COHESITY

Version 1.1

February 2025

Accelerate Anomaly Detection with Cohesity

ABSTRACT

Ransomware is the biggest concern for organizations across the globe, causing data loss, financial losses, and reputational damages. As attacks are continuing to grow in scale and sophistication, organizations need to leverage the power of machine learning and artificial intelligence to quickly detect and counter cyber-attacks.

This whitepaper provides a comprehensive overview of Cohesity's innovative approach to anomaly detection within the context of data management and security. It showcases the platform's ability to detect and respond to anomalies, enabling organizations to proactively protect their data and manage security risks to ensure recovery is possible, even at the largest scale.

Table of Contents

Figures

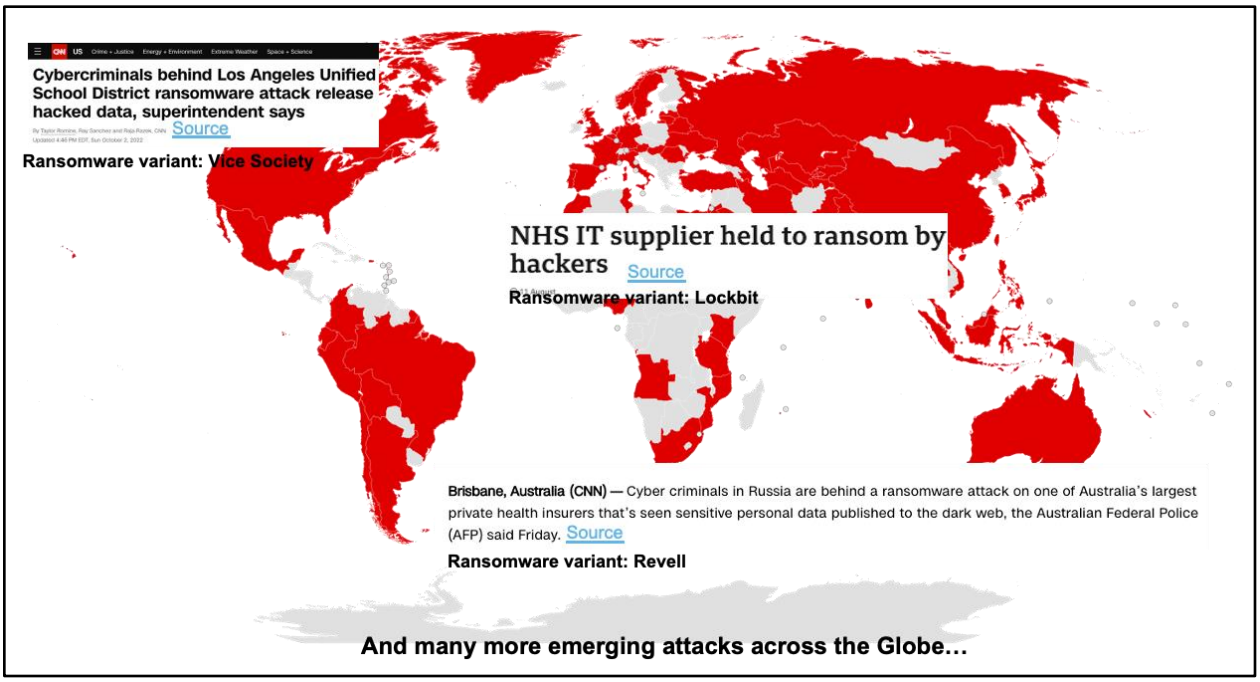
Tables

Introduction

In today's digital world, organizations face increasingly sophisticated cyber threats that can wreak havoc on their data, operations, and reputation. Among these threats, ransomware has emerged as a particularly dangerous and pervasive menace. Ransomware attacks have grown in scale, complexity, and impact, causing substantial financial losses and disruptions to businesses worldwide, often specifically targeting the organization's backup systems, which are meant to be a crucial safeguard against data loss and recovery from cyber-attacks.

According to Cybersecurity Ventures predictions, ransomware will affect a business, consumer, or device every 2 seconds by 2031. In 2021, the projected cost of damages worldwide due to successful attacks is $20 billion. However, this amount is expected to increase significantly to $265 billion annually by 2031. The estimated costs encompass a range of factors, such as downtime-related financial losses, productivity declines, and damage to reputation.

Figure 1: Rising State of Ransomware



To help address these risks and combat the impact of ransomware, organizations must adopt a multi-layered security approach. By staying informed about the latest attack techniques, organizations can better protect their systems, data, and operations from ransomware attacks.

Cohesity's best-in-class AntiRansomware solution offers to protect, isolate, detect, and rapidly recover your data to reduce downtime and ensure business continuity.

Refer to **Cohesity Data Cloud Security Hardening Best Practices Guide** and **Cohesity Ransomware Protection — Prepare and Recover Whitepaper** to learn more about the best practices to secure Cohesity Data Cloud and help customers respond and recover from today's growing ransomware attacks.

Need for Strong Anomaly Detection

As a growing trend in successful attacks and data breaches on businesses, the attackers have infiltrated critical data stores over weeks or months before eventually planting malware that encrypts the data. If the backup system is not capable of detecting and restoring all the organizational data promptly with a clean snapshot of data, then the victim must pay a ransom in hopes of receiving the decryption keys.

Guarding against the irreversible damage caused by ransomware requires timely detection of attacks and an ability to recover rapidly with a malware-free copy of data. As a result, organizations must adopt security measures to safeguard their critical data and systems.

The sooner the organization is aware that an attack has started, the faster the response plan can be initiated to minimize impact. Ideally, threats are discovered on primary systems prior to the beginning of a ransomware attack, but strong anomaly detection is a crucial last line of defense when other controls have failed.

Anomaly detection is one part of defense in depth. It's crucial to have a backup system that reliably and securely makes continuous or frequent backups, detects, and protects them from attack, and can immediately and safely put the data online at scale to support forensics and cyber recovery. Cohesity provides unique capabilities in those areas. Cohesity Data Cloud (SaaS) incorporates the leading security ecosystem technologies to provide actionable insights that help to safeguard your data from evolving threats to give security and IT teams an edge to speed up the investigation and response.

Below are the key capabilities offered from Cohesity Data Cloud (SaaS):

- Review the ransomware alerts generated using data signals collected from Cohesity snapshots, affected files, trends analysis.
- Analyze the results of threat detection scans that look for ransomware Indicators of Compromise (IOC) or lurking threats.
- Understand the user behavior by searching audit trails in near real-time.
- ML-based classification of critical data to understand which objects compromised due to ransomware contain sensitive data types and which types, PII, PHI, and so on.
- CyberScan App helps in identify vulnerabilities in backed up VMs that would let attackers in if recovered and help detect the presence of attackers before they plant ransomware.

For more details, see Security Center documentation.

The intelligence gathered from backing up the data, compressing it, deduplicating it, and indexing it provides additional signals that Cohesity uses to help flag an ongoing attack that might have gone unnoticed. The resulting anomaly detection allows for quick corrective action to lessen the damage and hasten the recovery. Integration with SIEM and SOAR platforms helps the security teams respond rapidly to attacks with automation workflows.

Factors That Govern Effective Anomaly Detection

To protect the organization's critical data from any anomalous activities, organizations must have robust detection mechanisms built with AI/ML algorithms and comprehensive investigation to respond to the anomaly alert in the quickest possible time. The detection platform should have the following capabilities:

Early Detection with adoption of advanced AI/ML based models to consume a broad range of signals from backup data to learn and improve the accuracy & performance.

Comprehensive Coverage to provide greater visibility across more workloads and detect the damage patterns caused by the latest ransomware variants.

Intelligent Investigation Dashboard to view and analyze the affected files with anomalous and clean snapshots, trend analysis, and instant recovery options with automated workflows.

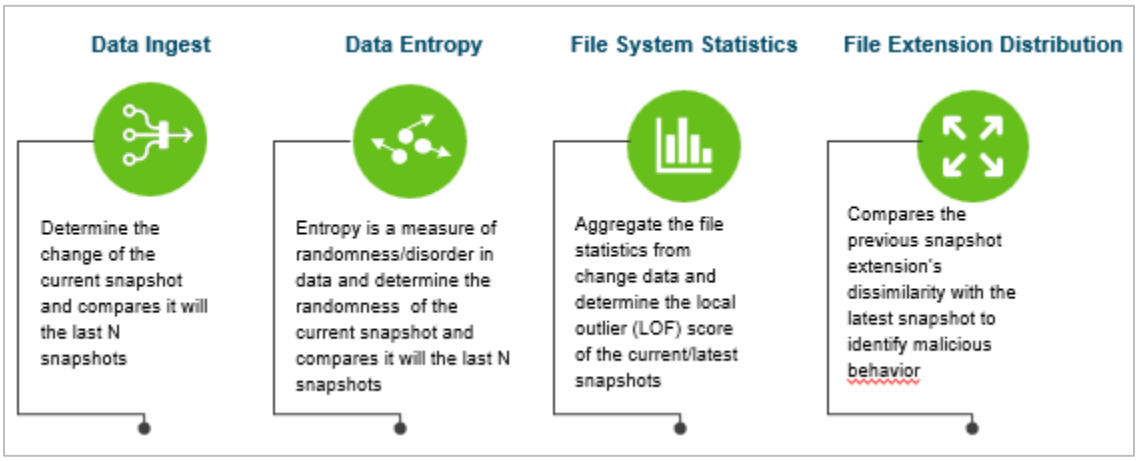
Broader Security Integration Ecosystem support with leading threat protection, data classification solutions to accelerate time to discovery, investigation, and remediation of ransomware attacks.

Machine Learning-based Detection Models

Machine learning has made a significant impact across domains, including cybersecurity. Its ability to detect sensitive data, fraud, malware, and ransomware has been well-tested and proven over the past decade. Machine learning has the capability to effectively detect hidden patterns, if supplied with adequate data. This has led to numerous applications in classifying malware and ransomware.

At Cohesity, we have developed a state-of-the-art Cloud platform to monitor backup and recovery. In addition to these functions, the platform serves as an integral part of detecting anomalies, setting tolerance for anomaly alerts, analyzing the threats, and accelerating recovery with a few clicks. Here is the snapshot of Cohesity-developed machine learning models' leverage for anomaly detection.

Figure 2: Cohesity-developed Machine Learning-based Models



To detect ongoing attacks and avoid false positives, Cohesity feeds multiple metrics into algorithms running in our Data Cloud control plane, including but not limited to:

- Content information per backup: Size of data written, size of data read, logical size.
- Entropy/data-reduction ratio per backup (post dedup & compression).
- Change tracking information per backup: Number of files added, files deleted, files updated, files unchanged.
- Aggregated stats across multiple backups: Max data written bytes, max source logical size bytes, number of successful runs, and so on.

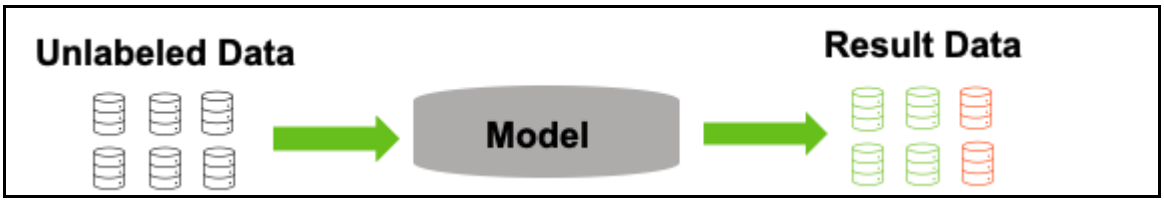
Cohesity's Anomaly Detection

Anomaly Detection

Anomaly Detection finds unusual patterns and sudden change rates such as file written, file read, logical size in the backup data that do not conform to the expected behavior.

Cohesity Anomaly Detection leverages the **Unsupervised Machine Learning** technique to identify any abnormal changes in backup data by getting telemetry data and signals.

Figure 3: Unsupervised Machine Learning Technique

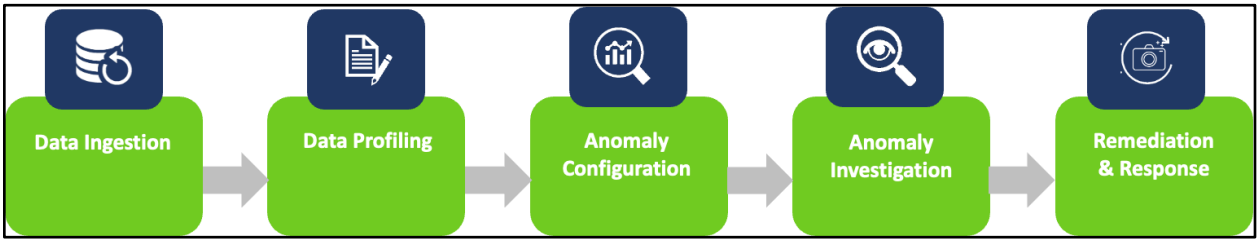


Cohesity Anomaly Detection Workflow

Cohesity data cloud offers a powerful automated anomaly detection in near real time as part of a AntiRansomware solution that continually tracks normal system activities to quickly spot irregularities and abnormal user behaviors that can signify a ransomware attack. Coupled with alerting, these capabilities don't just signal potential anomaly alerts but can also initiate remediation with pre-built workflows with instant recovery@scale from latest clean snapshot options that helps to minimize impact from ransomware.

Here's an overview of different phases of Cohesity's Anomaly Detection:

Figure 4: Anomaly Detection Learning phase



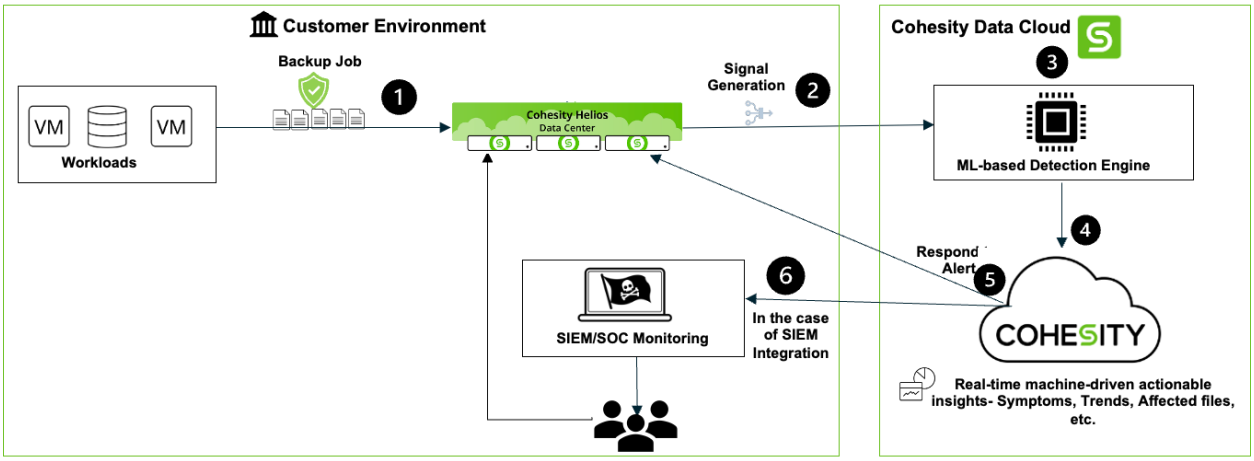
- **Data Ingestion**: Ingests data from supported workloads and backup details, protection group, metadata signals feeds (data size, entropy, and file stats).
- **Data Profiling**: Once the data is ingested, Cohesity performs data profiling to understand the characteristics (backup size thresholds, historical records, and metadata signals). This step helps establish a baseline and identify abnormal patterns and behaviors.

- **Anomaly Detection Configuration**: Analyzes the ingested data to identify any deviations from the established baseline and applies machine learning models to detect anomalies across different data sets and compute the anomaly strength.
- **Anomaly alert and Investigation**: When an anomaly is detected, Cohesity Helios generates alerts to notify the appropriate stakeholders and to the configured SIEM/SOAR solutions, so that both backup admin and the SecOps team can investigate the anomaly alert with trend analysis, sensitivity level, and the list of affected files.
- **Remediation and Response**: Based on the analysis and investigation, organizations can take appropriate actions to address the anomaly. This may involve instant recovery@scale from the latest clean snapshots with IMR and implement security measures. The actions can be automated from the SIEM/SOAR platform with Cohesity developed automation workflows.

It is important to note that Cohesity continuously monitors the data to identify new anomalies and the engine is tuned to maximize the accuracy, consistency, and speed of detection, and then tested on previously unseen data to continuously improve its performance.

The workflow below describes a high-level overview of the end-end detection workflow on Cohesity data cloud.

Figure 5: Anomaly Detection Workflow



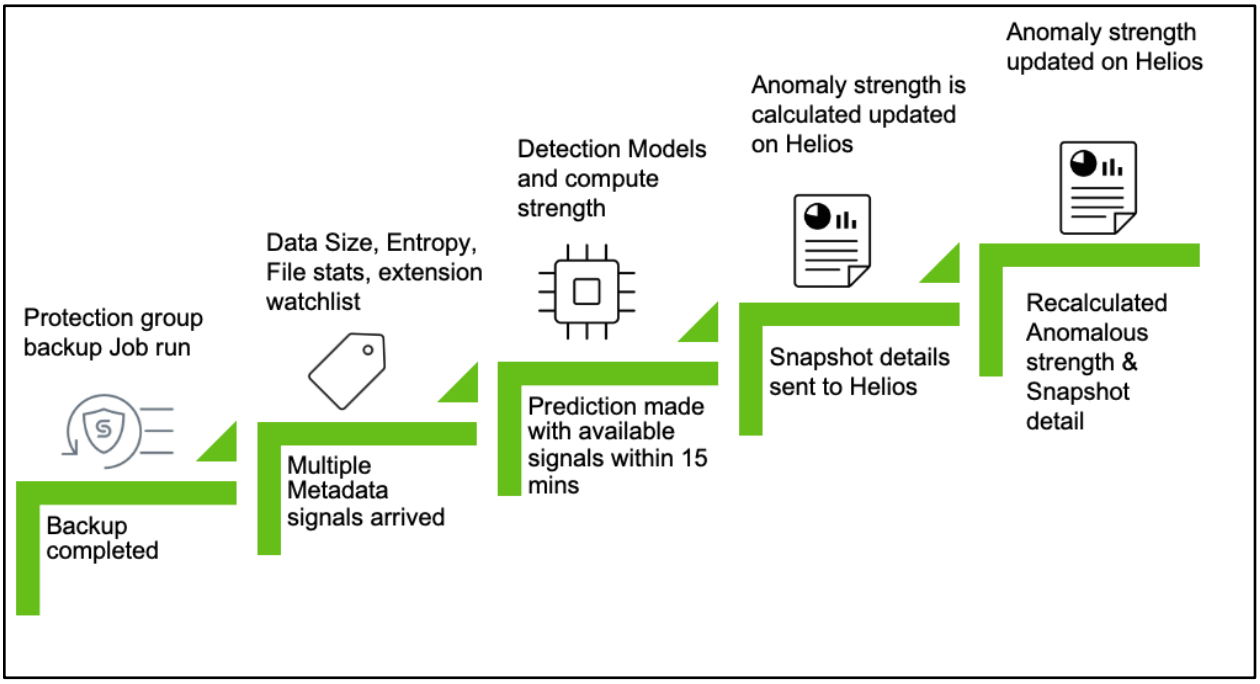
1. Each backup taken is a time-based snapshot, based on an associated Protection group policy, to back up data from a workload's source and store it on the cluster. A protection group uses the schedules and settings defined in the policy to determine when and how backups are captured, archived, or replicated.
2. Multiple metadata signals are generated from backup data and sent to the Cohesity-developed ML-based detection engine.
3. Using many historical backup samples and a size threshold collected by our services, machine learning models are used to build a detection engine that is trained to identify hidden signs of an attack across all recorded cases.

NOTE: The thresholds are based on calculated scoring by the ML models, not on the raw inputs; there are no thresholds tied to any lone metric like backup size, number of files, entropy ratio, etc.

4. The model's methodology varies based on the signals used and combines each of the signals as a historical collection of events instead of looking at them in isolation. We use a combination of unsupervised multivariate machine learning anomaly detection in conjunction with similarity detection techniques to identify potential anomalies. Multiple algorithms based on signal availability are assembled. Their anomaly scores are then assembled using weights (w_m) inferred from bayes analysis to arrive at the overall anomaly strength.
5. Based on the overall anomaly strength exceeding the defined threshold, Cohesity AntiRansomware dashboard generates anomaly alerts, which will be investigated based on trend analysis.
6. To accelerate the time to discovery, investigation and remediation, anomaly alerts can be integrated with leading SIEM/SOAR solutions. This enables the required action from the SOC central dashboard with Cohesity-developed automation workflows.

NOTE: After the backup, it will require at least 15 minutes before any anomaly is reported on Cohesity platform.

Figure 6: Detection Prediction Workflow



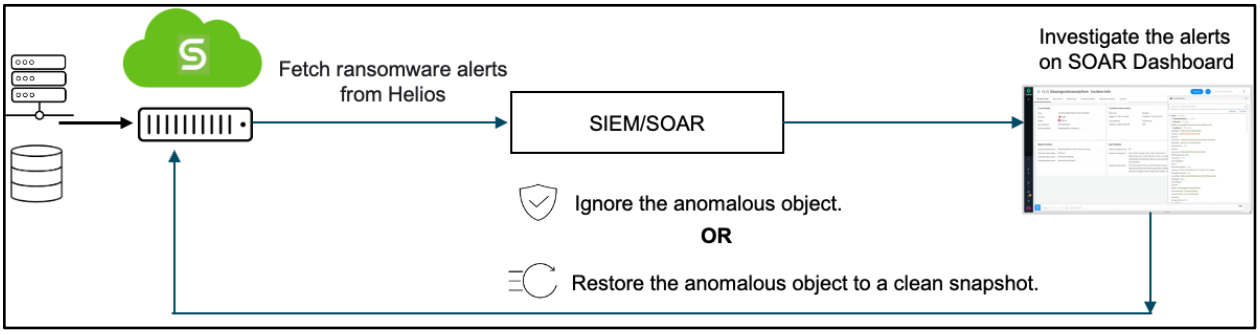
Integration with the Security Ecosystem

Cohesity provides capabilities to integrate anomalous alerts with leading SIEM/SOAR platforms, providing a unified threat visibility and enriching investigations. Rapid response and recovery also bring efficiency into SOC operations.

This integration enhances the overall security posture by combining anomaly detection with the capabilities of other security tools. Refer to Cohesity SIEM/SOAR Integration to learn more about how Cohesity integrates with different SIEM/SOAR platforms to support efficient response and recovery while enabling security and IT teams to work more closely together.

- Palo Alto Cortex XSOAR
- Cisco SecureX
- Microsoft Sentinel
- Crowdstrike
- ServiceNow (coming soon)
- Splunk (coming soon)

Figure 7: Closed-loop alert detection and remediation by integrations with SOC tools



Investigating the Anomaly Alert

Cohesity clusters managed on Helios are automatically connected to the AntiRansomware dashboard. Cohesity Data Cloud continuously monitors changes in the backup with signals and models for your organization. Based on the signals received, each model will compute the score/value and calculate the overall anomaly strength sum by ensembling models with associated weightage of each of the applicable models and the algorithm will take the decision to flag off the anomaly with the reason.

Following are the capabilities of Cohesity AntiRansomware Dashboard:

- Global dashboard for anomaly tracking
- Trend analysis across data ingested, data written, entropy, files change (added, modified, deleted), files unchanged information
- Lists of affected files within each object
- Single and bulk file download across anomalous and clean snapshot
- Single and bulk object restore (IMR) options across different sources and object types.
- Email alert notification
- Ongoing machine learning algorithms to improve detection.

For more details, refer to AntiRansomware investigation & response capabilities documentation.

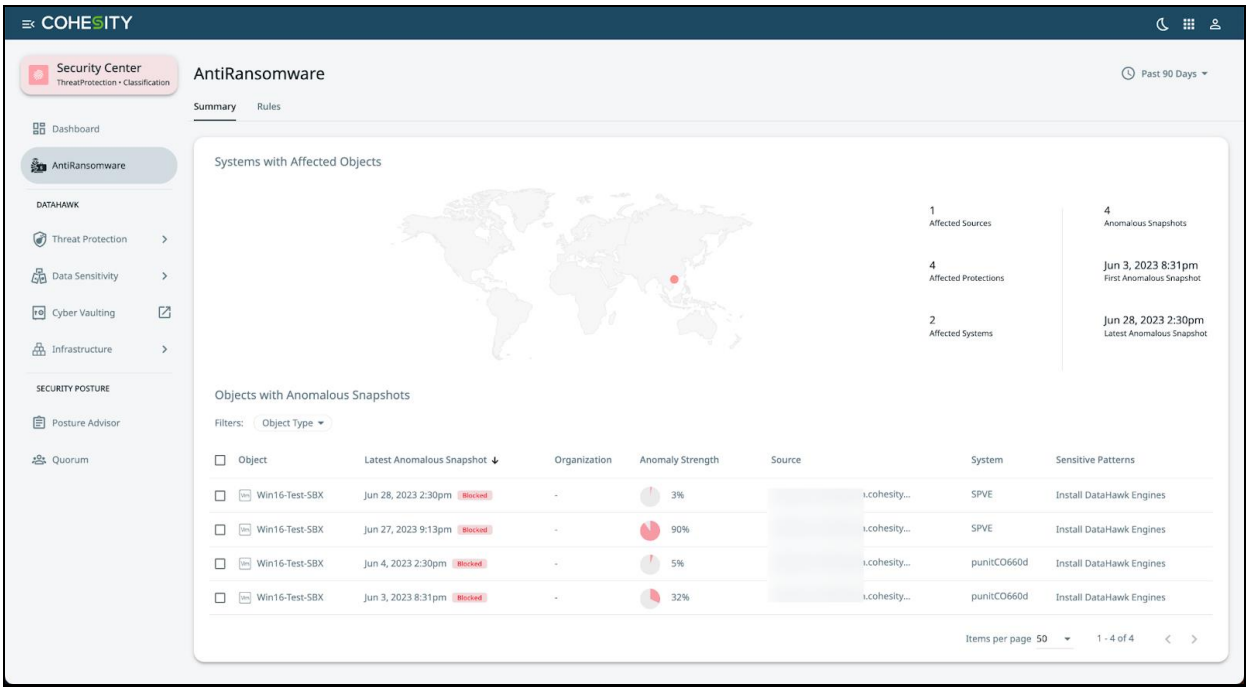
Global Dashboard

The **AntiRansomware Dashboard** shows the list of anomaly alerts in the anomalous objects of the selected time. You can click on the anomalous object to get more information in a separate details page organized into the following tabs:

- Snapshots
- Trends
- Affected Files

The dashboard contains the following sections:

- System with Affected Objects
- Objects with Anomalous Snapshots



You can view the anomalous objects detected on a map. Additionally, the following aggregate information is provided:

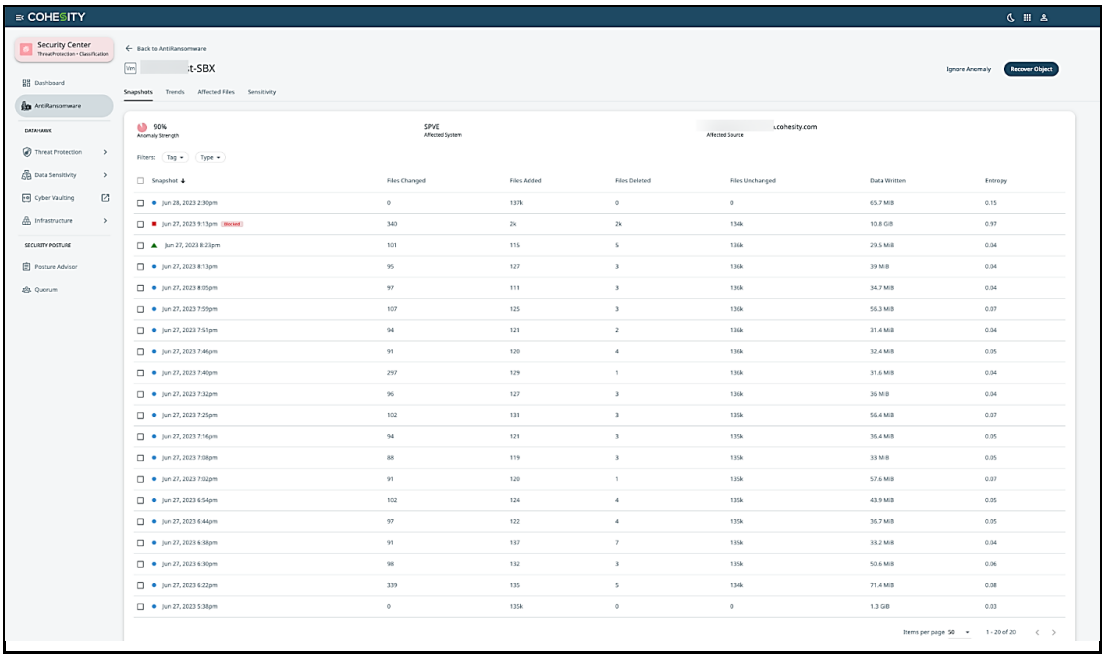
- Total number of affected sources
- Total number of affected protections
- Total number of affected systems
- Total number of anomalous snapshots
- Date and time of the first and latest detected snapshots

Below the map, you can see a listing of the detected anomalous objects with the following information:

- Object name
- Latest anomalous snapshot
- Anomaly strength (system determined)
- Source (object)
- System (Cohesity cluster)

Snapshot Analysis

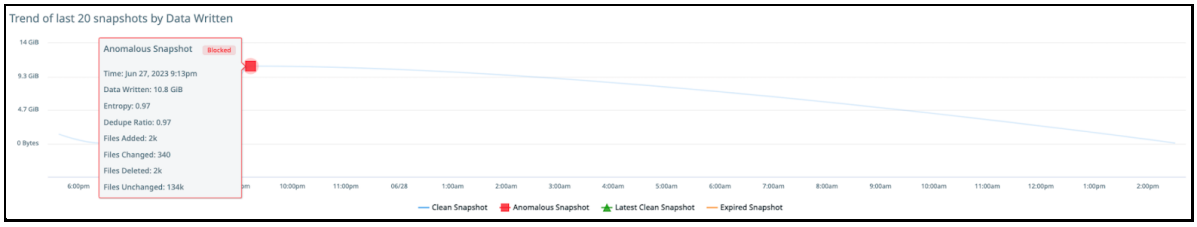
The **Snapshot** tab lists all the anomalous snapshots, non-anomalous snapshots, latest non-anomalous snapshots, tagged snapshots, and untagged snapshots.



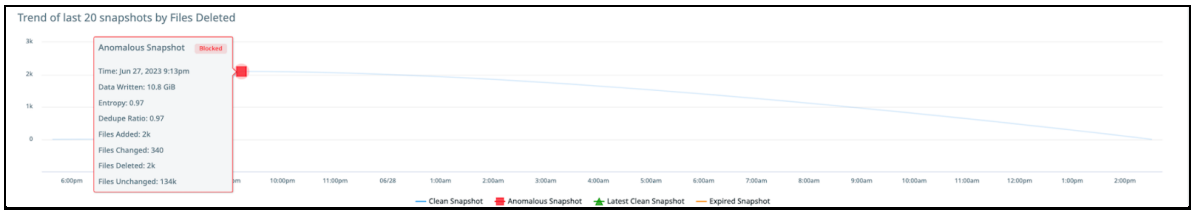
Trend Analysis

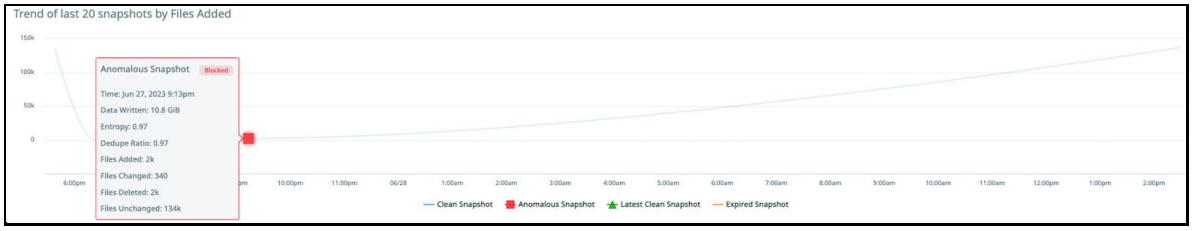
The **Trends** tab lists the trends of all the anomalous snapshots, clean snapshots, latest clean snapshots, and expired snapshots. It provides the following trends:

- Trend of the last 20 snapshots by **Data Written**

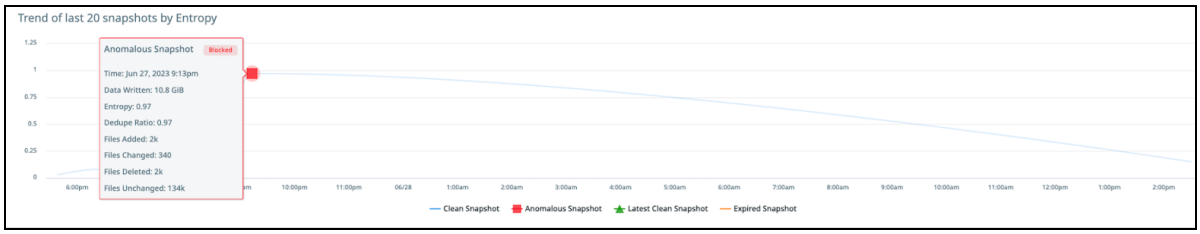


- Trend of the last 20 snapshots by **Files Changed, Deleted,** and **Added**

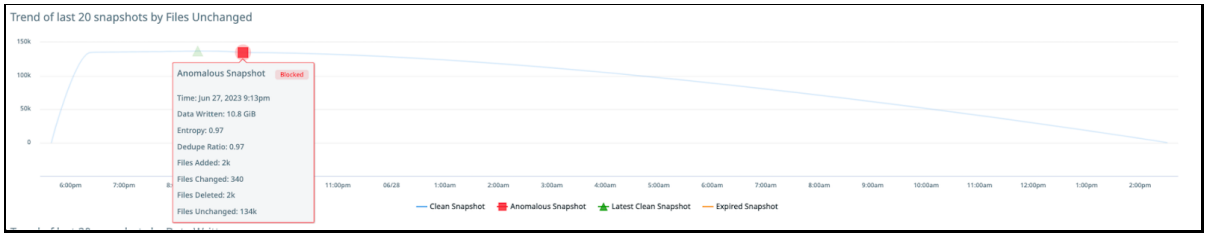




- Trend of the last 20 snapshots by **Entropy**

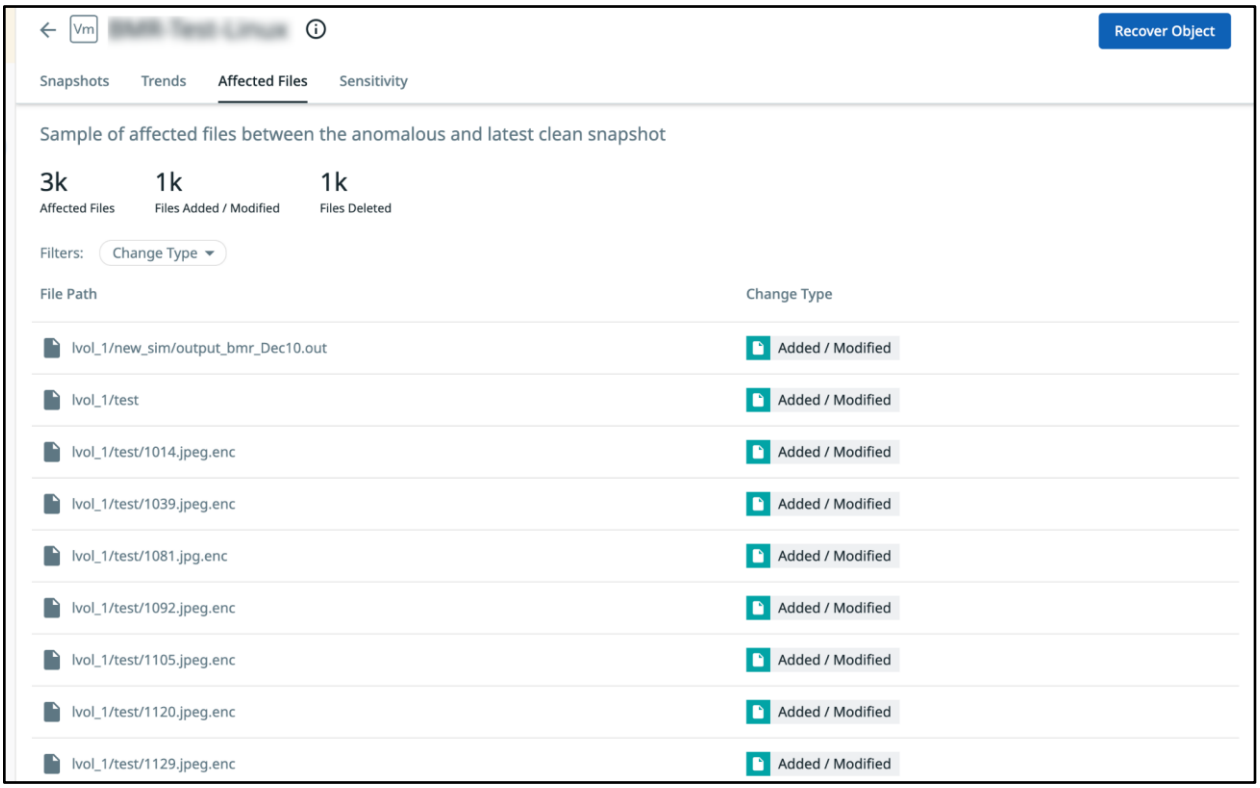


- Trend of the last 20 snapshots by **Files Unchanged**



Affected Files

The **Affected Files** tab shows the list of files that have changed between the clean snapshots and anomalous snapshots. It also shows whether the file is added, modified, or deleted under the **Change Type** column.



Take Actions on Alert

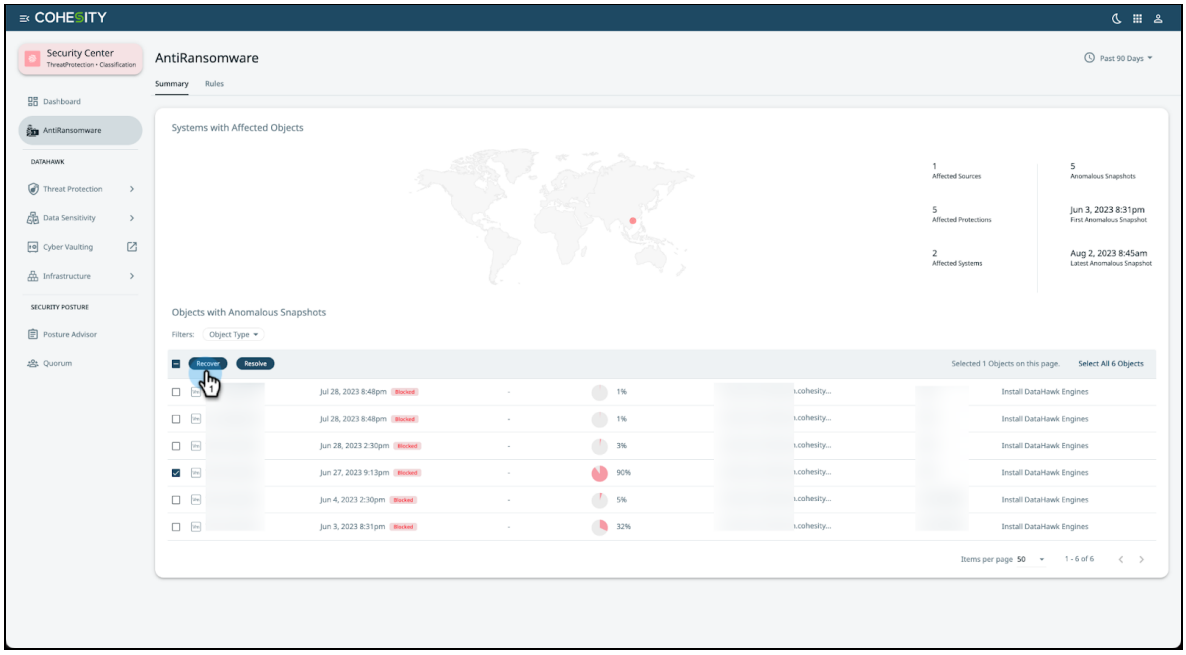
Investigate the triggered anomaly alert and take actions from Cohesity Helios dashboard or SOC integrated platforms:

- **Ignore Anomaly–**You can choose to ignore an anomaly if you do not want email notifications for the anomalous snapshot for the next 30 days. Ignoring the anomaly suppresses the alert on the object and removes the anomalous tag from all tagged snapshots. The alert will no longer be visible in the AntiRansomware dashboard.

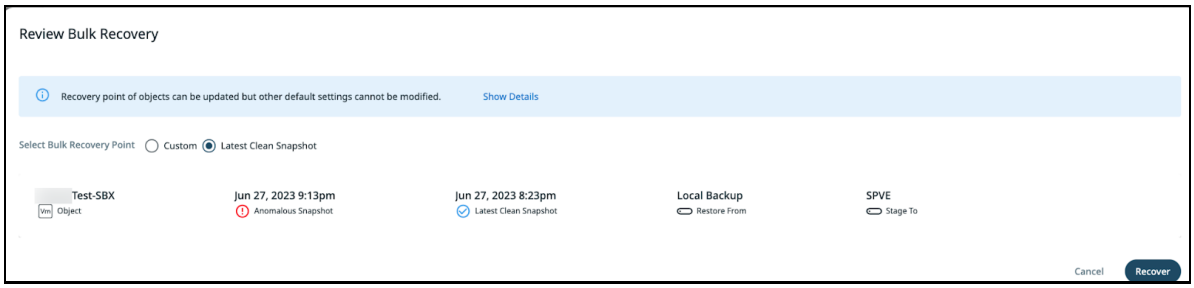
NOTE: If an anomaly is **ignored**, detection workflows will continue to evaluate the protection runs jobs but do not trigger an alert for 30 days or until the anomaly strength is greater than the ignored alerts.

- **Restore –** In case of an actual attack, it is essential to root out the problem and restore your data to the latest clean snapshot as quickly as possible. This can be done on a scale, and across more workloads, such as NAS, providing a more complete set of capabilities.

Cohesity CyberScan can help assess whether a VM has severe vulnerabilities that the attackers might use to immediately re-enter the system. Cohesity also apply leading ML-driven classification technology that leverages Natural Language Processing (NLP) methods to automatically to discover and classify large sets of data at scale to help minimize risk and improve security posture. For more details, refer to recover object documentation.



NOTE: You have the option to choose to recover from the latest **clean snapshot** or from a custom **clean snapshot.** By default, the latest clean snapshot is selected.



Ransomware Benchmarks

Ransomware analysis is crucial in detecting ransomware attacks. It helps to understand the behavior and extent of operations associated with a malicious variant to be prepared to respond to a ransomware attack.

Well-Known Ransomware Analysis Variants

Cohesity has built a simulation platform to test the different ransomware variants' behavior in an isolated lab environment. Subsequently, characterization of the different ransomware attacks is performed from the collected signals. A synthetic dataset will be constructed by injecting the anomalous patterns (due to ransomware attacks) into the time series of backup snapshots.

The resulting data produced by these experiments are used as positive samples in our machine-learning models for predicting ransomware activities. This helps Cohesity not only to make our models more accurate but also to come up with a quantification of our models' accuracy.

The table below shows a breakdown of few real ransomware strains that are analyzed in an isolated environment along with their detection performance:

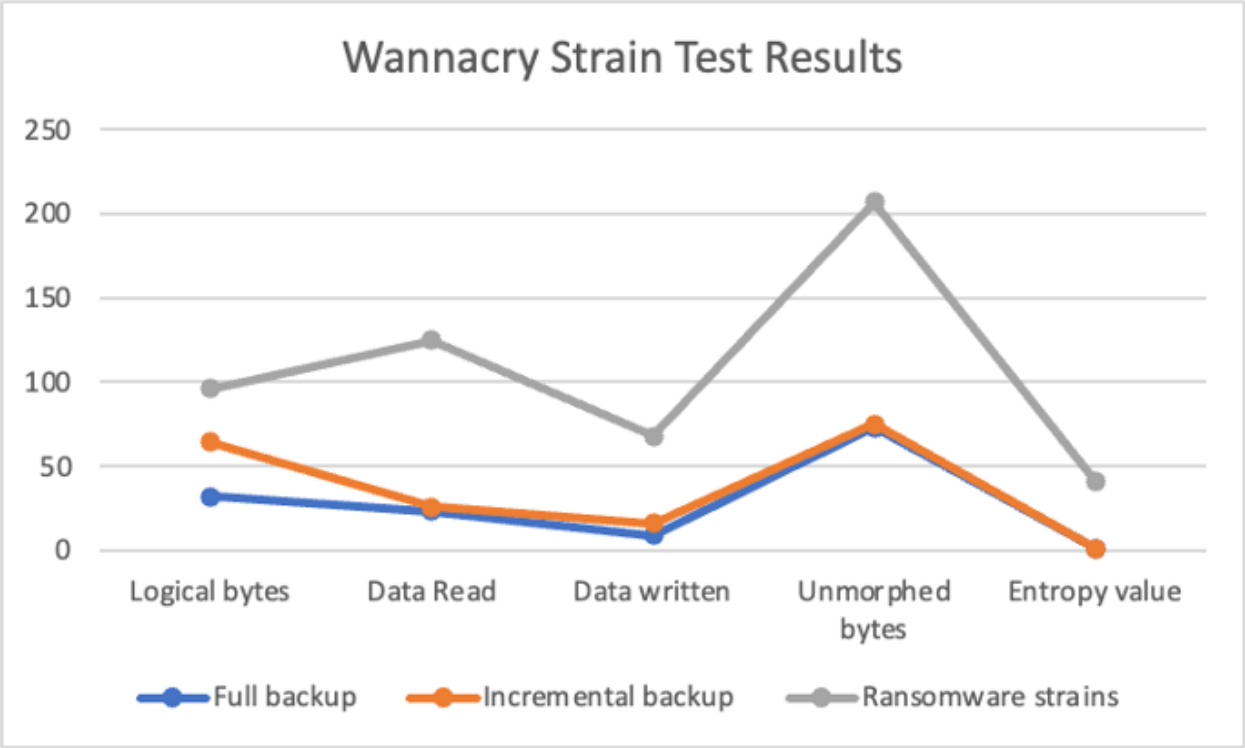
Table 1: Tested Well-known Ransomware Variants

Well Known Variant	Signals Analyzed	#Sample Test	Results Accuracy (Recall)
WannaCry	Entropy, File Stats, Data Written	488	100%
Locky	Entropy, File Stats, Data Written	495	100%
Radamant	Entropy, File Stats, Data Written	494	100%
Jigsaw	File Stats, Data Written	411	100%

Below are detailed statistics for one of the strains that are tested, and data collected from the experiment. Eg: the Wannacry Ransomware variant attempts to exploit vulnerabilities in the Windows SMBv1 server to remotely compromise systems and encrypt files.

Notice a significant increase in the signal strength and entropy value in the detection once the ransomware strain is executed after incremental backup. Most of the ransomware strain results in encrypting the data to make it unusable for the customers.

Table 2: Wannacry Strain Test Results



Supported Workloads

With growing IT needs, infrastructure is getting distributed and spread between on-premises datacenters and cloud services. Cohesity's anomaly detection capabilities are designed to support a wide range of workloads and data sources. The platform's versatility allows organizations to apply anomaly detection techniques to various data types and environments.

By applying anomaly detection techniques to these workloads, organizations can gain valuable insights into potential security threats, anomalous behaviors, and data issues, leading to improved incident response and proactive data protection.

For more details, refer to supported workloads and the backup type documentation.

Conclusion

Accurate and quick anomaly detection is a crucial component of defense against ransomware attacks, and it should complement and enrich the current security tools in an environment. Cohesity provides unique capabilities to help identify, protect, detect, and recover from cyber-attacks. Unmatched scalability allows for faster backup and restore performance as well as rapid analysis for faster and more accurate detection. An API-first approach and Pre-built app allows for tighter SIEM and SOAR integration, resulting in quick and precise response to minimize impact of attacks. Cohesity has a strong commitment to continue working with other top security solutions across the industry to give defending organizations the upper hand in a world of evolving threats.

Appendix A: Terminology

Table 3: Terminology

Terms	Description
Anomaly Detection	Finding patterns and unusual change rates in the backup data that do not conform to the expected behavior.
Compression ratio	Compression shrinks consumed storage capacity by reducing the size of bit or byte strings in a data stream using different algorithms.
Dedup ratio	The type of compression identifies redundant segments of data and replaces duplicate segments with a pointer. Deduplication happens before compression.
Entropy	A measure of randomness in data and is most helpful in identifying encrypted data.
Multivariate	Uses multiple variables to decide potential anomalies.
Supervised anomaly detection	Requires a data set to be trained with specific "normal" and "abnormal" labels.
Unsupervised anomaly detection	Detects anomalies in an unlabeled data set by comparing data points to each other, establishing a baseline "normal" outline for the data, and looking for differences between the points.
Univariate	Uses single variables to decide potential anomaly.
Snapshot tagging	A tag name can be applied to an anomalous snapshot if the snapshot has an anomaly strength above the defined threshold.

Your Feedback

Was this document helpful? Send us your feedback!

About the Authors

Karthick Radhakrishnan is Director, Technical Solution Engineering. In his role, Karthick focuses on managing Cohesity DataProtect and Security solutions.

Sagar Sethi is a Staff Technical Marketing Engineer at Cohesity. In his role, he focuses on various aspects of Data security to secure the Cohesity product design & solutions.

Other essential contributors included:

- Rob Young, Product Manager, Competitive Intelligence
- Robert Shields, Director, Product Marketing
- Jonathon Mayor, Field Technical Director
- Adaikkappan Arumugam, Director of Product Solutions
- Nagapramod Mandagere, Principal Engineer, Engineering
- Subash Babu, Staff Technology Editor
- Mary Juliya, Technical Editor

Document Version History

VERSION	DATE	DOCUMENT HISTORY
1.1	Feb 2025	Removed “File Watchlist”.
1.0	Aug 2023	First full 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.