

# COHESITY

Version 2.0

September 2020

# Archive Your Data Directly with Cohesity CloudArchive Direct

*Cohesity's Cost-effective CloudArchive Direct for NAS*

## ABSTRACT

*With the exponential growth of data, organizations need better, more cost-effective ways to archive their petabytes of data. Cohesity's CloudArchive Direct solution provides an efficient and innovative approach by streaming your data directly into lower-cost cloud storage without having to store a copy locally. What's more, the solution allows you to search and restore seamlessly while reducing TCO by storing only the metadata and index on the local cluster.*

# Table of Contents

- 1 Introduction to CloudArchive Direct ..... 3
- 2 Explore Cohesity’s Long-term Retention Solutions ..... 4
  - 2.1 Cohesity CloudArchive ..... 4
  - 2.2 Cohesity CloudArchive Direct..... 5
    - 2.2.1 Features and Benefits of CloudArchive Direct..... 6
  - 2.3 Compare CloudArchive with CloudArchive Direct..... 7
- 3 Use CloudArchive Direct for Data Migration & Analysis..... 9
  - 3.1 Set Up CloudArchive Direct..... 10
- 4 Recover Data with CloudArchive Direct ..... 11
- 5 Disaster Recovery with CloudArchive Direct ..... 12
- 6 Prerequisites ..... 13
- 7 Your Feedback..... 14
- 8 About the Authors..... 14
- 9 Document Version History ..... 14

# Figures

- Figure 1: Cohesity CloudArchive Archives Your Local Backups on External Targets ..... 4
- Figure 2: Cohesity CloudArchive Direct Archives Your NAS Data Directly to an External Target ... 5
- Figure 3: NAS Cloud Recover Decision Tree ..... 11
- Figure 4: Cloud Recover to Original Source & CloudRetrieve to New Cluster ..... 12

# Tables

- Table 1: Key Features and Benefits of CloudArchive Direct ..... 6
- Table 2: Features and Benefits Inherited from Cohesity DataPlatform..... 6
- Table 3: Compare CloudArchive Direct with CloudArchive ..... 7
- Table 4: Compare CloudArchive Direct’s Native and Non-native Formats ..... 9

# 1 Introduction to CloudArchive Direct

In today's technology-driven world, it is constantly becoming easier for organizations of all kinds to find themselves collecting, and having to store and protect, growing amounts of data — from sensitive HR data and healthcare records to large media and entertainment files. As many of those organizations store their growing data collections on network-attached storage (NAS), it has never been more important for them to focus on NAS data protection — how they store, protect, and retain all that data.

CloudArchive Direct is a policy-driven archival feature in Cohesity DataPlatform that was built specifically to address these challenges by streaming data directly to lower-cost storage on an External Target without storing local backups. And while Cohesity does not store the data, it indexes it and stores the metadata locally for fast search & recovery and offers options to compress and encrypt the data.

With these capabilities, Cohesity CloudArchive Direct solves many business needs, including:

- **Long-term data retention.** Most organizations are required to keep data for at least seven years, and many choose to keep it longer, accruing growing costs. With CloudArchive Direct, you can copy large datasets directly to cheaper storage easily. For example, hospitals often need to retain petabytes of patient records (EHR, PACS, etc.) for decades to meet regulatory requirements. Cohesity CloudArchive Direct can help you to reliably archive this dataset securely to the cloud.
- **Data migration.** As your data management infrastructure continually evolves, it is important to be able to migrate large datasets from one data center or silo to another. With CloudArchive Direct, you can copy your data more efficiently without consuming local storage and resources. For example, if a content company generates 1PB of data every week and requires these large datasets to be archived in the cloud once the content is published/broadcasted, they can use CloudArchive Direct to stream the data directly from their primary NAS storage to the cloud without saving it on the Cohesity cluster.
- **Data analysis.** As you accrue more and more data, it becomes difficult to track and analyze. CloudArchive Direct allows you to store the data in its native format in the External Target, enabling third-party applications to analyze the data directly from the External Target.

## 2 Explore Cohesity's Long-term Retention Solutions

Cohesity offers two similar (but importantly different) cloud-based solutions that help customers save space and cost by archiving their data to a lower-cost storage, typically in the cloud:

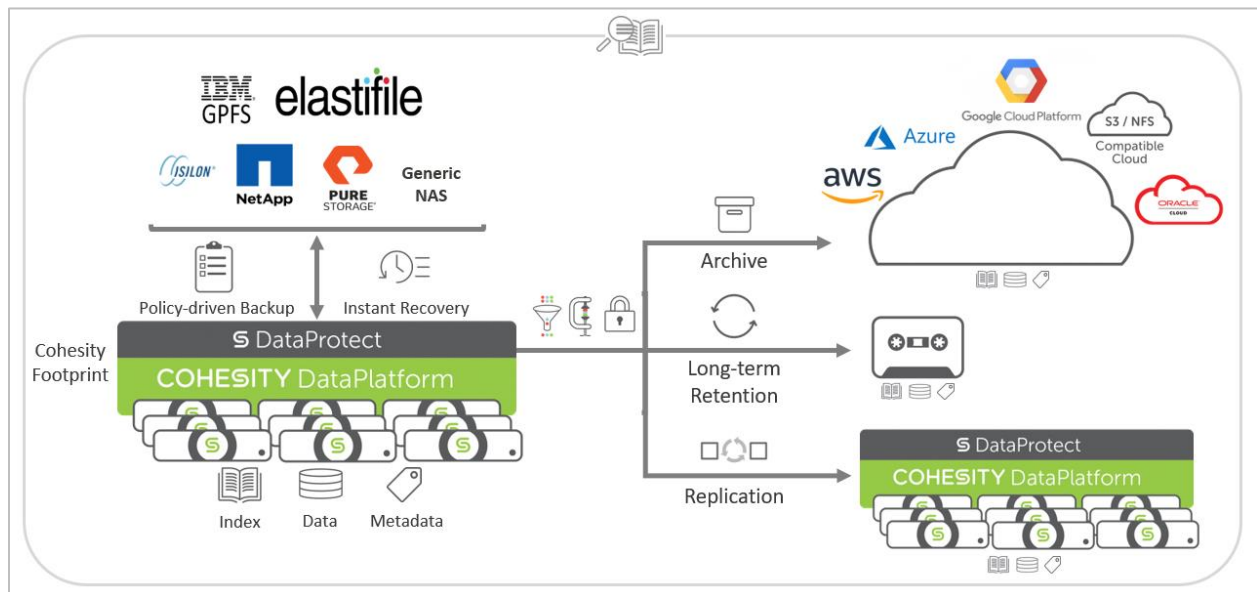
- [CloudArchive](#). Back up your data on a Cohesity cluster and copy it to the cloud for archival.
- [CloudArchive Direct](#). Stream your data directly to a storage target without creating local backups first.

### 2.1 Cohesity CloudArchive

Cohesity CloudArchive first backs up your data onto a Cohesity cluster and then copies your backups to an External Target — object storage from cloud vendors (AWS, Azure, GCP, and Oracle), S3-compatible storage, or NFS-mounted storage.

You can use CloudArchive if you have aggressive Recovery Time Objectives (RTOs) for data recovery, as CloudArchive offers nearly instant recovery of your data, by either recovering a complete storage volume to a Cohesity View or recovering selected files and folders.

Figure 1: Cohesity CloudArchive Archives Your Local Backups on External Targets



CloudArchive is very flexible. You can use it with [AWS](#), [Azure](#), [GCP](#), [NAS](#), and [S3-Compatible](#) cloud object storage.

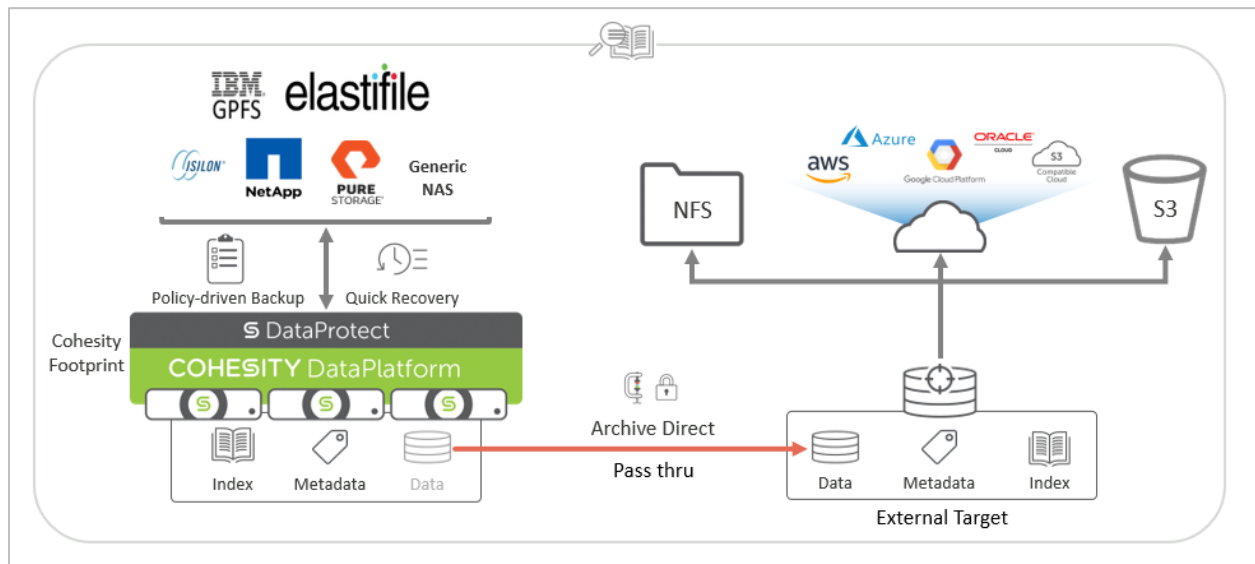
## 2.2 Cohesity CloudArchive Direct

For customers who need or prefer to archive their data directly without storing a copy on the cluster, Cohesity built the CloudArchive Direct solution that allows you to archive your NAS data directly from any NAS to any registered External Target. In this solution, the Cohesity cluster acts as a local cache to stream the dataset to the External Target, eliminating the need to store a full backup copy on the Cohesity cluster first. This approach dramatically reduces the capacity requirements on the local Cohesity cluster to a small cluster footprint, as only the metadata and index, and not the dataset itself, are stored on the local cluster. Storing the metadata and index on the local cluster enables much quicker search and recovery down the road. The External Target also stores the metadata and index, in addition to the data itself. On Cohesity DataPlatform, you can register any cloud storage (AWS, Azure, GCP, or Oracle), S3-compatible storage, or NFS-mounted servers as External Targets.

**NOTE:** For the list of all supported External Target types, see [Supported External Targets](#) in the online Help.

Depending on your specific business needs, you can choose to use CloudArchive Direct with the native NAS or a non-native format, that is, Cohesity's proprietary SpanFS™. Using the native format allows third-party tools to perform data analysis on your NAS data on the External Target, while using the non-native format allows Cohesity to compress and encrypt the data for dramatic savings in space and network traffic but does not support data access by third-party tools.

Figure 2: Cohesity CloudArchive Direct Archives Your NAS Data Directly to an External Target



### 2.2.1 Features and Benefits of CloudArchive Direct

CloudArchive Direct allows customers to manage, protect, and store their massive NAS datasets directly in the cloud, and offers a wide array of features, including storage efficiency, single-file recovery across billions of files, and more:

Table 1: Key Features and Benefits of CloudArchive Direct

CLLOUDARCHIVE DIRECT FEATURE	BENEFIT
<b>Low on-premises footprint</b>	Stores only the metadata and index on-premises, requiring only a minimal Cohesity cluster footprint.
<b>Flexible</b>	Archives your data in its native or non-native format, providing the flexibility you need to address your specific business needs, be they focused on space savings or data analysis.
<b>Metadata</b>	Available on-premises and in the cloud. And because Cohesity DataPlatform stores the metadata in the cloud, if your Cohesity cluster becomes unavailable in a disaster or other event, you will be able to recover it using any Cohesity cluster with Cohesity’s <a href="#">CloudRetrieve</a> (onto a new cluster) feature.
<b>Access to archived data</b>	You can use third-party applications to access your data directly from the External Target in its original format for data analytics. ( <a href="#">Native format only</a> )

There are also many benefits that CloudArchive Direct inherits from Cohesity DataPlatform, summarized in Table 2 below.

Table 2: Features and Benefits Inherited from Cohesity DataPlatform

COHESITY DATAPLATFORM FEATURE	BENEFIT
<b>Automated, policy-driven workflow</b>	Define simple and flexible policies to archive your NAS data directly to the cloud.
<b>Distributed and parallel upload</b>	Cohesity DataPlatform’s intelligent data-transfer logic creates an efficient archive plan and assigns archive streams across all nodes in a Cohesity cluster, performing distributed and parallel upload to the External Target, ensuring faster archives.
<b>Incremental indexing</b>	CloudArchive Direct indexes only the changed data between the last archive and the most recent archive, resulting in faster indexing and reduced impact on resources.

COHESITY DATAPLATFORM FEATURE	BENEFIT
<b>Faster and granular restores</b>	With file metadata stored and indexed on-premises and in the cloud, CloudArchive Direct keeps track of the archived data. This enables you to quickly search for a specific file from among hundreds of billions files in the cloud.
<b>Storage efficiency on External Target with Compression</b>	Compresses data to reduce NAS workload by 2-5x before sending it to the External Target, requiring much less cloud storage to store the entire NAS dataset. ( <i>Non-native format only</i> )
<b>Data security with encryption</b>	Encrypts the data while archiving in non-native format before sending the data to the cloud. This provides security for the data both at rest and in flight, using the industry standard 256-bit Advanced Encryption Standard (AES) algorithm. ( <i>Non-native format only</i> ).
<b>Mobility</b>	Recover to any other Cohesity cluster using <a href="#">CloudRetrieve</a> .

## 2.3 Compare CloudArchive with CloudArchive Direct

While the two solutions are very similar, each offers unique benefits, depending on your current operational infrastructure and future business needs. Table 3 below provides a side-by-side comparison of the benefits and trade-offs of each solution.

Table 3: Compare CloudArchive Direct with CloudArchive

FEATURE	CLOUDARCHIVE DIRECT	CLOUDARCHIVE
<b>Business needs addressed</b>	Reduced TCO	Local backups
	<ul style="list-style-type: none"> <li>Long-term data retention</li> <li>Security</li> <li>Compliance</li> </ul>	
<b>Local backup retention</b>	No	Yes
<b>Cluster space requirements</b>	Requires a fraction of the footprint of the data being archived.	Requires enough space for at least one copy of the backup.
<b>Indexed</b>		Yes
<b>Metadata</b>		Yes

FEATURE	CLOUDARCHIVE DIRECT	CLOUDARCHIVE
Data natively accessible in cloud	Yes	No; requires a Cohesity cluster.
Encryption	Yes	
Deduplication	No	Yes
Compression	Yes	
Incremental forever	Yes	Incremental with periodic full.
Replication	Cannot be replicated.	Yes, as there is a local backup copy.
Tape out	No	Yes
Number of archival External Targets supported per Protection Group	One	Multiple
Number of archival schedules supported per Protection Policy	Multiple (For the same target)	Multiple
Extended Retention (That is, daily backup retained for 30 days, weekly for 90 days, and monthly for 365 days)	Yes	Yes
Total copies (including the primary dataset)	Two (One primary and one in the External Target)	Unlimited. You can add as many backup copies, replication copies, archive copies, and copies on tape as you require.
Object Lifecycle Management	Not supported	Supports the cloud provider's native object lifecycle management.
Instant restore	No	Yes, to a Cohesity View
Clone	No	Yes
File and folder restore	Yes	



### 3 Use CloudArchive Direct for Data Migration & Analysis

You can archive your data directly to an External Target by enabling CloudArchive Direct in the Protection Group that contains your NAS source. When you do, Cohesity offers you the choice whether to do so in native or non-native format, each with its own benefits:

- **Native format (same as source).** Allows you to run analytical tools on the archived data but does not support compression or encryption.
- **Non-native format (Cohesity SpanFS).** Migrates the data in Cohesity’s web-scale file system, SpanFS, which allows you to take advantage of Cohesity’s compression and encryption features (enabled by default) as it streams to the External Target. However, as the data is in the Cohesity format, it cannot be analyzed by third-party tools.

Table 4: Compare CloudArchive Direct’s Native and Non-native Formats

COHESITY SETTINGS	NATIVE FORMAT	NON-NATIVE FORMAT
Protection Group: <b>Retain the original format</b> enabled	Enabled by default. It archives the data to the External Target in its native (source) format.	To use the non-native format, you must disable <b>Retain the original format</b> in the Protection Group.
External Target: <b>Encryption and Compression</b> enabled	Even if the External Target is configured for encryption and/or compression, these are ignored for data stored in its native format and Cohesity adds no at-rest encryption to the archived data.  However, Cohesity still applies in-flight encryption to the data as it streams it to the External Target for storage.	Cohesity encrypts and compresses the data as it streams it to the External Target, minimizing network bandwidth consumption as well as storage consumption on the External Target.  From data transfer to data storage, Cohesity ensures your data is secure with both in-flight and at-rest encryption.
External Target: <b>Source Side Deduplication</b> enabled	Ignored.	

COHESITY SETTINGS	NATIVE FORMAT	NON-NATIVE FORMAT
External Target: <b>Incremental Archival</b> enabled	Enabled by default. Even when this setting is disabled, however, it is ignored, because, for CloudArchive Direct, Cohesity DataPlatform always uploads the delta stream (the 'streaming difference') to the External Target, following an incremental forever method.	

In short, choose:

- **Native format** if you have business requirements to run data analytics tools on the data.
- **Non-native format** if you need to archive petabytes of sensitive data to an External Target, or if you have limited network bandwidth, limited cloud space, or elevated security requirements.

**NOTE:** When using the native format, due to limitations with External Target support for special characters, files are renamed to Cohesity's format while saving the data to the External Target, but the files remain readable in the original native format.

### 3.1 Set Up CloudArchive Direct

To enable CloudArchive Direct:

1. [Register or edit your NAS](#) as a source in Cohesity DataPlatform.
2. [Register an External Target](#) where you will archive your NAS data.  
When you register an External Target for CloudArchive Direct:
  - You can enable or disable the **Incremental Archive** option. However, CloudArchive Direct bypasses this setting and proceeds to perform incremental archives of the data.
  - Source-side deduplication is enabled by default, but CloudArchive Direct ignores this setting.
3. [Create a Protection Policy with archival scheduled after every run](#).
4. [Add or edit a Protection Group](#) to protect your NAS volumes.

**NOTE:** For more, see [NAS CloudArchive Direct](#) in the online Help.

## 4 Recover Data with CloudArchive Direct

Cohesity DataPlatform indexes and stores metadata on the cluster for improved search and accessibility of the files and folders that are archived in the External Target, thus enhancing the speed and simplicity of data recovery. You can search for individual files and folders, or an entire NAS volume, with their respective names, or you can browse the archived data index to find, select, and perform data recovery.

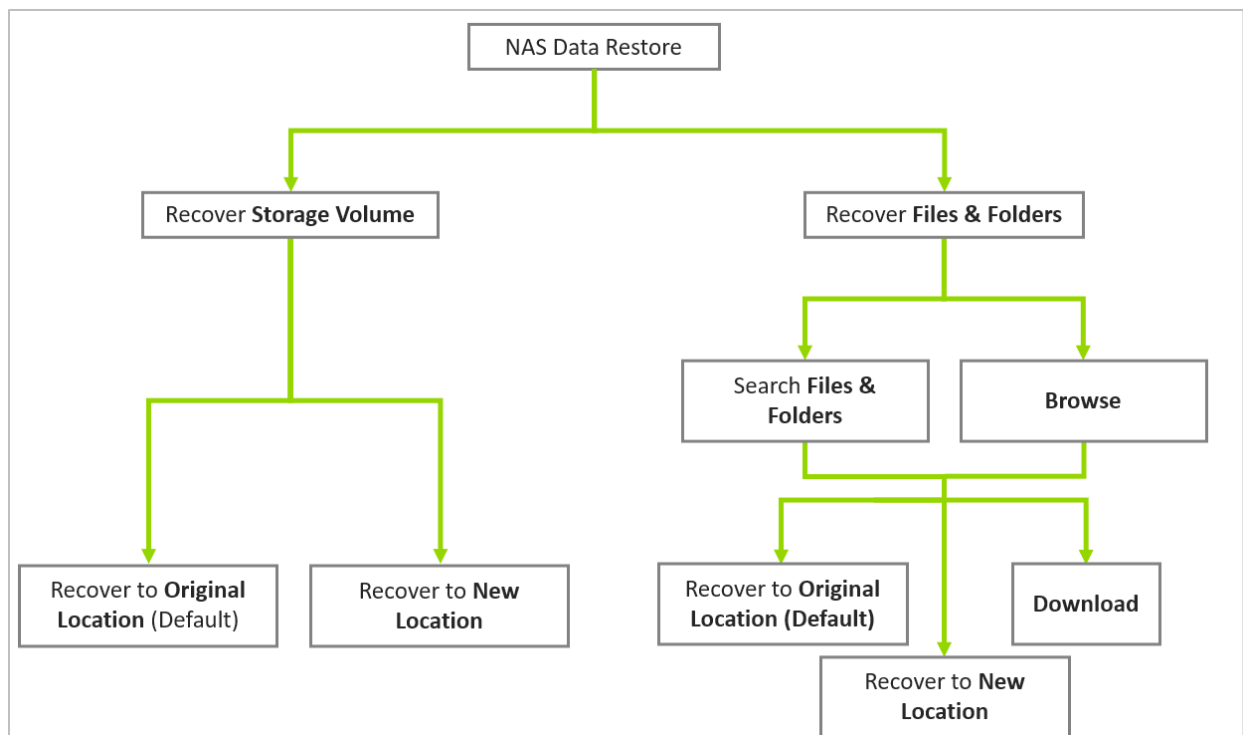
Once the data is archived to the External Target, you can choose to recover to your original NAS or to a new, alternate NAS.

Figure 3 below illustrates the many recovery choices and options you have with Cohesity.

### NOTES:

- There is no need to restore the entire file or volume onto the Cohesity cluster before restoring it to the NAS device.
- NAS volumes and files & folders can be recovered to their original location or to a newly specified location, which can be in the original or a different NAS source. You can also download files from any specific snapshot that was created by a Cohesity Protection Group.

Figure 3: NAS Cloud Recover Decision Tree



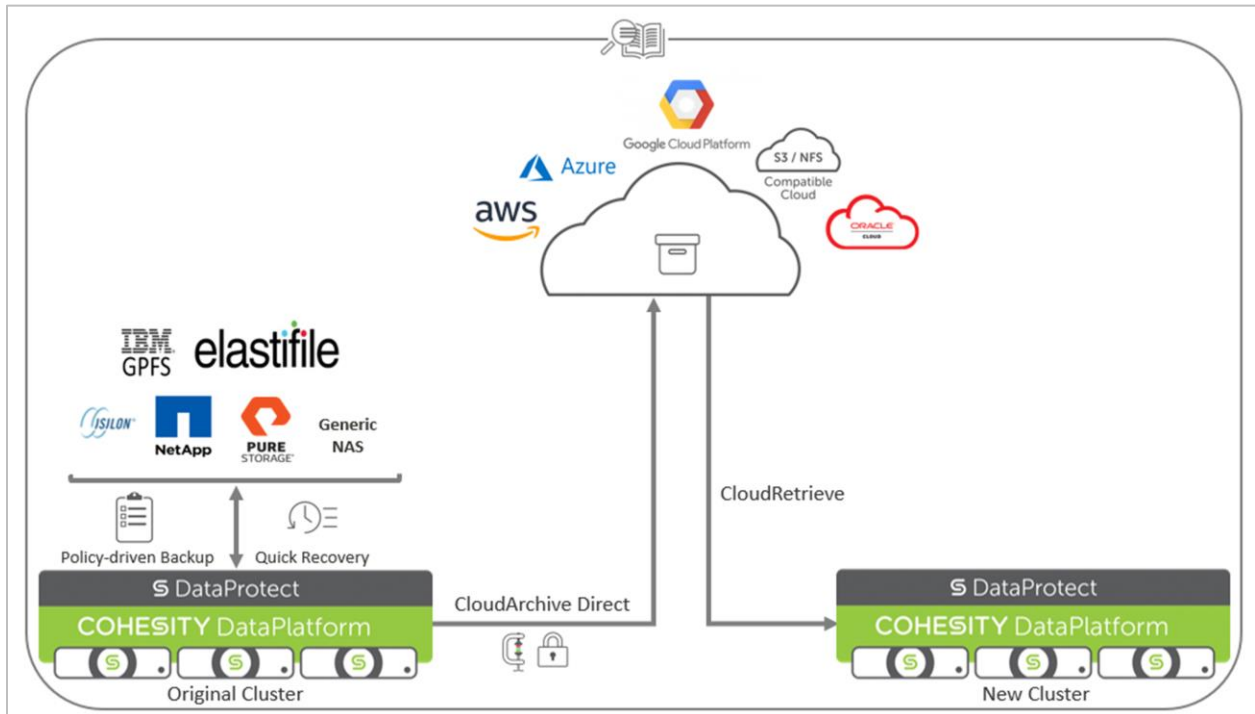
For details on recovering files & folders, see [Recover Files or Folders](#) in the online Help. For storage volume recovery, see [Recover Storage Volumes](#).

## 5 Disaster Recovery with CloudArchive Direct

Because CloudArchive Direct stores index and metadata together with the data in the External Target, you will be able to use Cohesity's CloudRetrieve feature to recover the archived data using a new Cohesity cluster in the event that your original cluster becomes unavailable due to disaster or some other unfortunate event. This acts as a cost-effective alternative for disaster recovery, geo-redundancy, and business continuity.

For more details, see [About CloudRetrieve](#) in online Help.

Figure 4: Cloud Recover to Original Source & CloudRetrieve to New Cluster



## 6 Prerequisites

Before you get started with this solution, ensure that you have met the prerequisites:

- The workload is in the NAS format.
- The workload is accessible via NFS or SMB/CIFS.
- The Cohesity Protection Policy is configured for:
  - Only one External Target for archival.
  - No archival to tape.
  - No remote clusters for replication.
  - Archive to run after every backup run.
  - No retries for backup runs.

## 7 Your Feedback

Was this document helpful? [Send us your feedback!](#)

## 8 About the Authors

Ruby Garg is a Technical Marketing Engineer at Cohesity. In her role, Ruby focuses on NAS data protection.

Other significant contributors included:

- Adaikkappan Arumugam, Sr. Manager, Technical Marketing
- Bart Abicht, Senior Technology Writer and Editor at Cohesity

## 9 Document Version History

VERSION	DATE	DOCUMENT HISTORY
1.0	May 2020	First full release
2.0	Sep 2020	Updated public release

## ABOUT COHESITY

[Cohesity](#) ushers in a new era in data management that solves a critical challenge facing businesses today: [mass data fragmentation](#). The vast majority of enterprise data — backups, archives, file shares, object stores, and data used for dev/test and analytics — sits in fragmented infrastructure silos that makes it hard to protect, expensive to manage, and difficult to analyze. Cohesity consolidates silos onto one web-scale [platform](#), spanning on-premises, cloud, and the edge, and uniquely empowers organizations to run apps on that platform — making it easier than ever to back up and extract insights from data. Cohesity is a [2019 CNBC Disruptor](#) and was named a [Technology Pioneer by the World Economic Forum](#).

Visit our [website](#) and [blog](#), follow us on [Twitter](#) and [LinkedIn](#) and like us on [Facebook](#).

© 2020. Cohesity, Inc. All Rights Reserved.

*Cohesity, the Cohesity logo, SnapTree, SpanFS, DataPlatform, DataProtect, Helios, and other Cohesity marks are trademarks or registered trademarks of Cohesity, Inc. in the US and/or internationally. Other company and product names may be trademarks of the respective companies with which they are associated. This material (a) is intended to provide you information about Cohesity and our business and products; (b) was believed to be true and accurate at the time it was written, but is subject to change without notice; and (c) is provided on an "AS IS" basis. Cohesity disclaims all express or implied conditions, representations, warranties of any kind.*

2000028-002-EN