COHESITY

Version 1.2

July 2024

Cohesity Cloud Tier Architecture Reference

Seamlessly Down- and Up-Tier Data to Leverage Low-Cost Object Storage

ABSTRACT

As the pace of data growth continues to increase unrelentingly in today's web-scale organizations, one of the greatest challenges facing IT administrators is optimizing the cost and managing that growth. Making matters more challenging, some of that data is accessed frequently while other data is rarely touched. Cohesity Cloud Tier provides a seamless way to down-tier your cold (infrequently accessed) data, and up-tier it back when it is hot (accessed frequently).

Table of Contents

Figures

Tables

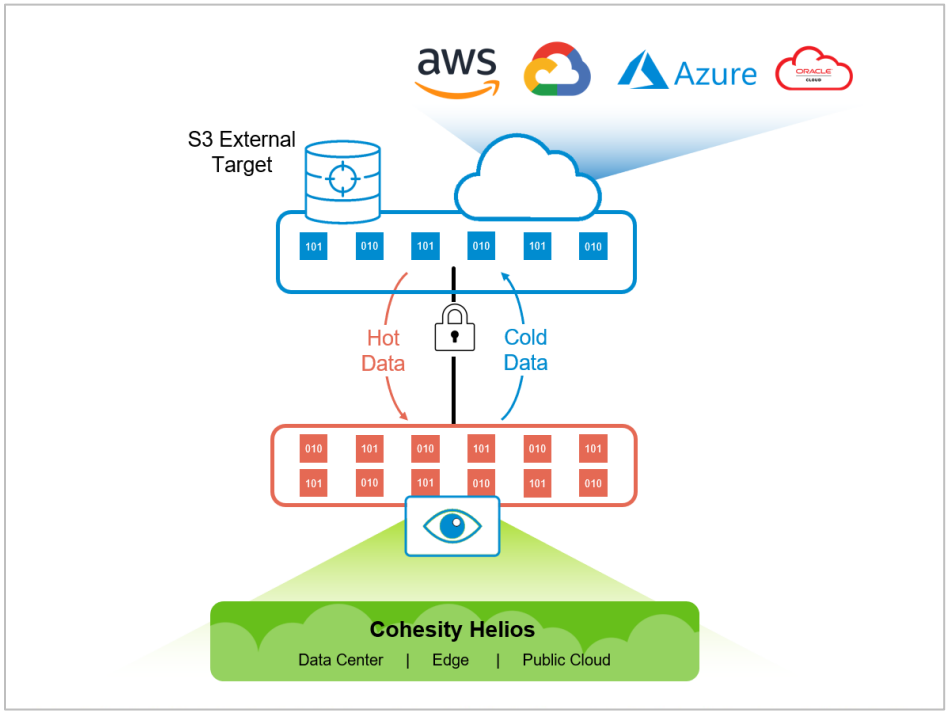
Introduction to Data Tiering

The performance, availability, and cost requirements for storing and accessing your data can change based on your business needs. But as your data continues to grow, it becomes difficult to manage and store the data at a lower cost. If you manage and store large amounts of data—for compliance, regulatory, or legal reasons—that are inactive for long periods of time, tiering those cold data blocks to lower-cost cloud storage can help you achieve significant savings.

Cohesity Cloud Tier allows you to seamlessly move infrequently accessed (cold) data to lower-cost storage in the cloud and move it back to the cluster when it is accessed again (hot data), back to the cluster when needed, reducing operating expenses and Total Cost of Ownership (TCO).

You can move infrequently accessed data to External Targets such as public cloud infrastructure providers (AWS, Azure, Google Cloud Platform) or any S3-compatible External Target. Once enabled, Cohesity Cloud Tier automatically moves cold data to the External Target.

Figure 1: Cloud Tier Moves Cold Data into the Cloud, and Hot Data Back



NOTE: Learn more about Cloud Tier terminology in Appendix A.

Key Considerations Before Enabling Cloud Tier

It is important to understand that tiering only moves cold data to lower-cost storage and, unlike archival, does not create a new complete copy of the data. Tiering extends the active file system into the External Target. Also, while deduplication savings extend to data tiered in the External Target, the metadata remains on the Cohesity cluster, in the flash tier (SSD), while the cold data itself is moved from HDDs in the cluster to the External Target in the cloud. In the event either is lost, the data will become unavailable.For this reason, Cloud Tier is not a replacement for long-term data retention.

Before enabling Cloud Tier, there are a few key facts you need to consider:

- Data in the External Target is not usable without the metadata from the cluster.
- Once enabled, tiering cannot be disabled. For this reason, you should not use it to expand cluster capacity temporarily.
- Latency might increase for operations that require access to cold data as I/O for cold data is serviced from the External Target.

The Difference Between Cloud Tier and CloudArchive

It is a common misconception that Cloud Tier and CloudArchive are similar. In fact, they are two separate features that address different needs.

- **Cloud Tier.** Extends the active file system into an External Target, and is focused on reducing your data center TCO.
- **CloudArchive.** A fully self-contained, off-site copy of your backup, intended as a long-term data retention solution to meet security, legal, and compliance requirements.

Table 1: Differences Between Cloud Tier and CloudArchive

PURPOSE	CLOUD TIER	CLOUDARCHIVE
Definition	The process of moving inactive orinfrequently accessed data (cold data) to an External Target while retaining the metadata on the cluster.	The process of moving a fully self-contained copy of a backup, with data, metadata, catalog, indexing, and deduplication fingerprint.
Business need	To reduce TCO.	• Long-term data retention • Security • Compliance
Trigger	Tiering threshold.	Protection Policy schedule.
Data type	Agnostic of the data type.	Applies only to data backups.
Granularity	Storage Domain	Protection Group
Space reclamation	Remains until/unless the userexplicitly deletes the data.	Automatic garbage collection, based on the retention period that is defined in the Protection Policy.
Limits	There are limits to the amount of data that can be tiered to an External Target. These limits are governed by the size of the clusterand other factors. For more, see the Appendix B: Impact of Deduplication on Cloud Tiering.	N/A

How Cloud Tier Works

Cloud Tier extends your active file system into an External Target and seamlessly moves cold (infrequently accessed) data to the External Target and back again to the Cohesity cluster when it is accessed.

The tiering of cold data to the External Target is based on a policy with two factors: tiering threshold and data policy. Tiering happens only if utilization exceeds the tiering threshold *and* there are data blocks that meet the data policy.

- **Tiering threshold**. The percentage of space utilization that is set to trigger the tiering of cold data.
- **Data policy**. Specifies the duration of time that the data must be inactive for it to be eligible for tiering.

Tiering Threshold

The tiering threshold and data policy can be set on the Cohesity cluster, or on an individual Storage Domain, or both.

When set on the cluster, all Storage Domains inherit the setting.

When set on an individual Storage Domain, it is not applied to any other Storage Domains.

When the threshold is set on both the cluster and a Storage Domain, the Storage Domain settings take precedence. This allows you to set a global threshold for your cluster and a different threshold for specific Storage Domains, giving you greater flexibility. (For example, on Storage Domains that are accessed frequently, you might disable Cloud Tier or set a very different threshold.)

NOTE: The Storage Domain-level Cloud Tier threshold can be set only if the Physical Quota is enabled for the Storage Domain. See Create or Edit Storage Domains in the online Help to enable Physical Quota for a Storage Domain.

Data Policy

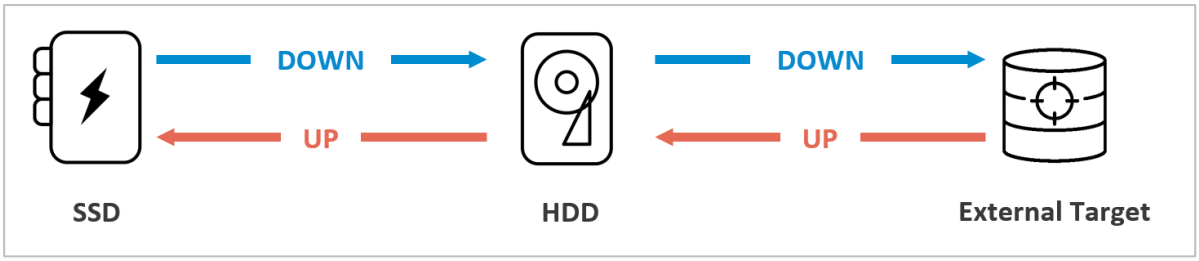
The Cloud Tier data policy specifies the duration of time that the data must be inactive (that is, not accessed) for it to be eligible for tiering. The data that is not accessed within the defined period will be eligible for tiering to the External Target. As with the Cloud Tier threshold, the Storage Domain's data policy takes precedence over the cluster's data policy.

Down-Tiering and Up-Tiering

There are two types of tiering:

- **Down-tiering.** The process of moving infrequently accessed data (cold data) from the Cohesity cluster to an External Target when the tiering threshold and data policy are met.
- **Up-tiering.** The process of moving data that is accessed frequently (hot data) from an External Target back to the Cohesity cluster.

Figure 2: Down-tier and Up-tier Data Flow



Down-tiering is a background process that triggers when the tiering threshold is exceeded and scans the cluster for cold data. The cold data (as defined in the data policy) is then moved to an External Target. This process continues on the cluster until the space utilization has returned to the tiering threshold, or until there is no more eligible cold data. This process stops once enough data has been down-tiered back to the defined threshold.

In this way, the threshold serves both as a marker for when usage is high enough to start down-tiering data, and also as a limit for how much data is down-tiered before the process stops, thereby never over-down-tiering.

NOTE: There is no guarantee tiering will return the cluster to the tiering threshold if there is not enough eligible cold data on the cluster to do so.

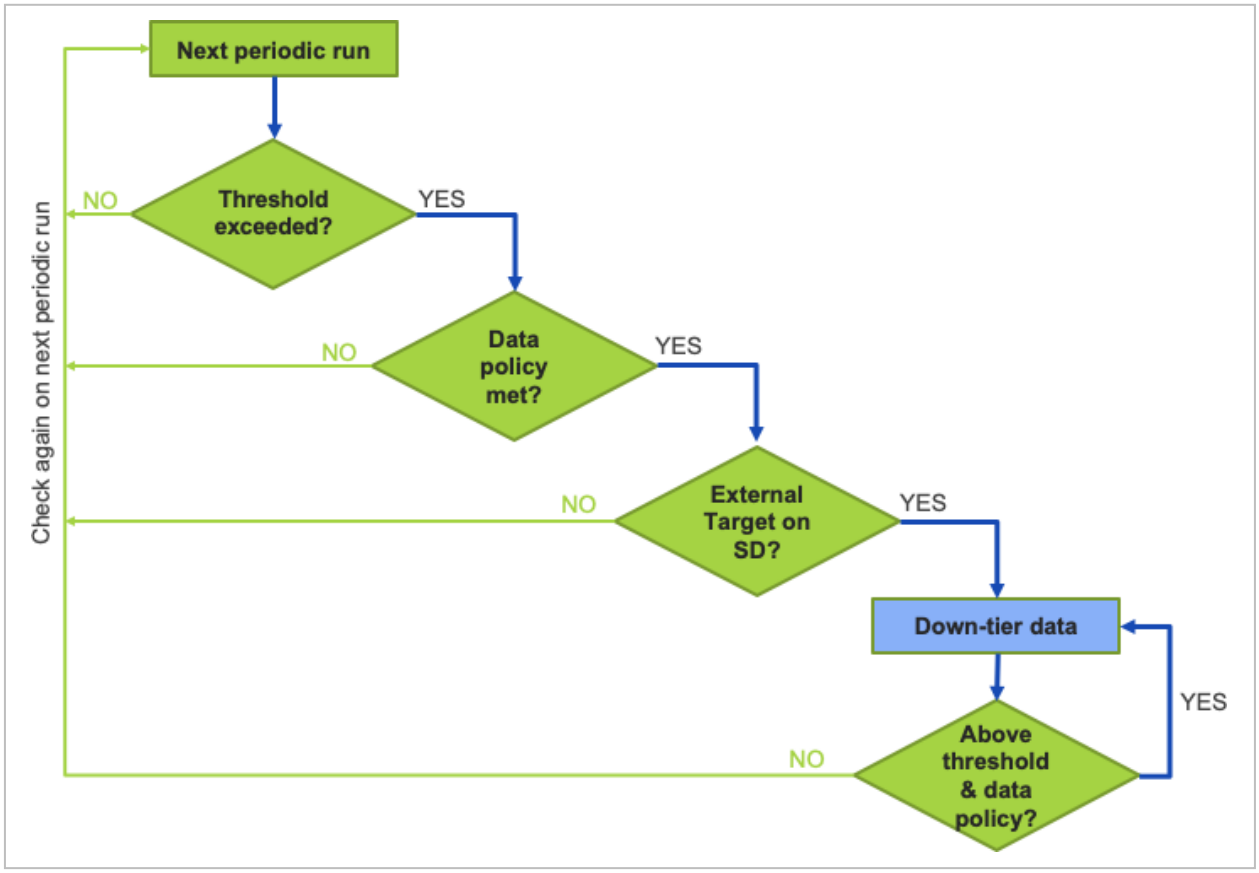
Once data is down-tiered, I/O operations for the down-tiered data are serviced directly from the External Target unless many I/O requests for the same data occur within a short period of time. In this case, down-tiered data that has become hot is up-tiered back onto the cluster.

NOTE: While down-tiering cold data onto an External Target is usually free, cloud vendors often impose egress charges for data that is up-tiered back to the cluster.

Down-Tiering Algorithm

As the background process scans your data (by default, every four hours), the system determines data utilization and age to decide which data blocks can be down-tiered to the External Target. Figure 3 below illustrates the sequence of decisions.

Figure 3: Down-tiering Algorithm Decision Tree



NOTE: Down-tiering is an active process that runs with a lower priority than foreground I/O (such as active workloads). As a result, during times of contention, down-tiering can occur at a reduced rate due toI/O prioritization.

IMPORTANT: Once you enable Cloud Tier on a cluster, you cannot disable it on that cluster, as it is not possible, once data has been tiering to an External Target, to completely evacuate the data from the External Target.

Default Tiering Settings on the Cluster

The default Cloud Tier thresholds — 80% utilization and data policy of two months — are preconfigured on every Cohesity cluster. The tiering of cold data only starts for those Storage Domains in the cluster that have External Targets registered, and only when the threshold conditions are met. The default Cloud Tier threshold on the cluster is inherited by the Storage Domains unless specifically configured on individual Storage Domains, where you can override both the threshold and the data policy.

The cluster threshold can only be overridden by a Storage Domain if a Physical Quota is set on that Storage Domain. (This is because, without a specific Physical Quota, Storage Domains can utilize the full capacity of the Cohesity cluster.) When set, Storage Domain thresholds take precedence over cluster thresholds.

NOTE: If the cluster has no Storage Domains with registered External Targets, no cloud tiering takes place, even when thresholds are met.

Impact of Fault Tolerance on Down-Tiered Data Consumption

Being a distributed system, Cohesity is resilient to both hardware and software failures. This fault tolerance is achieved by employing fault tolerance technologies like Replication Factor (RF) and Erasure Coding (EC) for data that is stored in the cluster. When the data is down-tiered to an External Target, the fault tolerance of that data is determined by two factors:

- Storage Domain fault tolerance setting on the cluster from which it was down-tiered.
- The provider of the External Target used for tiering.

The amount of storage you consume in the External Target is directly affected by the fault tolerance setting of the Storage Domain. Therefore:

- If the External Target is AWS, Azure, or GCP, RF1 is used as the fault tolerance setting — that is, only one copy of the data is stored in the cloud, regardless of the Storage Domain fault tolerance setting.
- If the External Target is *not* on AWS, Azure, or GCP, the down-tiered data mirrors the fault tolerance level (in RF or EC) of the Storage Domain (with RF2, RF3, etc.), that is, it provides the same level of fault tolerance as the Storage Domain.

Table 2: Cloud Tier Fault Tolerance in Public vs S3-Compatible Clouds

STORAGE DOMAIN FAULT TOLERANCE	CLOUD TIER FAULT TOLERANCE	
	OTHERS	AWS/AZURE/GCP
RF2	RF2	RF1
RF3	RF3	RF1
EC X:1	RF2	RF1
EC X:2	RF3	RF1
EC X:3	RF4	RF1

When you use vendors other than AWS, Azure, or GCP, the storage consumed in the External Target is the same as the data that needs to be tiered. On the other hand, when you use AWS, Azure, and GCP (which only store one copy of the data), it often consumes less space in the External Target than the data that needs to be tiered.

Example

The following table illustrates an example of the storage capacity utilized on Storage Domain and External Target when the user data is 100 GB for various Storage Domain fault tolerance settings.

NOTE: For the sake of simplicity, these examples assume that deduplication and compression are disabled on the Storage Domain. However, in real-world scenarios, with deduplication and compression *enabled*, Storage Domain utilization is significantly reduced.

Table 3: Example: Fault Tolerance & Data Usage in Storage Domain & External Target for 100 GiB

USER DATA (GiB)	STORAGE DOMAIN			CLOUD TIER PROVIDER			
	NUMBER OF DISK FAILURES TOLERATED	FAULT TOLERANCE	USAGE (GiB)	OTHERS		AWS/AZURE/GCP	
				FAULT TOLERANCE	USAGE (GiB)	FAULT TOLERANCE	USAGE (GiB)
100	1	RF2	200	RF2	200	RF1	100
		EC 2:1	133				
		EC 3:1	125				
	2	RF3	300	RF3	300		
		EC 2:2	200				
		EC 4:2	133				
		EC 5:2	129				
	3	RF 4	400	RF4	400		
		EC 3:3	150				
		EC 4:3	143				
		EC 5:3	137				
		RF2	200				

Security & Storage Efficiency of Down-Tiered Data

When down-tiering data to an External Target, it is crucial to ensure that the data that is stored outside the Cohesity cluster is not compromised. Cohesity employs software-based, AES 256-bit encryption. The down-tiered data is always encrypted, for both in-flight and at-rest, and this cannot be disabled, even if encryption is not enabled for that data on the Cohesity cluster or in the individual Storage Domain. The encryption key used for the down-tiered data is securely stored in the Cohesity cluster. Tiering employs the native Cohesity encryption capability. For details, see Cohesity Security Features in the online Help.

To reduce the storage cost and network utilization, down-tiered data is always compressed in addition to being encrypted regardless of the cluster and Storage Domain settings. For deduplication, however, the down-tiered data in the External Target reflects the same deduplication settings as in the Storage Domain on the cluster. In other words, if it's deduped on cluster, it's deduped in the External Target, and when it's not deduped on the cluster, it's not deduped in the External Target, either.

Table 4 below illustrates how deduplication, compression, and encryption are applied to down-tiered data based on various Storage Domain settings for each.

Table 4: Efficiency & Security Settings in Storage Domain & Cloud Tier

EXAMPLE	STORAGE DOMAIN			DOWN-TIERED DATA ON EXTERNAL TARGET
	DEDUPLICATION	COMPRESSION	ENCRYPTION	
Storage Domain 1	Yes	Yes	Yes	• Deduplicated • Compressed • Encrypted
Storage Domain 2	No	No	No	• Not deduplicated • Compressed • Encrypted
Storage Domain 3	Yes	No	Yes	• Deduplicated • Compressed • Encrypted
Storage Domain 4	No	Yes	No	• Not deduplicated • Compressed • Encrypted

Supported External Target Types

Because I/O operations for the down-tiered data are serviced directly from the External Target, it is crucial that the type of storage in the External Target provides low latency to your data access requests. Otherwise, the user might start seeing delays when accessing down-tiered data. For this reason, not all object storage types are supported. For example, AWS Glacier Vault is not supported because the data retrieval SLA is minutes to hours instead of seconds or less.

WARNING: Do not apply an Object Lifecycle Policy, like transition or deletion, to the object storage you use for tiering. Applying a policy with these actions can result in data loss or unresponsive operations. As Cloud Tier stores metadata on the cluster and the associated cold data on the External Target, a transition or deletion action on an object can lead to data unavailability.

Table 5 below lists the supported External Targets that can be registered with Cohesity for Cloud Tier.

Table 5: Supported External Targets for Cohesity Cloud Tier

PROVIDER	CLASS OF STORAGE	FAULT TOLERANCE
AWS	S3 Intelligent-Tiering	RF1
	S3-Standard	RF1
	S3-GOV	RF1
	S3-C2S	RF1
Azure	Hot Blob-Standard	RF1
	Hot Blob-Gov	RF1
Google Cloud Platform	Multi-Regional	RF1
	Regional	RF1
Oracle	Object Storage	Mirrors Storage Domain
Other External Targets	S3 Compatible	Mirrors Storage Domain

Cloud Tiering Scenarios

Now that we've covered the basic concepts of cloud tiering, we can explore some different scenarios of Cloud Tier thresholds set on the cluster and Storage Domains. We'll look at various factors, including Storage Domain quotas, number of Storage Domains, and External Targets that are set to different values.

IMPORTANT: The parameter values provided for the following scenarios are arbitrary values and are used for example purposes only.

Storage Domains without Physical Quota

If a Storage Domain has no Physical Quota set, it inherits the Cloud Tier thresholds from the cluster.

There are several use cases for such a scenario:

- Cluster with only one Storage Domain
- Cluster with one Storage Domain, different data policy
- Cluster with more than one Storage Domain, each with a registered External Target
- Cluster with more than one Storage Domain, some with and some without an External Target

Scenario 1: Cluster with One Storage Domain

The simplest scenario is Cloud Tier enabled on a Cohesity cluster with a single Storage Domain that has no Physical Quota.

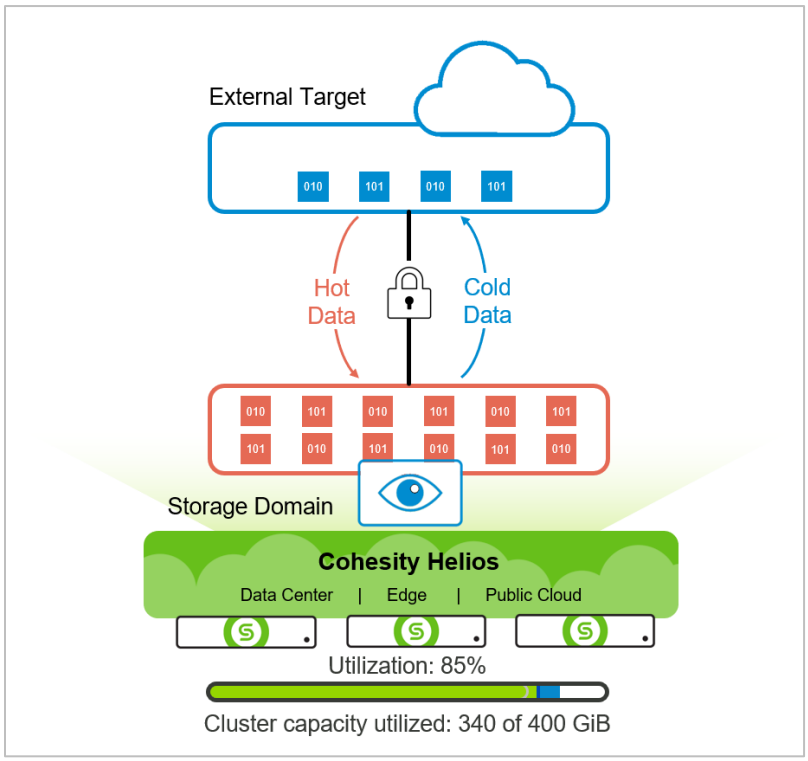
In our example, the Cohesity cluster has:

- A total capacity of 400 GiB
- One Storage Domain with External Target registered
- No Physical Quota set on the Storage Domain
- Total cluster utilization of 85%, or 340 GiB
- Cloud Tier threshold of 80%
- Data policy of 2 months

Table 6: Scenario 1 Parameters: Cluster with One Storage Domain, No Quota

CLUSTER STORAGE CAPACITY	STORAGE DOMAINS	STORAGE DOMAIN QUOTA	THRESHOLD		DATA POLICY	EXTERNAL TARGET SELECTED	TIERED DATA
			CLUSTER	STORAGE DOMAIN			
400 GiB	1	Not enabled	80%	Inherited	2 months	Yes	20 GiB

Figure 4: Scenario 1: Cluster-level Cloud Tiering



Scenario 1 Outcome:

When the background process finds that the tier threshold is exceeded, it tiers any cold data that meets the data policy, starting with the oldest data, until the cluster utilization returns to the threshold or runs out of eligible data. If there are enough eligible data blocks, tiering continues until the total utilized capacity of the cluster is brought back to the cluster threshold (80%) by down-tiering 20 GiB of cold data to the External Target.

NOTE: To determine how much space will be consumed in the External Target for the down-tiered data, see Impact of Fault Tolerance on Down-Tiered Data Consumption above.

Scenario 2: Cluster with One Storage Domain, Different Data Policy

Our next scenario is very similar — Cloud Tier enabled on a Cohesity cluster with a single Storage Domain that has no Physical Quota but has a different data policy at the Storage Domain level.

In our example, the Cohesity cluster has:

- A total capacity of 400 GiB
- One Storage Domain with External Target registered
- No Physical Quota set on the Storage Domain

- Total cluster utilization of 85%, or 340 GiB
- Cloud Tier threshold of 80%
- Data policy:
  - Cluster: 2 months
  - Storage Domain: 3 months

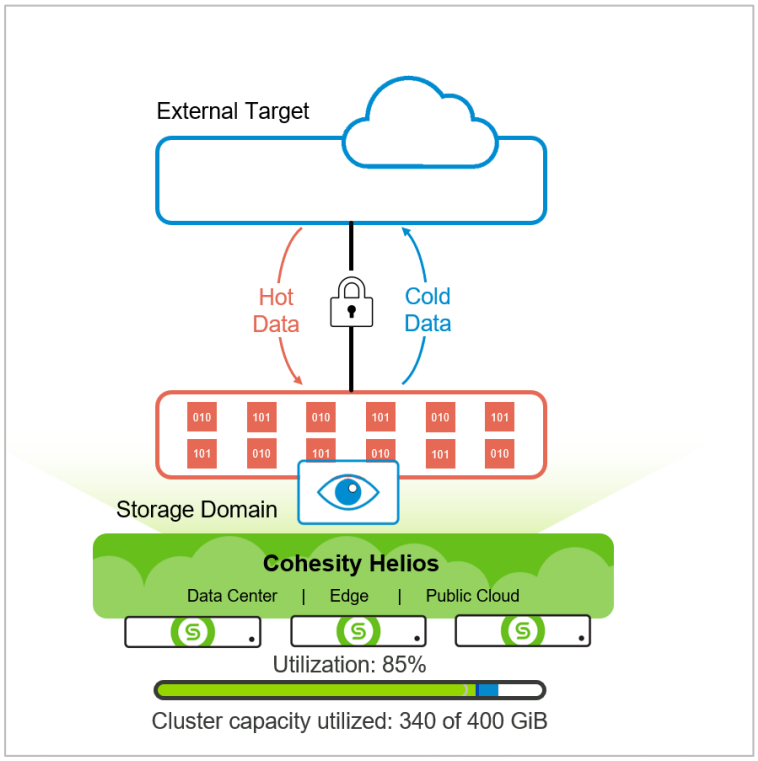
Table 7: Scenario 2 Parameters: Cluster with One Storage Domain, Different Data Policy

CLUSTER STORAGE CAPACITY	STORAGE DOMAINS	STORAGE DOMAIN QUOTA	THRESHOLD		DATA POLICY		EXTERNAL TARGET SELECTED	TIERED DATA
			CLUSTER	SD	CLUSTER	SD		
400 GiB	1	Not enabled	80%	Inherited	2	3	Yes	20 GiB

Table 8: Scenario 2 Data: Utilized and Eligible Data in Storage Domain

NAME	UTILIZED CAPACITY (GiB)	ELIGIBLE COLD DATA (GiB)	AGE OF DATA (DAYS)
SD1	340	0	70

Figure 5: Scenario 2: Cluster-level Cloud Tiering, Different Data Policy



Scenario 2 Outcomes:

The background process finds no data that exceeds the Storage Domain's data policy of 3 months, and the Storage Domain Cloud Tier settings supersede the cluster Cloud Tier settings. Even though the Cloud Tier threshold and cluster data policy (2 months) are exceeded, as the data in the Storage Domain is only 70 days old and therefore not eligible per the *Storage Domain's data policy* (3 months). As a result, no data is down-tiered in this scenario.

NOTE: To determine how much space will be consumed in the External Target for the down-tiered data, see Impact of Fault Tolerance on Down-Tiered Data Consumption above.

Scenario 3: Cluster with Several Storage Domains, Each with External Target

Our next scenario has a cluster with three Storage Domains with no Physical Quota set, and each Storage Domain has an External Target selected.

In our example, the Cohesity cluster has:

- A total capacity of 400 GiB
- 3 Storage Domains with External Targets selected
- No Physical Quota on the Storage Domains
- Total cluster utilization of 85%, or 340 GiB
- Cloud Tier threshold of 80%
- Data policy of 2 months

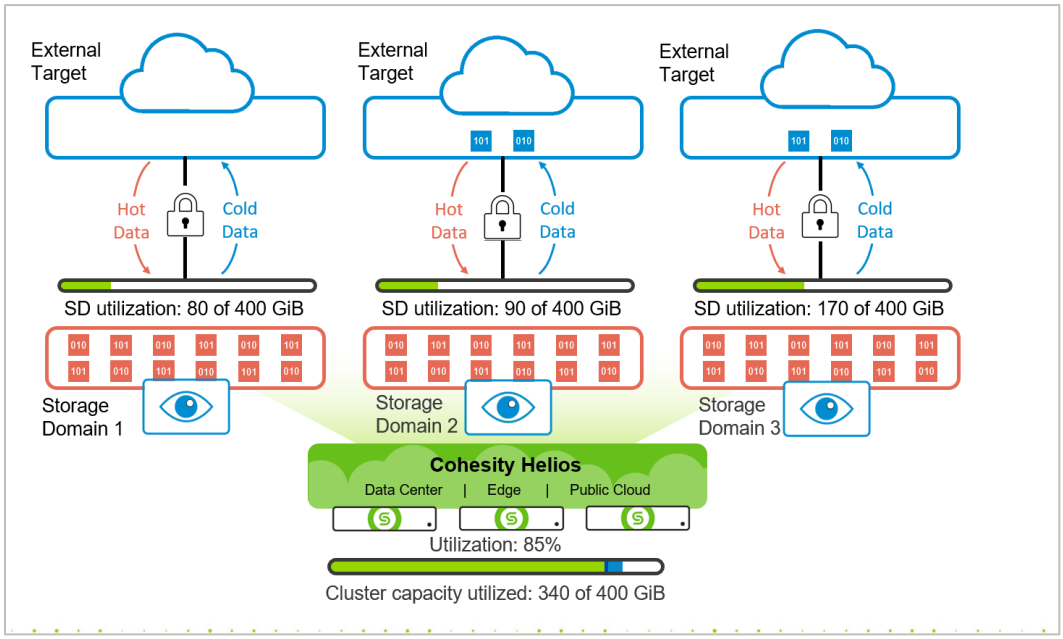
Table 9: Scenario 3 Parameters: Cluster with 3 SDs, No Quota

CLUSTER STORAGE CAPACITY	STORAGE DOMAINS	STORAGE DOMAIN QUOTA	THRESHOLD		DATA POLICY	EXTERNAL TARGET SELECTED	TIERED DATA
			CLUSTER	STORAGE DOMAIN			
400 GiB	3	Not enabled	80%	Inherited	2 months	Yes, one for each SD	20 GiB

Table 10: Scenario 3 Data: Utilized and Eligible Data in Each Storage Domain

NAME	UTILIZED CAPACITY (GiB)	ELIGIBLE COLD DATA (GiB)	AGE OF DATA (DAYS)
SD1	80	10	70
SD2	90	10	80
SD3	170	10	90

Figure 6: Scenario 3: Cluster with 3 SDs, No Quotas



Like the previous scenario, the tiering process runs and locates the data that is eligible to be tiered. Because none of the three Storage Domains have a Physical Quota set, they inherit the cluster Cloud Tier threshold, and the tiering process must determine which data to move first. Tiering is agnostic of the Storage Domains from which it moves data, and moves the oldest data blocks first, regardless of the Storage Domain.

Scenario 3 Outcomes:

Of the 30 GiB of eligible data, only 20 GiB of eligible data needs to be tiered to return to the threshold. Because of their varying ages, the first data to be down-tiered is 10 GiB from SD3. The next oldest data is 10 GiB from SD2. Finally, having moved 20 GiB, the process leaves the 10 GiB of the eligible data on SD1 because the tiering threshold is no longer exceeded.

NOTE: To determine how much space will be consumed in the External Target for the down-tiered data, see Impact of Fault Tolerance on Down-Tiered Data Consumption above.

Scenario 4: Cluster with Several Storage Domains, Not All with External Target

Our next scenario has a cluster with three Storage Domains with no Physical Quota set, but not every Storage Domain has a registered External Target.

In our example, the Cohesity cluster has:

- A total capacity of 400 GiB
- 3 Storage Domains

- SD1 and SD3 have External Targets, but SD2 does not
- No Physical Quota on the Storage Domains
- Total cluster utilization of 85%, or 340 GiB
- Cloud Tier threshold of 80%
- Data policy of 2 months

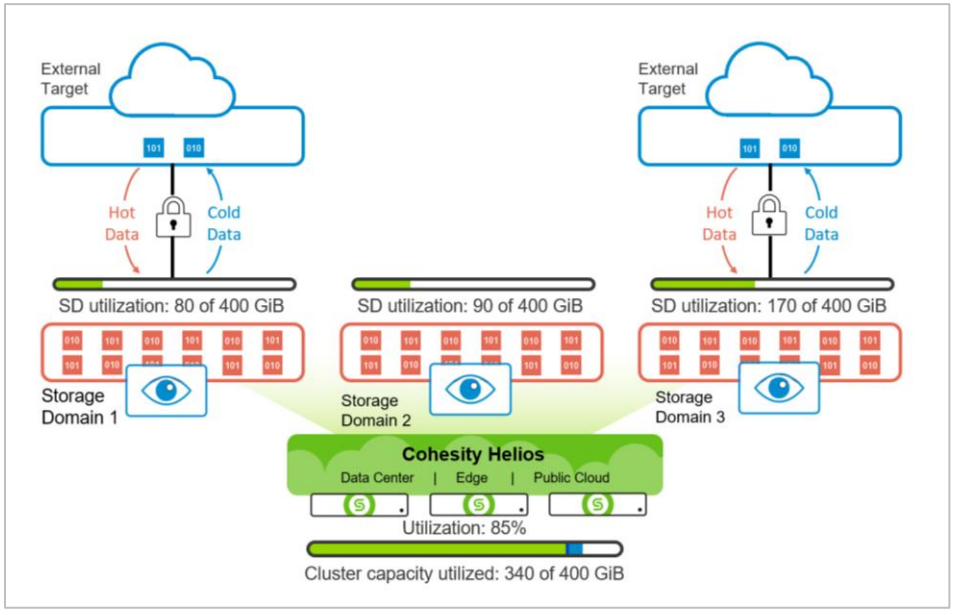
Table 11: Scenario 4 Parameters: Cluster with 3 SDs, No Quota, Some ETs

CLUSTER STORAGE CAPACITY	STORAGE DOMAINS	STORAGE DOMAIN QUOTA	THRESHOLD		DATA POLICY	EXTERNAL TARGET SELECTED	TIERED DATA
			CLUSTER	STORAGE DOMAIN			
400 GiB	3	Not enabled	80%	Inherited	2 months	Yes, except for SD2	20 GiB

Table 12: Scenario 4 Data: Utilized and Eligible Data in Each Storage Domain

NAME	UTILIZED CAPACITY (GiB)	ELIGIBLE COLD DATA (GiB)	AGE OF DATA (DAYS)
SD1	80	10	70
SD2	90	10	80
SD3	170	10	90

Figure 7: Scenario 4: Cluster with 3 SDs, No Quota, Some ETs



Like the previous scenario, the tiering process runs and locates the data that is eligible to be tiered. Even though there is no Physical Quota set and all three Storage Domains inherit the cluster Cloud Tier threshold, only two of the three have External Targets selected, and the eligible cold data from SD2 cannot be tiered.

The tiering process must determine which data to move first between the two Storage Domains that have External Targets. Tiering is agnostic of the Storage Domains from which it moves data, and moves the oldest data blocks first, regardless of the Storage Domain.

Scenario 4 Outcomes:

Of the 30 GiB of eligible data, only 20 GiB of eligible data needs to be tiered to return to the threshold. Because of their varying ages, the first data to be down-tiered is 10 GiB from SD3. Even though the next oldest data is 10 GiB from SD2, it cannot be down-tiered because SD2 does not have an External Target selected. Finally, having moved only 10 GiB, the process moves 10 GiB of the eligible data from SD1, even though it's younger than the cold data on SD2.

NOTE: To determine how much space will be consumed in the External Target for the down-tiered data, see Impact of Fault Tolerance on Down-Tiered Data Consumption above.

Storage Domains with Physical Quota

You can configure a custom Cloud Tier threshold and data policy (age at which data is considered 'cold') for individual Storage Domains. When they differ from the cluster, the Storage Domain settings take precedence. However, you can only configure the Cloud Tier threshold in a Storage Domain if that Storage Domain also has a Physical Quota set.

There are several use cases for such a scenario:

- Cluster with only one Storage Domain with Quota
- Cluster with more than one Storage Domain with Quotas, each with a registered External Target
- Cluster with more than one Storage Domain, only some with Quotas, only some with registered External Target

Scenario 5: Cluster with One Storage Domain with Quota

Consider a Cohesity cluster that has one Storage Domain with one External Target registered. The Storage Domain Quota is set to 200 GiB and the Storage Domain threshold is set to 90% (180 GiB).

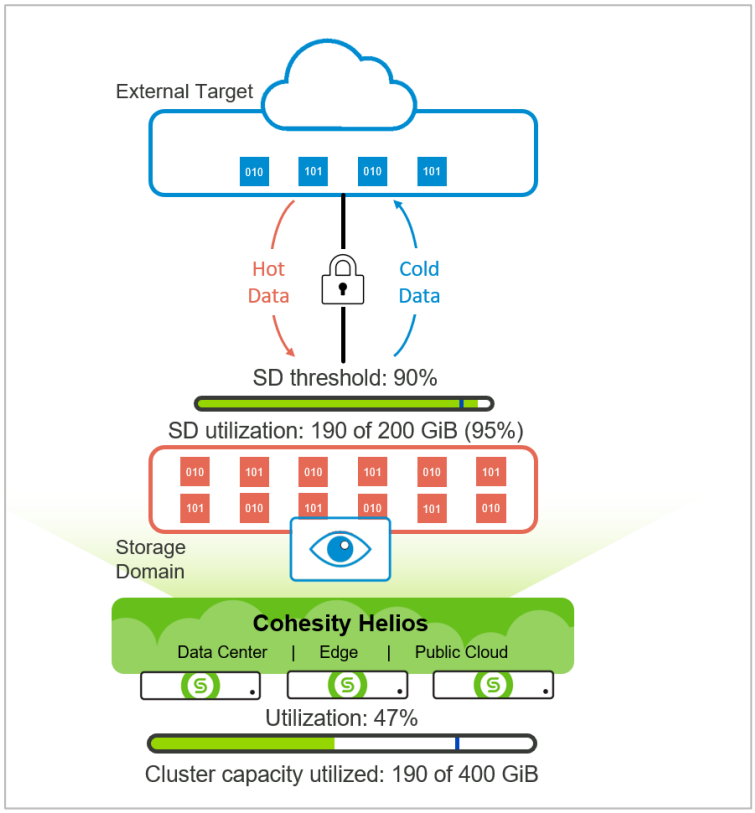
In our example, the Cohesity cluster has:

- A total capacity of 400 GiB
- One Storage Domain with registered External Target
- Storage Domain with a Physical Quota of 200 GiB
- Total cluster utilization of 47%, or 190 GiB
- Cloud Tier threshold:
  - Cluster: 80% of 400 GiB, or 340 GiB
  - Storage Domain: 90% of 200 GiB, or 180 GiB
- Data policy of 2 months

Table 13: Scenario 5 Parameters: Cluster with One SD, with Quota

CLUSTER STORAGE CAPACITY	STORAGE DOMAINS	STORAGE DOMAIN QUOTA	THRESHOLD		DATA POLICY	EXTERNAL TARGET SELECTED	TIERED DATA
			CLUSTER	STORAGE DOMAIN			
400 GiB	3	200 GiB	80%	90%	2 months	Yes	20 GiB

Figure 8: Scenario 5: Single Storage Domain with Physical Quota



Scenario 5 Outcomes:

- Because the Storage Domain has a Physical Quota, its Cloud Tier threshold (90%) takes precedence over the cluster threshold (80%).
- When the background process finds that the Storage Domain tier threshold is exceeded, it tiers any cold data that meets the data policy, starting with the oldest data, until the Storage Domain utilization returns to the threshold or runs out of eligible data. If there are enough eligible data blocks, tiering

continues until the total utilized capacity of the Storage Domain is brought back to its threshold (90%) by down-tiering 20 GiB of cold data to the External Target.

NOTE: To determine how much space will be consumed in the External Target for the down-tiered data, see Impact of Fault Tolerance on Down-Tiered Data Consumption above.

Scenario 6: Multiple Storage Domains with Quotas

Our next scenario has a cluster with three Storage Domains with Physical Quotas, and each Storage Domain has an External Target selected.

In our example, the Cohesity cluster has:

- A total capacity of 400 GiB
- 3 Storage Domains with registered External Targets
- Physical Quotas on the Storage Domains
- Total cluster utilization of 85%, or 340 GiB
- Cloud Tier threshold:
  - Cluster: 80% of 400 GiB, or 320 GiB
  - Storage Domain 1: 90% of 100 GiB, or 90 GiB
  - Storage Domain 2: 70% of 100 GiB, or 70 GiB
  - Storage Domain 3: 60% of 200 GiB, or 120 GiB
- Data policy of 2 months

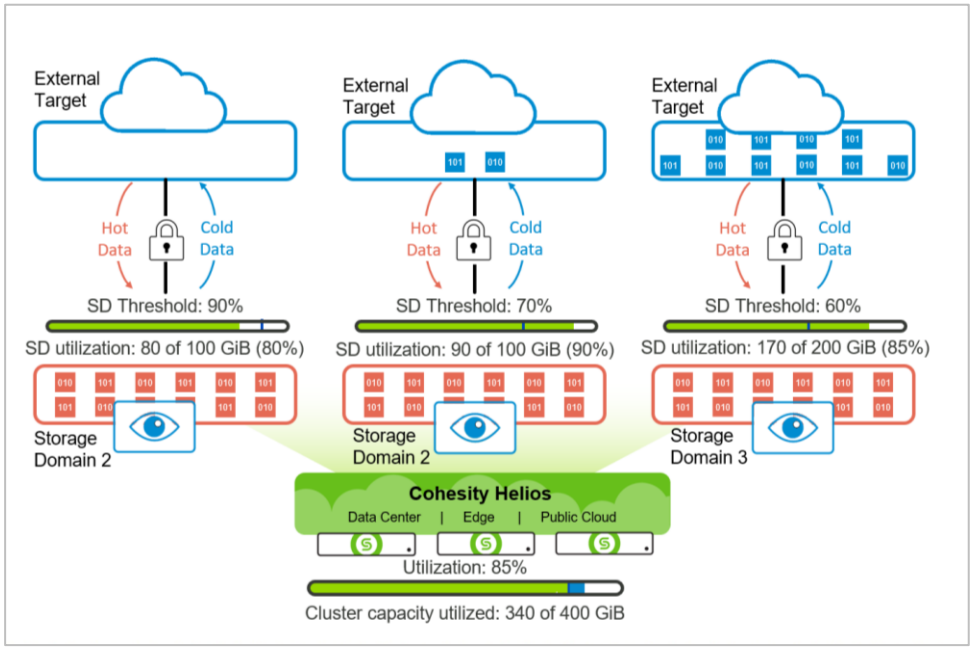
Table 14: Scenario 6 Parameters: Cluster with 3 SDs, with Quotas

CLUSTER STORAGE CAPACITY	STORAGE DOMAINS	STORAGE DOMAIN QUOTA (GiB)	THRESHOLD		DATA POLICY	EXTERNAL TARGET SELECTED	TIERED DATA
			CLUSTER	STORAGE DOMAIN			
400 GiB	3	SD1 – 100 SD2 – 100 SD3 – 200	80%	SD1– 90% SD2– 70% SD3– 60%	2 months	Yes, one for each SD	60 GiB

Table 15: Scenario 6 Data: Utilized and Eligible Data in Each Storage Domain

NAME	UTILIZED CAPACITY (GiB)	STORAGE DOMAIN THRESHOLD	ELIGIBLE COLD DATA (GiB)	AGE OF DATA (DAYS)
SD1	80	90% of 100 GiB	10	70
SD2	90	70% of 100 GiB	10	80
SD3	170	60% of 200 GiB	75	61

Figure 9: Scenario 6: Cluster with 3 SDs, with Quotas



Like the previous scenario, the tiering process runs and locates the data that is eligible to be tiered. When the tiering process starts, only SD2 and SD3 have exceeded their Cloud Tier thresholds.

- No data is down-tiered from SD1, as it has not met its threshold.
- Per its threshold, SD2 needs to down-tier 20 GiB of data but there is only 10 GiB of data that is eligible per the data policy. 10 GiB is down-tiered from SD2, even though the threshold remains exceeded.
- Per its threshold, SD3 needs to down-tier 50 GiB of data. Even though there is 75 GiB of data that is eligible per the data policy, only 50 GiB is down-tiered from SD3, as that returns it to its threshold.

Scenario 6 Outcomes:

- When a Storage Domain has a Physical Quota set, that takes precedence, and data is down-tiered according to each Storage Domain's threshold and data policy.
- While the comparative age of the data determines the order of data blocks to move when Storage Domains have no Physical Quota set when Storage Domains do have a Physical Quota, Cloud Tier operates independently on each, and the comparative age of the data is not a factor.
- Tiering only returns utilization to the Cloud Tier threshold when there are enough eligible cold blocks.

> **NOTE**: To determine how much space will be consumed in the External Target for the down-tiered data, see Impact of Fault Tolerance on Down-Tiered Data Consumption above.

Scenario 7: Multiple Storage Domains with Mixed Quotas

Our next scenario has a cluster with three Storage Domains, one of which has a Physical Quota set and another does not, and one of the three Storage Domains has no External Target.

In our example, the Cohesity cluster has:

- A total capacity of 400 GiB
- 3 Storage Domains, only 2 with registered External Targets
- Physical Quota on only 1 of the Storage Domains
- Total cluster utilization of 85%, or 340 GiB
- Cloud Tier threshold:
  - Cluster: 80% of 400 GiB, or 320 GiB
  - Storage Domain 1: 90% of 100 GiB, or 90 GiB
  - Storage Domain 2: No External Target, no threshold
  - Storage Domain 3: Inherited, 80% of 400 GiB, or 320 GiB
- Data policy of 2 months

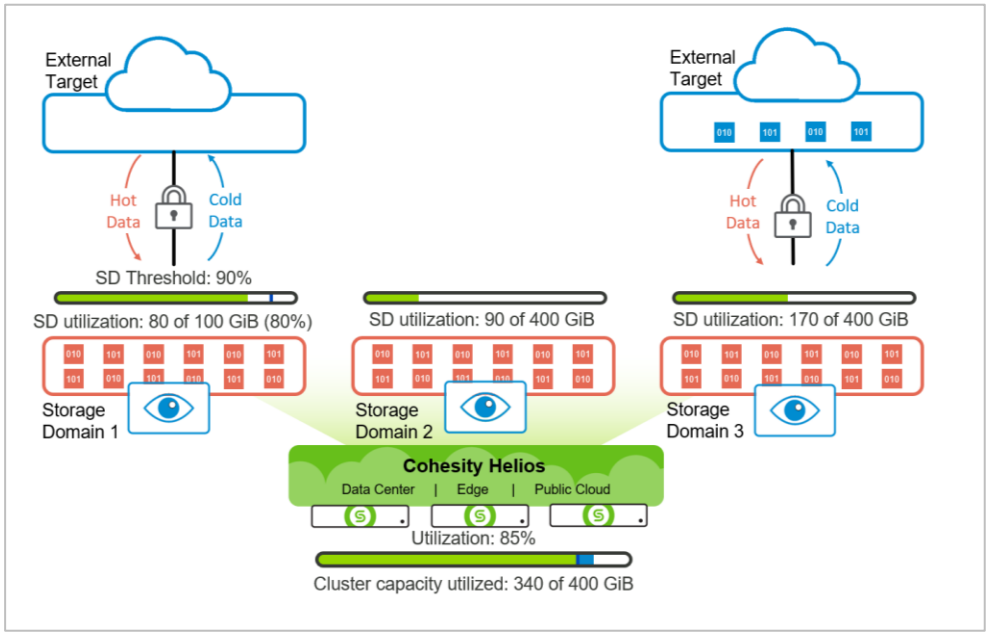
Table 16: Scenario 7 Parameters: Cluster with 3 SDs, Mixed Quotas, Some ETs

CLUSTER STORAGE CAPACITY	STORAGE DOMAINS	STORAGE DOMAIN QUOTA (GiB)	THRESHOLD		DATA POLICY	EXTERNAL TARGET SELECTED	TIERED DATA
			CLUSTER	STORAGE DOMAIN			
400 GiB	3	SD1 – 100 SD2 – Not set SD3 – Not set	80%	SD1 – 90% SD2 – n/a SD3 – Inherited	2 months	Yes, except for SD2	20 GiB

Table 17: Scenario 7 Data: Utilized and Eligible Data in Each Storage Domain

NAME	UTILIZED CAPACITY (GiB)	STORAGE DOMAIN THRESHOLD	ELIGIBLE COLD DATA (GiB)	AGE OF DATA (DAYS)
SD1	80	90% of 100 GiB	10	70
SD2	90	n/a	10	80
SD3	170	Inherited - 80% of 400 GiB	65	90

Figure 10: Scenario 7: Cluster with 3 SDs, 1 with Quota, 1 without External Target



The tiering process runs and locates the data that is eligible to be tiered. When the tiering process starts, only SD1 and SD3 are evaluated, as SD2 has no registered External Target.

- No data will be down-tiered from SD1, as it has not met its threshold.
- Because SD3 has no Physical Quota set and inherits the cluster threshold, 80%, its utilization is calculated based on the total cluster utilization.
- As total utilization on the cluster is 340 GiB and the cluster threshold is 80%, SD3 needs to down-tier 20 GiB of data. Even though it has 65 GiB of eligible cold data, SD3 down-tiers only 20 GiB, as that returns the cluster to its threshold.

Scenario 7 Outcomes:

- When a Storage Domain has a Physical Quota set, that takes precedence, and data is down-tiered according to that Storage Domain's threshold and data policy.
- When a Storage Domain has no Physical Quota set, the utilization and Cloud Tier threshold of the cluster is used to determine whether data needs to be down-tiered.
- Tiering only returns utilization to the Cloud Tier threshold when there are enough eligible cold data blocks. In fact, the cloud tiering process stops once data utilization returns to the threshold. In doing so, the Cloud Tier threshold serves both as a marker for when usage is high enough to start down-tiering data, and also as a limit for how much data is down-tiered before the process stops.

> **NOTE**: To determine how much space will be consumed in the External Target for the down-tiered data, see Impact of Fault Tolerance on Down-Tiered Data Consumption above.

Enable Cloud Tiering

By default, Cloud Tier is enabled at the cluster level in Cohesity, and the default tiering threshold is 80%. However, Cloud Tier will not activate until it is enabled in at least one Storage Domain that has a registered External Target. When you configure the Cloud Tier threshold for the cluster, that value becomes the default threshold for all Storage Domains where Cloud Tier is enabled.

You can also set the Cloud Tier threshold for individual Storage Domains. Storage Domain thresholds take precedence over the cluster threshold; if you set a different Cloud Tier threshold for a Storage Domain, it overrides the cluster threshold for that Storage Domain.

NOTE: You cannot override the cluster threshold in a Storage Domain unless you have also set a Physical Quota on that Storage Domain. See Enable Cloud Tier on a Storage Domain below for details.

Cloud Tier Workflow

To enable cloud tiering, you need to enable it in a Storage Domain with a registered External Target. When you enable Cloud Tier in a Storage Domain, you must select the External Target where eligible cold data will be down-tiered. There you can also change the default data policy (that is, the age at which data becomes eligible for tiering) of 2 months. If you set a Physical Quota on that Storage Domain, you can also change the Cloud Tier threshold, which is 80% by default.

- To get started, enable Cloud Tier on a Storage Domain.
- To set the default Cloud Tier threshold across the cluster for any Storage Domains that do not override it, you can configure the cluster-level threshold in your cluster settings.

Enable Cloud Tier on a Storage Domain

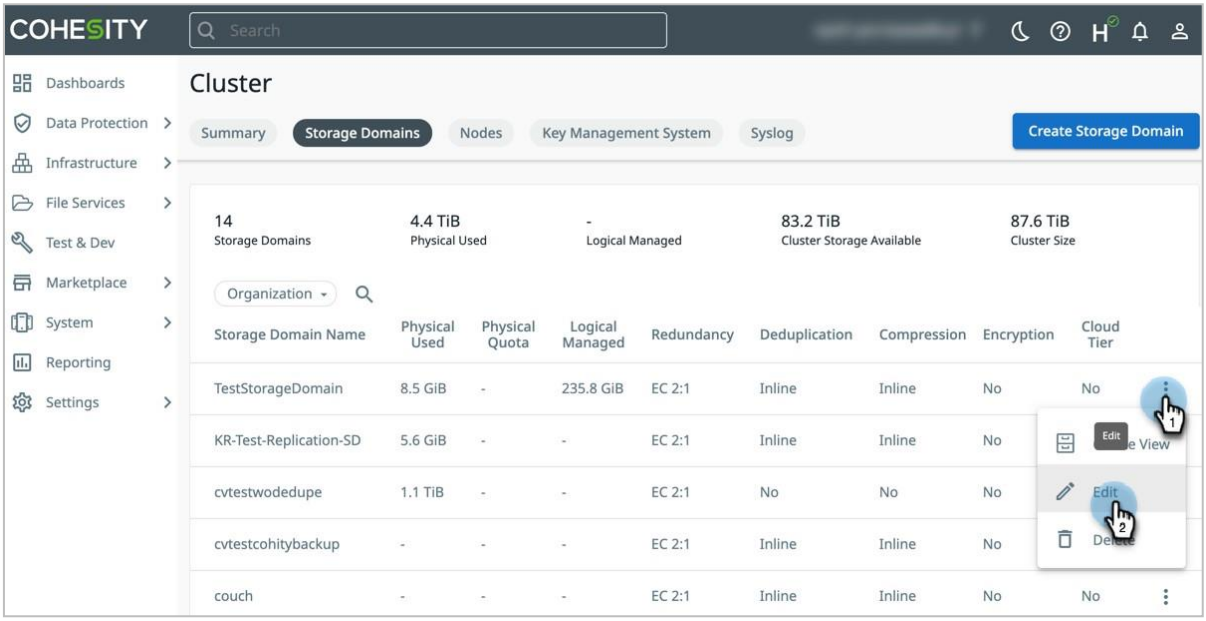
The first thing to do is to enable Cloud Tier on at least one Storage Domain that has a registered External Target. Without that, there is no cloud tiering. When you enable Cloud Tier for a Storage Domain, you can customize the Cloud Tier threshold and data policy for that Storage Domain, or leave them to inherit the cluster defaults.

To enable and configure Cloud Tier on a Storage Domain:

1. Log in to Cohesity and select **Settings**> **Summary**.

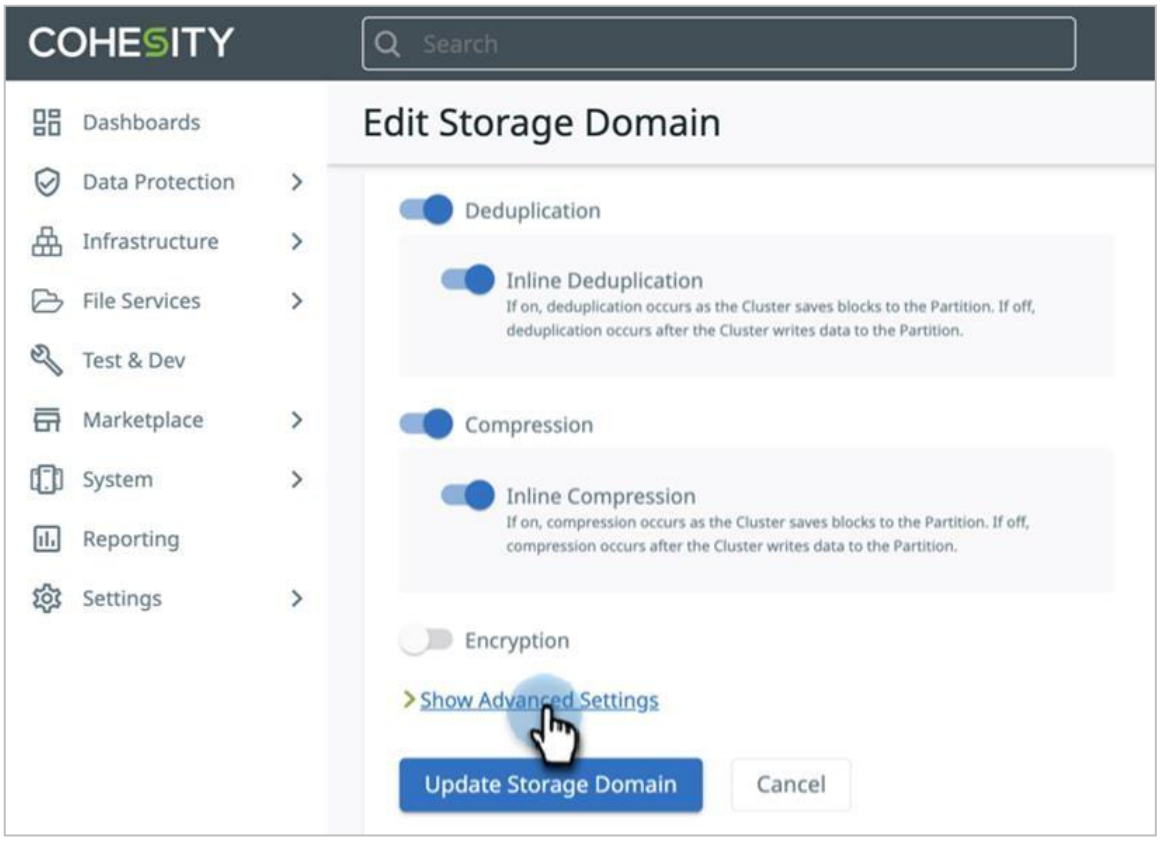
2. Click the **Storage Domains** tab.

3. In the row for the Storage Domain, click the **Actions** (⋮) menu and select **Edit**.



TIP: To create a new Storage Domain, click Add Storage Domain in the top-right of the page. For details, see Create or Edit Storage Domains in the online Help.

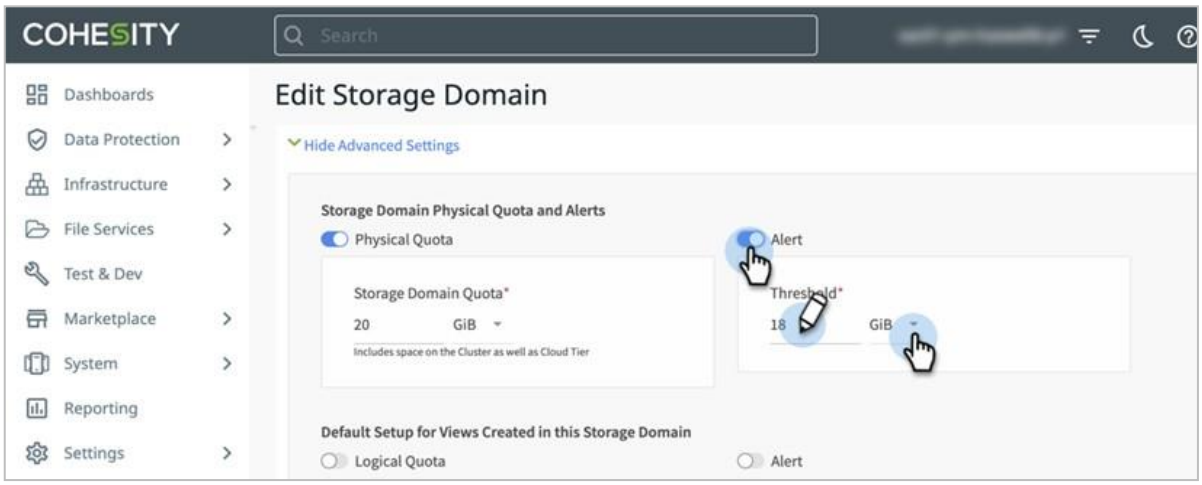
4. Scroll down and click **Show Advanced Settings**.



5. Enable **Physical Quota** and set the storage capacity for this Storage Domain. The capacity must be a whole number.

NOTE: This is optional, but you need to enable the Physical Quota if you plan to override the cluster Cloud Tier threshold for this Storage Domain.

6. Optionally, enable **Alert** and specify a **Threshold** to trigger an Alert when the Storage Domain reaches that threshold.

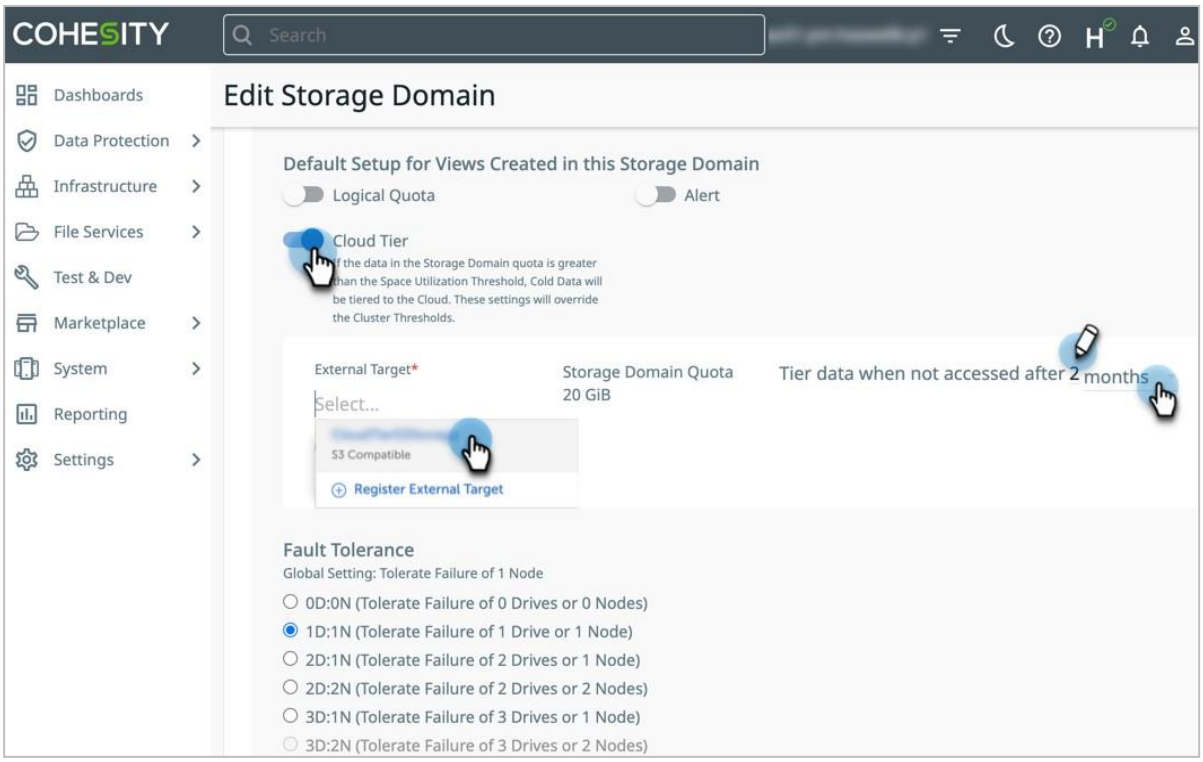


7. Enable **Cloud Tier**. Select the **External Target** where data will be down-tiered, and set the period for **Tier data when not accessed after** (the data policy, or 'coldness threshold'). By default, the data policy is 2 months, but you can adjust it to meet your specific business needs.

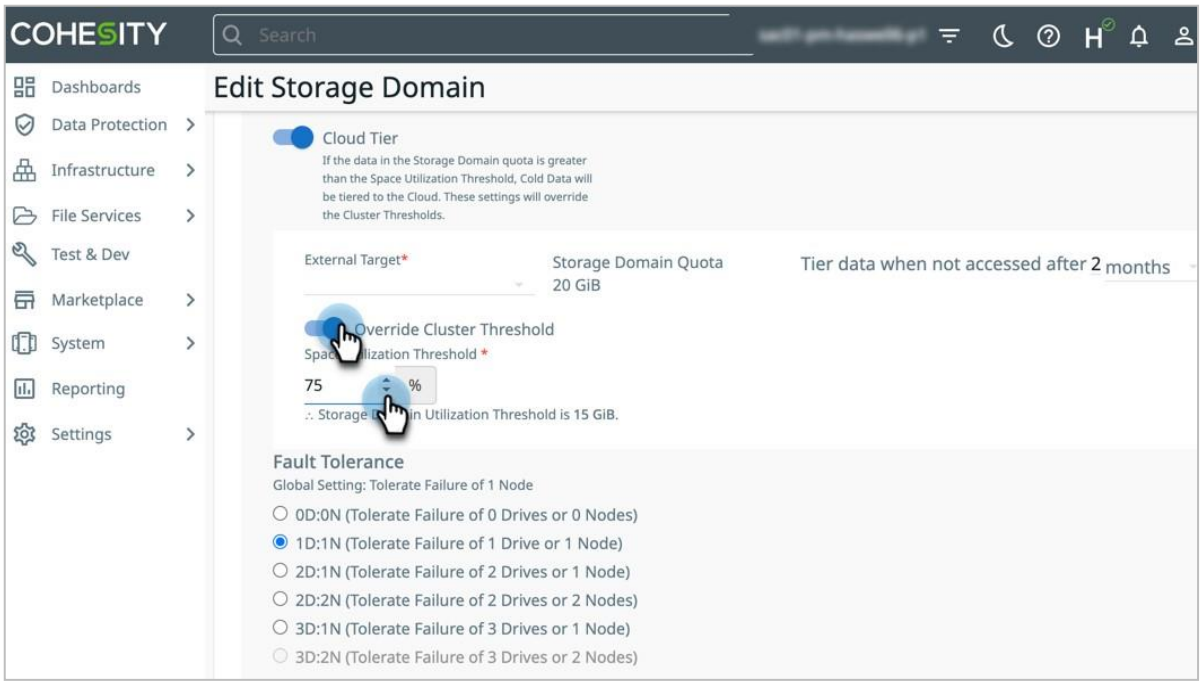
NOTES:

- When registering an External Target to use with Cloud Tier, for **Purpose**, select **Tiering** instead of the default, **Archival**.
- To register a new External Target, click **Register External Target** in the drop-down menu. For details, see Register an External Target in the online Help.

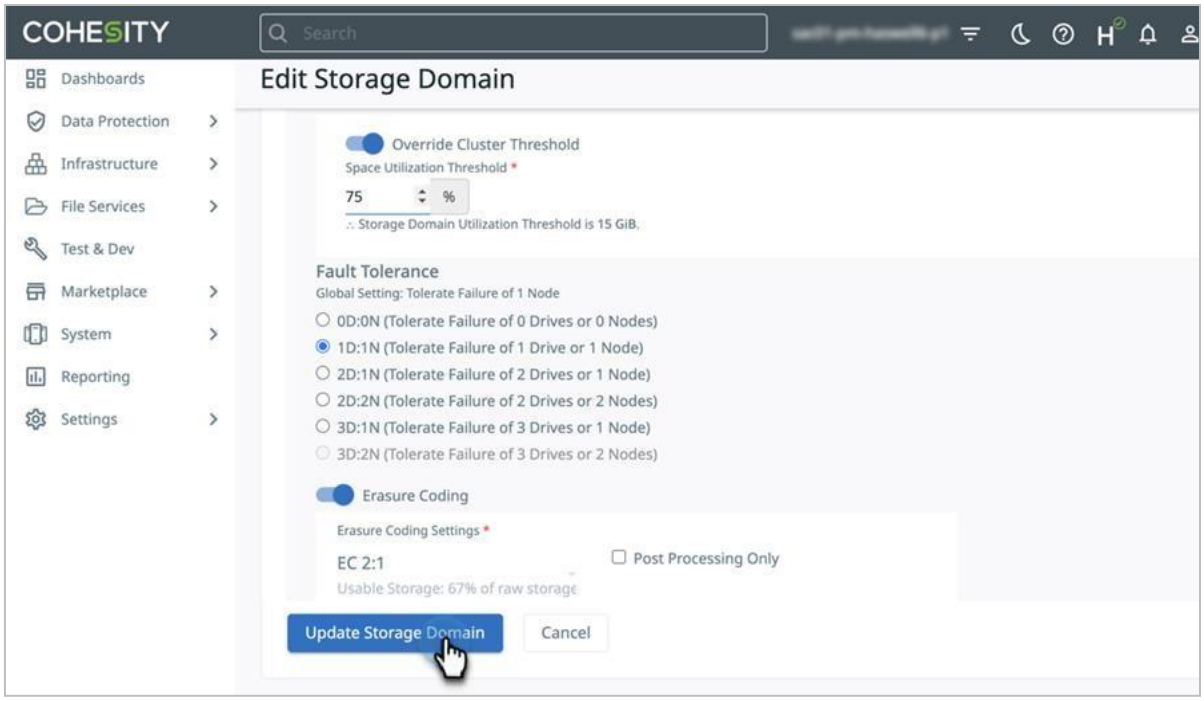
NOTE: The same External Target can be selected for each Storage Domain where you enable CloudTier.



8. Optionally, enable **Override Cluster Threshold** (80% by default) to choose a different tiering threshold for the Storage Domain.



9. Click **Update Storage Domain**.



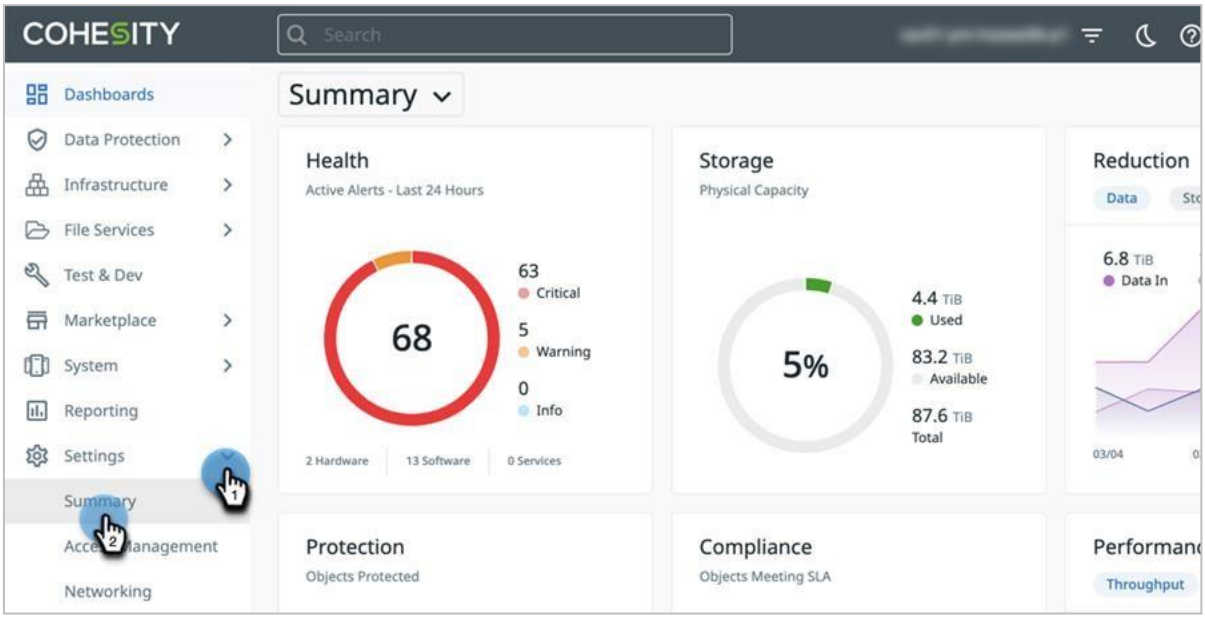
Repeat these steps for all Storage Domains where you want to enable Cloud Tier.

Configure Cluster-Level Cloud Tier Threshold

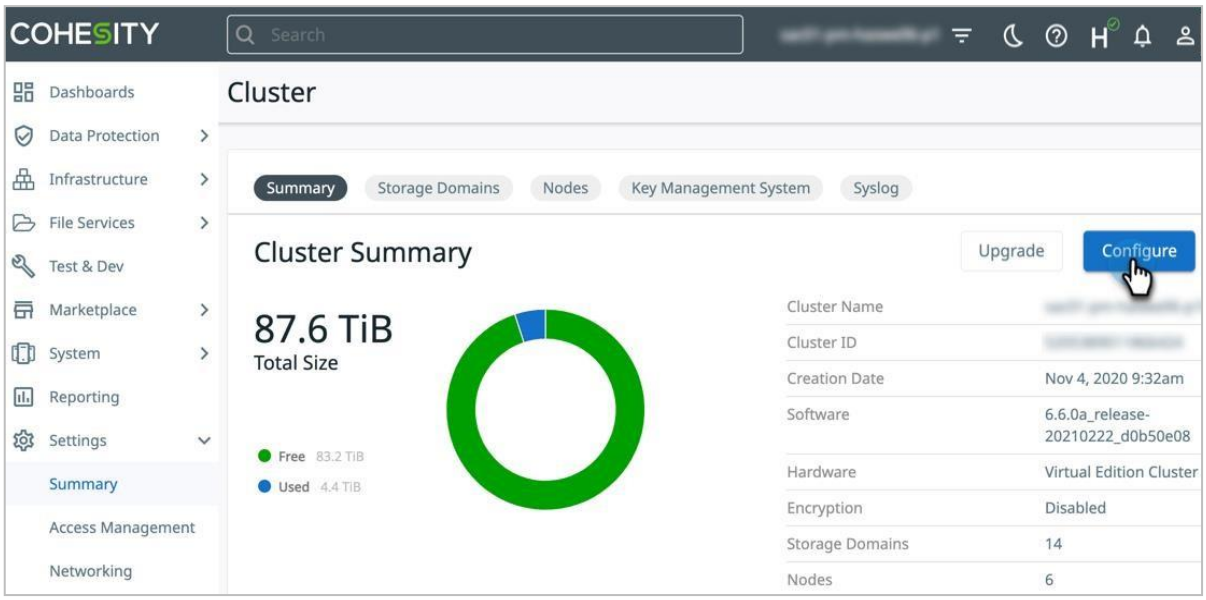
Once there is at least one Storage Domain with Cloud Tier enabled on a cluster, cloud tiering will begin once the cluster utilization threshold is exceeded, and eligible cold data is identified. However, if you do not override the cluster threshold in a Storage Domain, it will inherit the cluster-level Cloud Tier threshold, which is 80% by default. You can change this threshold to adjust it to your specific needs, and all Storage Domains that are not set to override it will inherit it.

To change the cluster Cloud Tier threshold:

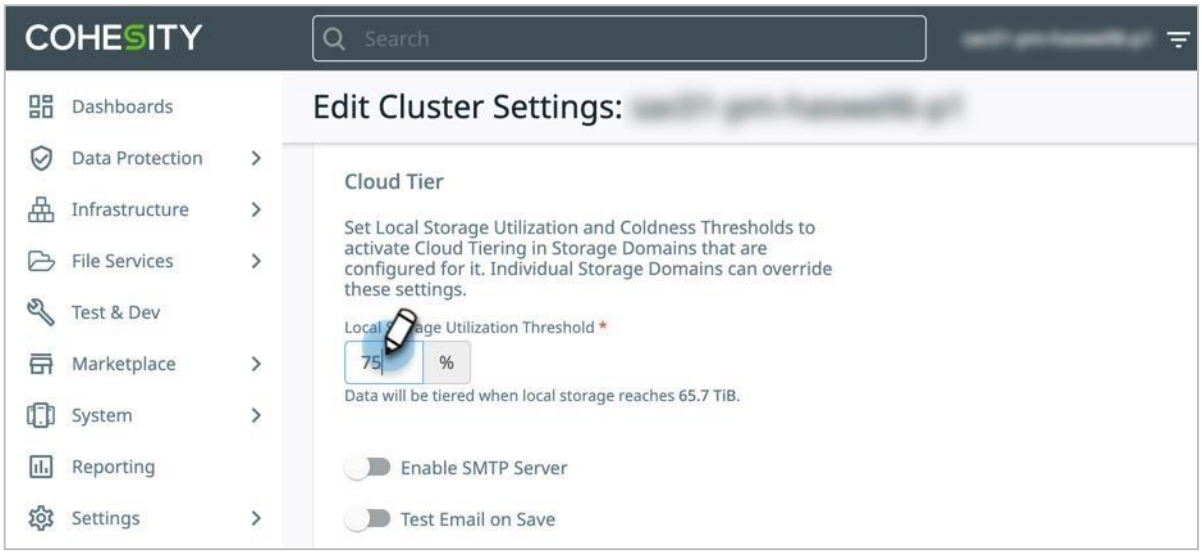
1. Log in to Cohesity and select **Settings** > **Summary**.



2. Click **Configure**.



3. Under **Cloud Tier**, in the **Local Storage Utilization Threshold** field, set the threshold value in percentage to activate Cloud Tiering in Storage Domains that are configured for it.



4. Click **Save**.

NOTE: In addition to the existing Storage Domains that are not set to override it, all future Storage Domains that you create on the same cluster will inherit this new Cloud Tier threshold unless and until you configure them to override it. See Enable Cloud Tier on a Storage Domain above.

Report on Down-Tiered Data

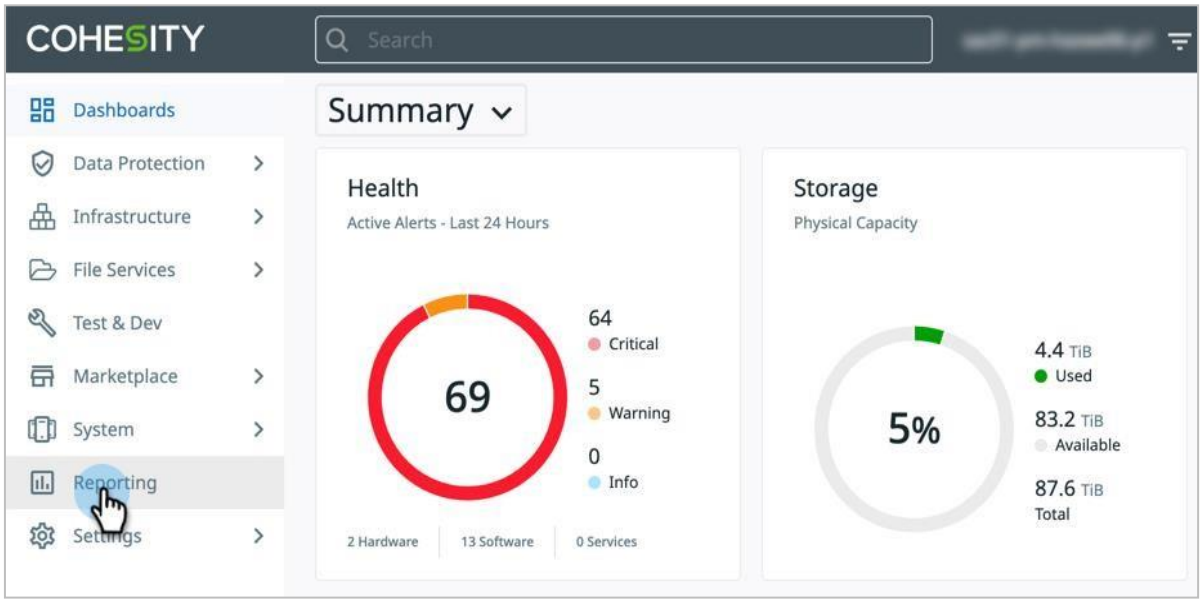
If Cloud Tier is enabled on your cluster, Cohesity gives you insight into the amount of cold data that has been down-tiered to cloud storage from the cluster, at two levels.

- At the cluster level, see the Storage Consumed by Storage Domains report to compare how much Cloud Tier-enabled Storage Domains consume with the logical data they actually manage.
- For an individual Storage Domain, see the Summary page for that Storage Domain to compare the percentage of data on the cluster with down-tiered data on the External Target.

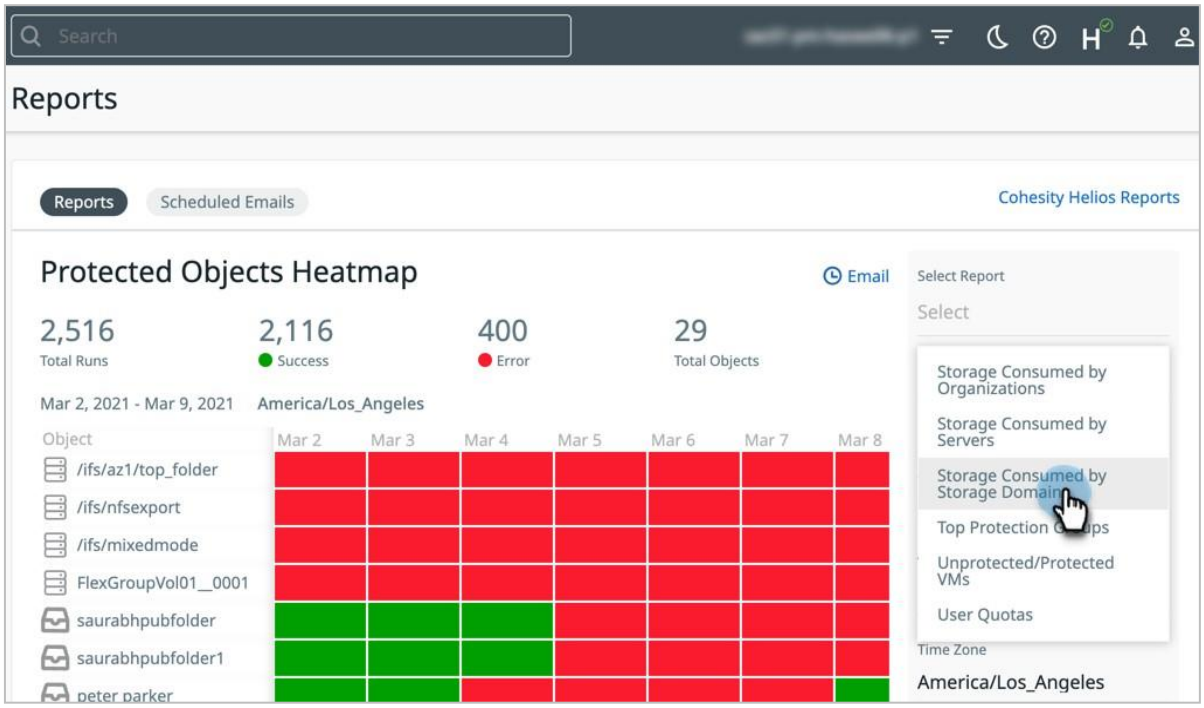
View Cluster Cloud Tier Report

To see how much of your data has been down-tiered to the cloud across the whole Cohesity cluster:

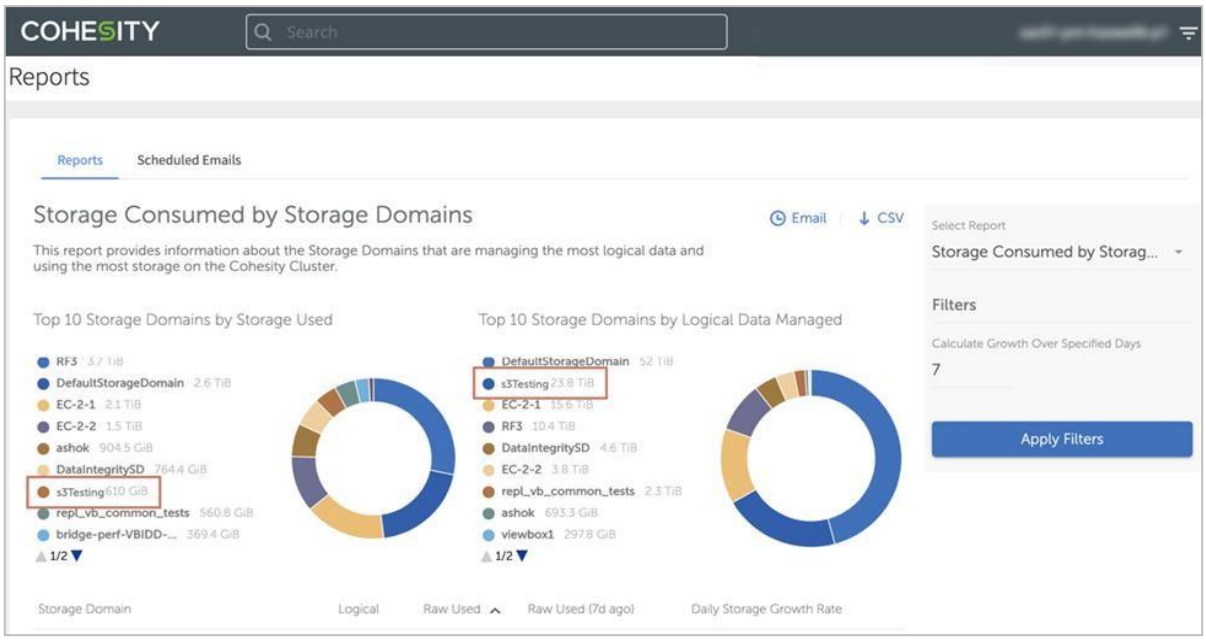
1. Log in to Cohesity and select **Reporting**.



2. From the drop-down list on the right, select the **Storage Consumed by Storage Domains** report.



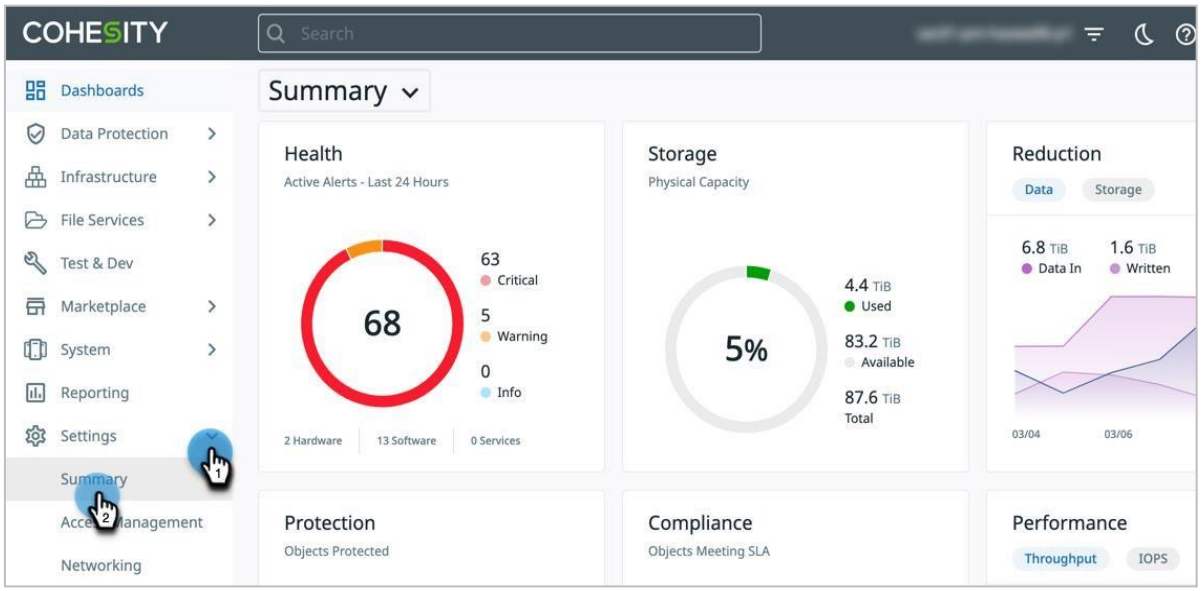
3. At the top of the report, there are summaries for the **Top 10 Storage Domains by Storage Used** and **Top 10 Storage Domains by Logical Data Managed**. Find your Cloud Tier-enabled Storage Domains to compare how much storage they consume with the logical data they actually manage.



View Storage Domain Cloud Tier Report

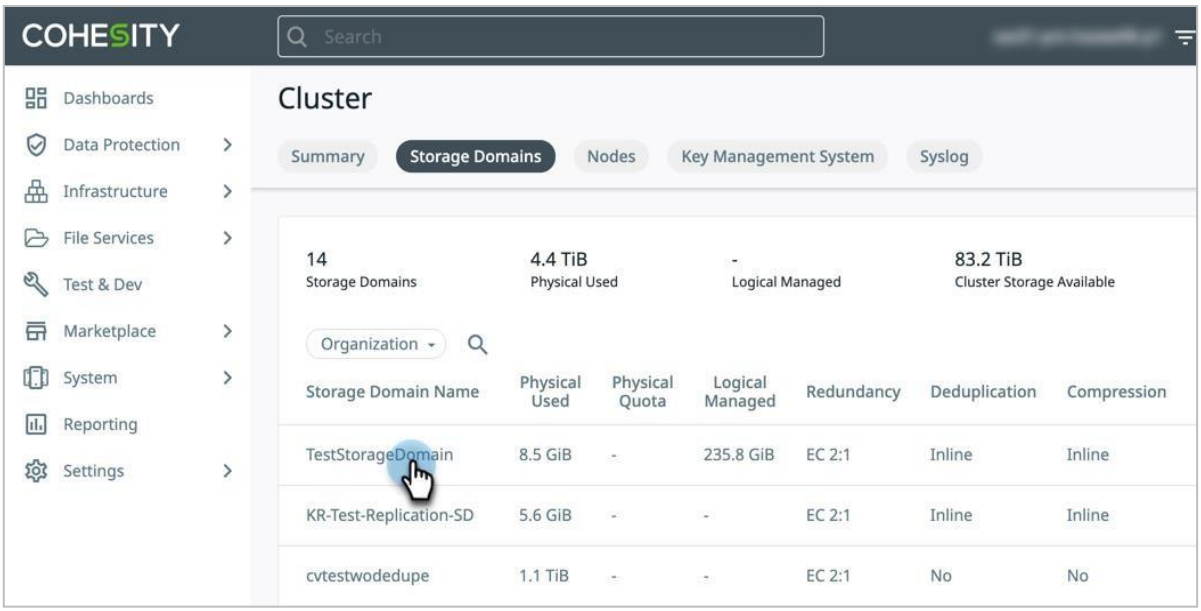
To see how much data is stored in the cloud for a specific Storage Domain:

1. Log in to Cohesity and select **Settings** > **Summary**.

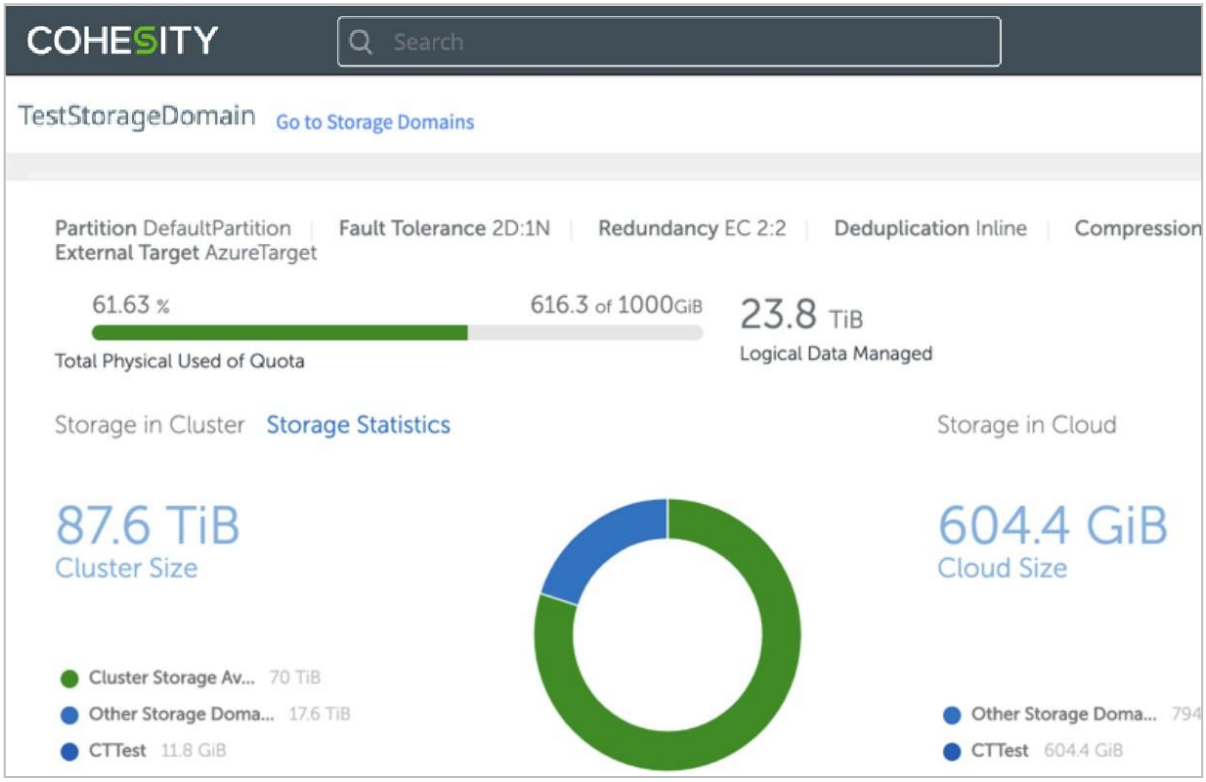


2. Click the **Storage Domains** tab.

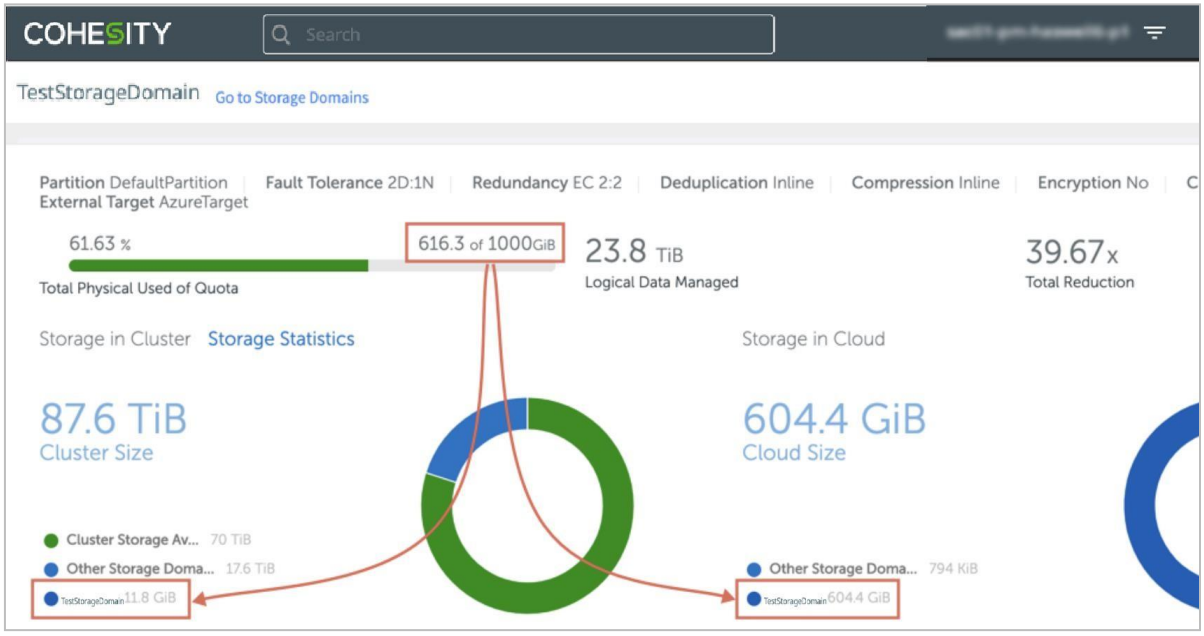
3. Select the Storage Domain you need.



4. The Storage Domain Summary page displays how much data is stored in the cluster and how much in the cloud.



Note that the **Total Physical Used of Quota** value displayed is the sum of data in the Storage Domain and the data down-tiered to the External Target.



Appendix A: Terminology

There are several terms that are especially important to understand as you learn about Cloud Tier in Cohesity.

Table 18: Cloud Tier Terminology

TERMINOLOGY	DESCRIPTION
Data Policy (Age of data)	The amount of time since the last access that defines data as 'cold.' Data that is not accessed within the defined period is eligible for tiering to the External Target. The default value is 2 months.
Cold Data	Data that has not been accessed (read or written) within the defined data policy.
Warm Data	Existing data that has been accessed or new data that was written within the defined data policy.
Hot Data	Data that was moved to the External Target and is now accessed frequently to qualify for up-tiering back to the cluster.
Physical Quota	The physical capacity set for a storage Domain.
Threshold	Percentage of capacity consumed (across the cluster or in a Storage Domain with Physical Quota) before data is down-tiered to the External Target. The default value is 80%.
Up/Down-Tiering	The process of moving data between tiers of storage within Cohesity (SSD and HDD) and in External Targets in the cloud.

Appendix B: Impact of Deduplication Ratio on Cloud Tiering

Tiering is not unlimited. Typically, tiering supports up to a 1:8 ratio, meaning that the amount of data that can be tiered to the External Target can be no more than 8x the capacity of the source cluster. The deduplication settings for the Storage Domain are applied to the External Target. When deduplication is enabled, the tiering ratio is reduced, because the metadata to manage the deduplicated data continues to grow on the Cohesity cluster.

To determine the ratio of data that can be tiered, divide 8 by the deduplication rate. For example, if the deduplication rate is 2x, divide 8 by 2, and the tiering ratio is 1:4. If the deduplication rate is 4x, the tiering ratio is 1:2.

Your Feedback

Was this document helpful? Send us your feedback!

About the Authors

Adaikkappan Arumugam is Senior Manager, Tech Marketing, Solutions Engineering, & Tech Pubs at Cohesity. In his role, Adai focuses on connecting the technical expertise of Cohesity's developer and product management staff with the needs and feedback from Cohesity's customers, support staff, and sales enablement staff.

Other essential contributors included:

- Bart Abicht, Senior Technology Writer and Editor at Cohesity
- Jay White, Principal Engineer, Technical Field Enablement
- Prathibha Mohan, Technical Writer
- Zhihuan Qiu, Technical Director, Data Path

Document Version History

VERSION	DATE	DOCUMENT HISTORY
1.2	July 2024	Republishing
1.1	Mar 2021	Minor update with key considerations
1.0	Oct 2019	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.