix.

CloudArchive Terminology

It is important to understand the following terms as you learn how CloudArchive works.

Table 2: CloudArchive Terminology

TERM	DEFINITION	NOTES
Cohesity Data Cloud	Data Cloud consolidates secondary data and applications, including backups, files, objects, test/ dev, and analytics on a single, software-defined	

TERM	DEFINITION	NOTES
	platform. Inspired by web-scale architecture. Cohesity Data Cloud is a scale-out solution based on a unique distributed file system, SpanFS.	
Archive	Completely self-contained copy of the backup (data and metadata) that is stored outside the Cohesity cluster.	
Archive Chain	The set of a Full Archive and the Incremental Archives that depend on it and the preceding Incrementals.	If the Full Archive is lost for any reason, the entire archive chain becomes unusable. If an Incremental Archive is lost, the restore points that follow it are lost as well.
CloudRetrieve	The process of retrieving an archived protection group and its job run details from an External Target to a different cluster. Used for geo-redundancy and disaster recovery.	CloudRetrieve cannot be performed on the same cluster that performed the archive.
Cluster	An instance of Data Cloud.	
Deduplication Chain	The set of a Reference Archive and all the archive chains that depend on it for deduplication. This includes the Scheduled Full and Incremental Archives for each archive chain in the deduplication chain.	These dependencies determine when Data Cloud can retire and eventually delete Reference Archives.
External Target	Any storage to which data is sent outside the source Cohesity cluster.	Archive to Cloud, Tape, NFS, and replication targets are all External Targets in Data Cloud.
Full Archive	A full copy of the protection group that is archived.	
Incremental Archive	An archive that records just the changed data since the most recent archive.	
Protection Group	Defines *operational* requirements, such as which source objects to protect, the Protection Policy to use, and operational considerations like indexing,	Each protection group has a schedule of job runs, and each archive is a collection of those job runs.

TERM	DEFINITION	NOTES
	exclusions and inclusions, alerts, app consistency, and more.	
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.	
Scheduled Full Archive	A Full Archive that runs at regular intervals (configurable, 90 days by default).	The Scheduled Full Archive does not send the same amount of data, as it is deduplicated against the Active Reference Archive. In those cases when there is no Active Reference Archive, the data sent for the Scheduled Full is deduplicated only with itself and not against any other archive. For example, if the Active Reference Archive size is 100GB and the Scheduled Full deduplication usage is 60%, then only 40GB is sent. If there is no Active Reference Archive, then the size of the Scheduled Full is 100GB.
Source-Side Deduplication	The process of eliminating redundant copies of data to reduce storage use before sending over the network. Redundant data blocks are replaced with a pointer to the unique data copy, which ensures that only unique instances of data are transferred over the network and retained on storage media.	Reduces storage as well as network bandwidth requirements and, in doing so, saves time and money.
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.	

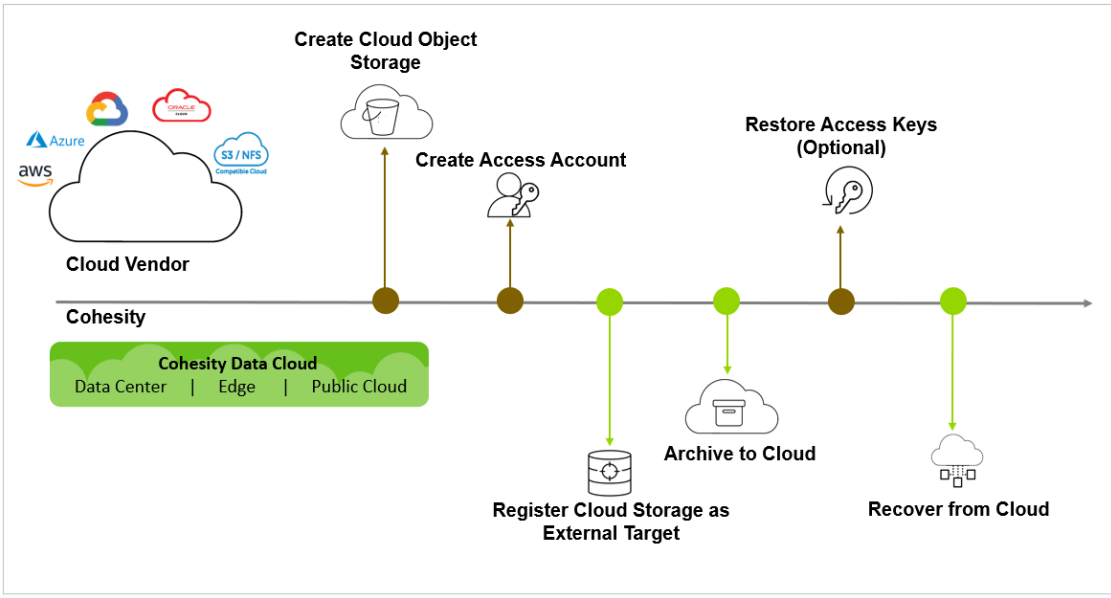
TERM	DEFINITION	NOTES
Reference Archive	The Full Archive against which all subsequent Incremental Archives (in the archive chain) and Scheduled Full Archives *as well as their* Incrementals are deduplicated.	All Reference Archives are full archives. A new Reference Archive is created when Data Cloud detects that deduplication with it is below 50%. **NOTE**: 50% is the default threshold. This is internally configurable, but changing this value only delays *when* (and not whether) the full data set is sent.
Retired Archive	A Reference Archive that is no longer used for deduplication.	

CloudArchive High-Level Workflow

At the highest level, leveraging CloudArchive involves several sequential tasks:

1. Create cloud object storage with the cloud provider of your choice.
   a) Create a user and assign the necessary permissions to the object storage for Data Cloud to access it.
2. Register your cloud object storage to Data Cloud as an External Target.
3. Archive your data to the cloud.
   a) Create a Cohesity Protection Policy.
   b) Create a Cohesity Protection Group.
4. Recover your data from the cloud.

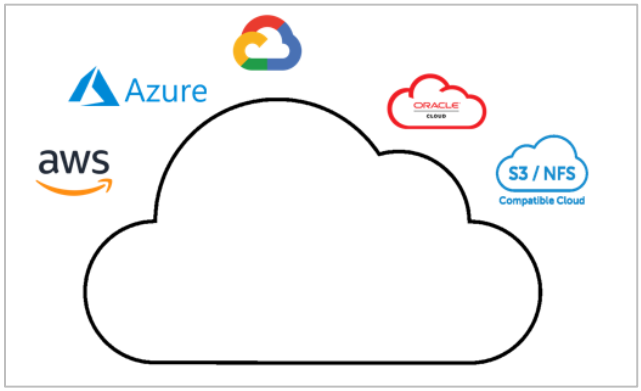
Figure 2: Leverage Cloud Storage with Data Cloud



Create Your Cloud Object Storage

The first thing you'll do is create a bucket or vault or blob with your cloud storage vendor. Though the process is slightly different for each vendor, it always involves creating the cloud object storage and a user account that has access to it. Finally, you'll need to capture the access key that gives that account access.

Figure 3: Create Your Cloud Object Storage



Connect Your Cloud Object Storage

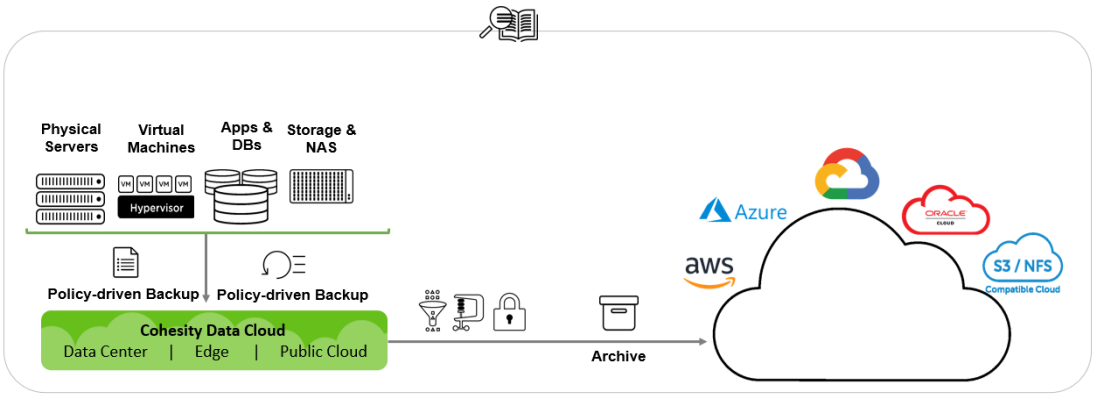
Next, you need to connect that new cloud object storage to Data Cloud by registering it as an External Target in Data Cloud. For this, you'll need the container name, access key, and geographic region.

Figure 4: Register Cloud Object Storage with Data Cloud

Archive Your Data to the Cloud

With your cloud storage now registered with Data Cloud, the next step is to archive your data by creating a Protection Policy (which reflects your business needs, like frequency and archival retention requirements) and running a protection group (where you define operational requirements, such as which data objects to protect, the Protection Policy to use, indexing, alerts, and SLA requirements).

Figure 5: Archive Data to Cloud Object Storage



Recover Your Data from the Cloud

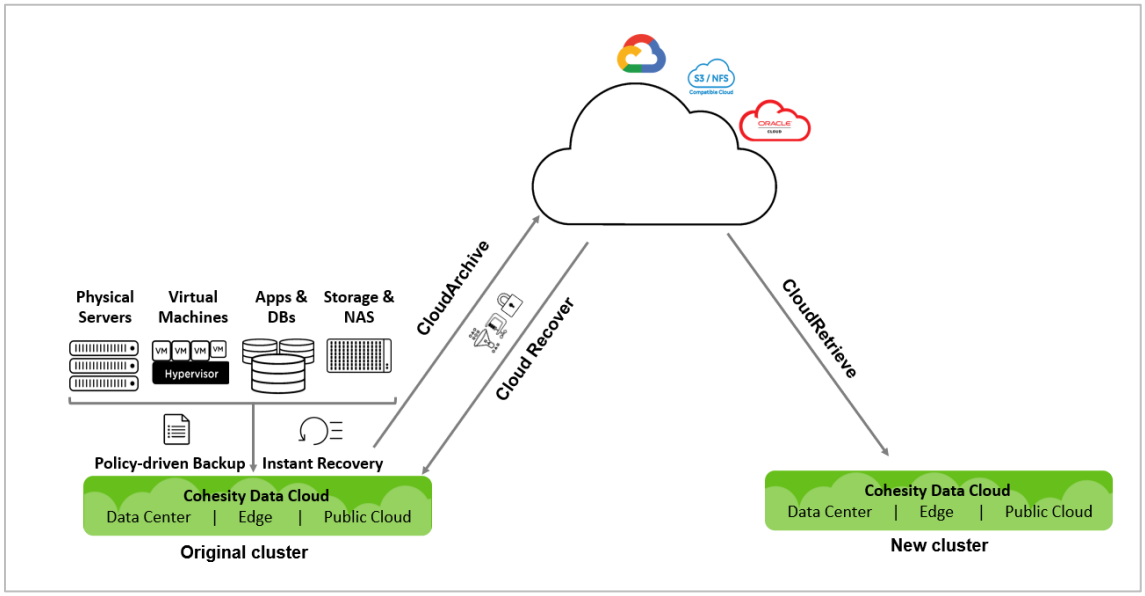
In most organizations, customers use on-premises storage for data that only has a short retention period, but store data with long-term retention requirements in the cloud. When a need for some or all of that cloud-stored data arises, the challenge is to locate, identify, and recover it quickly and reliably.

Data Cloud includes an indexing engine that enables rapid search and recovery of files and virtual machines from archives stored both on-premises and in the cloud. As virtual machines and physical servers are backed up, Data Cloud's indexing engine opens the underlying files and indexes the metadata. This enables extremely fast, wild-card search results that are then used for granular recovery.

Once your data is archived with CloudArchive, when you need to access it again, you'll be able to get it back using Cloud Recover (to your original cluster) or CloudRetrieve (to a new cluster).

- **Cloud Recover to source cluster**: Recover entire objects (VMs, databases, NAS, etc.), or individual files and folders, to your original cluster.
- **CloudRetrieve to new cluster**: Retrieve your previously archived data onto an entirely new cluster, for disaster recovery and geo-redundancy.

Figure 6: Recover Data from the Cloud — Cloud Recover and CloudRetrieve



In the next chapter, we cover the individual steps that are involved in each of these tasks. Following that, we walk through the specific procedures for connecting your cloud storage vendor to Data Cloud, archiving your data to your cloud object storage, recovering, and restoring your data from your cloud object storage.

Leverage Your Cloud Storage with Data Cloud

The following sections provide a quick overview of the sequence of actions that you will be undertaking to set up your cloud storage as an External Target in Data Cloud. We will dive into the step-by-step instructions for your cloud storage vendor in the next chapter.

Create and Register Cloud Object Storage

Start by setting up your cloud object storage. Note that the same cloud object storage can be registered multiple times in the same cluster as different External Targets, as well as in multiple clusters.

1. Create a storage container on your cloud platform, and an account that can access it.
2. Get the container name and access key from your cloud platform.
3. Using the cloud object storage information and access key, register the cloud object storage as an External Target in Data Cloud.

> **IMPORTANT**: Customers should never manually edit, change, or delete Data Cloud archives directly in cloud object storage.

When you register your External Target in Data Cloud, you will be able to enable or disable:

Table 3: External Target Options

FEATURE	DESCRIPTION
Encryption	By default, Data Cloud writes the data into External Targets in encrypted format in real time. You can disable it, but Cohesity recommends you leave it enabled in almost all cases, except when the data is already encrypted. You can choose to keep your encryption key in the cloud with your archive, or, for additional security, to manage it manually. **NOTE**: If you choose the manual option, you will need to download the key after registering the External Target and store it outside the Cohesity cluster.
Compression	Reduces the impact on data transfers and data storage. Useful except when the data format doesn't compress well, such as with databases and large image files.
Source-Side Deduplication	The process of eliminating redundant copies of data to reduce storage use. Redundant data blocks are replaced with a pointer to the unique data copy, which ensures that only unique instances of data are sent across the network and retained on storage media, and dramatically reduces the impact on bandwidth and storage utilization. Cohesity strongly recommends it in all cases.

FEATURE	DESCRIPTION
Incremental Archival	An archive that records just the changed data since the most recent archive. This allows you to return to any restore point without having to create, transfer, and keep a backup copy of your whole dataset each time. Cohesity strongly recommends this setting in all cases. If this option is not enabled, it will send a full archive on every archive run.
Bandwidth Throttling	If needed, you can throttle the upload and download bandwidth that is consumed by network traffic between Cohesity Data Cloud and an External Target. You can also limit bandwidth throttling to a specific time range, if there are particular days and times when it is needed. **NOTE**: If an archive is still running when bandwidth throttling switches to 0 throughput (that is, a blackout), the run is paused until the throughput value is greater than 0. When it resumes, it does so from the point where it paused.

NOTE: The same cloud object storage can be registered multiple times in the same cluster as different External Targets, as well as in multiple clusters.

Required Cloud Vendor Fields

To register your cloud object storage as an External Target, Cohesity Data Cloud requires the following fields:

- Container Name
- Region
- Storage Account Name
- Storage Access Key ID
- Secret Access Key

Each cloud provider has slightly different terminology for these fields:

Table 4: Required Cloud Vendor Fields

CLOUD PROVIDER	CONTAINER NAME	REGION	STORAGE		SECRET ACCESS KEY
			ACCOUNT NAME	ACCESS KEY ID	
AWS	Bucket Vault	Region	IAM User	Access key ID	Secret access key
Azure	Blob	Location	Storage account	Access key	n/a

CLOUD PROVIDER	CONTAINER NAME	REGION	STORAGE		SECRET ACCESS KEY
			ACCOUNT NAME	ACCESS KEY ID	
GCP	Bucket	Location	Service Account (Client Email Address)	Client Private Key	n/a

Configure Your Policy-based Archive

Once Cohesity Data Cloud registers your cloud object storage as an External Target, you will create a Protection Policy to define your business needs. The Protection Policy allows you to incorporate the cloud storage External Target that you created earlier 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 you can 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 cloud object storage).

Protect Your Data

Protection groups combine operational requirements with the business requirements that are defined in a protection policy.

In the job, 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. See the all the advanced protection group settings in the Appendix.

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 job can have only one policy.

Recover Data from Your Archive

When the time comes to recover your archived data, Cohesity Data Cloud gives you three options:

- Restore entire data objects (VMs, databases, NAS, etc.).
- Recover individual files and folders.
- Retrieve your data onto an entirely new Cohesity cluster (for disaster recovery, etc.).

For instructions, see Recover Data from CloudArchive below.

Manage Your Cloud Storage Access Keys

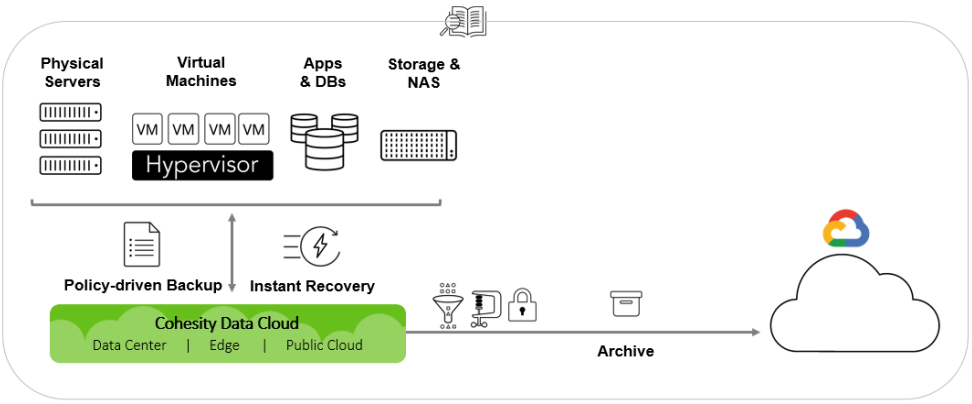
Most organizations have a corporate policy for changing passwords and access keys at regular intervals. You must update the access key to your cloud storage and then update your Cohesity External Target with the new key.

See instructions on rotating GCP service account private key at the end of the next chapter.

Connect Google Cloud Platform to Data Cloud

Cohesity's CloudArchive enhances Platform and DataProtect by adding seamless connectivity to Google Cloud Platform (GCP) storage as an extension of the data center infrastructure. Customers are using CloudArchive to reduce their reliance on tape for cost-effective long-term data retention.

Figure 7: Cohesity CloudArchive with Google Cloud Platform



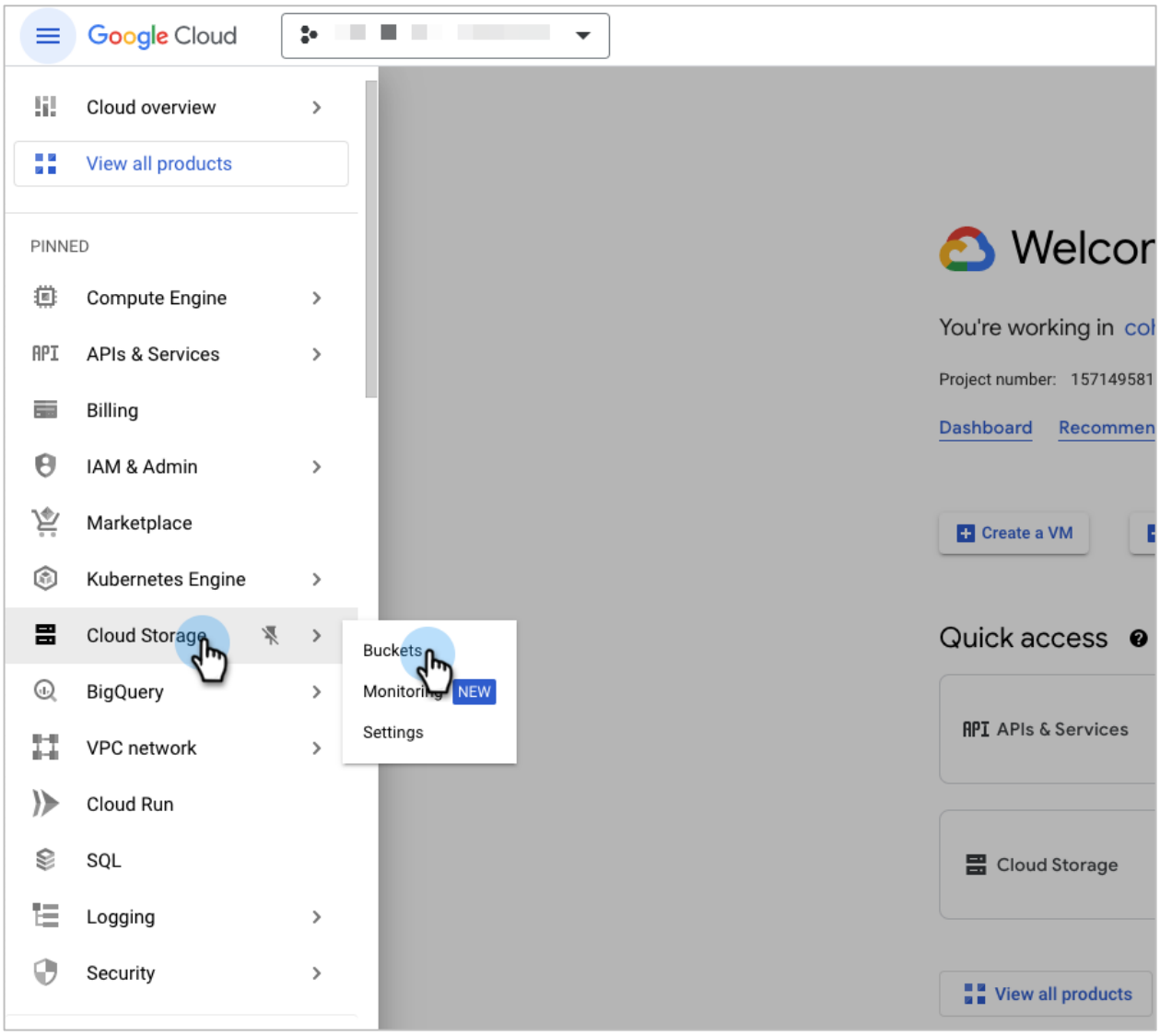
Create Your GCP Bucket for CloudArchive

Cohesity supports GCP Multi-Regional, Regional, Nearline, and Coldline cloud storage. Choose one for archiving your data in the steps below.

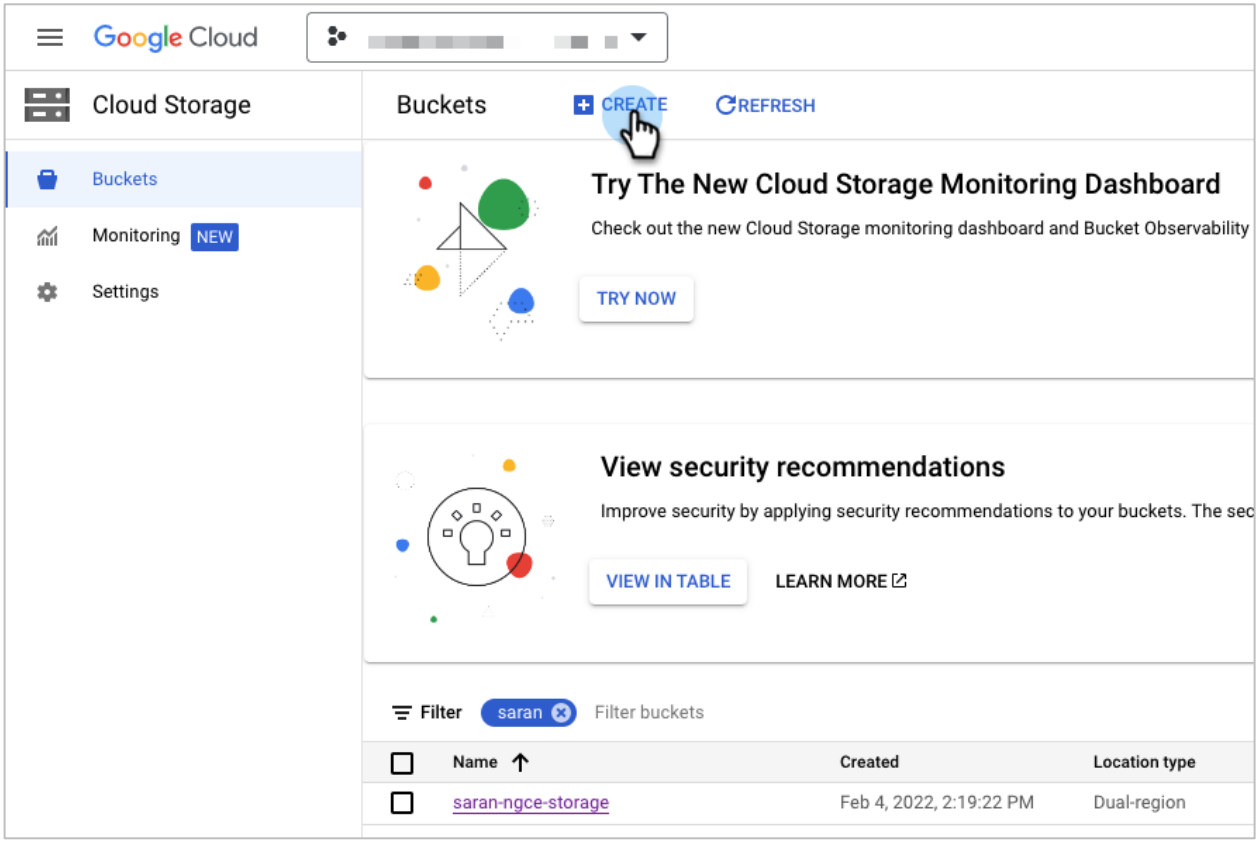
To create a GCP Bucket:

1. Sign in to your Google Cloud Platform console at: https://cloud.google.com/.

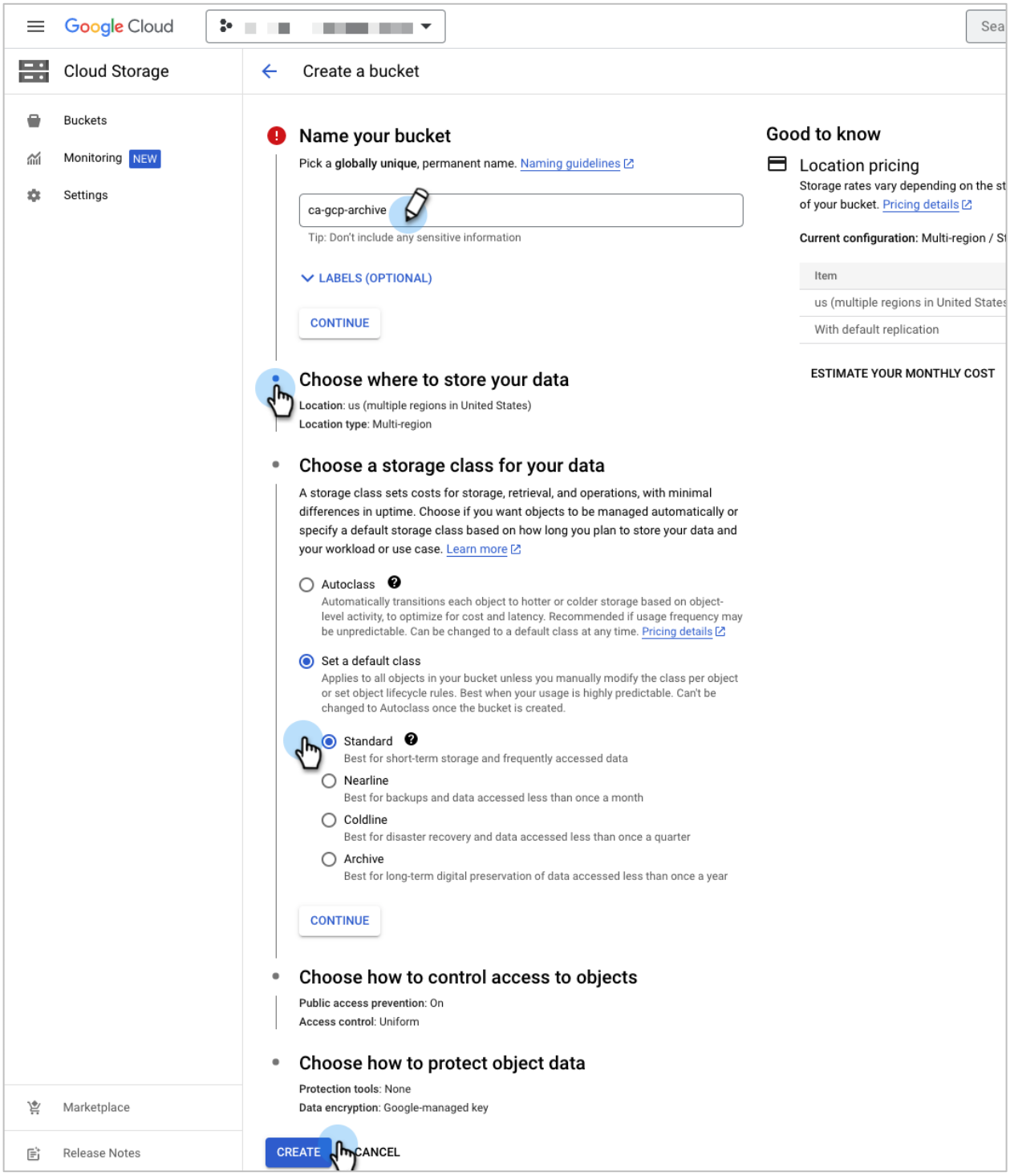
2. Select **Menu > Cloud Storage > Buckets**.



3. Click **CREATE**.



4. Enter a **Name** for your GCP bucket. Choose **Location** and the **storage class** (Standard, Nearline, or Coldline). Click **CREATE**.



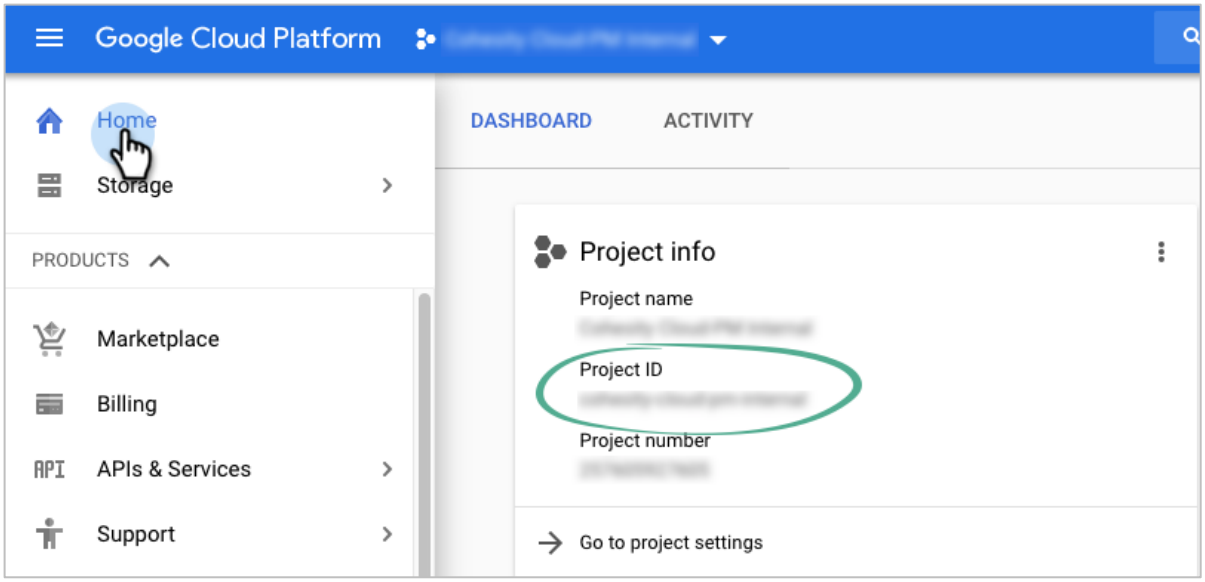
Your new bucket appears in the **Bucket details** window.

Create a GCP User and Capture Client Private Key

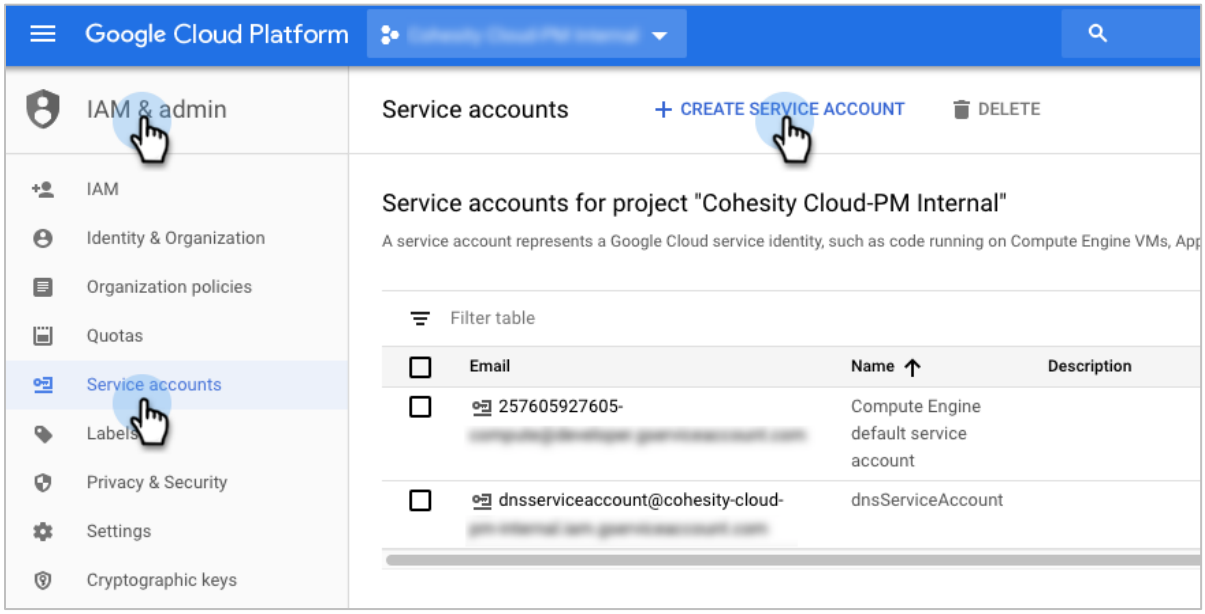
Next, you need to add an IAM (Identity and Access Management) user (a “service account” in GCP), a Project ID, and capture the access key for that user.

1. Sign in to your Google Cloud Platform console.

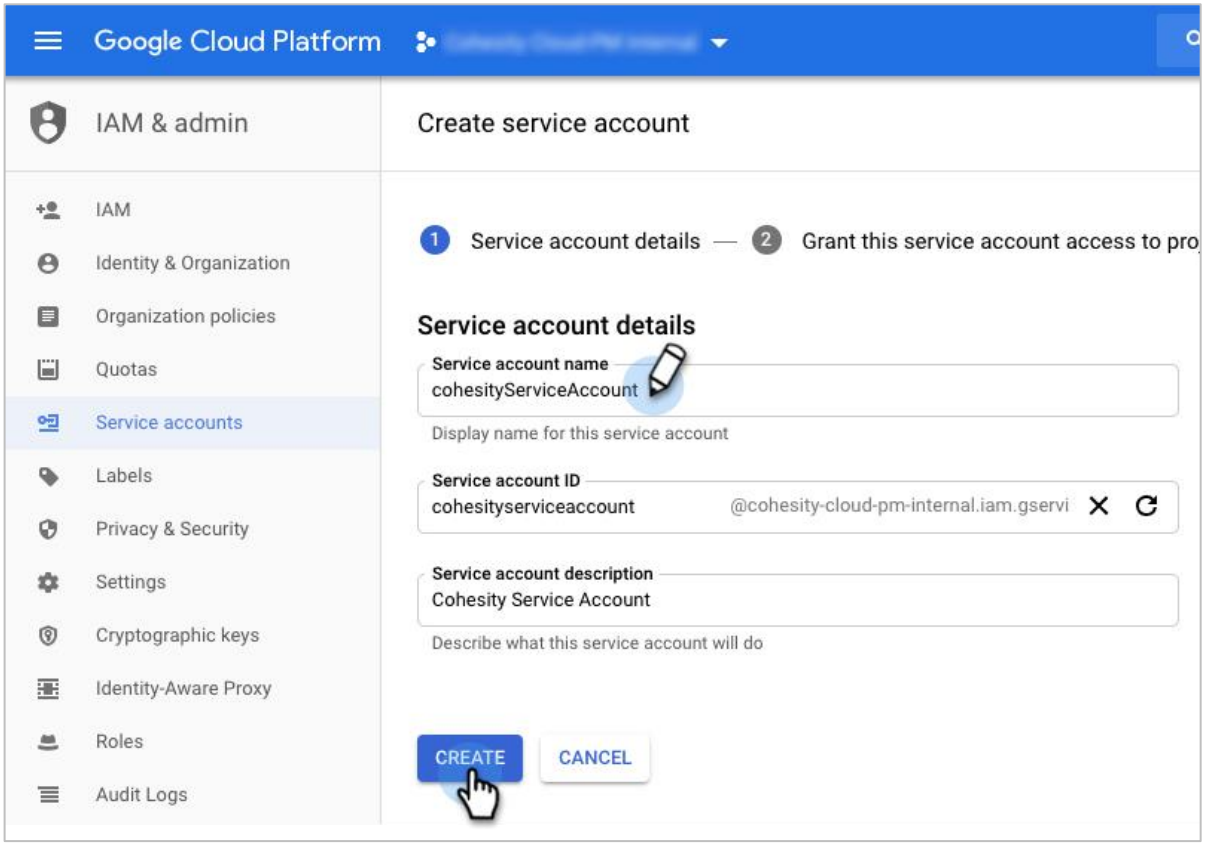
2. From the **Project info** tile, capture the **Project ID**.



3. Go to **Menu > IAM & admin > Service accounts** and click **CREATE SERVICE ACCOUNT**.

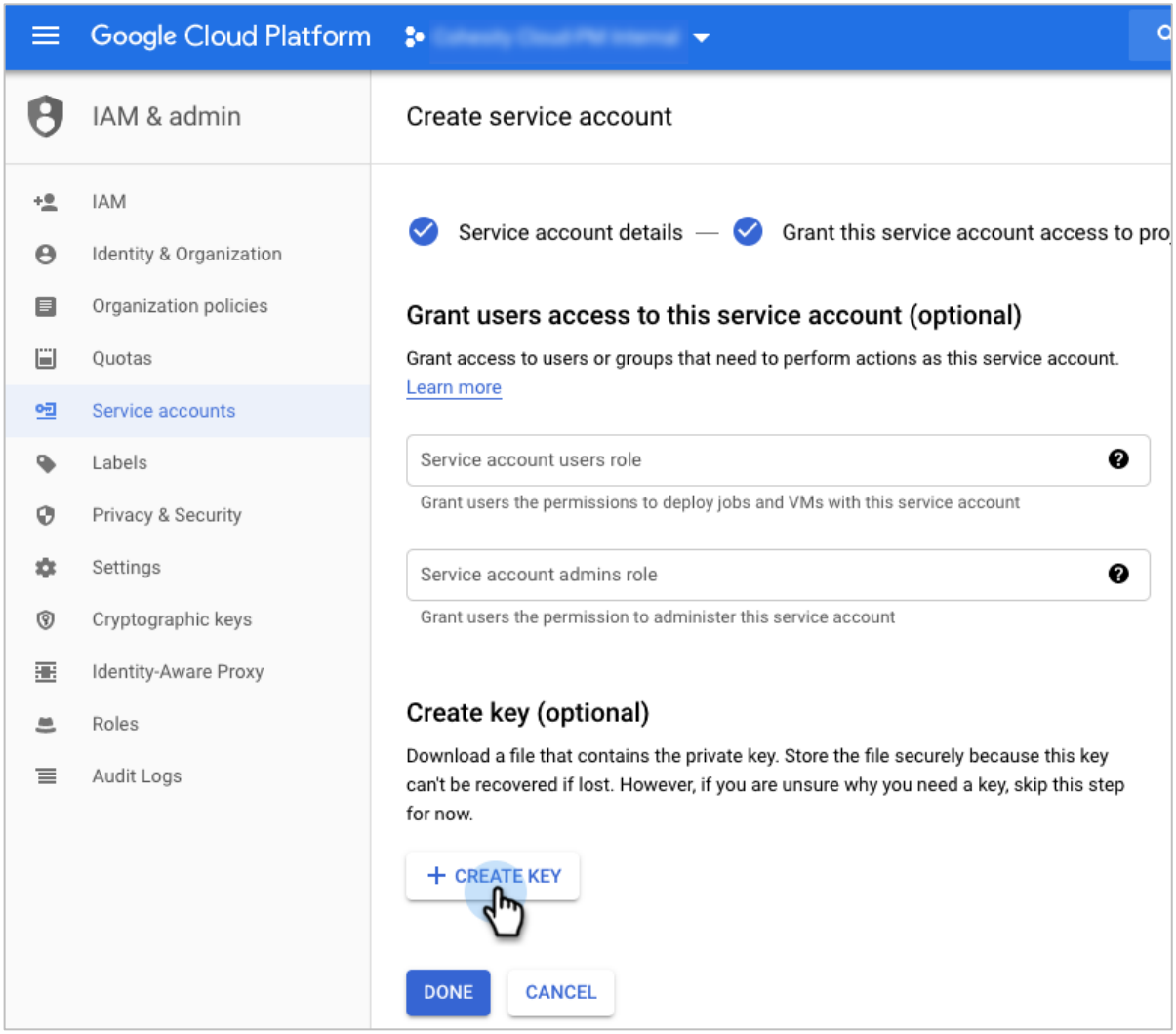


4. Enter a name for your service account and click **CREATE**.

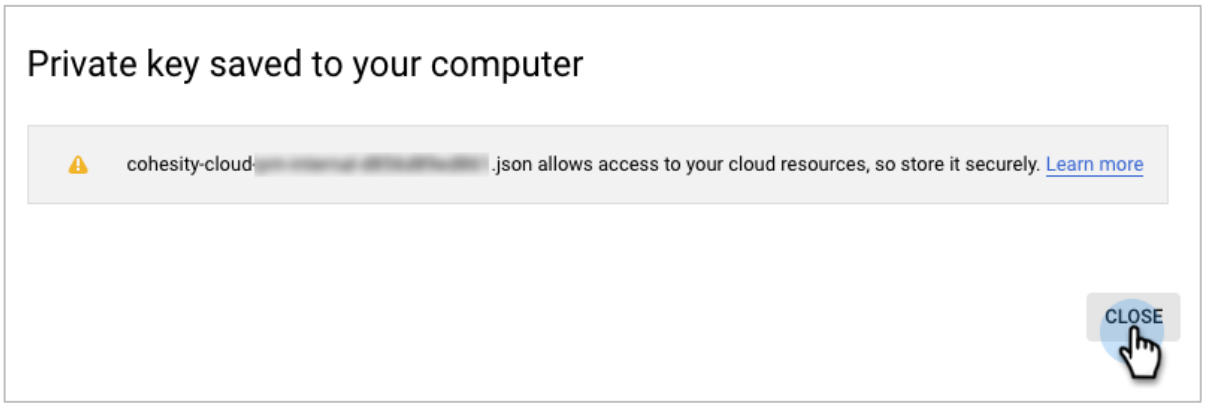


5. On the next screen, give the user **Owner** permissions, and click **CONTINUE**.

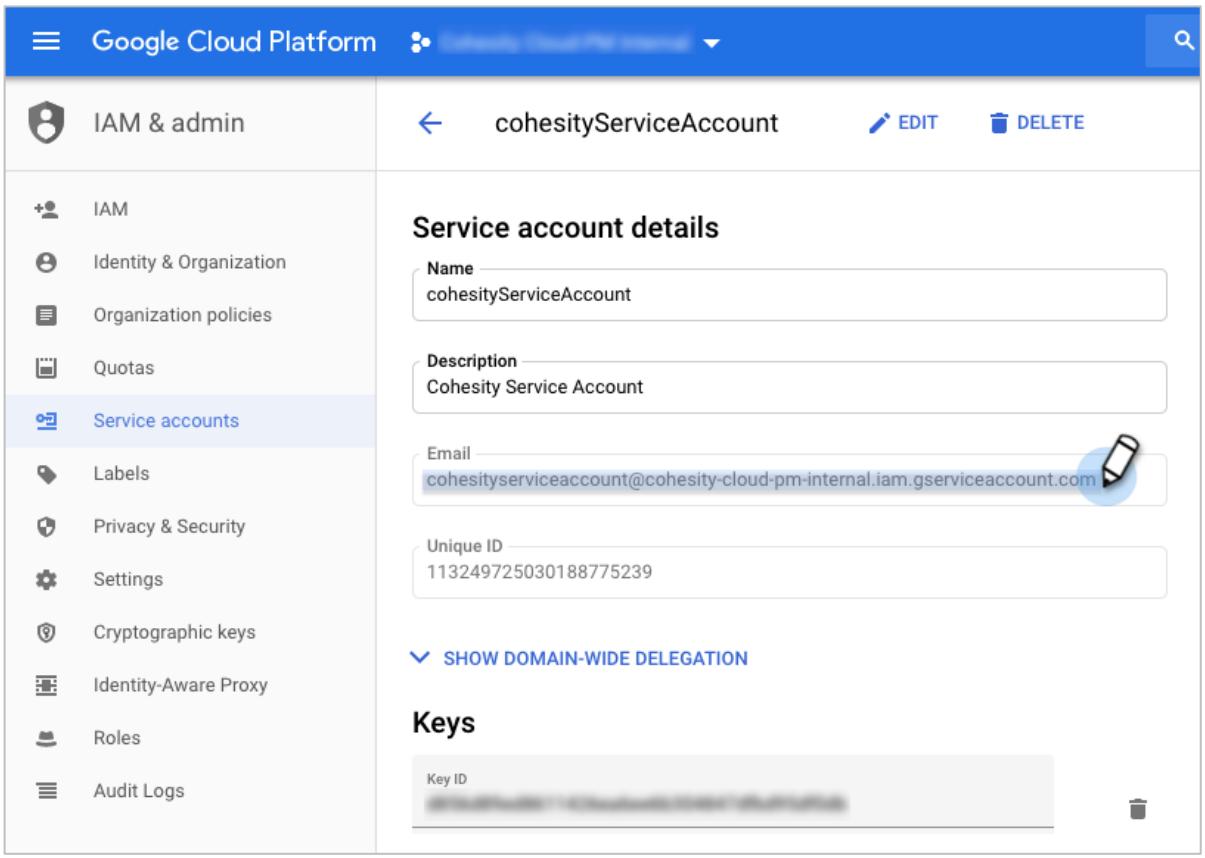
6. On the next screen, click **CREATE KEY**.



7. In the window that opens, leave the default key type, and click **CREATE**. The key file is saved to your computer. Click **CLOSE**.



8. Go to **Menu > IAM & admin > Service accounts**.
9. Click the account you just created.
10. Capture the service account **Email** address.



You are now ready to register your bucket with Cohesity in the next section.

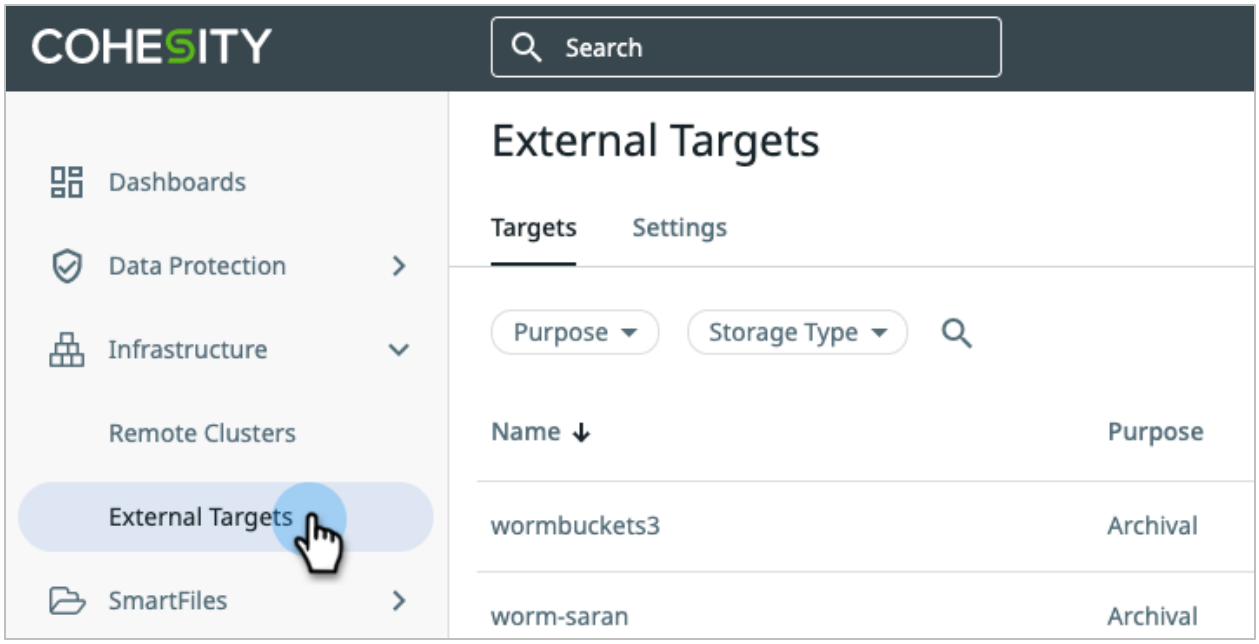
Register GCP Bucket with Data Cloud

Now that you have the bucket that you need, you're ready to connect it to Data Cloud (whether your cluster is on-premises, Cloud Edition, or Virtual Edition).

To register an External Target with your cluster:

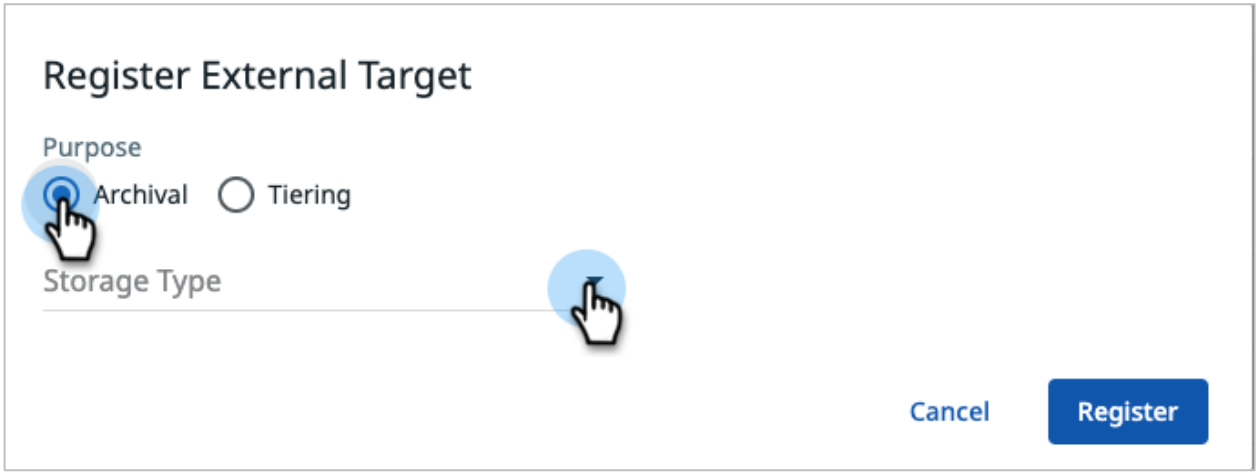
1. Log in to Data Cloud.

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

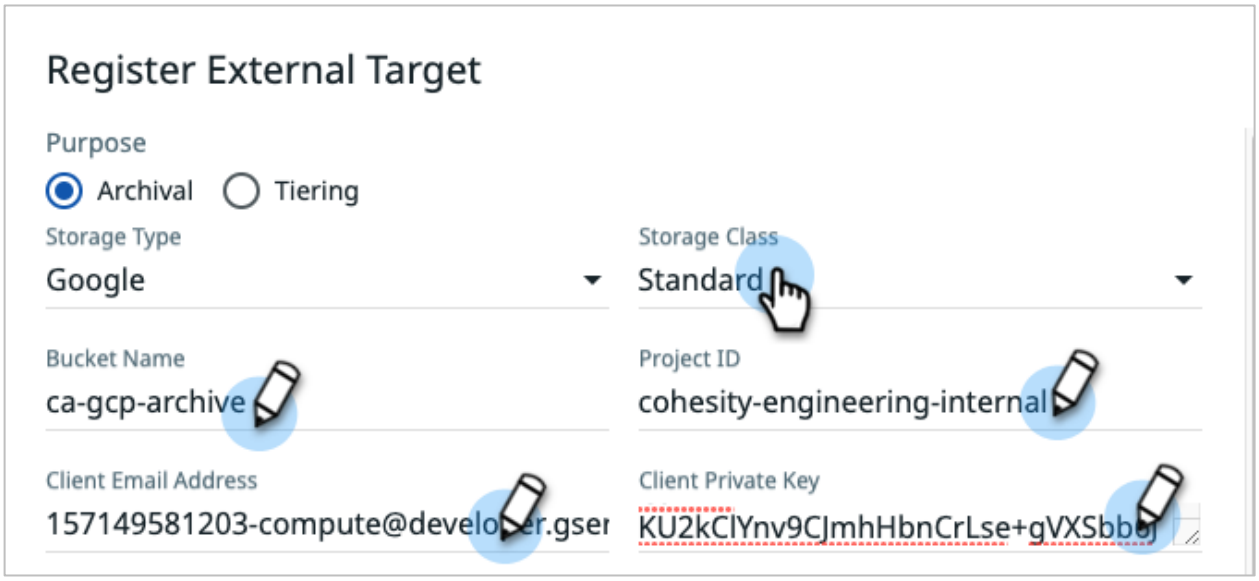


3. Click **Add External Target**.

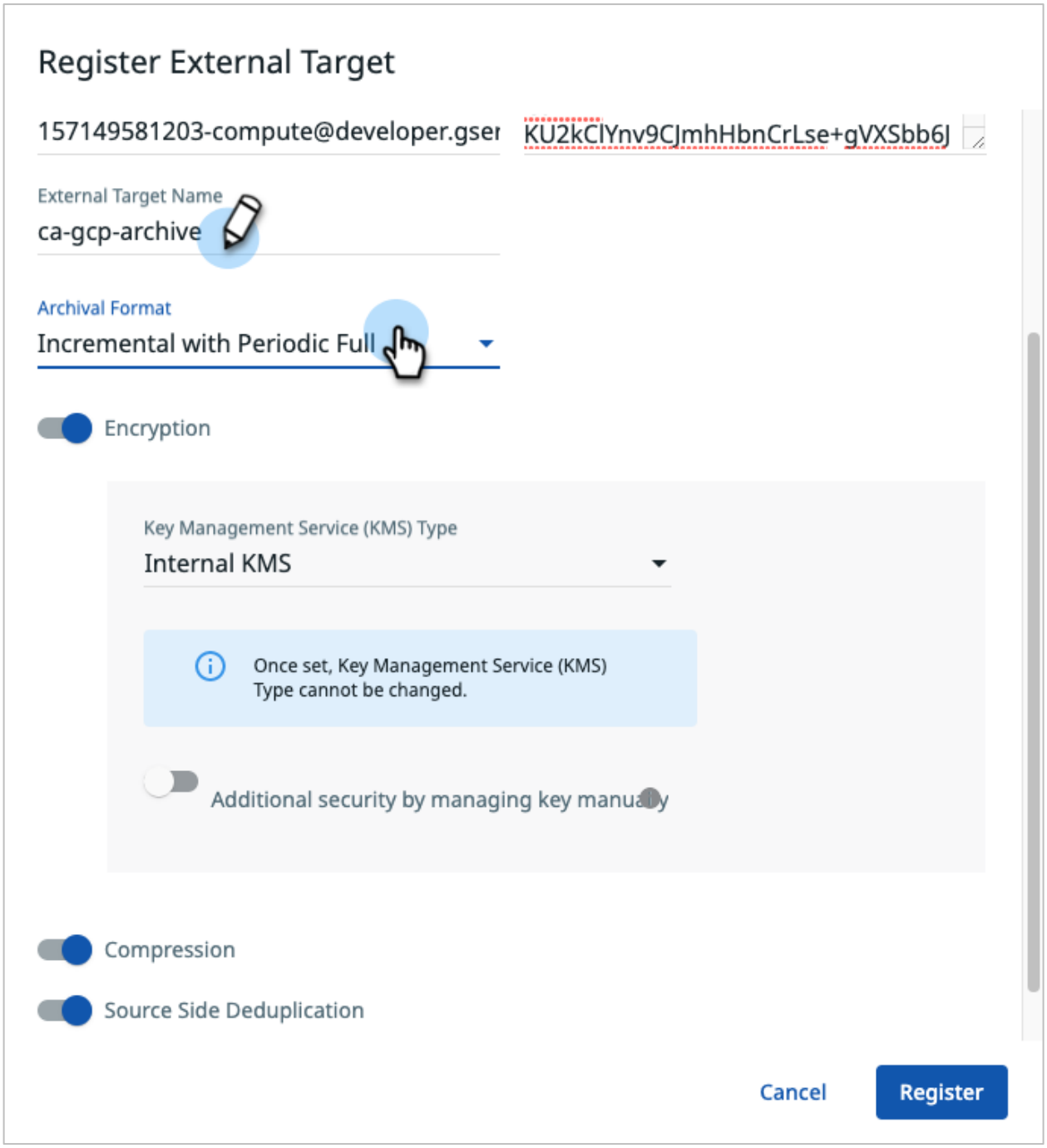
4. Select the purpose as Archival and Storage Type as Google.



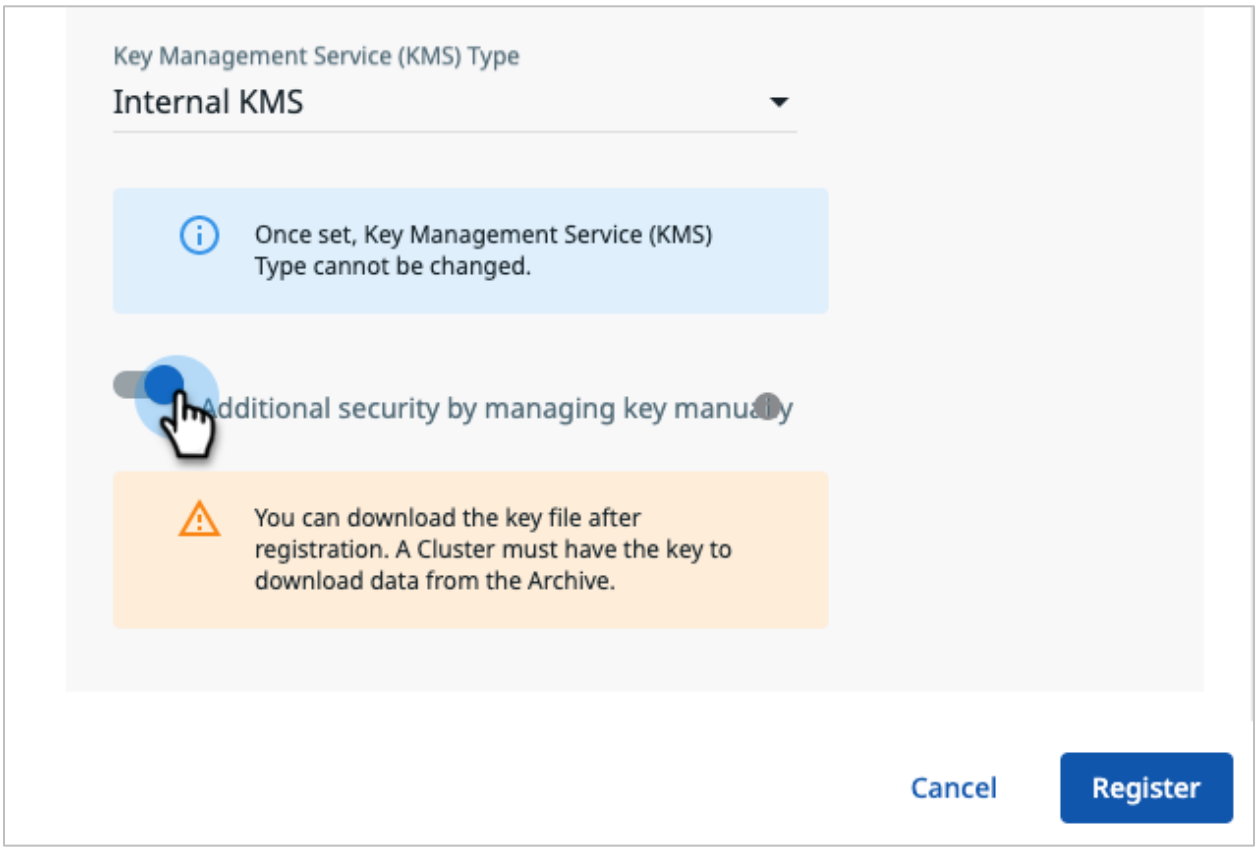
5. Select **Storage Class** (Stanadard,Coldline,Nearline), enter **Bucket Name**, **Project ID**, **Client Email Address**, and paste the **Client Private Key** from the key file that you downloaded.



6. On the same screen, enter a name of the external target. Review the default settings, by default, **Encryption**, **Compression**, and **Source Side Deduplication** are enabled, while **Additional security by managing key manually** and **Bandwidth Throttling** are disabled. Select the Archival format as **Incremental with Periodic Full** (Incremental Forever archival is not supported for google).

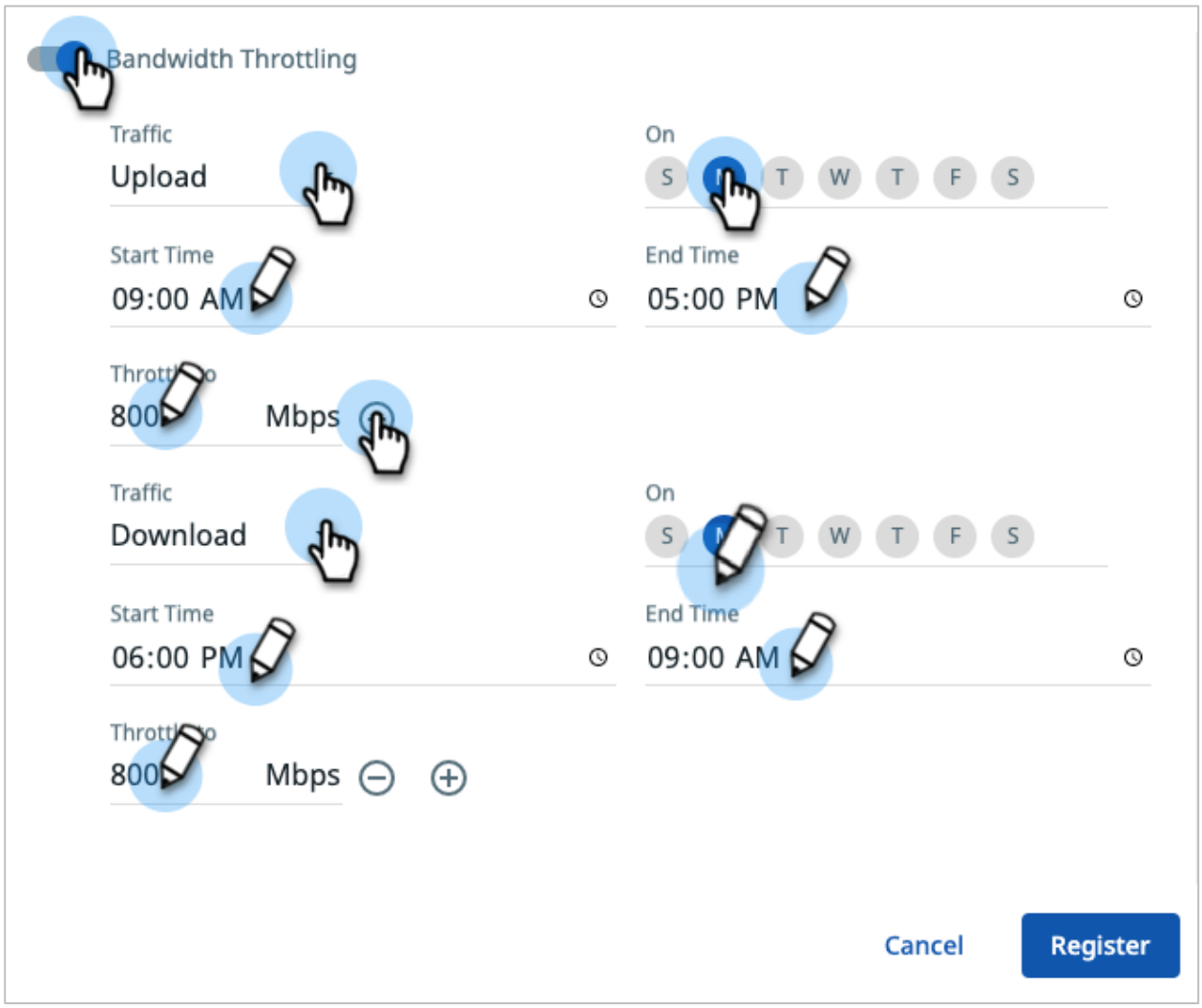


a. If you want to enable manual key management for extra security, turn it on it here:



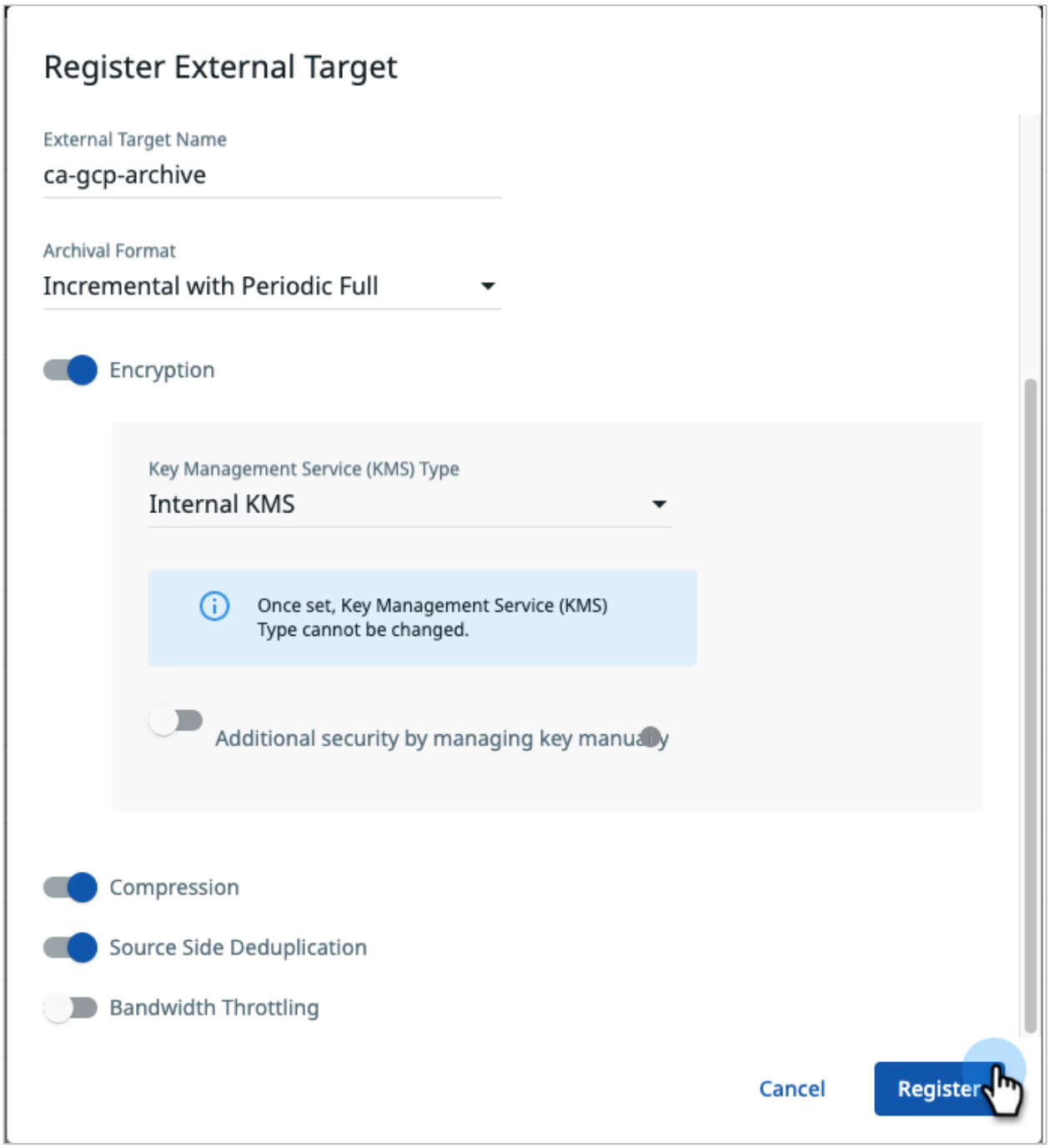
IMPORTANT: With this option on, a cluster must have the key to access data from the archive. You can download the key file (only once) after you register your bucket. This key is required when you use CloudRetrieve. If you do not have it, you will still be able to recover data to its original cluster, but you will not be able to retrieve it onto a new cluster (in a disaster-recovery scenario, for example).

b. Enable **Bandwidth Throttling** if needed. You can throttle upload and download speeds separately and apply throttling all the time or only specific days and times.



NOTE: For more on Encryption, Compression, Source Side Deduplication, Incremental Archival, and Bandwidth Throttling, see Create, Register Cloud Object Storage above.

7. Click **Register**.



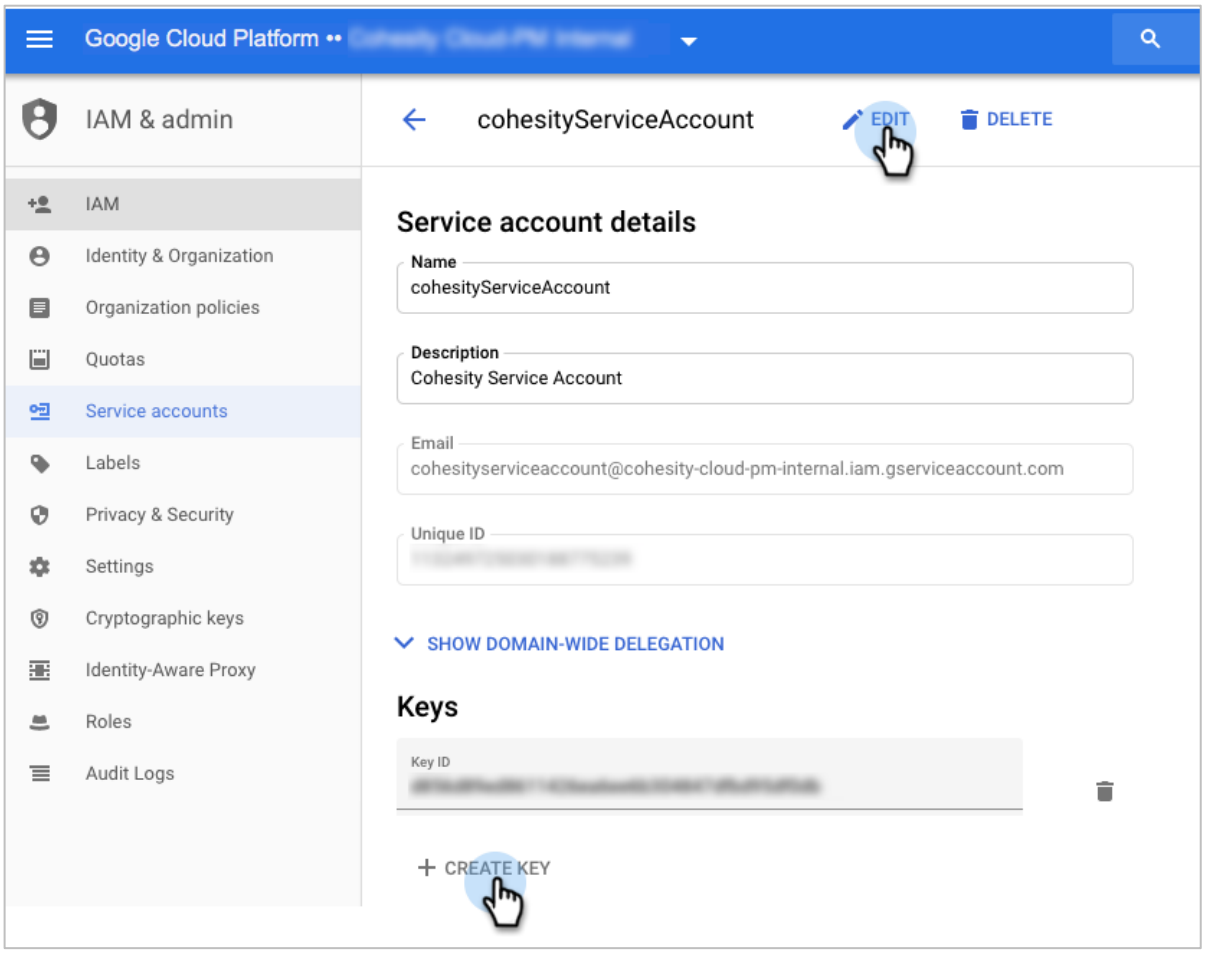
Your bucket is now an External Target in Data Cloud, and is available to select when you create a Cohesity Protection Policy for use in protection groups.

Rotate GCP Service Account Private Key

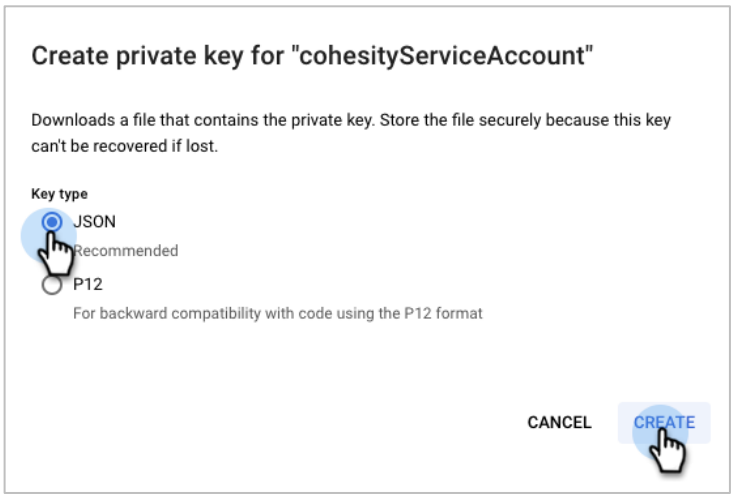
For security, it is important to rotate the access key to your External Target in GCP. Depending on your corporate policy for changing keys and passwords, when the time comes, you will have to rotate your GCP target's access keys and update your Cohesity External Target with the new keys.

To rotate the access keys:

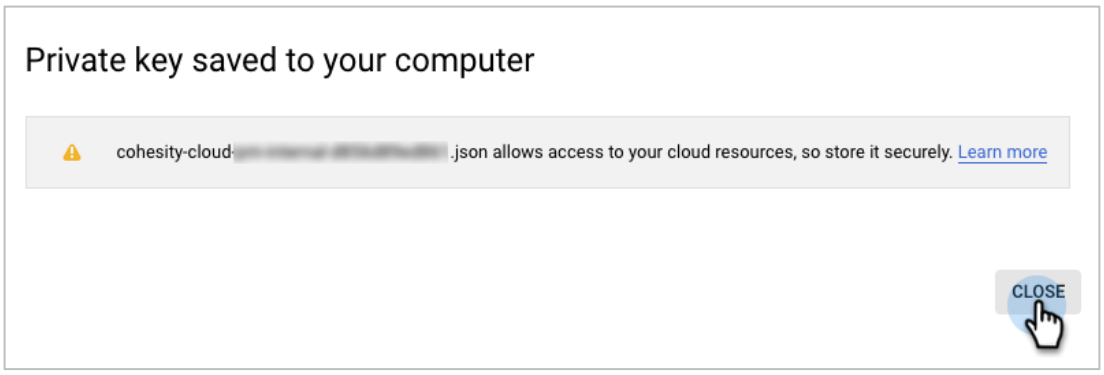
1. Sign in to your Google Cloud Platform console.
2. Go to **Menu > IAM & admin > Service accounts**.
3. Select the service account and click **EDIT**, then click **CREATE KEY**.



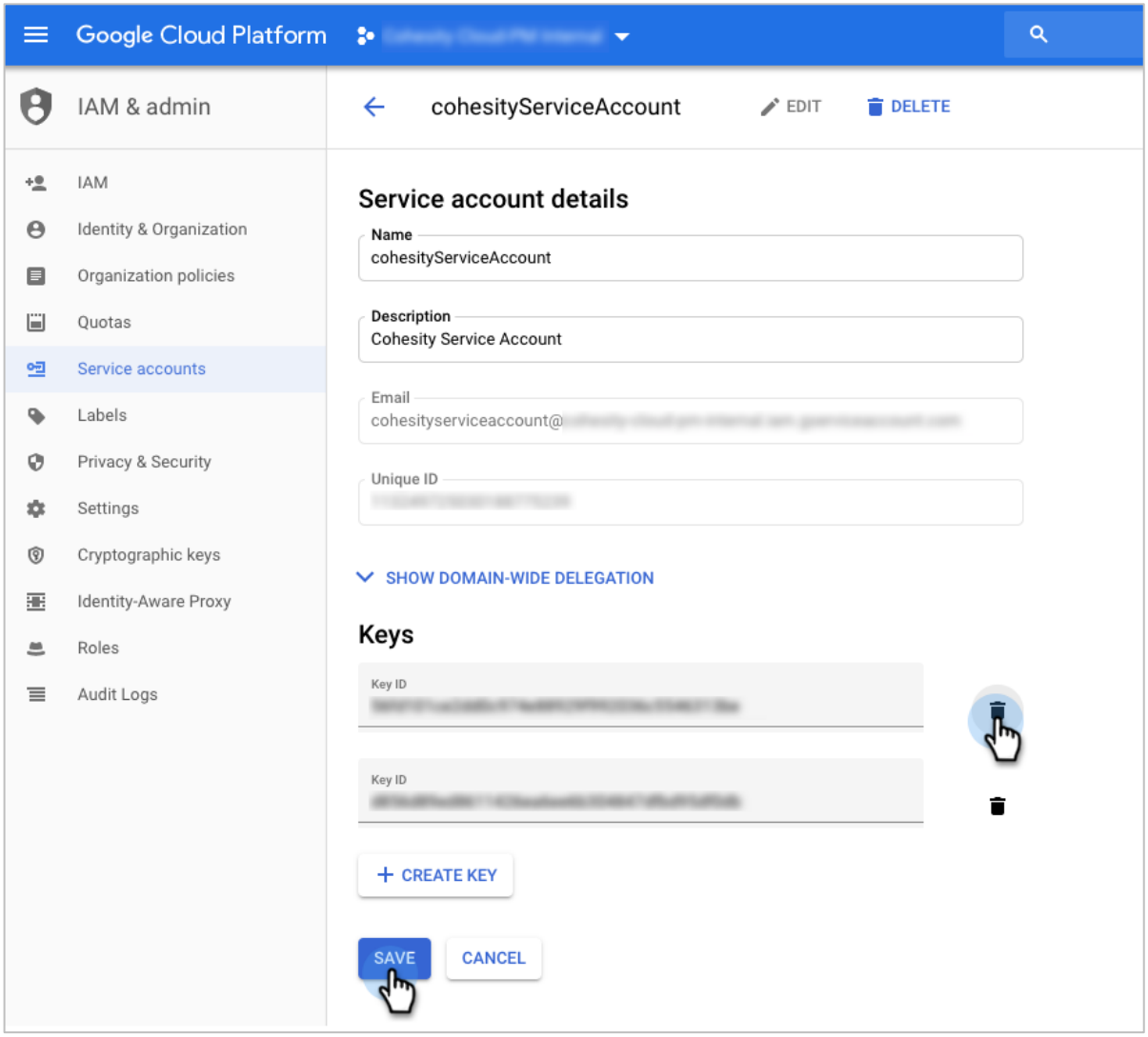
4. Select **JSON** and click **CREATE**.



5. Your new private key is saved to your computer. Click **CLOSE**.



6. Delete the old key and click **SAVE**.



With this, you have rotated your GCP keys!

Create a Protection Policy

In Cohesity Data Cloud, protection groups use protection policies. Protection policies reflect *business* needs, such as backup and archival frequency, retention requirements for each, as well as Recovery Point Objectives and Recovery Time Objectives, while a protection group defines *operational* requirements, such as which source objects to protect, the protection policy to use, and operational considerations like indexing, exclusions and inclusions, alerts, app consistency, and more. This process of combining business needs (policy) with the objects to protect (source data) and the operational requirements (job) provides rich flexibility to customers.

A protection policy defines:

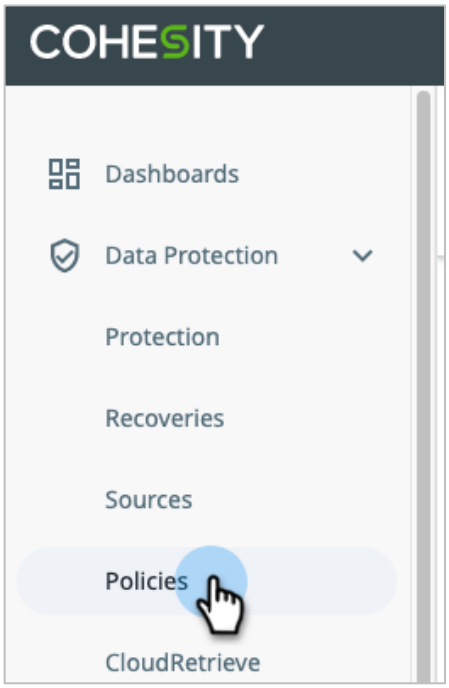
- How source data (like virtual and physical servers, databases, unstructured data, etc.) will be backed up and then archived.
- Where and how frequently they will be archived.
- How long the archives will be retained.

This list addresses parameters that affect CloudArchive operations. For the complete list of protection policy parameters, see Create or Edit a Policy in the online Help.

In the protection policy, you can select the cloud-based External Target you just created and registered as an External Target.

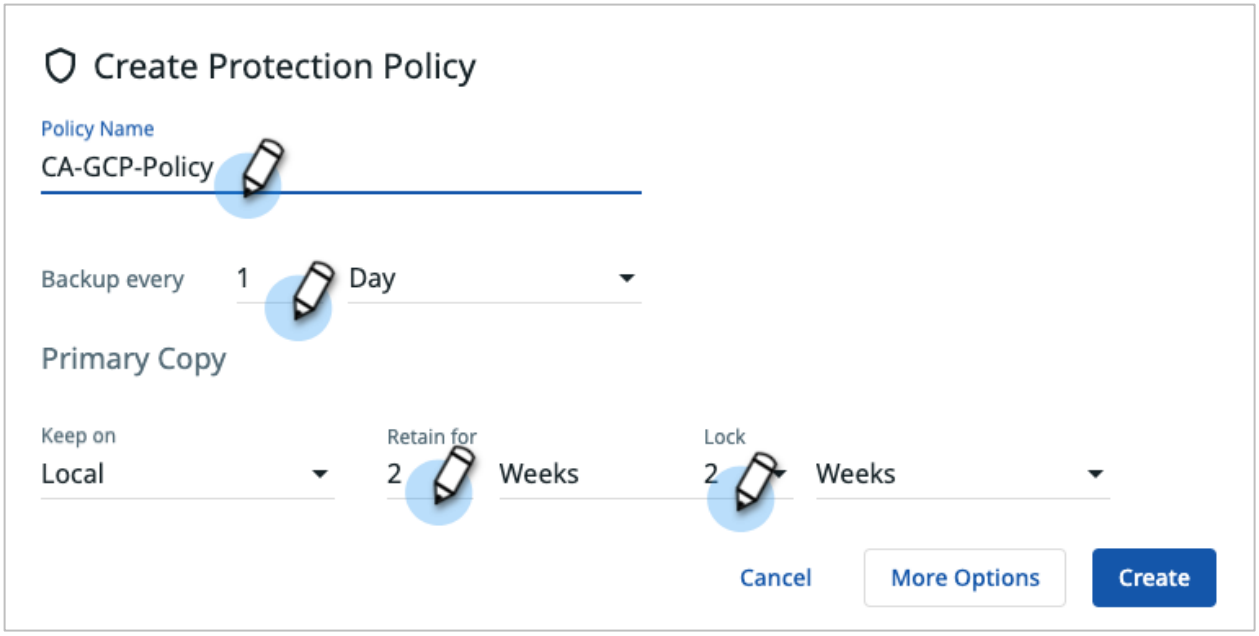
To create a protection policy:

1. Log in to Data Cloud.
2. Click **Protection** > **Policies**.



3. Click **Create Policy**.

4. In the form that opens, enter a **Policy Name** and select **More Options**.

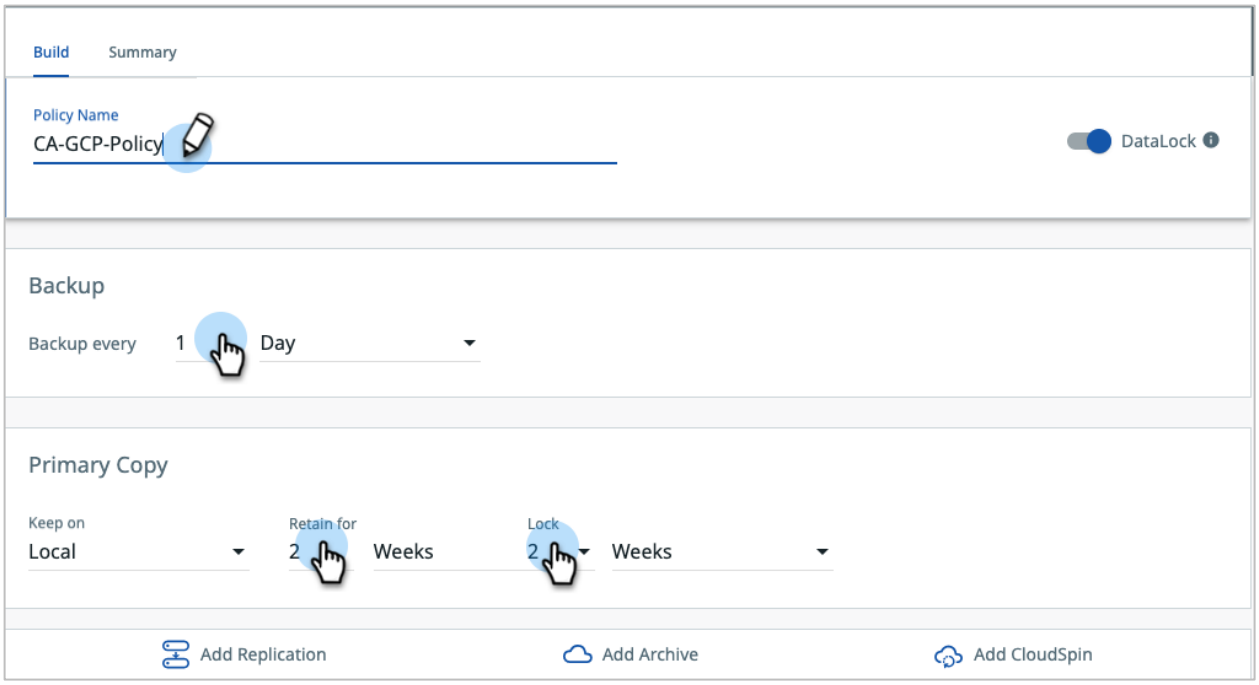


5. Add a **DataLock** for compliance and regulatory requirements, to ensure that your protected data, including local backups, archives, and replication, cannot be modified until the DataLock expiration.

   Once applied, a DataLocked Snapshot will be deleted only after its retention period expires. A DataLock prevents all users, including those who have the Data Security role in Data Cloud, from modifying or deleting any Snapshots that were generated by the protection groups that use this policy. Only users with the Data Security role can add, modify, or remove a DataLock from a Policy. See online Help for more information.

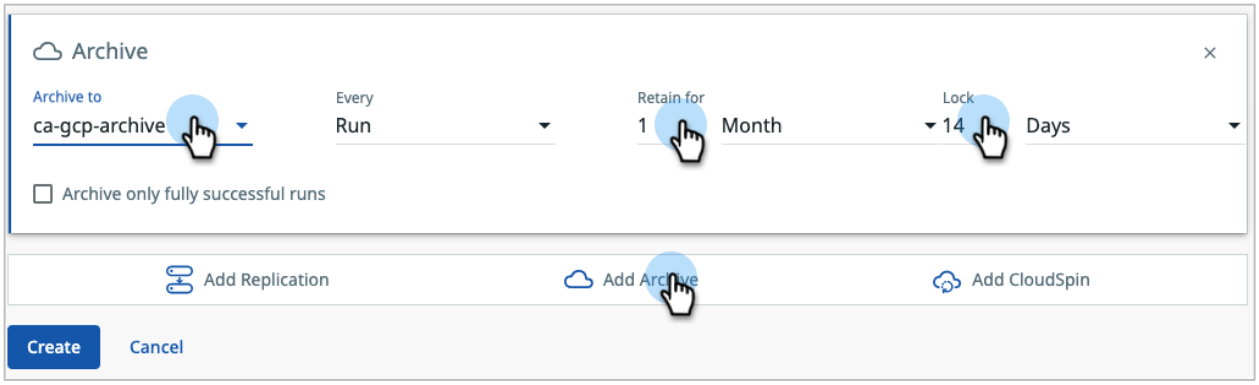
NOTE: You can also add a legal hold to a specific protection group run (a *Snapshot*) to preserve it for legal reasons. See Apply Legal Hold to Completed Job Run below.

6. Under **Backup**, set the **Backup** interval (every day, by default) and **Retain for** period (90 days, by default) for your local cluster.



7. Click **Add Archive**, and for **Archive to**, select the External Target you just created. Set the **Archival** interval (every day, by default) and **Retain for** period. You can also enable **Archive only fully successful Runs** in the checkbox below.

   Click **Add Archive** again if you need additional archival schedules.



NOTE: You can add multiple archival schedules that use the same or different External Targets, as well as the same or different intervals and retention periods, to a given Protection Policy. When you add more schedules and send them to the same External Target with different retention and schedule times, the schedules rationalize among themselves and only the necessary archive is sent, with the longest retention.

For example, if you add these three archival schedules to the same External Target:

- Once a day, retain for 90 days.
- Once every 7 days, retain for 180 days.
- Once every 30 days, retain for 365 days.

Then:

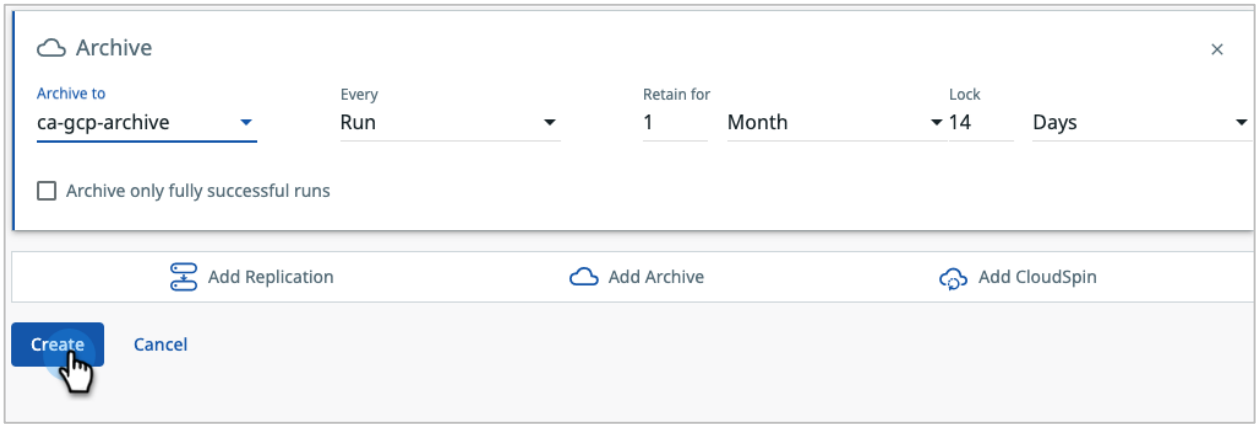
- On Day 7, only one archive is sent, meeting both Schedule 1 and Schedule 2 (and retained for 180 days, per Schedule 2, as it is the longer of the two).
- On Day 30, only one archive is sent, meeting both Schedule 1 and Schedule 3, but is retained for 365 days, to meet the Schedule 3 retention requirement.

By contrast, if you send the archives to different External Targets, then:

- On Day 7, per Schedule 1, the archive is sent to the first External Target and retained for 90 days. Per Schedule 2, the archive is also sent to the second External Target and retained for 180 days.
- On Day 30, per Schedule 1, the archive is sent to the first External Target and retained for 90 days. Per Schedule 3, the archive is *also* sent to the third External Target and retained for 365 days.

When you use multiple schedules with different External Targets, the schedules don't rationalize, and you accrue network and storage usage for each scheduled run.

8. Click **Create**.



Your new Policy can now be used in protection groups. For the complete list of Protection Policy parameters, see online Help.

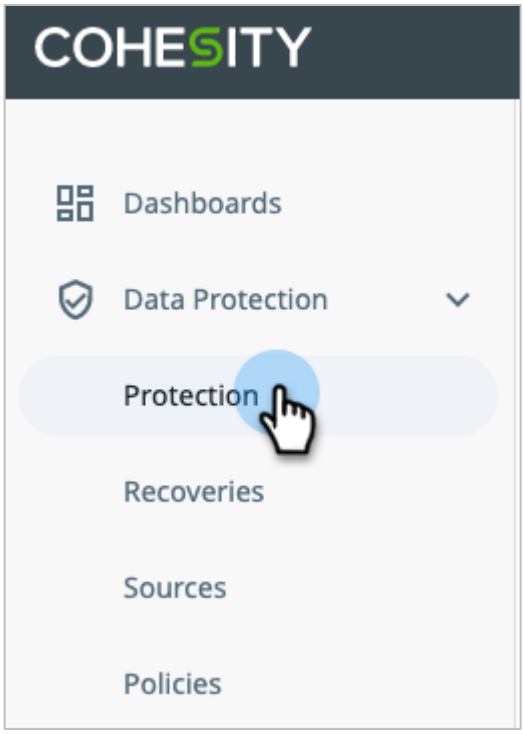
Create a Protection Group

Protection groups combine operational requirements with the business requirements that are defined in a protection policy. Multiple protection groups can use the same protection policy, but each group can have only one policy. Protection groups protect specific source objects, such as virtual servers, physical servers, Views, SQL servers, Oracle databases, Remote adapters, Pure Storage Volumes, or network-attached storage (NAS).

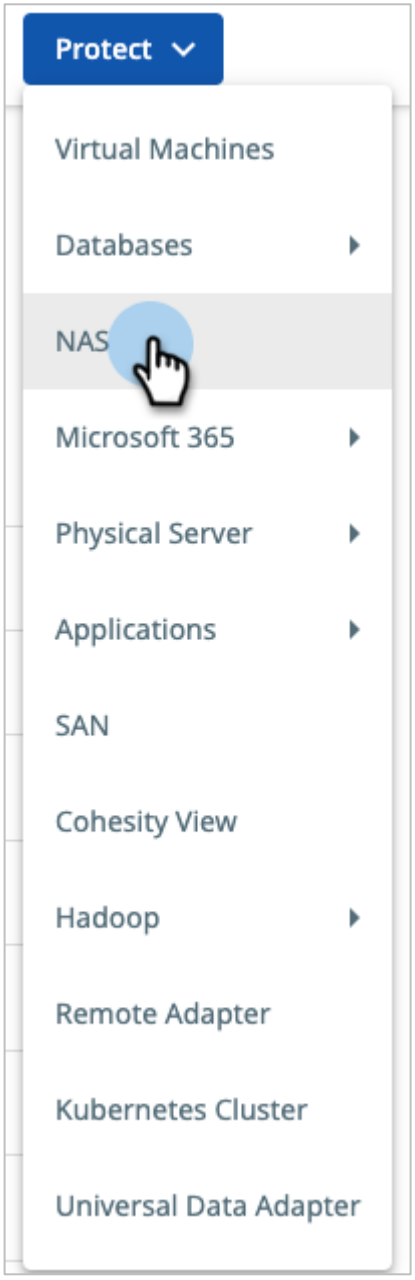
For this example, we look at the steps to create a protection group for NAS data, but the steps to protect other source objects are very similar.

To create a protection group:

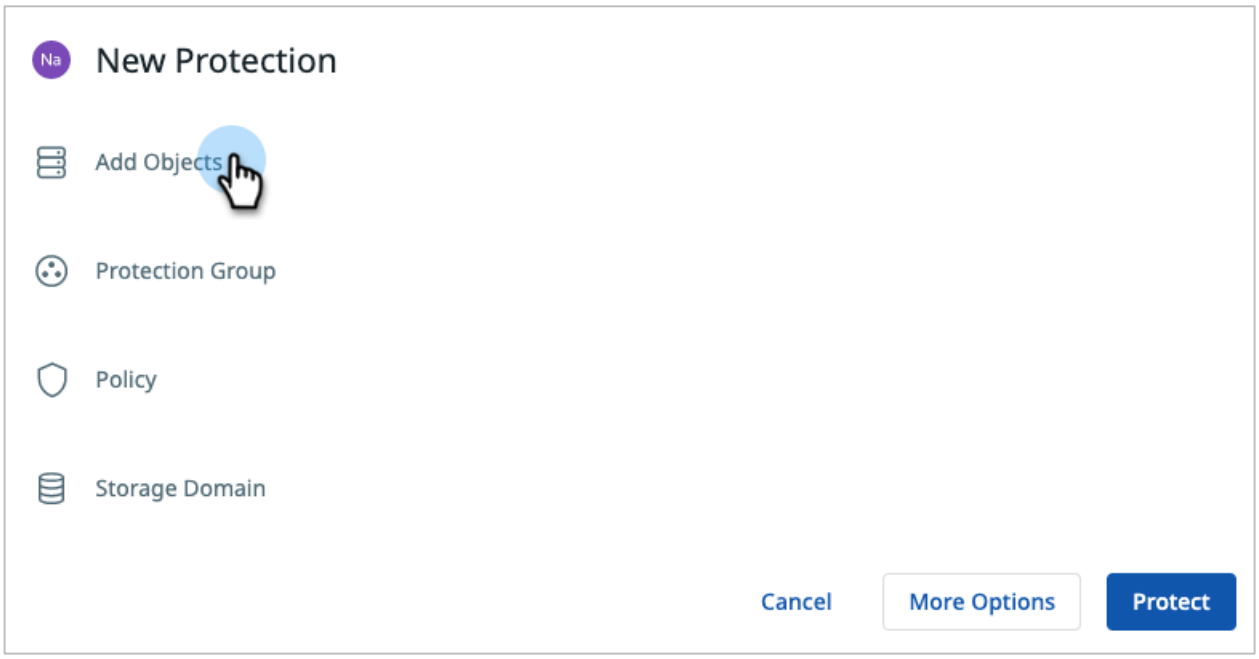
1. Log in to Data Cloud.
2. Click Data **Protection > Protection**.



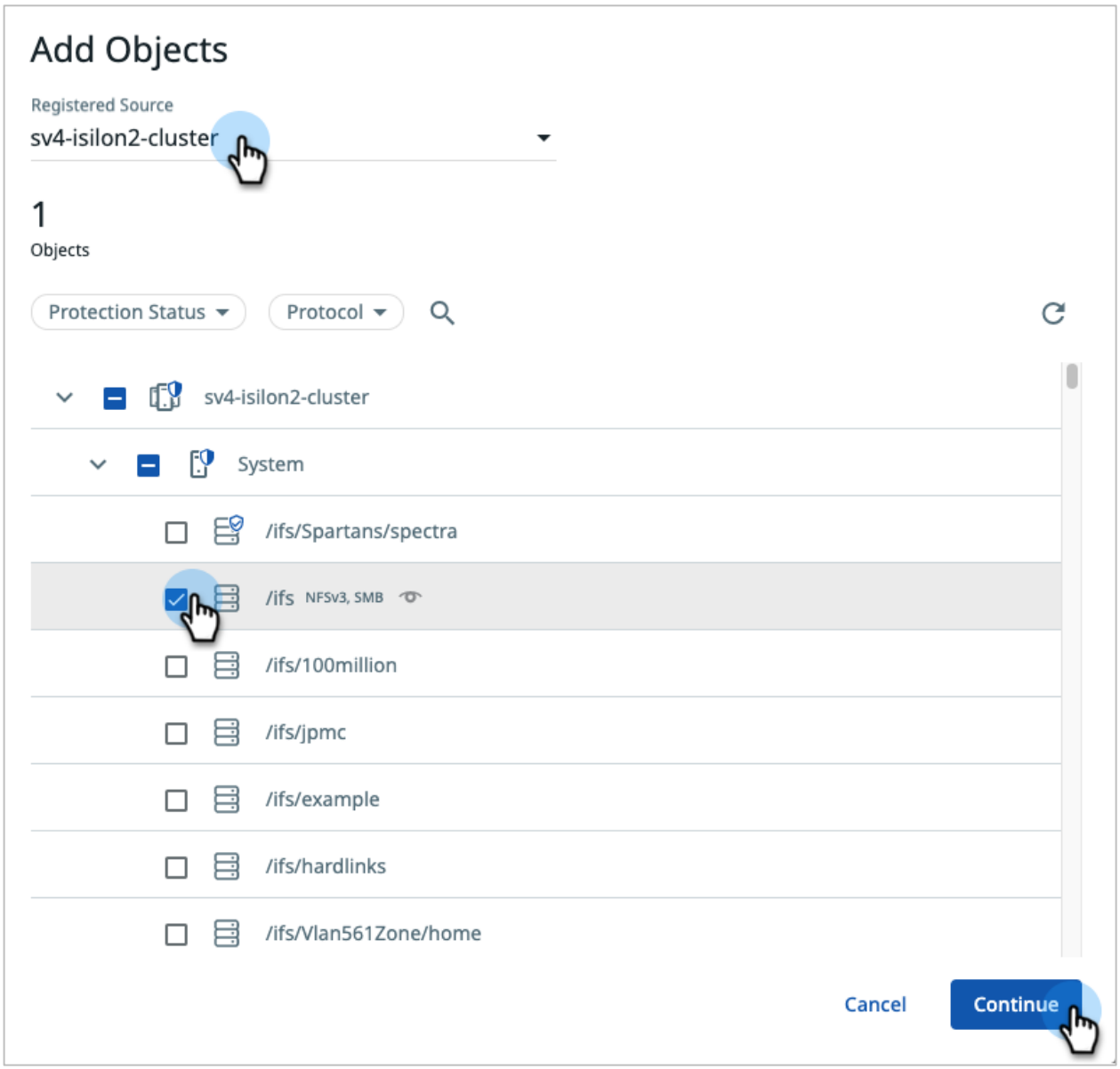
3. Click **Protect** and choose the type of data to protect.



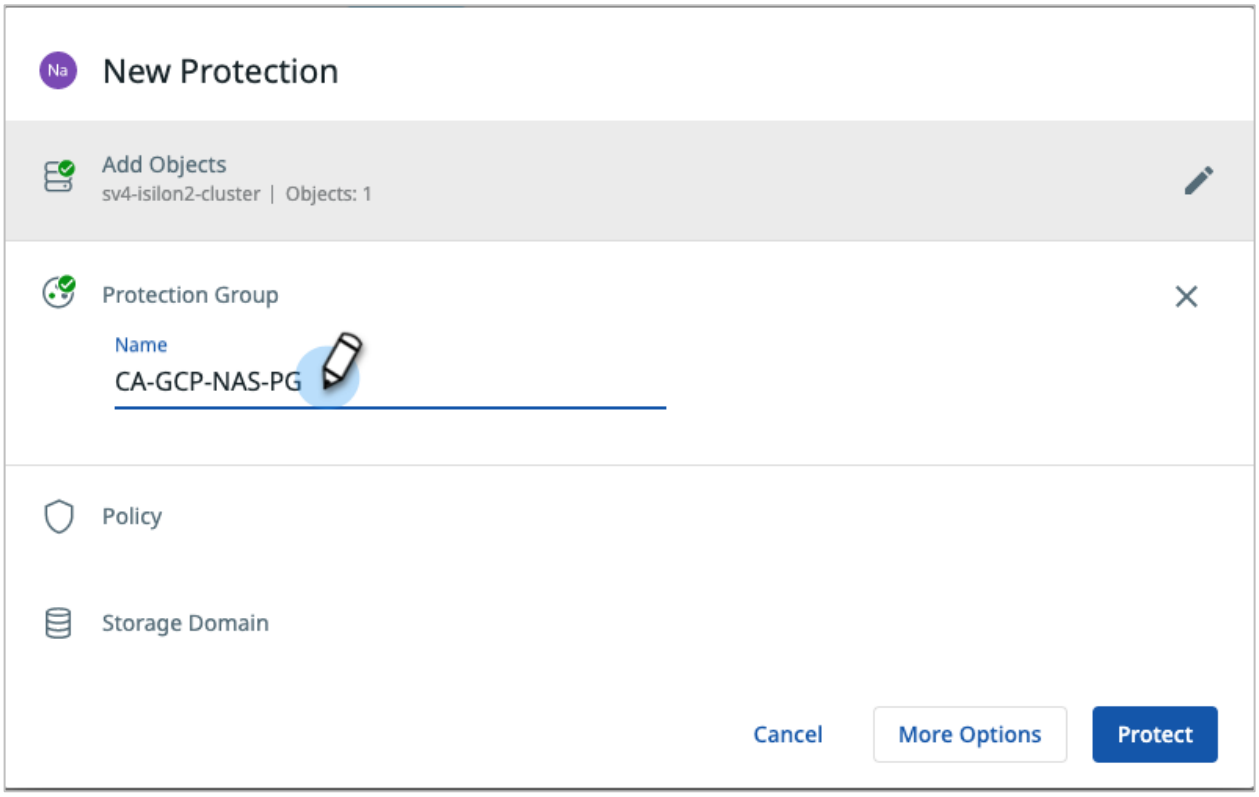
4. Click Add Objects to select a source to protect.



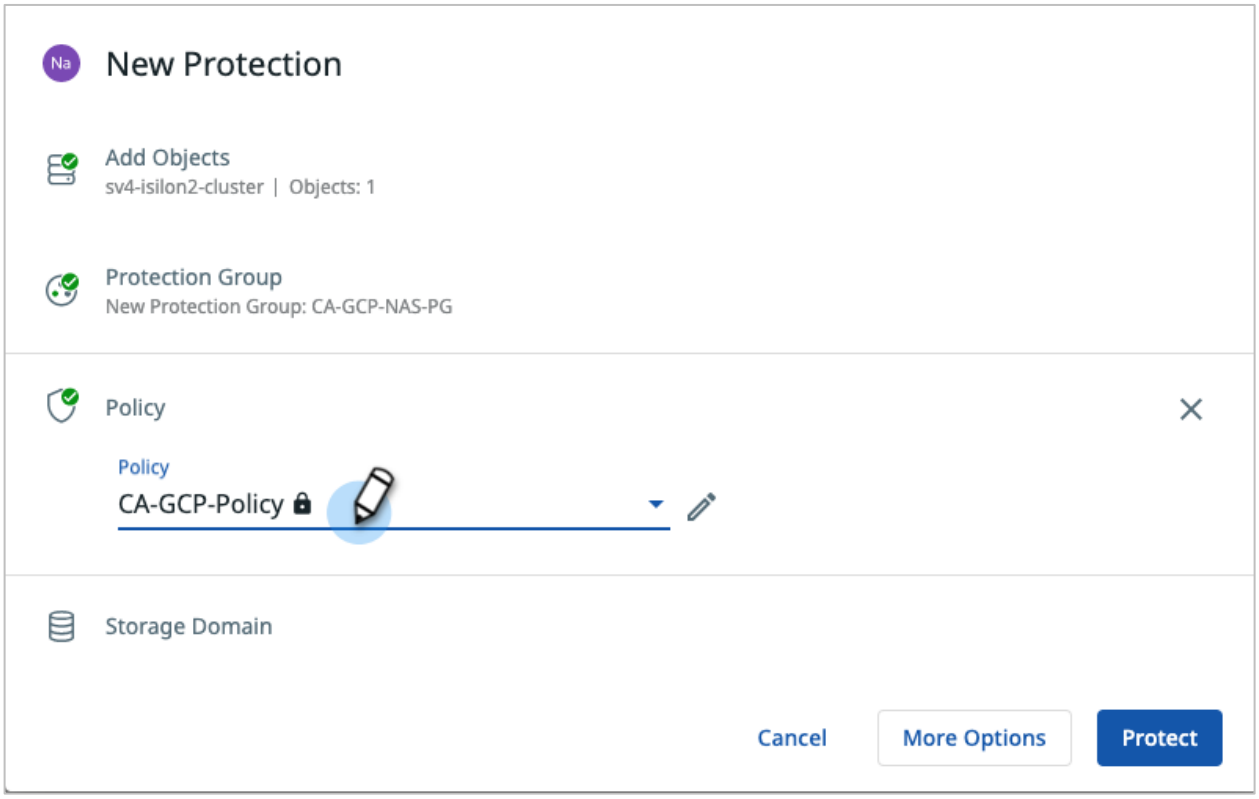
5. Select the specific objects you wish to protect with this Protection Group, then click **Continue**.



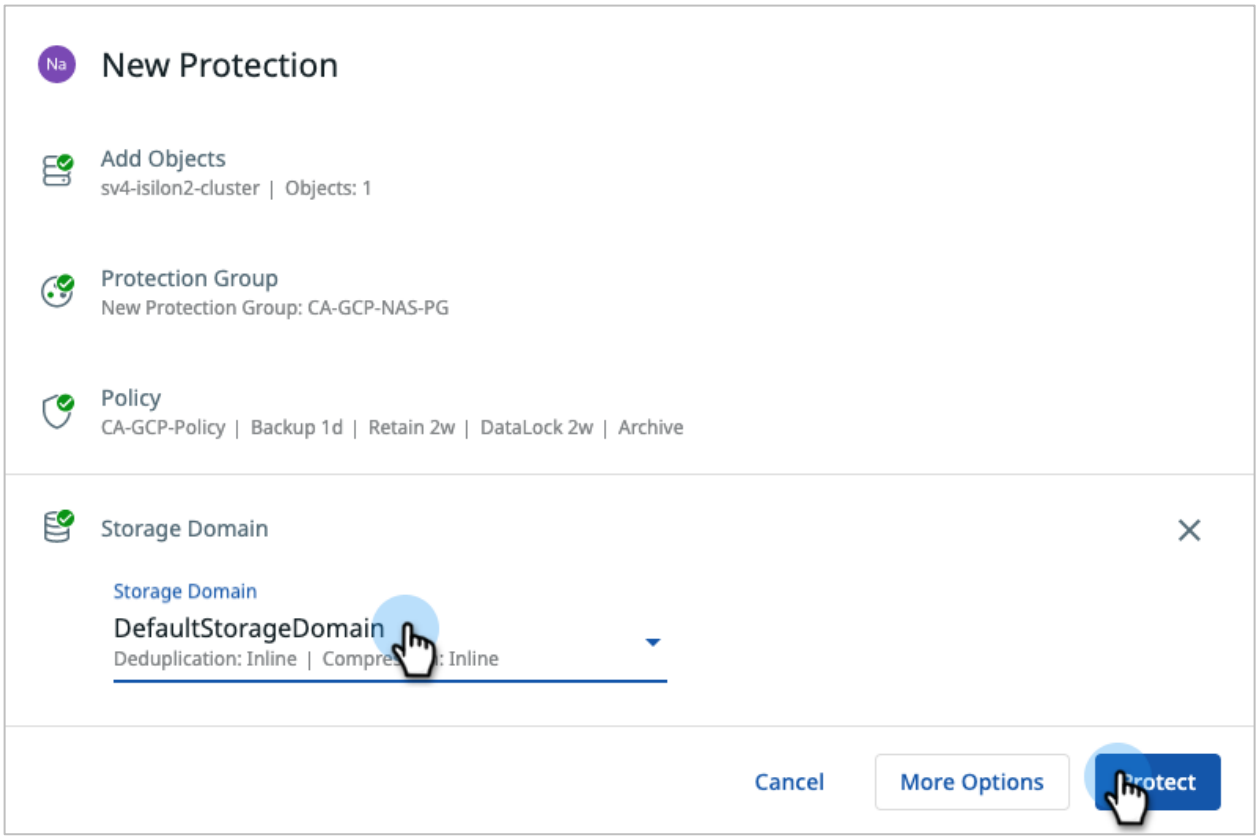
6. Name the **Protection Group**.



7. Select a **Policy**.

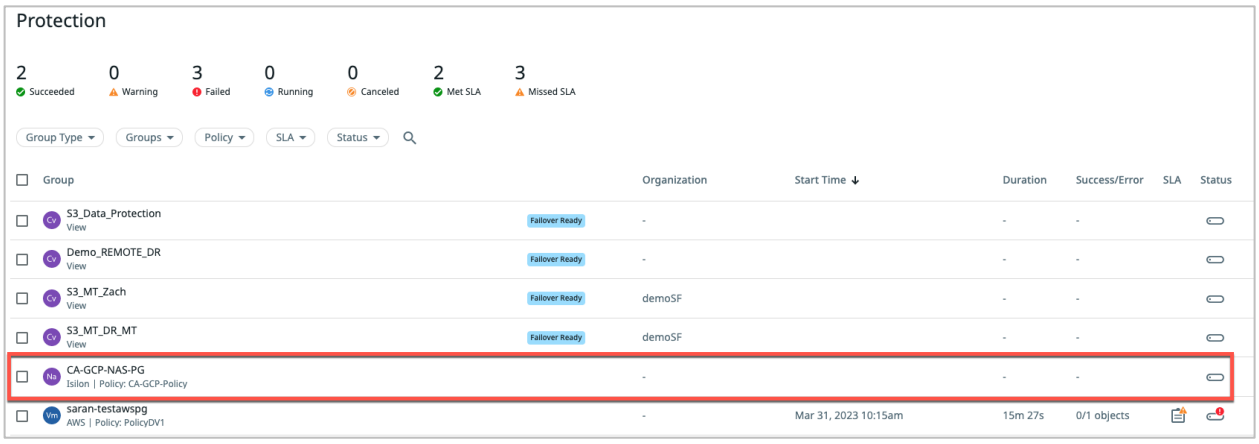


8. On the same screen, select a **Storage Domain**. Click **More Options** If you need to change any of the **Advanced** settings. When you're done, click **Protect**.



NOTE: See the complete list of Advanced settings and the job types that contain them in the Appendix.

9. In the top navigation, select **Protection > Protection Groups** to verify that your new Group is in the list.



Your new Protection Group is now active and running. To manage Protection Groups, see online Help.

Apply Legal Hold to Completed Job Run

Only users who are assigned the Data Security role can put a legal hold on existing Snapshots (protection group job runs), to preserve them for legal purposes. Once a legal hold is applied, the retention period is ignored and the Snapshot is preserved until the legal hold is removed. Legal hold Snapshots can only be deleted by a user with the Data Security role.

NOTE: A legal hold can be added to both regular and DataLocked Snapshots.

You can add a legal hold to a protection group job run or to individual objects in a protection group job run:

- If you add a legal hold to a protection group job run, it applies to all the Snapshot objects that were backed up by that protection group job run, and the legal hold is propagated to replicated and archived objects.
- If you add a legal hold only to selected objects in a protection group job run, the legal hold is propagated to archived objects, but not to replicated objects. You must manage the legal hold status on the remote replication cluster manually.

  NOTE: A legal hold prevents Snapshots from being deleted until the legal hold is removed. Using a legal hold for long periods of time can result in the cluster running out of space.

To add or remove a legal hold from a protection group job run, see Adding a Legal Hold to a Snapshot in the online Help.

The Difference Between Legal Hold and DataLock

While both a legal hold and DataLock are features that empower the Data Security role in Data Cloud to prevent backed up and archived data from being deleted, they differ in purpose and function.

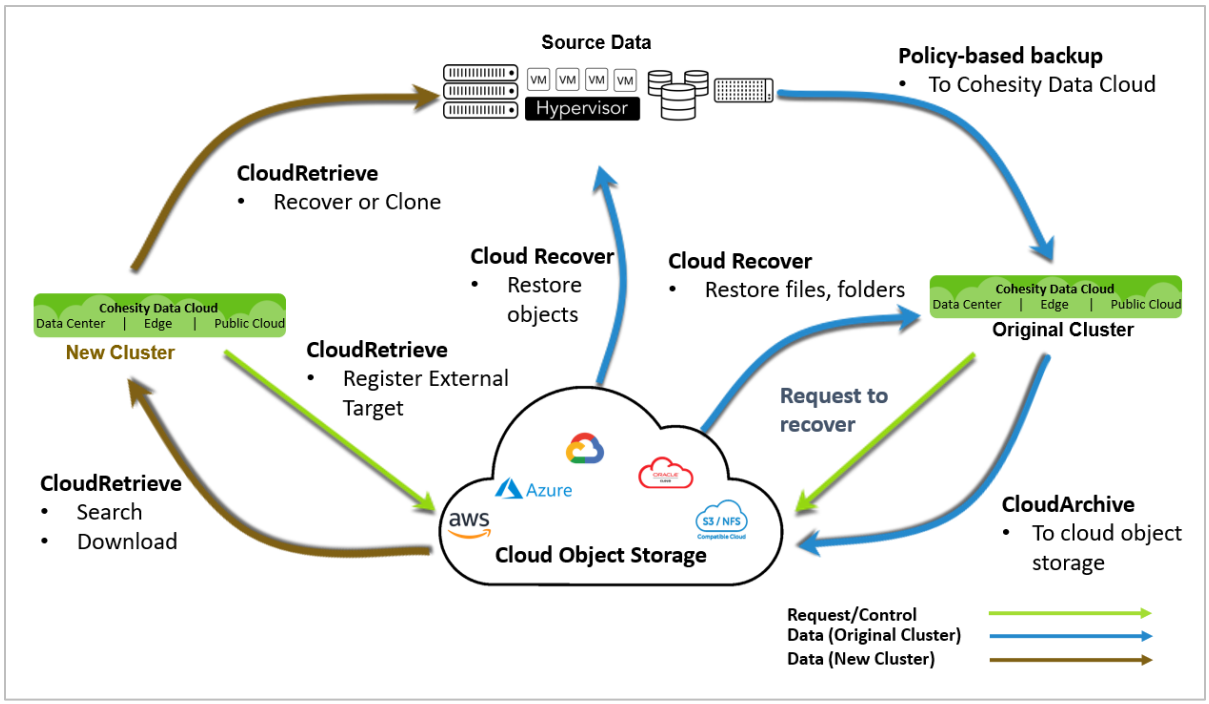
Table 5: The Difference Between Legal Hold and DataLock

PURPOSE	LEGAL HOLD	DATALOCK
Business need	Reactive: Set on a specific Snapshot (i.e., job run), usually prompted by legal requirements.	Planned: Set on all job runs that use a protection policy with DataLock, usually for compliance.
Expiration period	No expiration. Removal managed by the user.	Defined in the protection policy.
Granularity	Set on individual job runs and at the Object Level.	Applies to all job runs of any protection groups that use a policy with DataLock.
Deletion	Can be deleted to recover storage space, but only by a user with the Data Security role.	Cannot be deleted before the DataLock expiration date, even by a user with the Data Security role.

Recover Data from CloudArchive

Cohesity Data Cloud provides two ways to get your data back from cloud storage: Cloud Recover and CloudRetrieve.

Figure 8: Cohesity Protection, CloudArchive, Cloud Recover, and CloudRetrieve



- **Cloud Recover**: Recover entire objects (such as VMs, databases, NAS, etc.) or individual files and folders back onto the Cohesity Data Cloud that archived them.

  NOTE: When you recover a complete object (such as a VM or database), it is restored to its original location once it is downloaded to the Data Cloud from the cloud, and restored via the Instant Volume Mounting capability in Cohesity Data Cloud.

- **CloudRetrieve**: CloudRetrieve allows you to extract your protection group and its metadata, including job run details, from the archive in the cloud, so you can search it and recover the data you need onto a new or different cluster. This approach involves several steps:

  a. Register the External Target containing your archived data.

  b. Search the archive in the cloud.

  c. Select and download metadata for the archived protection groups.

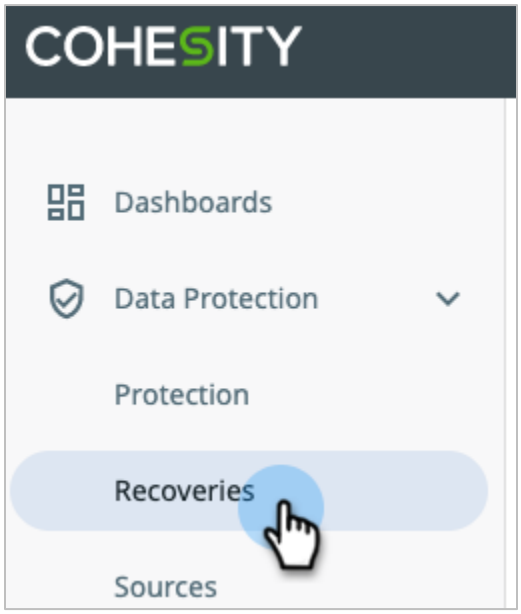
  d. Recover objects from the downloaded protection group job run.

But first, let's start with recovering data onto your original Data Cloud cluster.

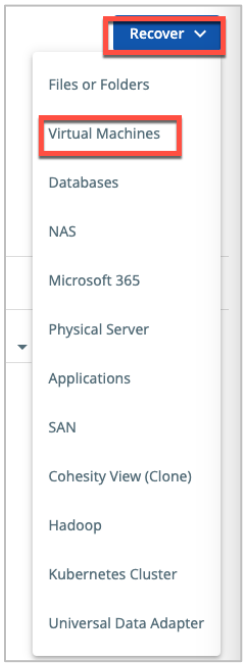
Recover Your Data to Original Cluster

To locate and recover a file, a folder, or an entire virtual machine to the original cluster:

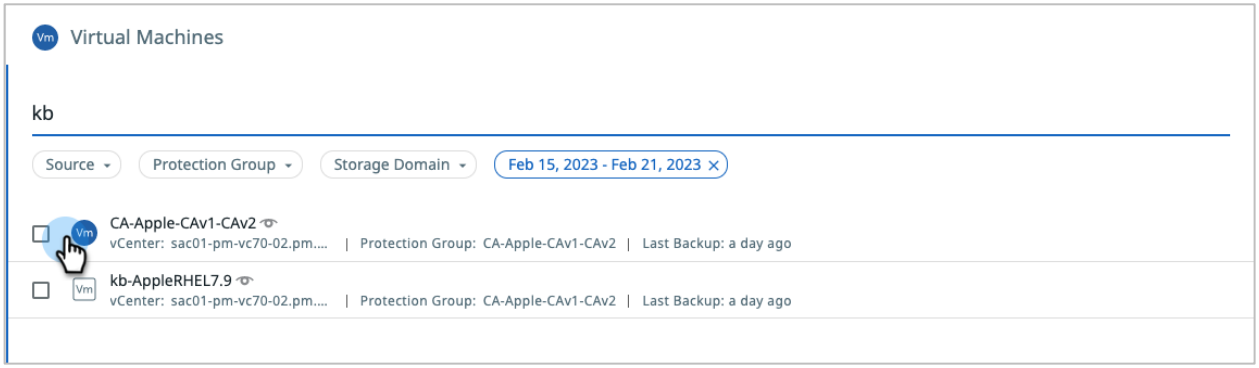
1. Log in to **Data Cloud**.
2. Select **Data Protection** > **Recoveries**.



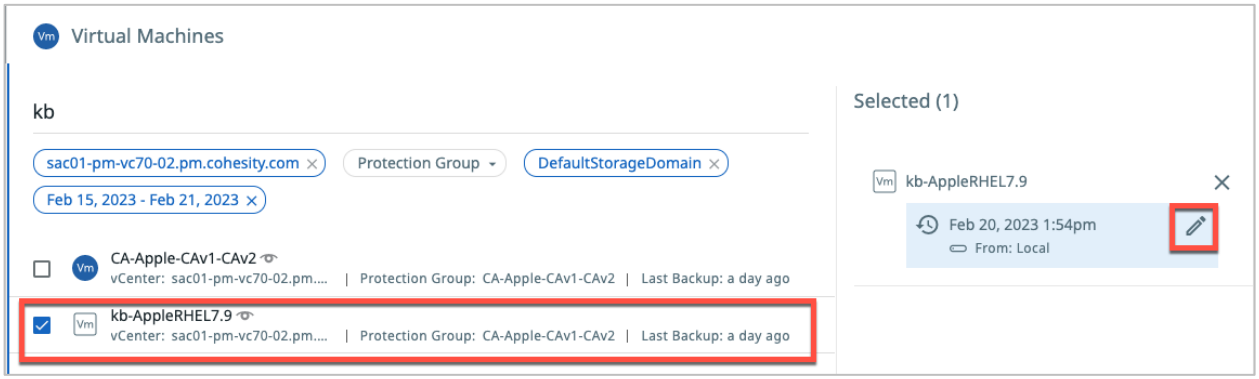
3. Click **Recover** and select the type of object you seek — a file or folder, VMs, physical server, and more.



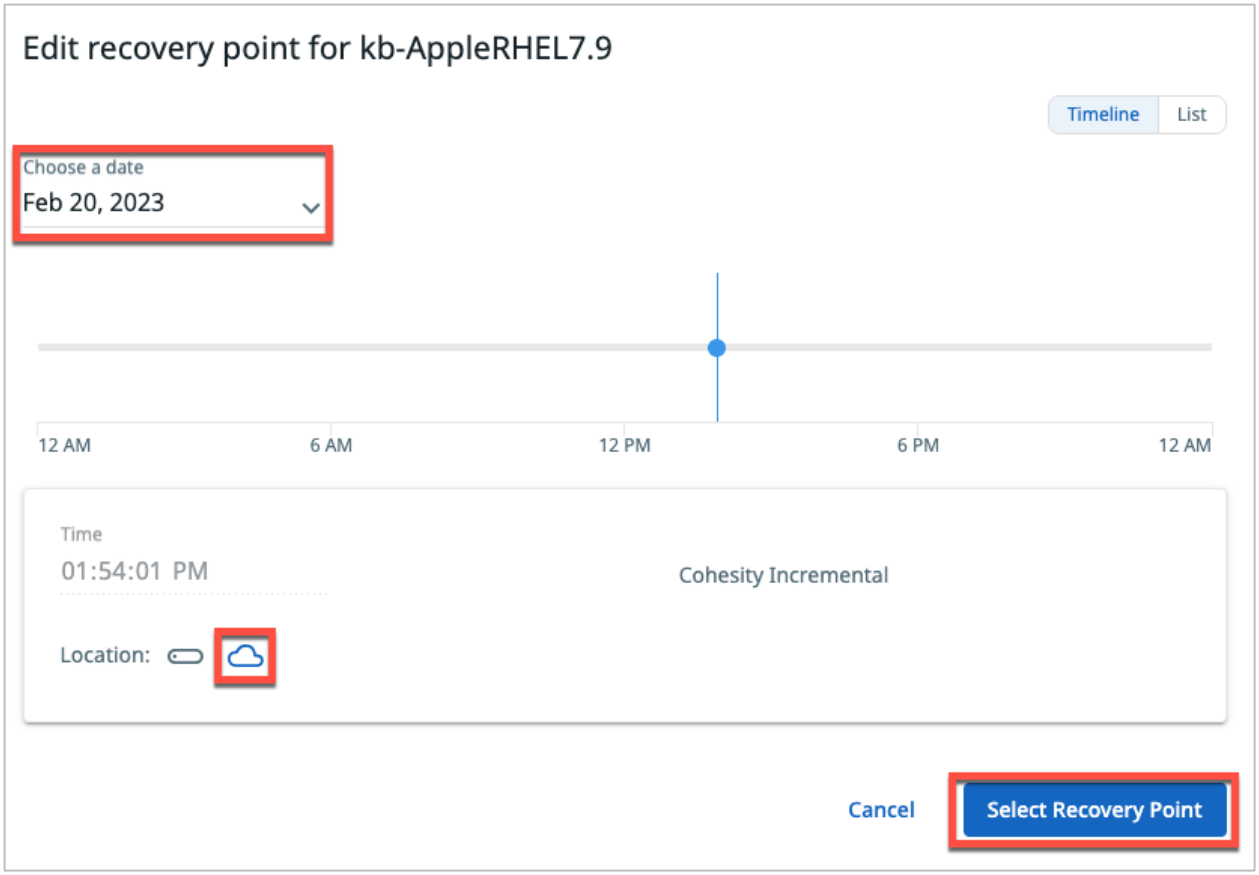
4. To retrieve a list of virtual machines, for example, select **VMs** and enter part or all of the VM names:



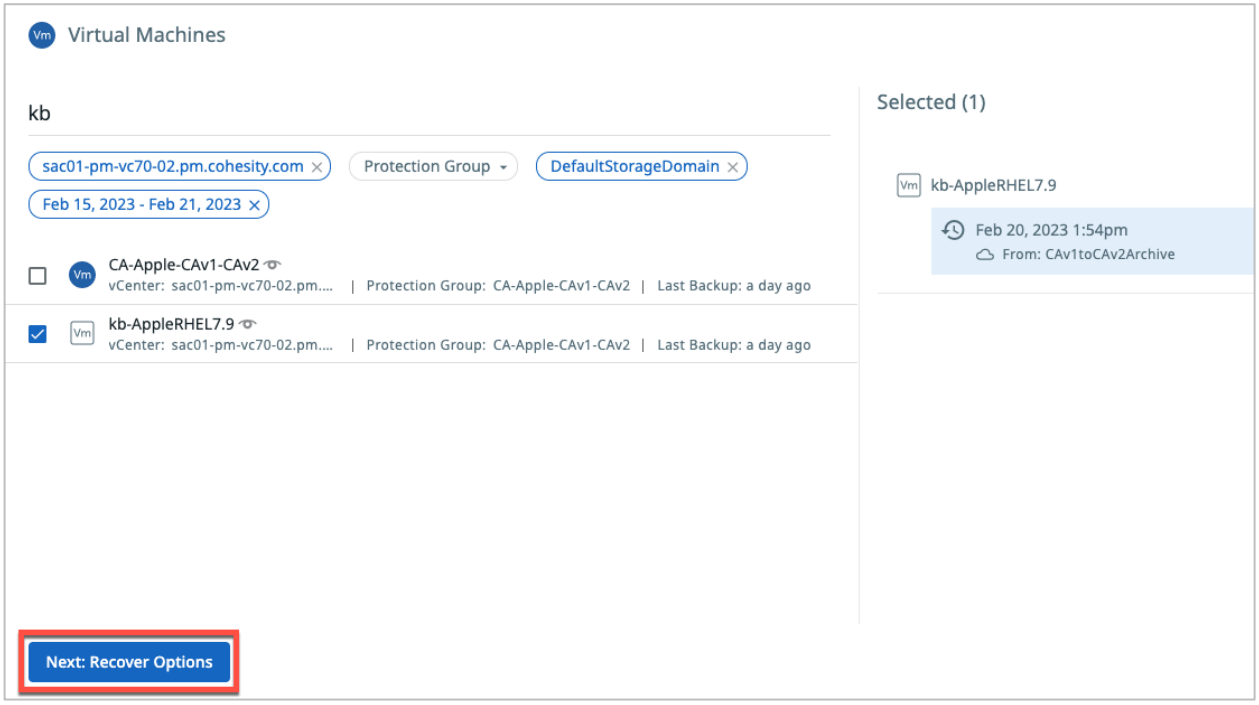
5. Select the VMs you need, or select an entire Protection Group to recover all the VMs it archived, and then click **Edit** to select a recovery point.**Next: Recover Options**.



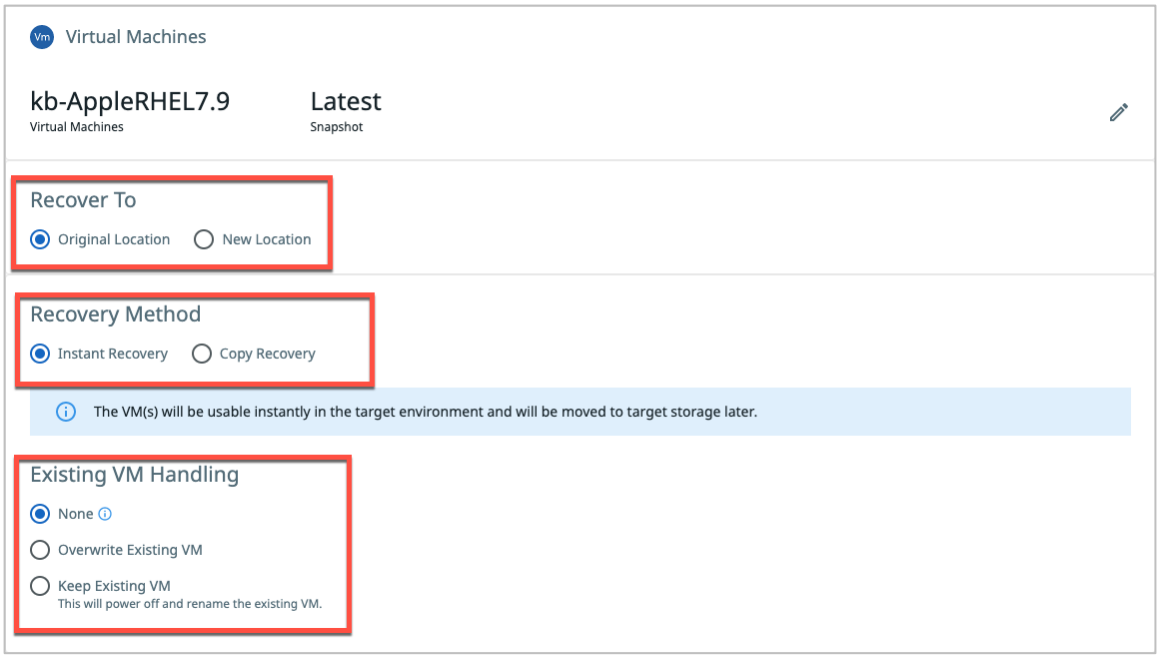
6. Choose a date. To recover from the external target, change the location from local to cloud. Click **Select Recovery Point**, **Location**, **Networking Options** and **Additional Options.**



7. Click **Next: Recover Options**.



8. Choose a **Recovery Location** and Recovery Method and then specify how to handle the existing VM.



9. In the Recovery Options, attach Network, **Rename** Recovered VMs with appropriate **Prefix** and **Suffix**. Select the **Power State** of the VM and also enter a **Task Name**. Click **Recover** to start the recovery process.

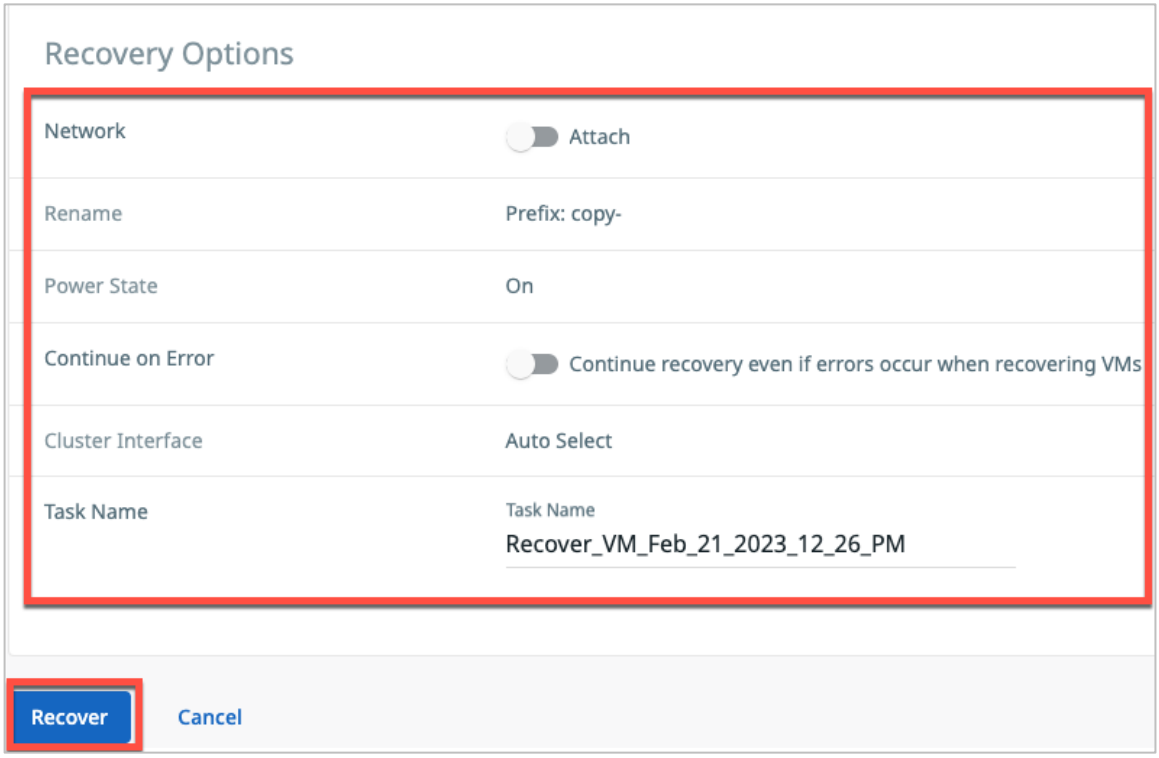


Table 6: Recover Task Options

RECOVERY OPTIONS	DETAILS
Recover to a new location	Specify this option to recover the VM files (such as the VMDK files) to their original datastores and create new instances of the VMs in the original location in the original source. For more, see Recover to Original Location in the online Help.
Keep original	For each recovered VM, keep the original virtual Network Interface Cards (vNICs) and attach them to the original network connections. **NOTE**: This option is only supported when VMs are recovered back to their original location.
Start Connected	For each recovered VM, connect to the original or new network when the VM reboots. **IMPORTANT**: If this option is not selected, the VMs are not connected to any network on reboot.
Detach network	For each recovered VM, the virtual Network Interface Card (vNIC) is removed from the VM.
Leave recovered VMs powered off	The recovered VMs remain powered off after they are created. **TIP**: Cohesity recommends this option if you are recovering from a storage domain that has CloudTier enabled.
Continue recovery even if errors occur when receiving VMs	With this option, if one of the VMs cannot be created, Cohesity Data Cloud will still attempt to create the other VMs.

NOTE: This example is for recovering a VM. The recovery options vary by protection group type.

10. Click **Finish** to start the recovery process.

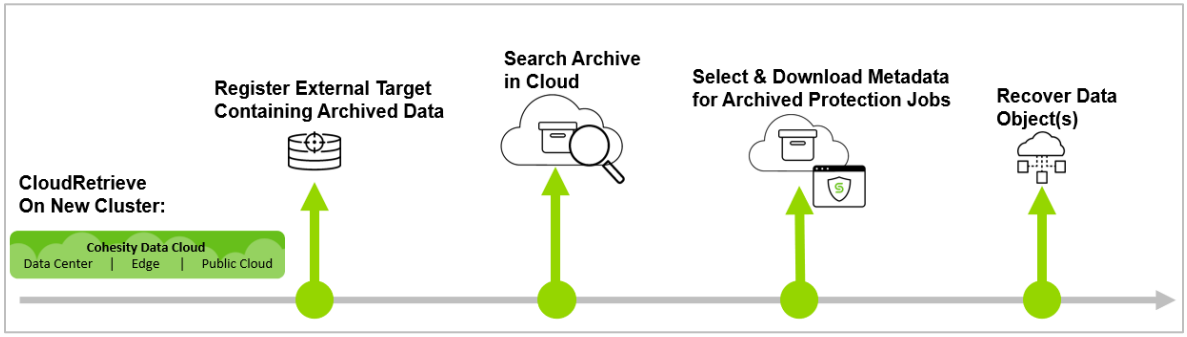
For more on the many capabilities and choices in our recovery process, see Recovery in the online Help.

CloudRetrieve Your Data to New Cluster

CloudRetrieve provides the ability to download data that was archived from a cluster to an alternate (non-original) cluster. In other words, you have Cluster A, which archives data to an External Target, but you need to download that archived data to Cluster B, for geo-redundancy or disaster recovery.

When you need to recover data from cloud storage to a different Cohesity cluster, there are several steps:

Figure 9: CloudRetrieve Workflow



The sections below describe the steps to:

1. Register the External Target containing your archived data to the new cluster.

2. Enter the retrieve parameters (cluster name, date range, protection group name) to search the archive in the cloud. (The search can take from minutes to several hours, depending the data-retrieval SLA for the class of storage you are using from your cloud provider. For example, some storage classes have a standard retrieval SLA of up to several hours.)

   NOTE: If your External Target is protected by a manually managed key, before you can search it, you will need to upload the External Target's access key.

3. From your search results, select and download the metadata (the job run details) for the archived protection groups onto the new cluster, so that you can review Job Run details and choose just the specific you need to recover or clone.

   NOTE: In this step, you are prompted to select a date range, and if you know exactly which job run (Snapshot) you need, you can also choose to download it along with the metadata, to be able to recover your data objects as soon as it completes.

4. After the metadata download completes, select the necessary job run from the archived protection group to recover or clone your objects.

Register External Target Containing Archived Data

To register your cloud object storage as an External Target on the new cluster:

1. Log in to a cluster other than the cluster that archived your data, or stand up a new cluster.

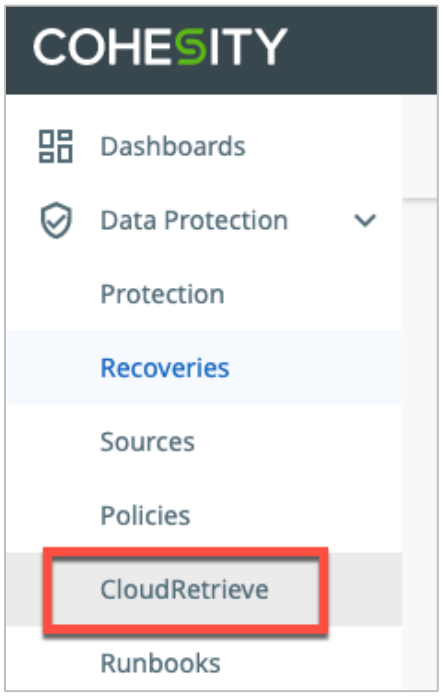
2. Log in to Data Cloud on your new cluster.

3. Follow the steps in Register GCP Bucket with Cohesity Data Cloud to register your archived cloud storage to the new cluster.

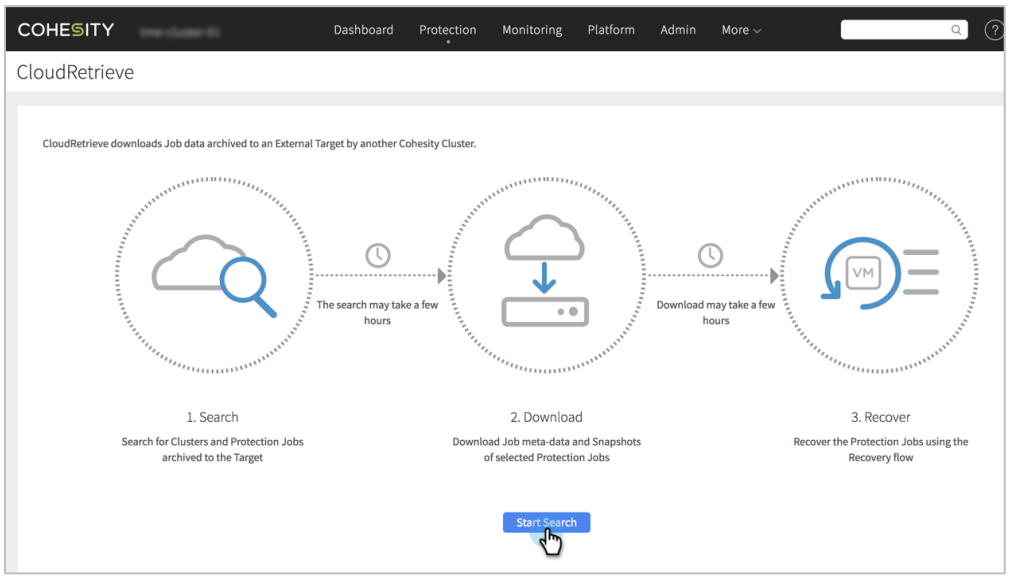
Search Archived Data in the Cloud

To submit a search request for a list of archived clusters and protection groups:

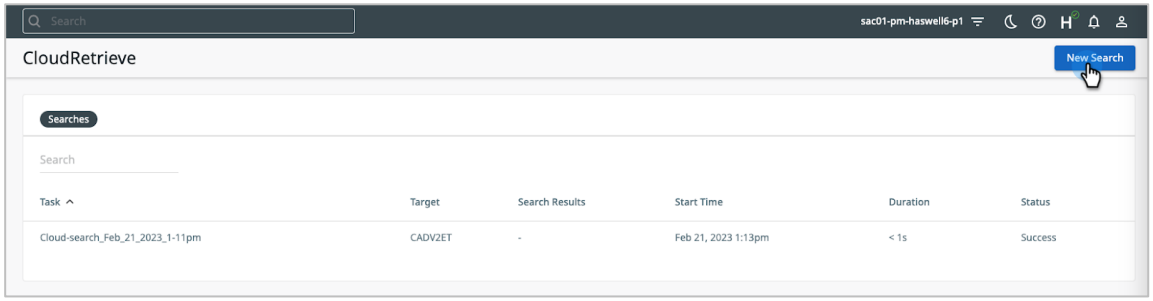
1. Log in to Cohesity Data Cloud on the new cluster.
2. Select **Data Protection > CloudRetrieve**.



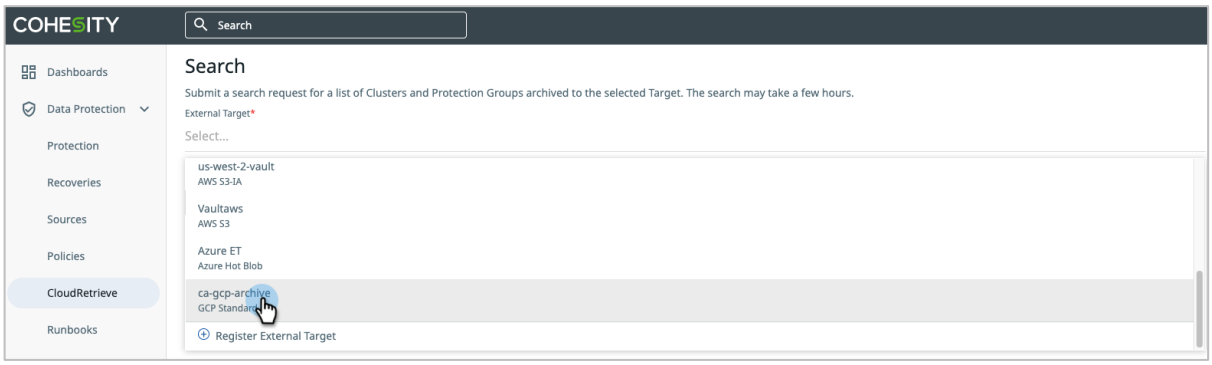
3. If this is the first time you have used CloudRetrieve, the **CloudRetrieve** summary screen appears. Click **Start Search**.



a. If this is not your first visit, the list of downloaded jobs appears. In that case, click **New Search**.



4. Select your **External Target** from the drop-down list.

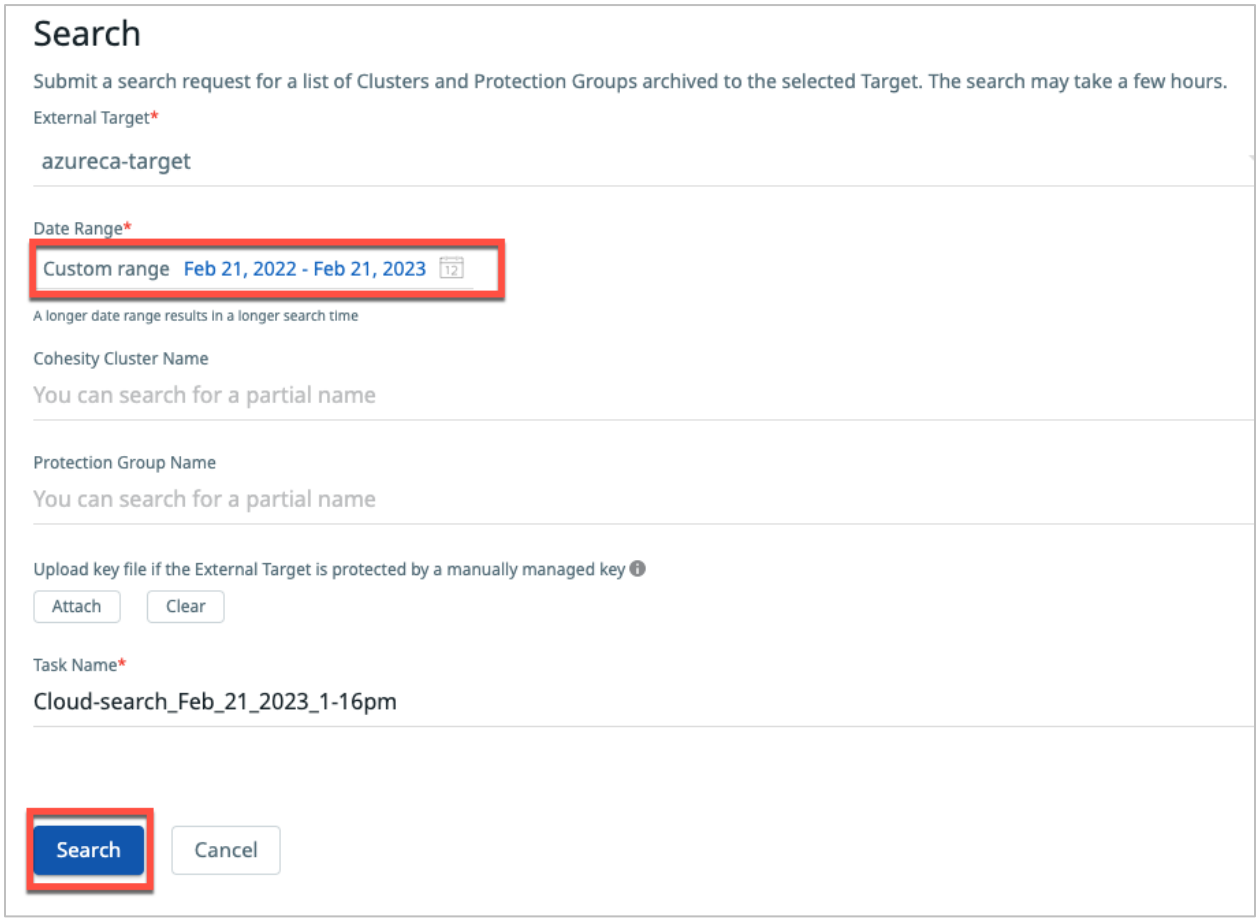


NOTE: If you skipped the first step and have not yet registered your External Target, you can register it here. To do so, click **Register External Target** from the drop-down menu and follow the steps in Register GCP Bucket with Cohesity Data Cloud.

5. In the form that opens, enter the required and optional fields based on your requirements and then click **Search**:

Table 7: CloudRetrieve Search Options

FIELD	DESCRIPTION	NOTES
Date Range (required)	Select a Date Range (past year by default) to limit the scope of your search.	
Cohesity Cluster Name (optional)	To narrow your search to a specific cluster, enter a cluster name. This is especially helpful if the same cloud storage is used with more than one cluster. To broaden your search to match more than one cluster, use a partial name (for example, 'Acme' instead of 'Acme_Raleigh').	**IMPORTANT**: Wildcard characters (like '*****') are NOT supported. If you enter search terms for both **Cluster Name** *and* **protection group Name**, your search must find matches for the protection group *within* clusters that match. If your search is too narrow, try entering a search term for just **Cluster Name** *or* **protection group Name**, or leave one or both empty.
Protection Group Name (optional)	To narrow your search to a specific protection group, enter a job name. This is especially helpful if the same cloud storage is used for more than one protection group. To broaden your search to match more than one protection group, use a partial name (for example, 'NAS' instead of 'NAS-Bronze').	
Upload key file (optional)	If your External Target is protected by a manually managed key, click **Attach**.	
Task Name (required)	By default, Cohesity Data Cloud uses the current timestamp to name the task automatically (for example, 'Cloud_search_***<CurrentTime>***'). Cohesity recommends you replace the automatic **Task Name** with terms that will make it easy to identify (for example, '***<ExternalTarget>***_From_***<SourceCluster>***_***<Purpose>***').	



6. Wait while the search runs.

> **NOTE**: The search can take from minutes to several hours, depending on the data-retrieval SLA for the class of storage you are using from your cloud provider. For example, some storage classes have a standard retrieval SLA of up to several hours.

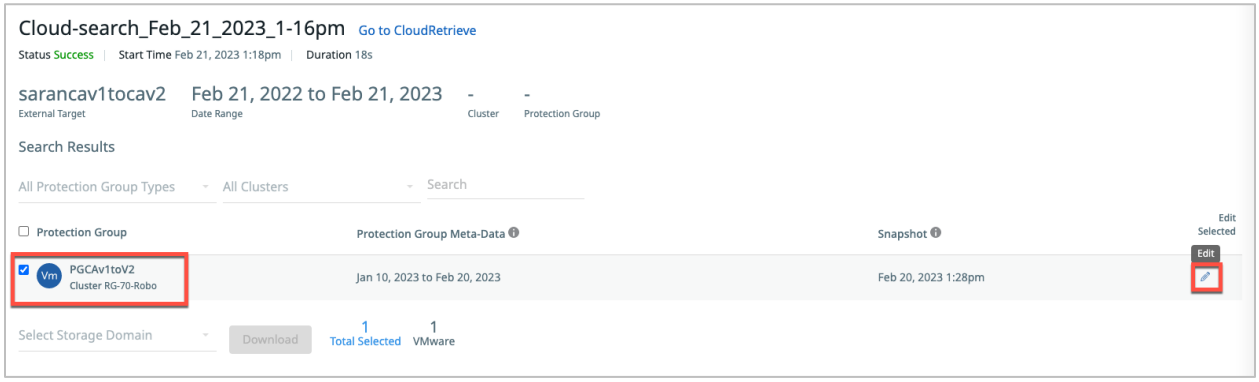
The success of a CloudRetrieve search does not guarantee that the search found any matches. It means only that the search operation completed successfully. If your search results came up empty, broaden your search with partial names for the cluster and/or Job, leave them blank, and/or extend the date range.

Select and Download Metadata for the Archived Protection Groups

Once you have your search results, choose the protection groups to download to your new cluster. After the download, you can recover your data from the downloaded archive. See Figure 8 above.

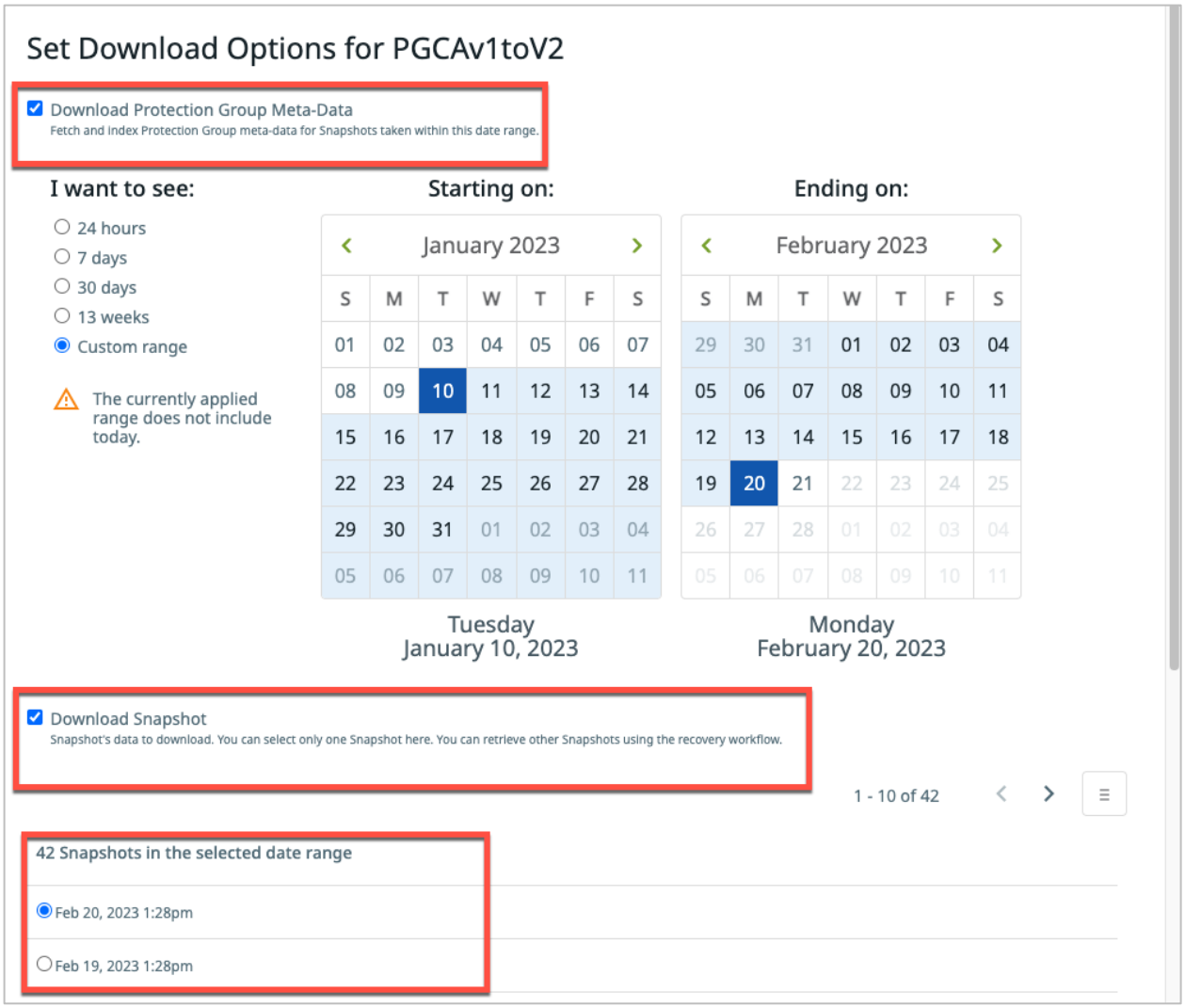
When your search completes:

1. Select the Protection Group(s) you wish to recover from the search results and click **Edit**.



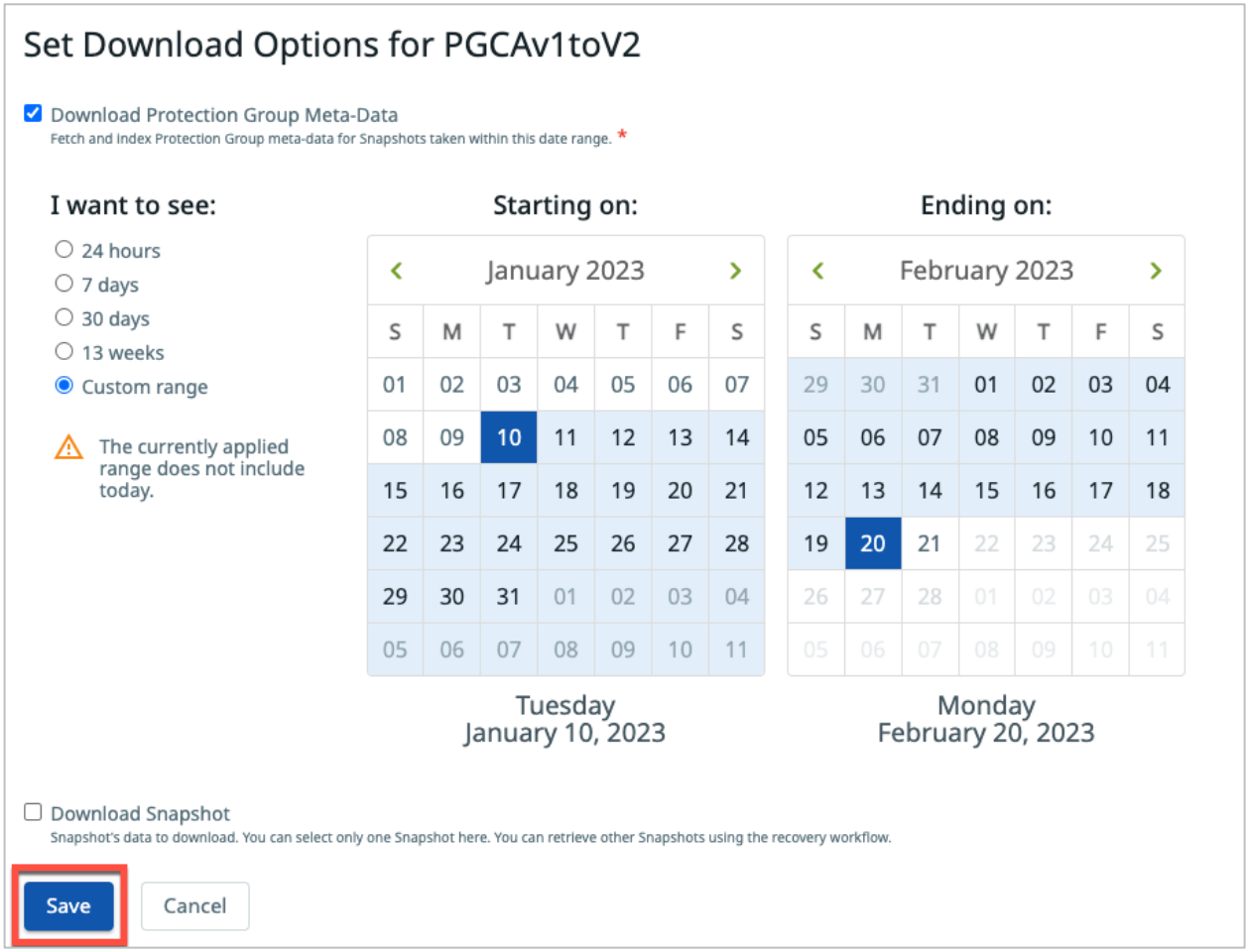
2. In the form that opens, you can choose to **Download Job Meta-Data** (that is, the details of each job run in the archived protection group), **Download Snapshot** (a specific job run), or both.

3. In the form that opens, you can choose to **Download Job Meta-Data** (that is, the details of each job run in the archived protection group), **Download Snapshot** (a specific job run), or both.

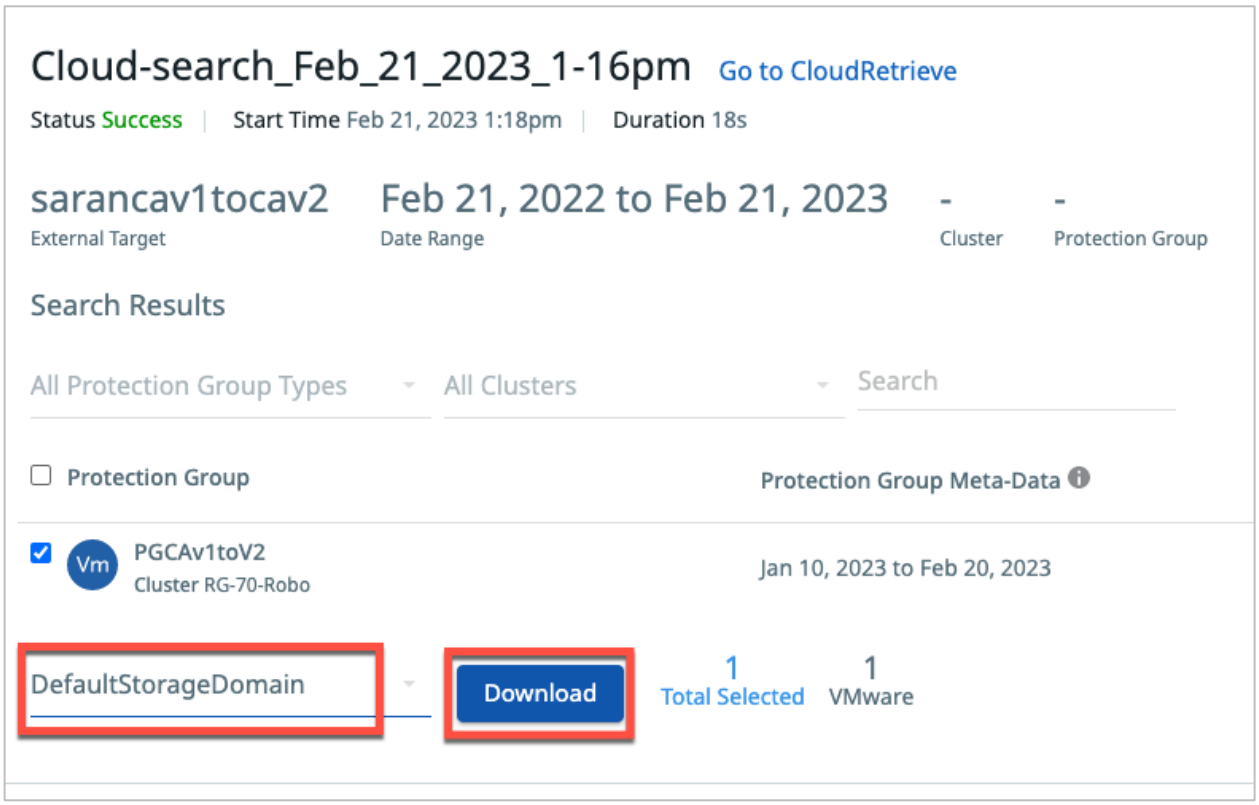


NOTE: If you are not certain which Snapshot contains the objects you need to restore, Cohesity recommends you deselect **Download Snapshot**. Once you have the job metadata, you will be able to review the details of each Snapshot in the Protection Group, to help you narrow the download to just the specific data you need.

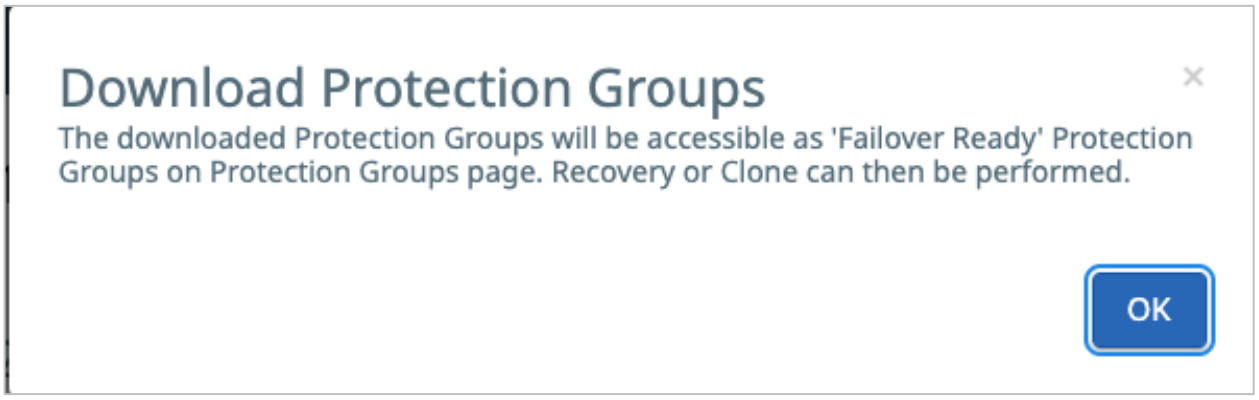
4. Make your choices and click **Save**.



5. Select the **Storage Domain** and click **Download**.

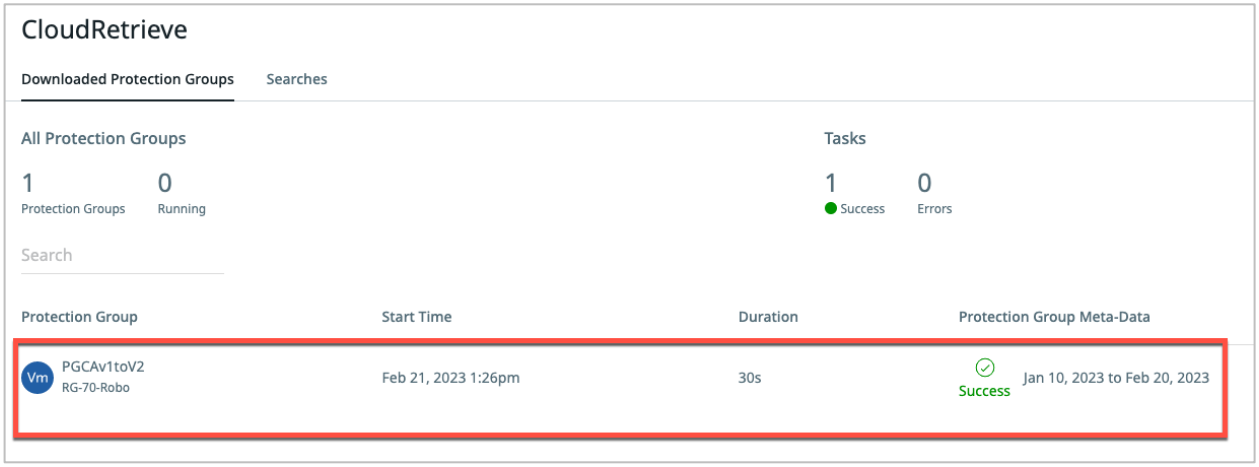


6. The downloaded Protection Group(s) will be accessible as **Failover Ready** under **Protection Groups**.



Wait for the download to complete.

7. Go to **Protection > CloudRetrieve** to monitor the progress of your download.



The Protection Group is now available on your new Cohesity cluster and can be used to recover your archived data.

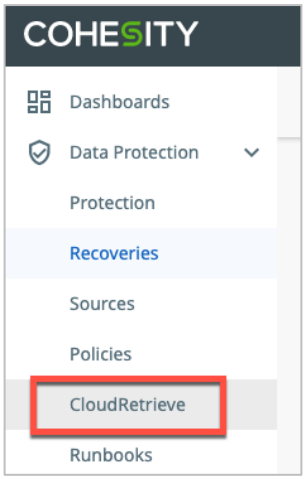
NOTE: CloudRetrieved Snapshots are not expired automatically by the new cluster. Once you have recovered the data you need, if you need to reduce your cloud storage expenses, you will have to delete the archived data from your cloud object storage manually. Do NOT do this if the original cluster is still intact

Recover Source Objects from Retrieved Archive on New Cluster

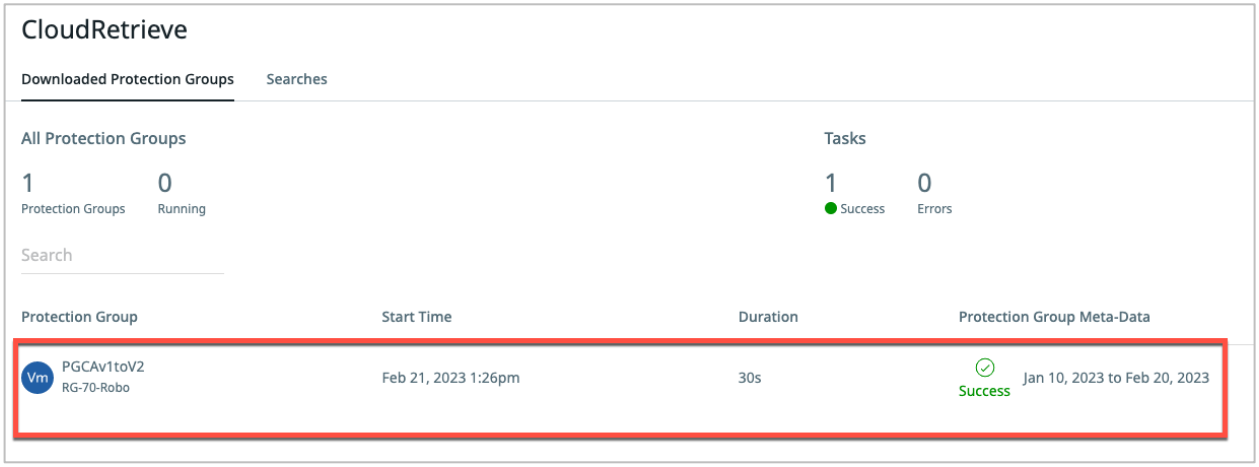
Now that you have downloaded the archived job runs metadata onto the new cluster, you can recover whole objects or individual files from the downloaded archive.

To recover an entire data object from a CloudRetrieved archive:

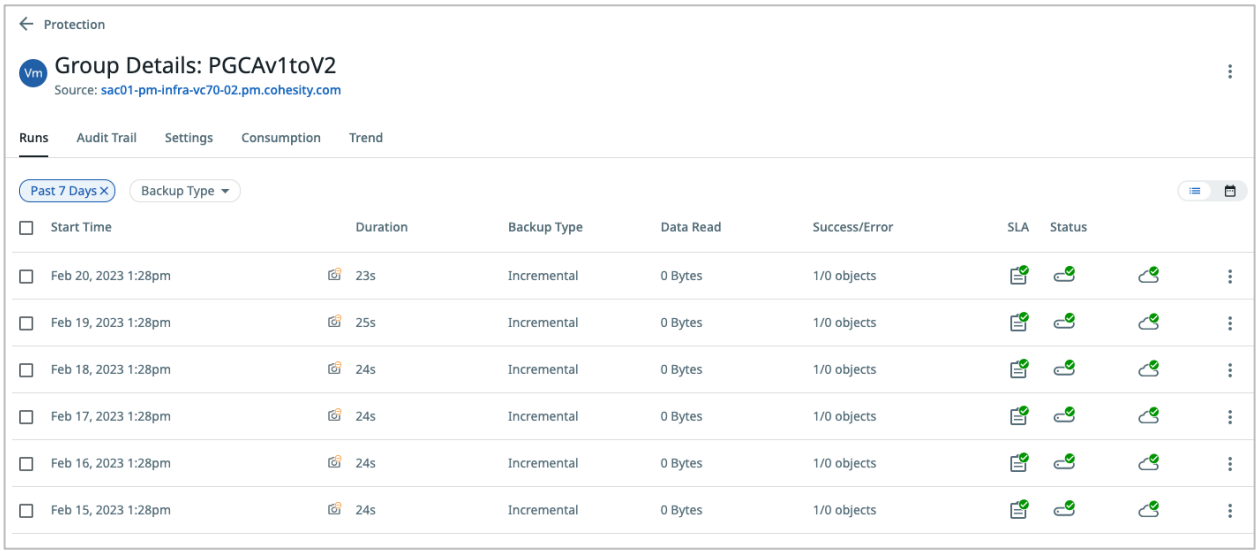
1. Log in to Data Cloud on the new cluster.
2. Select **Data Protection > CloudRetrieve**.



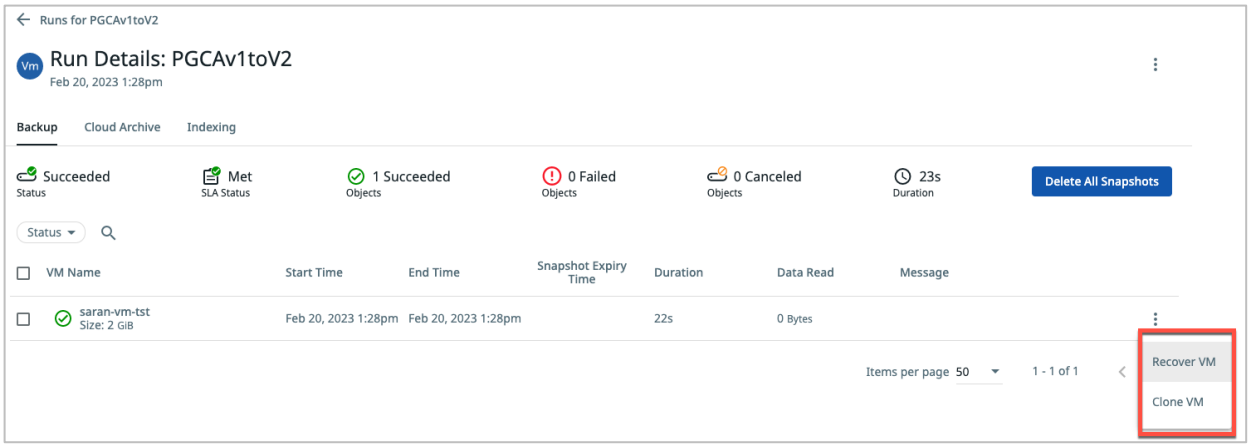
3. On the **Downloaded Protection Groups**. tab, find the Protection Group you retrieved and click it.



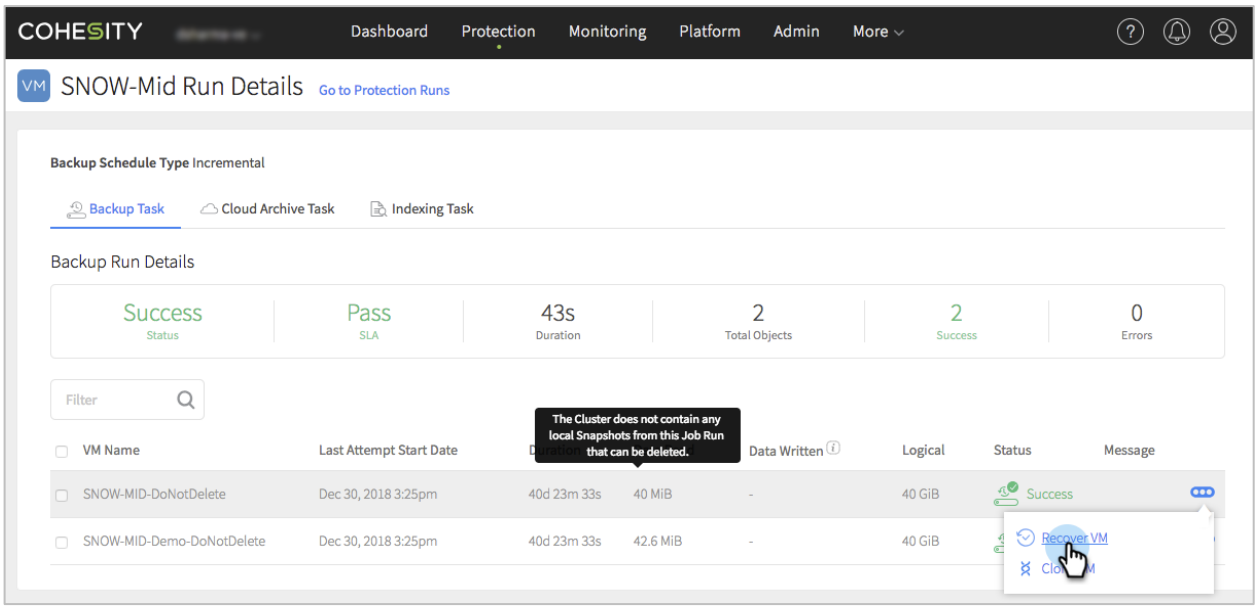
4. When the list of job runs in the retrieved archive appears, inspect the details for each Run (**SLA**, **Schedule Type**, **Logical**, **Data Read**, **Success/Error**, and **Run Status**) and click the most appropriate job run.



5. In the list of data objects included in that job run, find the object you need to recover (for example, a particular VM), hover over the Action menu on the right, and select **Recover VM** or **Clone VM**.



6. Edit the **Task Name** and **Recover** as fields, if necessary, and then follow the rest of the standard procedure for recovery above to complete your recovery task.



7. Edit the **Task Name** and **Recover** as fields, if necessary, and then follow the rest of the standard procedure for recovery above to complete your recovery task.

   See About CloudRetrieve in the online Help for more.

Appendix: Protection Group Advanced Settings

Protection groups combine operational requirements with the business requirements that are defined in a protection policy. See the all the advanced protection group settings, and the job types that include them, in Table 8:

Table 8: Protection Group Advanced Settings

FIELD	DESCRIPTION	APPLICABLE JOB TYPE
Pause Future Runs	Once enabled, no runs will be scheduled	All job types
End Date	Toggle on **End Date** and select the date on which the protection group stops capturing Snapshots. A job run that starts prior to this date will run until completion even if it completes after this date.	All job types
QoS Policy	Select **HDD** or **SSD**. **Backup HDD**: The Cohesity cluster writes the data directly to an HDD drive for this protection group. **Backup SSD**: The Cohesity cluster writes the data directly to an SSD drive for this protection group. Only specify this policy if you need fast ingest speed for a small number of protection groups. Cohesity recommends **HDD** (the default).	All job types
Pre & Post Scripts	Edit this option to run scripts on the protected server before and/or after a protection group runs. If configured, the scripts are run every time an object is backed up by a job run.	Physical Server, MS SQL, Oracle Database, NAS
Skip Files on Errors	Toggled on by default. The protection group continues to run even if it encounters errors on files, such as permissions errors. If files are skipped, the job run details page indicates a warning status and provides additional information. If toggled off, the protection group stops when it encounters an error.	NAS **NOTE**: This setting is always enabled automatically for file-based Physical Server backups.
Use Isilon Change List	Leverages the Isilon Changelist API to directly discover changed files/directories for faster incremental backup. Cohesity needs to keep one extra snapshot on Isilon after each backup, which will be deleted by the next successful backup.	Isilon

FIELD	DESCRIPTION	APPLICABLE JOB TYPE
File DataLock	Enable DataLock in Compliance or Enterprise mode.	
Exclusions and Inclusions	Everything is included by default. Toggle on Exclusions and Inclusions if you want to exclude or include locations. By creating exclusion and inclusion rules, you can limit the protection group to a specific set of files and directories and therefore minimize the disk space used to store the data. Cohesity automatically excludes the following NetApp system files: .vtoc_internal and .bplusvtoc_internal files **.copy-offload** directory and **.tokens** file **WARNING**: Always specify forward slashes (/) even for Windows systems. For Windows, do not specify the drive letter and colon in front of directory path.	Virtual Server, NAS, Microsft365
Indexing	Indexing is required for file recovery. The Cohesity Cluster will scan all the files in the protection group and create an internal index that can be used later by a Recover task to locate files by name. When creating a volume-based SQL job, indexing is not turned on automatically. Cohesity recommends turning indexing on because indexing is required to restore .mdf, .ldf and .ndf files from the cloud.	Virtual Server, Physical Server, MS SQL, MicrostfOffice 365, NAS
Cancel Runs at Quiet Time Start	Cancel in-progress Protection Runs at the start of quiet times (as defined in the associated Protection Policy).	All job types
Alerts (optional)	Select one or more of the following settings if you want Alerts to be created for the following triggers: **Success**: Create an Informational Alert when a protection group completes successfully. Emails are not sent when Informational Alerts are created. **Failure**: Create a Critical Alert if the protection group fails to complete. Emails are sent when Critical Alerts are created. **SLA Violation**: Create a Warning Alert if the protection group takes longer than the time period specified in the SLA field. Emails are sent when Warning Alerts are created.	All job types

FIELD	DESCRIPTION	APPLICABLE JOB TYPE
Priority	Select a priority for the protection group execution. Cohesity supports concurrent backups, but if the number of jobs exceeds the ability to process jobs, they are executed in priority order: **High** first, then **Medium**, and then **Low**.	All job types
SLA	The Service-Level Agreement (SLA) defines how long the administrator expects a job run to take. **Incremental**: Enter the number of minutes you expect an incremental backup job run to complete. An incremental backup captures only the differences (changed blocks) since the last job run. **Full**: Enter the number of minutes you expect a full backup job run to complete. A full backup captures the entire object (all blocks).	All job types
Description	Specify a description for the protection group.	All job types

Use these settings when you are setting up your protection group.

Your Feedback

Was this document helpful? Send us your feedback!

About the Authors

Saran Ravi is a Staff Technical Solutions Engineer at Cohesity. In his role, Saran focuses on Cloud and Kubernetes.

Other essential contributors include:

- Adaikappan Arumugam, Director, Product Solutions
- Subash Babu, Staff Technology Editor
- Sai Krishna Mukundan, Director, Product Management
- Dayanand Sharma, Product Manager
- Radhani Guturi, Cloud Engineering Director
- Praveen Yarlagadda, CloudArchive Lead Engineer
- Kevin Hill, Cloud Solutions Architect

Document Version History

VERSION	DATE	DOCUMENT HISTORY
2.1	July 2024	Republishing
2.0	Mar 2023	Updated to Cohesity version 7.0
1.0	Feb 2019	First full release
0.3	Dec 2018	Changes based on feedback
0.2	Nov 2018	First full draft
0.1	Oct 2018	Original document

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.