

Archive Your Data Directly with Cohesity CloudArchive Direct

Technical Overview of Cohesity's Cost-effective CloudArchive Direct

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

Introduction to CloudArchive Direct.....	4
Key Terms	4
Explore Cohesity's Long-term Retention Solutions	5
Cohesity CloudArchive	5
Cohesity CloudArchive Direct.....	5
<i>Differences Between Legacy CloudArchive Direct or CAD v1 and CloudArchive Direct v2.....</i>	7
<i>Features and Benefits of CloudArchive Direct</i>	8
<i>Cohesity Data Movement.....</i>	9
Compare CloudArchive with CloudArchive Direct	10
Configure CloudArchive Direct	12
Set Up CloudArchive Direct.....	12
Recover Data with CloudArchive Direct	13
Disaster Recovery with CloudArchive Direct.....	15
Prerequisites and Limitations	16
Prerequisites	16
Limitations	16
Your Feedback	17
About the Authors.....	17
Document Version History.....	17

Figures

Figure 1: Cohesity CloudArchive Archives Your Local Backups on External Targets	5
Figure 2: Legacy CloudArchive Direct Archives Your NAS Data Directly to an External Target.....	6
Figure 3: CloudArchive Direct Archives Your NAS Data and VMs Directly to an External Targets.....	7
Figure 4: Data Migration.....	10

Figure 5: NAS Cloud Recover Decision Tree..... 13

Figure 6: VM Cloud Recover Decision Tree 14

Figure 7: Cloud Recover to Original Source and CloudRetrieve to New Cluster..... 15

Tables

Table 1: Key Feature Differences Between Legacy CloudArchive Direct Or CAD V1 And CloudArchive Direct v2..... 7

Table 2: Key Features and Benefits of CloudArchive Direct 8

Table 3: Features and Benefits Inherited from Cohesity Platform..... 9

Table 4: Compare CloudArchive Direct with CloudArchive 10

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) and run most of their workloads on VMWare infrastructure, it has never been more important for them to focus on NAS and VMWare data protection—how they store, protect, and retain all that data.

CloudArchive Direct (CAD) is a policy-driven archival feature in Cohesity 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 and 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.

Key Terms

This document covers the following key terms to explain CloudArchive Direct:

- **CloudArchive Direct or CloudArchive Direct v2 (CAD v2):** In this document, the term CloudArchive Direct or CAD v2 defines the latest version of CloudArchive Direct released with Cohesity version 6.7.
- **Legacy CloudArchive Direct:** In this document, the term Legacy CloudArchive Direct defines the older version of CloudArchive Direct introduced in Cohesity version 6.4.1 onwards.

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 lower-cost storage, typically in the cloud:

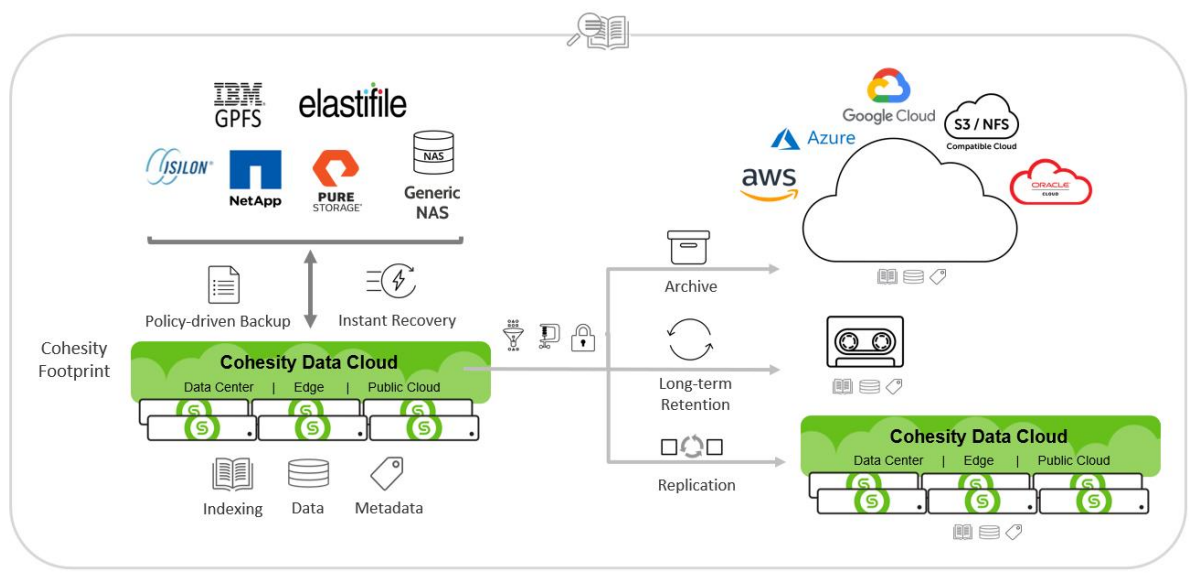
- **CloudArchive**. Back up your data on a Cohesity cluster and copy it to the cloud for archival.
- **CloudArchive Direct**. Stream your data directly to a storage target without creating local backups first.

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, that can be restored to any Cohesity cluster, of any type, regardless of the origin of the data (using CloudRetrieve).

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



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

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 and VMware Virtual Machine (VM) instances directly 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's platform, you can register AWS cloud storage, S3-compatible storage, or NFS-mounted servers as External Targets.

NOTE:

- VMware support is only available with Cohesity version 6.7 onwards.
- With legacy CAD, AWS, Azure, GCP, or Oracle cloud storage is supported.
- 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, dedupe, and encrypt the data for dramatic savings in space and network traffic but does not support data access by third-party tools. With Cohesity version 6.7 onwards, CloudArchive Direct does not support backing up NAS data in native NAS format.

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

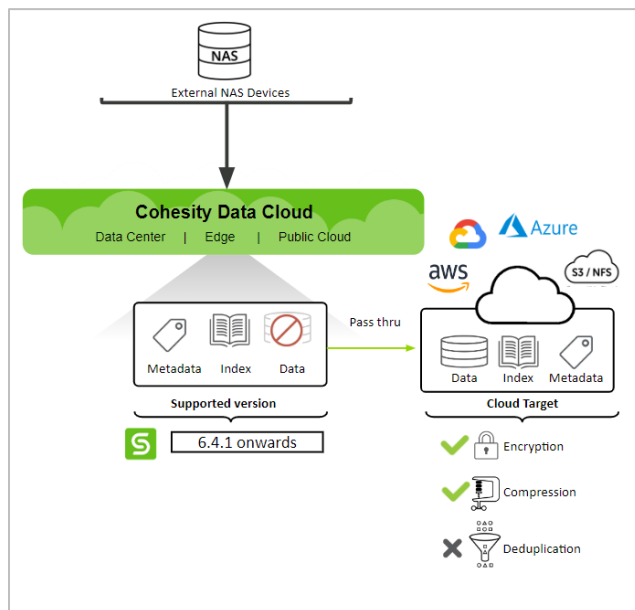
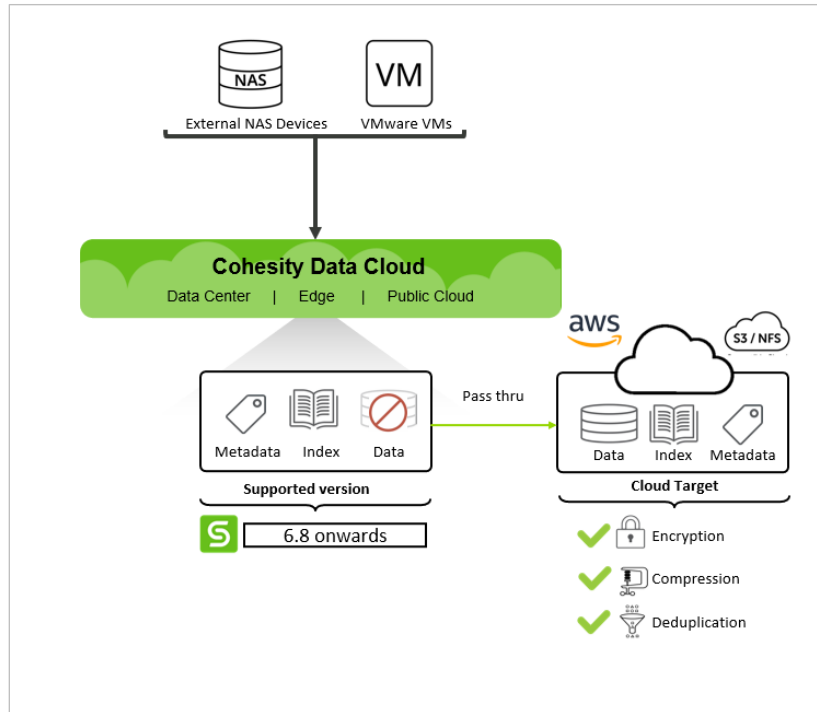


Figure 3: CloudArchive Direct Archives Your NAS Data and VMs Directly to an External Targets



Differences Between Legacy CloudArchive Direct or CAD v1 and CloudArchive Direct v2

Legacy CloudArchive Direct or CAD v1 was introduced in Cohesity version 6.4.1. CloudArchive Direct v2 is available from Cohesity version 6.8 onwards.

Table 1: Key Feature Differences Between Legacy CloudArchive Direct Or CAD V1 And CloudArchive Direct v2

LEGACY CLOUDARCHIVE DIRECT OR CAD v1	CLOUDARCHIVE DIRECT v2
Uses incremental forever with periodic metadata full method for archiving. While data is incremental forever, by default every 90 days a full metadata archive will be sent to the external target (this period is configurable).	Uses incremental forever upload of data and metadata. No periodic full metadata upload required.
From a source, only a single object can be selected for backup per Protection Group.	From a source, supports multiple objects per Protection Group.
No Life Cycle Management support	Life Cycle Management provides multiple data movement policies to downtier the data to lower storage classes.
No deduplication support	Supports deduplication
Supports only NAS sources	Supports NAS and VMware Sources

LEGACY CLOUDARCHIVE DIRECT OR CAD v1	CLOUDARCHIVE DIRECT v2
Can backup data in Native or Non-Native format	Always backs up data in Non-Native format
Enabled at Protection Group level	Enabled at Protection Policy level
Supports AWS,Azure,GCP, and Oracle Cloud storage	Supports only AWS

Features and Benefits of CloudArchive Direct

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

Table 2: Key Features and Benefits of CloudArchive Direct

CLOUDARCHIVE DIRECT FEATURE	BENEFIT
Low on-premises footprint	Stores only the metadata and index on-premises, requiring only a minimal Cohesity cluster footprint.
Metadata	Available on-premises and in the cloud. And because Cohesity 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 CloudRetrieve (onto a new cluster) feature.
Cost Storage Efficiency	Incremental archival at block level. Only changed blocks will be archived instead of a complete file.
Cohesity Data Movement (Life Cycle Management (LCM) support)	Based on your business SLAs, Cohesity Data Movement feature allows the data to be downtiered to a lower storage class.

There are also many benefits that CloudArchive Direct inherits from Cohesity's platform, summarized in Table 3 below.

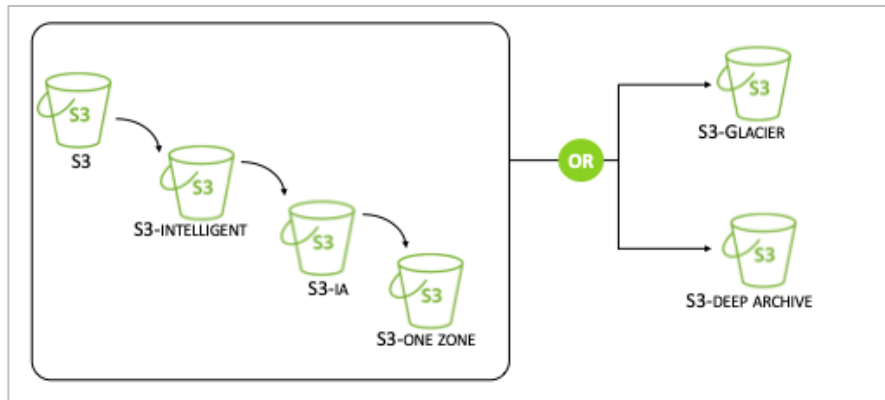
Table 3: Features and Benefits Inherited from Cohesity Platform

COHESITY PLATFORM FEATURE	BENEFIT
Automated, policy-driven workflow	Define simple and flexible policies to archive your NAS data directly to the cloud.
Distributed and parallel upload	Cohesity 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.
Incremental indexing	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.
Faster and granular restores	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 of files in the cloud.
Storage efficiency on External Target with Compression and deduplication	Compresses and deduplicates data to reduce workload by 2-5x before sending it to the External Target, requiring much less cloud storage.
Data security with encryption	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.
Mobility	Recover to any other Cohesity cluster using CloudRetrieve .

Cohesity Data Movement

From Cohesity version 6.7 onwards, Cloud Archive Direct supports Data Movement, a policy-based data LCM in external targets to downtier the data to lower storage classes. This allows Cohesity to directly own the LCM of data in an external target. When using data movement, data must always be downtiered. Data movement does not support uptiering data. When it comes to S3 Glacier and S3 Deep Archive a choice must be made. It is not possible to downtier from S3 Glacier to S3 Deep Archive as AWS does not allow downtiering from S3 Glacier to S3 Deep Archive. Once you have downtiered to either S3 Glacier or S3 Deep Archive you cannot downtier any further.

Figure 4: Cohesity Data Movement



As a best practice, Cohesity never dntiers the metadata and index. It always stores the metadata and index in the original tier to speed up the recovery operations (E.g.: Data down-tiered from S3 > S3 One Zone, will always have metadata retained in S3).

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 4 below provides a side-by-side comparison of the benefits and trade-offs of each solution.

Table 4: Compare CloudArchive Direct with CloudArchive

FEATURE	CLOUDARCHIVE DIRECT	CLOUDARCHIVE
Business needs addressed	Reduced TCO	Local backups
	Long-term data retention Security Compliance	
Local backup retention	No	Yes
Cluster space requirements	Requires a fraction of the footprint of the data being archived.	Requires enough space for at least one copy of the backup.
Indexed	Yes	
Metadata	Yes	
Data natively accessible in cloud	Yes	No, requires a Cohesity cluster
Encryption	Yes	

FEATURE	CLOUDARCHIVE DIRECT	CLOUDARCHIVE
Deduplication	Yes	
Compression	Yes	
Incremental forever	Yes	Yes Subject to source supportability
Replication	Cannot be replicated.	Yes, as there is a local backup copy
Tape out	No	Yes (Supported only with Legacy Cloud Archive)
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	
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.
Cohesity Data Movement (Object Lifecycle Management)	Supported	
Cloud Native Object Lifecycle Management	Not Supported	Supported (Legacy CloudArchive)
Instant restore	No	Yes, to a Cohesity View
Clone	No	Yes
File and folder restore	Yes	
WORM Support	No	Supported (Legacy CloudArchive) 6.8 onwards

Configure CloudArchive Direct

Set Up CloudArchive Direct

To enable CloudArchive Direct:

1. [Register or edit your NAS VMware vcenter](#) as a source in Cohesity.
2. [Register an External Target](#) where you will archive your NAS and VMware data.
When you register an External Target for CloudArchive Direct:
 - Select an archival format (Incremental Forever, Incremental with Periodic Full).
3. [Create a Protection Policy with archival scheduled after every run](#).
4. [Add or edit a Protection Group to protect your NAS or VMware workloads](#).

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

Recover Data with CloudArchive Direct

Cohesity 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 or VM, with their respective names, or you can browse the archived data index to find, select, and perform data recovery.

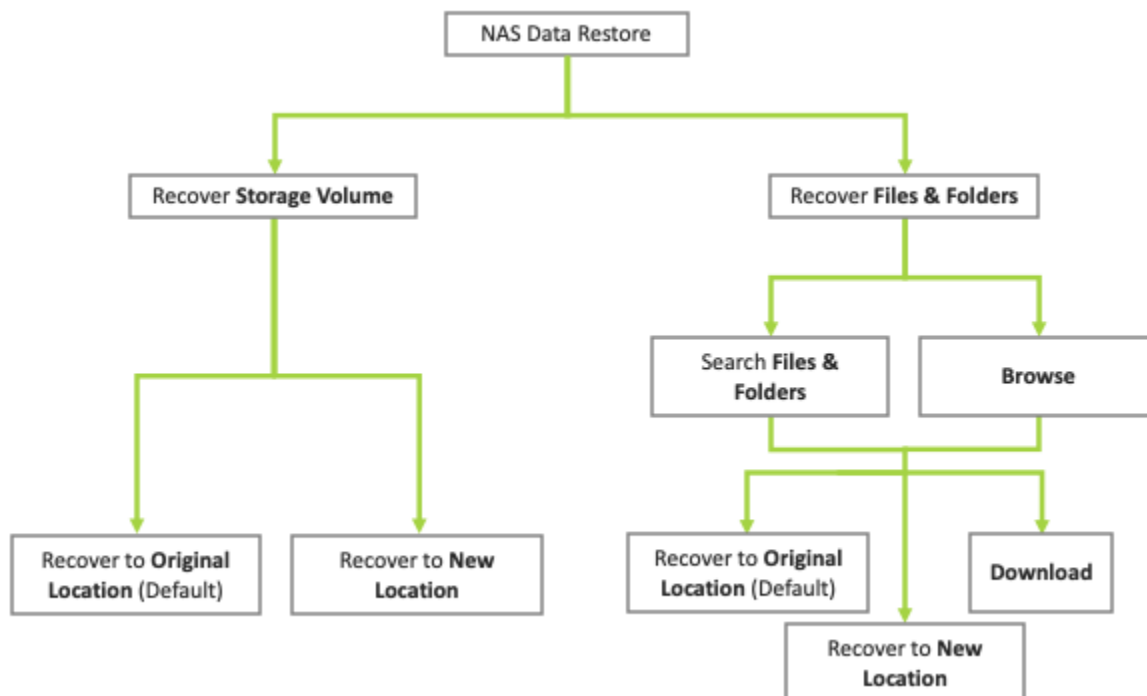
Once the data is archived to the External Target, you can choose to recover to your original source or to a new, alternate source.

Figure 5 below illustrates the many recovery choices and options you have with Cohesity.

NOTES:

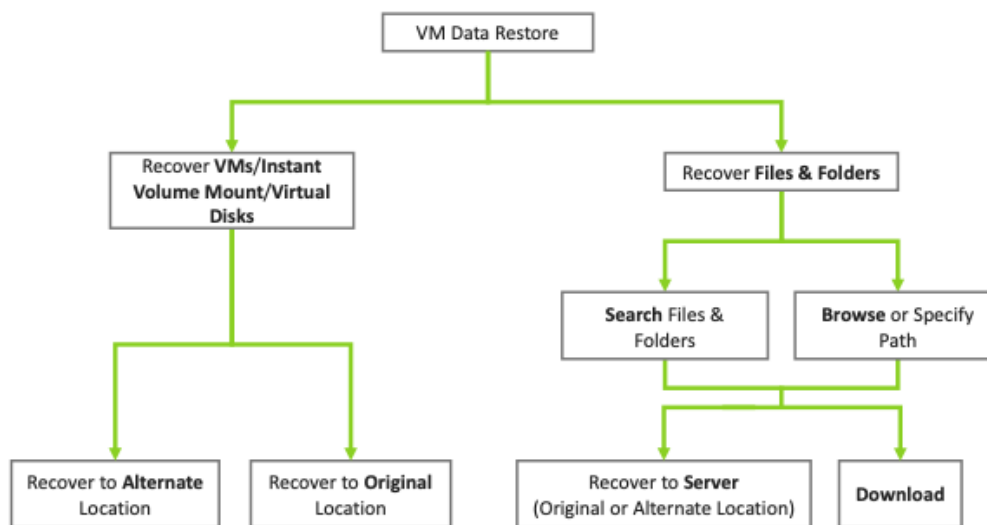
- There is no need to restore the entire file or volume onto the Cohesity cluster before restoring it to the NAS or VMware source.
- You can recover a NAS or VM and its files and folders to their original location or to a newly specified location, which can be in the original or a different source. You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

Figure 5: NAS Cloud Recover Decision Tree



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

Figure 6: VM Cloud Recover Decision Tree



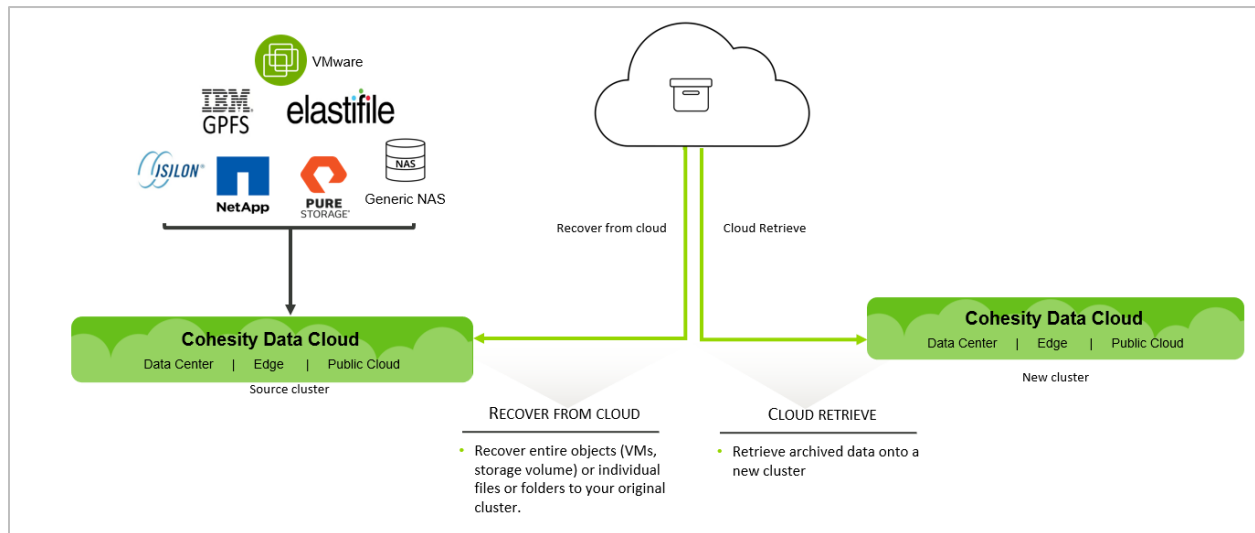
For details on recovering VM, Instant Volume Mount, Virtual Disks, files and folders, see [Recovery](#).

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 7: Cloud Recover to Original Source and CloudRetrieve to New Cluster



Prerequisites and Limitations

Before you get started with this solution, there are a few things to consider.

Prerequisites

In order to archive your data directly to ExternalTarget, ensure:

- The workload is in the NAS or VM format.
- External target is configured for Incremental Forever archival format.

Limitations

Be sure to consider these limitations as you develop your data protection strategy with CloudArchive Direct:

- If the source NAS is Dell EMC Isilon, Isilon ChangeList is not supported with CloudArchive Direct.
- No recovery to a Cohesity View.
- If the data was archived via CloudArchive Direct and your Cohesity cluster becomes unavailable, you can stand up a new Cohesity cluster to recover the metadata and index, but you will have to recover the data itself to the original or a new source.
- Do not apply object lifecycle policy actions like *transition* or *expiration* to the External Targets that you use for CloudArchive Direct, as applying an object lifecycle policy with these actions can result in data loss or unresponsive operations.
- From Cohesity version 6.7 onwards the data is always archived in CAD v2 format and Legacy CloudArchive Direct is deprecated. Contact support if you want to enable Legacy CloudArchive Direct.
- After version upgrade to 6.7, existing CloudArchive Direct jobs will NOT get converted to v2 format. You need to create new protection jobs and retire the old ones.

Your Feedback

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

About the Authors

Saran Ravi is a Sr. Technical Solution Engineer at Cohesity. In his role, Saran focuses on Cloud and Cohesity Cloud Services Portfolios.

Other significant contributors include:

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

Document Version History

VERSION	DATE	DOCUMENT HISTORY
4.2	July 2024	Republishing
4.1	Mar 2023	Rebranding updates
3.0	Nov 2021	Updated for 6.7 release
3.0	Nov 2021	Updated for 6.7 release
2.1	Sep 2020	Updated for 6.5.1 release
1.1	May 2020	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](#).

© 2024. Cohesity, Inc. All Rights Reserved. The information supplied herein is the confidential and proprietary information of Cohesity and may only be used (a) by the intended recipients and (b) in conjunction with validly licensed Cohesity software and services. Find the terms of Cohesity licenses at www.cohesity.com/agreements.

Cohesity, the Cohesity logo, SnapTree, SpanFS, DataPlatform, DataProtect, Helios, the Helios logo, DataGovern, SiteContinuity, DataHawk, 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.