TECHNICAL WHITE PAPER

How It Works: Rubrik NAS Cloud Direct

Alpika Singh, Manan Trivedi, Sherif A. Louis

RWP-0645

Table of Contents

INTRODUCTION

Welcome to the "How It Works: Rubrik NAS Cloud Direct" whitepaper. This document is designed to acquaint readers with the functionalities, architectural design, and operational processes related to safeguarding unstructured data, whether on-premises or in the cloud, using Rubrik Security Cloud. The insights provided here will benefit those assessing, planning, or deploying the technologies covered in this guide.

AUDIENCE

This guide is intended for individuals looking to gain a deeper understanding of Rubrik NAS Cloud Direct with Rubrik Security Cloud, as well as the technical architecture supporting these functionalities. It is particularly useful for architects, engineers, and administrators who oversee data protection, compliance, governance, and cloud or enterprise infrastructure, as well as for those involved in security or governance roles.

OBJECTIVES

This white paper is designed to serve as a clear and concise technical reference on the architecture and workflows utilized by Rubrik's NAS Cloud Direct. Upon completing this document, readers will be equipped to address the following inquiries pertaining to Rubrik's protection capabilities for unstructured data:

- What functions does Rubrik NAS Cloud Direct perform?
- Which challenges does Rubrik NAS Cloud Direct address?
- What is the architectural design of Rubrik NAS Cloud Direct, and what are the reasons behind it?
- In what manner does Rubrik NAS Cloud Direct function?
- How does Rubrik NAS Cloud Direct stand out from alternative solutions?

CHALLENGES

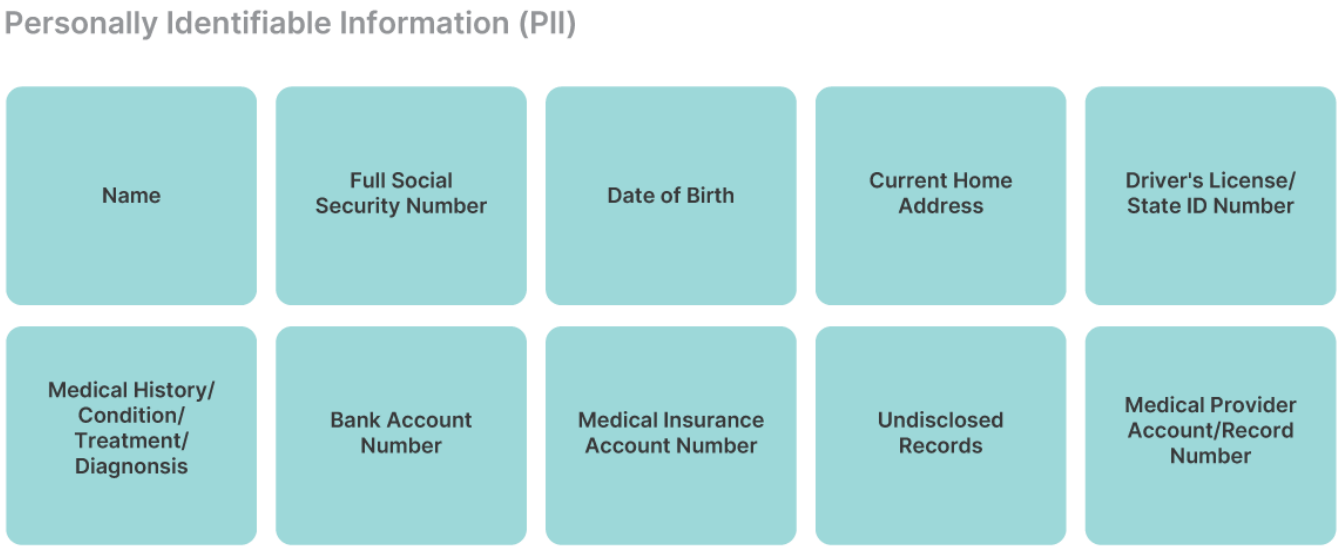
Forbes highlights an exponential growth in data generation, projecting an escalation to 150 zettabytes by 2025 with an annual growth rate of 23%. It is critical to understand that a substantial 80% of this burgeoning data exists in unstructured formats, encompassing documents, text files, images, videos, and websites. This unprecedented increase exerts considerable demands on unstructured data storage systems, including NAS and object storage, compelling the need for innovative management strategies.

The challenges of managing, securing, and ensuring the recoverability of such a vast data expanse—spanning on-premises, edge, and cloud environments—are magnified by the potential for disruptions arising from cyber threats and natural disasters. These threats jeopardize business continuity and impede critical operations, potentially resulting in delayed interventions.

In addressing this complex landscape, legacy data protection solutions encounter several pressing concerns:

Sensitive data protection

Enterprises deal with extremely sensitive information, such as financial records, design files, and personal data. Protecting this confidential information is crucial for complying with regulations like HIPAA and GDPR. The 2022 Data Breach Report by ITRC lists the top 10 Personally Identifiable Information (PII) attributes commonly stolen during cyber attacks.



Legacy solutions often fail to securely back up sensitive data and provide insights into potential exposures, risking privacy and compliance.

Complex data and application ecosystem

Modern enterprises consist of diverse and complex unstructured data ecosystems across hybrid infrastructure. For example, a healthcare system using an EHR application like Epic involves various data sets from ancillary applications to provide patient care, audit, and financial information. These include:

- WebBLOB share, which contains unstructured patient information.
- Caboodle database.
- Cache/IRIS database.
- Clarity database.

These applications are deployed across on-premises and cloud environments, increasing the cyberattack surface area. Legacy backups struggle to cover these varied applications and environments while ensuring data consistency and integrity.

Data volume and scalability

The volume of data generated across industries is enormous and continuously expanding. In the latest Rubrik Zero Labs report, Rubrik observed that a typical global organization has 278.4 terabytes of data, 80% of which is unstructured. Traditional solutions lack the scalability to manage this exponential growth effectively, compromising performance and accessibility as they are typically designed for smaller data volumes.

Recovery and business continuity

Enterprises cannot afford extended downtime. Architectural patterns such as geo-disperse replication and infrastructure duplication are typically used to solve this problem but can quickly spiral out of control in terms of cost and complexity, contributing to data sprawl and increasing the attack surface area. A reliable backup solution that can employ different protection methodologies depending on the criticality of the data is essential to controlling data sprawl, costs, and downtime.

Security concerns

As the frequency of cyberattacks increases, gaining insights into backed-up data is becoming increasingly important. Traditional backup solutions are insufficient as they lack sensitive data discovery and anomaly detection capabilities. This means organizations are unaware of sensitive data exposure and have no visibility into any changes made to data over time.

THE RUBRIK APPROACH

Rubrik NAS Cloud Direct (NAS CD) revolutionizes data protection for petabyte-scale unstructured data. It delivers exceptional performance, offering ten times the speed of NDMP while centralizing NAS and object data management and enhancing sensitive data visibility. NAS Cloud Direct secures Unstructured Data Storage Systems with immutable backups and isolates credentials while simultaneously detecting malicious activity and uncovering sensitive data exposure. It enables rapid, efficient recovery and archival with the ability to quickly search through billions of files and recover to on-premises, cloud, or alternate targets. With NAS Cloud Direct, organizations can overcome unstructured data growth challenges, combat ransomware threats, and ensure rapid NAS and object data recovery and archiving.

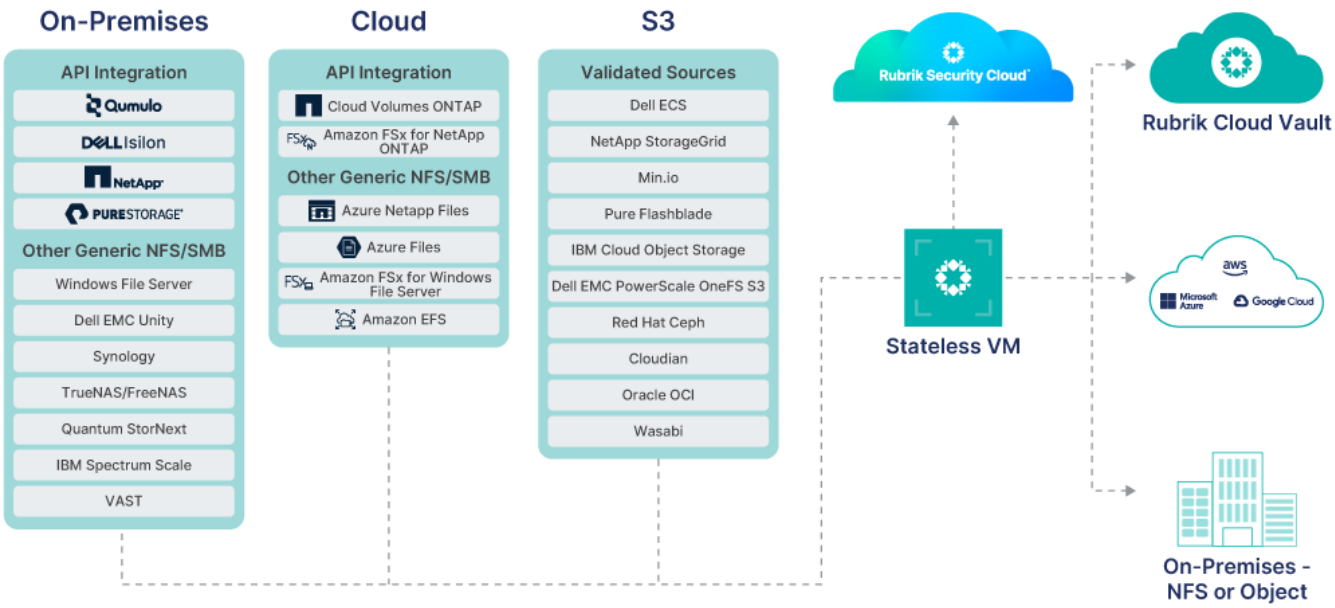


Figure 1 – Rubrik Security Cloud

Rubrik NAS Cloud Direct is a SaaS-based offering. Each customer's control plane is hosted in a Rubrik-owned, isolated tenant environment, which also hosts the index database and compute resources that manage backup jobs and policies. The solution includes a stateless virtual machine (VM) deployed within the customer's infrastructure.

It is engineered for high-speed data transfers to optimize network bandwidth during backup, archive, and replication tasks, without impacting production file services. NAS Cloud Direct uses a simplified policy-based model to determine a data source's backup frequency, data retention, and optional replication target (either local or cloud-based).

Underlying the NAS Cloud direct data-management services is a set of engines specifically designed to overcome traditional data-management tools' typical performance and scale limits, including:

- A scanning engine can discover and index hundreds of thousands of files per second.
- A highly scalable index engine engineered and optimized specifically for metadata-based tracking of billions of new and changed files.
- A data mover that uses parallel data streams to maximize available network bandwidth.

NAS Cloud Direct is integrated with Rubrik security tools such as Anomaly detection and Sensitive data Monitoring capabilities to investigate anomalies, detect sensitive data exposure, and ensure rapid recovery after a cyber attack.

Key Features

Rubrik NAS Cloud Direct provides data security and performance for unstructured data at a petabyte scale. The key features of Rubrik NAS CD include:

VOLUME

Rubrik's NAS CD capabilities are engineered for the rigorous demands of modern, unstructured data, where scalability and performance are non-negotiable. The system stands out for its ability to scan billions of files swiftly—across any file workload and at any given instance—without compromising on speed or efficiency. This advanced functionality is made possible by Rubrik's robust scanning and indexing engine, which has been finely tuned to discover and index hundreds of thousands of files per second. Rubrik provides unparalleled visibility into data ecosystems by harnessing the power of metadata-based tracking. This granular tracking enables you to proactively manage vast datasets, as it is optimized to precisely monitor the lifecycle of billions of new and modified files.

SPEED

Rubrik's NAS CD is designed to optimize speed and minimize operational disruption in data-heavy environments. Rubrik has developed NFS/SMB clients optimized for backups and recovery, performing less number of read and write operations compared to standard SMB/NFS clients. By leveraging parallel processing, Rubrik is proficient in scanning, indexing, and moving NAS data concurrently across multiple streams. This approach maximizes network throughput and ensures that data movements are completed swiftly and efficiently. Additionally, the system is latency-aware and engineered to operate at line speed while mitigating any impact on production environments. Such capabilities are critical for maintaining system performance even during intensive data management tasks. Rubrik's truly incremental forever backups transform data protection processes by dramatically reducing backup windows. This method captures only the delta changes post-initial full backup. Rubrik also performs inline compression using the LZ4 algorithm on the backup data, significantly driving down the volume of data transferred and stored during subsequent backups, ultimately enhancing operational efficiency and reducing storage costs.

CYBER RESILIENCE

Rubrik empowers your organization with robust cyber resilience through its Zero Trust Data Protection framework. With built-in Multi-Factor Authentication (MFA) your data is shielded from unauthorized access at every level. Leveraging logically air-gapped protection, Rubrik ensures an additional layer of defense against cyber threats. Rubrik offers Role-Based Access Control (RBAC) capabilities to ensure administrators have the appropriate level of access. RBAC can also restrict administrators' access to different applications and

workloads, enhancing security and control. Anomaly Detection capabilities enable the swift identification of ransomware-impacted files, facilitating surgical recovery and minimizing downtime. Furthermore, the solution's Sensitive Data Monitoring feature enables the classification of petabytes of data, allowing for the precise identification of sensitive information and ensuring compliance with regulatory requirements. Rubrik equips your technical team with the tools necessary to fortify your data infrastructure against evolving cyber threats while maintaining operational efficiency and regulatory compliance.

ARCHITECTURE

Rubrik NAS Cloud Direct (NAS CD) enables organizations to backup their unstructured data by rapidly copying data and metadata from any source unstructured data storage systems (NFS/SMB/S3) to either a cloud object store and any tier of their choice or store it in on-prem S3 object storage and/or NFSv3 storage; or both and avoid any vendor lock-in. The architecture powering NAS CD consists of three components:

RUBRIK SECURITY CLOUD (RSC)

RSC provides organizations with a centralized platform to secure their data across enterprise, cloud, and SaaS applications, implementing zero-trust data security. This comprehensive security approach includes data protection, anomaly detection, sensitive data monitoring, and recovery. Rubrik Security Cloud capabilities adhere to the zero trust principle, ensuring that users, admins, and network traffic are not trusted without strong authentication via single sign-on (SSO), Multi-Factor Authentication (MFA), and Role-based Access Control (RBAC). Additionally, it offers SSO access to the NAS CD user interface using Security Assertion Markup Language (SAML). It restricts access to backups to authorized users only and provides capacity, compliance, protection, and anomaly detection reports for effective planning and compliance purposes.

STATELESS VM

This VM is downloaded from RSC and deployed near customer-source unstructured data storage systems storage, with connectivity to both the Rubrik-managed control plane and the customer data center or cloud tenant as backup/archive destinations. The VM can be deployed in various hypervisor or cloud environments, including:

- VMWare ESX
- Nutanix AHV
- Windows Hyper-V
- Linux KVM
- AWS EC2
- Azure VM

RUBRIK MANAGED NAS CLOUD DIRECT CONTROL PLANE

Rubrik provides customer-specific isolated NAS CD Control Plane tenants to ensure the dedicated and secure management of individual customer environments. It further consists of three components:

Cloud Slab

Cloud Slab is the center of persistence for NAS CD's cloud services, providing essential states for various critical components. At its core, it ensures the integrity and accessibility of crucial data layers, including metadata, indexes, and the blob data store. The NAS CD Cloud Slab ensures seamless continuity and efficiency in data management operations by maintaining this foundational data. Additionally, it serves as the repository for vital job information, recording crucial details such as job success, failure, and retry attempts. This centralized repository facilitates real-time monitoring and analysis, enabling swift and informed decision-making. Moreover, it ensures resiliency and consistency across all operations, safeguarding against data loss or corruption and upholding the highest data integrity standards.

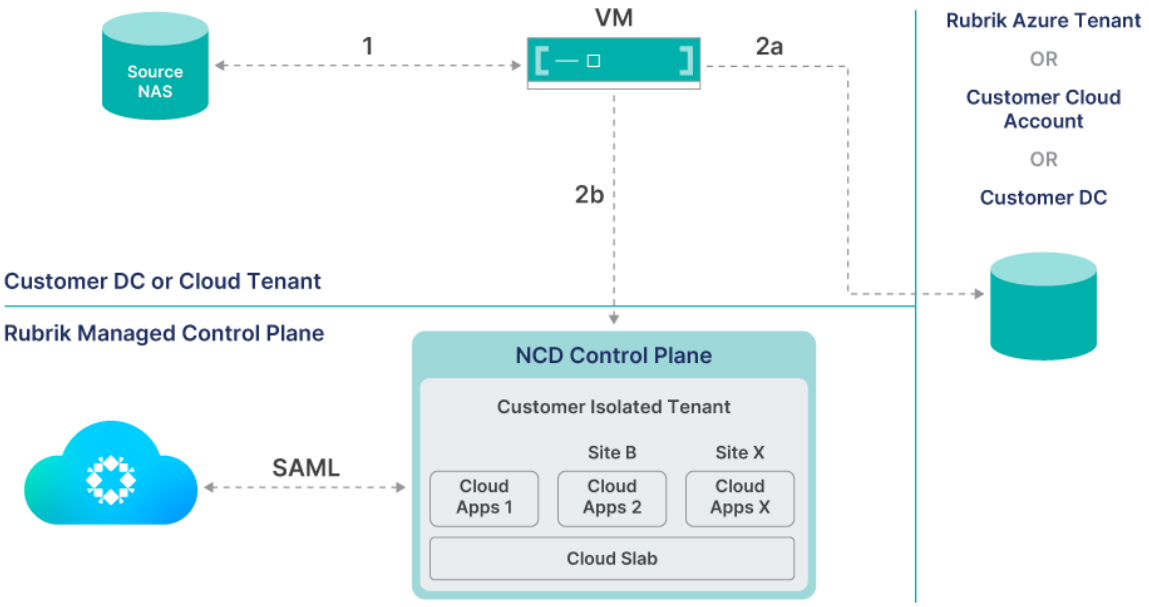
INDEX DB (CATALOG)

Index Database is a comprehensive catalog, housing essential metadata and indexing information critical for efficient data management. This centralized repository maintains a detailed inventory of data assets stored within the NAS CD environment, facilitating rapid search, retrieval, and organization of data. By capturing metadata such as file attributes, access permissions, size information, hard links, and the most recent time that the file was accessed, modified (data), and changed (metadata), the Index Database plays a pivotal role in facilitating data protection and recovery operations.

Cloud Apps

Cloud Apps serve as the comprehensive management plane for Rubrik NAS Cloud Direct (NAS CD), encompassing critical functionalities essential for efficient data management. This layer of software, housed within the Rubrik-managed isolated customer tenant, provides a centralized hub for overseeing NAS CD operations. At its core, Cloud Apps integrate a user-friendly interface (UI), facilitating intuitive interaction and control over NAS CD workflows. The Data Lifecycle Manager ensures seamless adherence to policy-defined data retention and lifecycle management strategies, guaranteeing data integrity and compliance. Moreover, the Job Scheduler optimizes task execution, maximizing resource utilization and operational efficiency. Acting as the central coordinator, Cloud Apps streamline communication and orchestration across various components within the NAS CD ecosystem.

All communication between the VM and other components is encrypted and conducted over HTTPS. The VM handles only outbound traffic, and inbound connections are not allowed.



NAS CD orchestrates initiating backup processes through a series of streamlined steps.

1. First, the VM connects to the designated source unstructured data storage systems, SMB share, NFS export, or S3 bucket, enabling seamless access to the target data repository. Leveraging advanced scanning algorithms, the VM scans the filesystem, comprehensively analyzing the data landscape. During this process, the VM identifies and flags any data categorized as "new," modified, or deleted since the last backup operation. Subsequently, the VM efficiently reads and extracts this newly identified data, preparing it for secure transmission to the backup destination. Those operations (scan, index, and copy) are done in parallel.

2a. Once the stateless virtual machine (VM) in Rubrik NAS Cloud Direct (NAS CD) identifies and extracts new backup data, it seamlessly transitions to the next stage of the backup process: data transfer. Leveraging optimized protocols and algorithms, the VM efficiently moves the newly extracted backup data to its designated destination, either cloud object storage or on-premises storage infrastructure. This transfer process ensures secure and reliable data replication, maintaining data integrity and accessibility while adhering to defined backup policies and configurations.

2b. Following the transfer of new backup data, the stateless VM orchestrates the movement of associated metadata to the NAS CD Control Plane. This critical step ensures the centralization of metadata, enabling comprehensive management and oversight of backup operations.

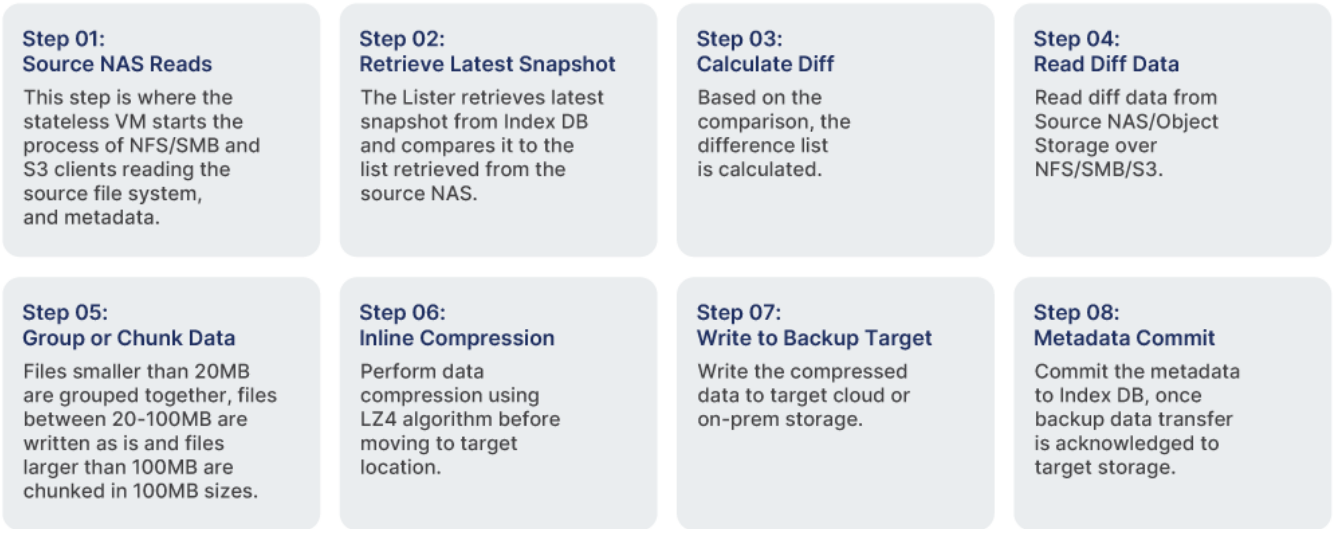
3. Upon completing the backup process, the stateless VM commits metadata and job status updates to the NAS CD Control Plane. This final step ensures the synchronization of critical information across the entire NAS CD ecosystem, enabling centralized monitoring and management of backup operations.

NAS CD maintains a 1-1 relationship between the source and target. A full backup is performed during the initial backup of a source to a target. Subsequent backups will be incremental, provided that the source and target destinations remain unchanged.

These scanning, indexing, and data movement steps occur in parallel. The scanning engine scans some data, compares it against the index to identify changes, updates the index, writes the data, and commits the index. This approach ensures faster handling of voluminous unstructured data and optimized speed.

HOW IT WORKS

BACKUP



Once the NAS source onboarding to NAS CD has been completed, you can assign a backup policy that defines the policy type (backup or archive), backup frequency, data retention period, and backup destination. The backup process comprises the following three phases, regardless of whether it is a First-Full or subsequent incremental backup:

Phase 1: The Scan Phase

The scan phase ascertains what Rubrik needs to ingest via NFS/SMB or S3 clients.

STEP 01: SOURCE UNSTRUCTURED DATA STORAGE SYSTEMS/OBJECT LISTS AND READS

The stateless VM uses a custom-built NFS/SMB client to establish direct connections to shares/exports to read both the file system data and metadata without the need for traditional mounting processes, eliminating unnecessary overhead and enhancing performance. For on-premises S3 objects, the stateless VM utilizes an S3 client to read the contents of the buckets. Leveraging multi-threaded dynamic work allocation optimizes resource utilization by intelligently distributing tasks across threads in real-time. Furthermore, the platform seamlessly rebalances workloads from other threads, ensuring no resources remain idle and maximizing throughput. With built-in latency monitoring capabilities, NAS Cloud Direct dynamically scales resources up or down to maintain optimal performance levels, seamlessly adapting to fluctuations in workload demands.

For full backups, NAS CD proceeds directly to phase 2, the move phase.

STEP 02: RETRIEVE LATEST SNAPSHOT

The Lister component retrieves the most recent snapshot from the Index database. Once retrieved, this snapshot is meticulously compared to the list of files and metadata obtained directly from the source unstructured data storage systems share. This comparison process is crucial for identifying any changes, updates, or additions within the source unstructured data storage systems since the last snapshot was taken.

STEP 03: CALCULATE DIFF

After retrieving the latest snapshot from the Index Database and comparing it to the list obtained from the source unstructured data storage systems, the stateless VM produces the differences. This crucial stage involves reconciling the two datasets to identify any variations or updates between them. Any changes found during the comparison process are analyzed and processed to determine the necessary actions. Then, it combines these differences to create a comprehensive record of changes, ensuring that modifications are accurately captured and accounted for.

STEP 04: READ DIFF DATA

The diff data is then retrieved from the source unstructured data storage systems or object storage leveraging various protocols like NFS, SMB, or S3 using the custom-built NFS/SMB client.

Phase 2: Move Phase

The move phase utilizes different techniques to move the backup data as fast as possible across the network bandwidth.

STEP 05: GROUP OR CHUNK DATA

Before the backup data transfer, files are intelligently grouped and processed based on size to optimize efficiency and performance. Files smaller than 20MB are grouped into bundles, maximizing network throughput and reducing overhead associated with processing individual small files, especially for reducing cloud target overhead in terms of API costs, object size limitations, and target network throughput and IOPS limitations. Files between 20MB and 100MB are transmitted as-is. Files larger than 100MB are chunked into 100MB segments to facilitate smoother transmission and mitigate potential network latency issues.

STEP 06: INLINE COMPRESSION

Next, inline data compression is performed using the LZ4 algorithm. This algorithm is renowned for its high-speed compression and decompression capabilities, which minimize the amount of data transferred over the network, optimize bandwidth utilization, and reduce storage requirements on the destination side.

STEP 07: WRITE TO THE BACKUP TARGET

The stateless VM establishes direct connections to the designated backup target storage (NFSv3 or object), which could be cloud-based or on-premises storage infrastructure, and then writes the compressed data. NAS CD uses the native Golang SDK client for writing to the target storage (AWS, Azure, GCP and S3 compatible storage and NFS/SMB storage). It leverages the multi-threaded dynamic work allocation described before.

NAS CD offers another unique value by allowing direct writing to any storage tier. This is possible because NAS CD maintains its own metadata tracking, eliminating the need to check what is already written in the destination bucket before writing new objects.

Phase 3: Index Phase

STEP 08: METADATA COMMIT

Upon acknowledgment of the successful transfer of backup data to the target storage, the stateless VM commits the associated metadata to the Index Database. This critical step ensures metadata synchronization with the corresponding backup data, facilitating comprehensive management and retrieval of stored information. By committing metadata to the Index Database post-transfer, NAS CD maintains data consistency and integrity, enabling efficient indexing and search capabilities.

These steps occur in parallel, with data processed in batches to ensure faster handling of voluminous unstructured data and optimized speed.

RESTORE

Rubrik NAS CD provides a robust and efficient recovery workflow designed to streamline the restore process for unstructured data storage systems, ensuring minimal downtime and maximum data integrity. The data restore process comprises the following two phases:

Phase 1: Initiation of Restore Process

The initiation of the restore process phase ascertains what Rubrik needs to be recovered to the destination NAS share.

STEP 01: RESTORE ISSUED

The recovery process is initiated by issuing a restore from a specific snapshot at a specific time through the Rubrik NAS CD UI or using APIs. This snapshot-based approach ensures precise recovery of data from the desired point in time.

STEP 02: LIST OF FILES/OBJECTS CREATED

Based on the restore timestamp, the restore crawler generates a detailed list of directories, files and objects that need to be restored from the time period/snapshot specified by the user. It also identifies the references/indexes for these files and objects in the target storage.

STEP 03: READ PLAN TABLE CREATED

The read or restore plan table is then created, which contains the list of objects and blobs that will be read from the target storage based on the detailed list created in the previous step.

Phase 2: Restore Completed

The restore phase restores data as fast as possible across the network bandwidth.

STEP 04: DATA RESTORE

The data restore can be performed on the source, or alternate unstructured data storage systems share.

For file data being restored to the source or an alternate unstructured data storage system, the process includes the following:

If restoring from an offline tier (AWS Deep Archive or Azure Archive): The data is first rehydrated to a hot tier.

Restore Directory Tree: The directory structure is recreated to mirror the original hierarchy. This is done at the same time during the rehydration phase.

Restore Large Files: Large files (data and permissions) are restored immediately as we have a direct reference to them.

Restore Small Files: To optimize the restoration process, small files (data and permissions) are restored after initiating the restoration of large files. Small files are grouped into 100 MB chunks for efficient processing. A comprehensive list of all groups containing the small files to be restored is generated, and the restoration of these small files is then initiated.

Apply Directory Permissions: Permissions are applied to the directories to maintain access controls.

For objects being restored to the source or alternate object storage:

Large Objects: Restored along with their associated metadata to ensure integrity and accessibility.

Small Objects: Similarly restored with metadata to maintain complete data context and usability.

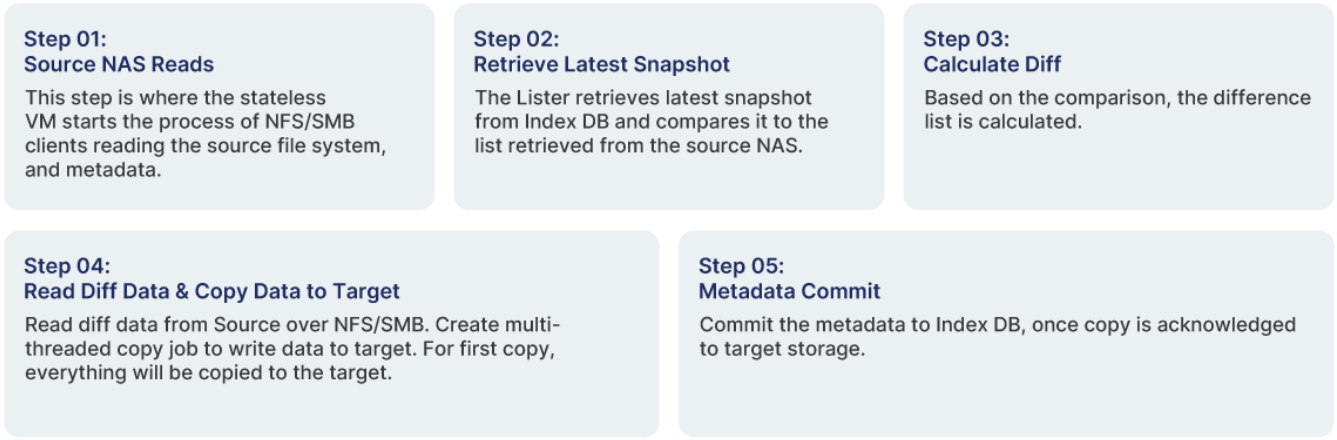
Once the data restore is initiated, administrators can monitor the progress through the Rubrik NAS CD UI, ensuring transparency and control over the restore.

Restoring data from offline/archive tier

Once the restoration plan is created, NAS CD reviews whether any of these objects are stored in an offline tier. Objects situated in offline storage tiers (such as AWS Deep Archive or Azure Archive) are rehydrated to the backup storage service's default online tier. This ensures they are accessible for reading during the data restoration phase. Objects that are in colder tier but still online tiers (like AWS S3 Infrequent Access/Glacier or Azure Blob Storage Cool/Cold tiers) do not undergo tier changes but are instead read directly from their current storage tier.

Rehydrating or moving offline objects to an accessible state can be time-consuming, often requiring several hours. Therefore, necessary objects are prepared for access before initiating the data restoration stage to ensure a seamless restoration process. At the start of the data restoration, it's possible that the initial set of files may not immediately be accessible. In such cases, the system checks every 10 minutes to determine if these objects have been successfully moved to an online tier, allowing the restoration to continue once access is confirmed. During this phase, NAS Cloud Direct also creates the required directory structure for the data. This process of making offline objects available for access is called 'rehydration' by Azure and a 'restore process' by AWS.

COPY



The NAS CD copy function is designed to efficiently sync data from source unstructured data storage systems or object stores to a target environment, whether on-premises or in the cloud. This process ensures data consistency, optimizes performance, and minimizes the operational impact on production environments. It also emphasizes that the file-to-file copy must be conducted using the same protocol, i.e., NFSv3 to NFSv3, or SMB to SMB, to ensure compatibility.

It also supports copying data from source unstructured data storage systems to various storage destinations, including S3 buckets (all tiers), Azure containers (all tiers), GCP containers (all tiers), and any S3-compliant buckets, whether on-premises or in the cloud.

The data copy process comprises the following six steps:

STEP 01: SCHEDULING AND INITIAL SCANNING

NAS CD schedules a copy job, initiating an initial scan of the source file system using NFS or SMB clients. The NFS and SMB clients are the same lightweight clients used for backup and recovery processes. Following the scan captures, the stateless VM captures the metadata and creates an index.

STEP 02: HANDLING INITIAL AND SUBSEQUENT COPIES

First Copy: For the initial copy, NAS CD moves directly to *step 4* and copies everything from the source directory to the target. This is because there is no existing index against which to compare, so the entire dataset needs to be transferred.

Secondary Copies: Lister retrieves the latest scan from the Index DB for subsequent copies. It then compares this scan to the current state of the source unstructured data storage systems to identify any changes or new data since the last backup. This differential approach ensures that only modified or new files are copied, optimizing the transfer process.

STEP 03: CALCULATE THE DIFF

It then merges the new scan results with the previous index to produce a list of differences. This differential list indicates which files have been added, modified, or deleted since the last copy. This step is crucial for minimizing the amount of data that needs to be transferred.

STEP 04: READING DIFFERENTIAL DATA AND COPY

The differential data is then read from the source unstructured data storage systems over NFS or SMB protocols, depending on the protocol used by the source file system. The stateless VM creates a multi-threaded copy job to enhance performance. This parallelization of the data transfer process significantly speeds up the operation. The VM writes the data to the target NAS and sets the necessary permissions on the target array to ensure access controls are correctly enforced. To minimize the impact on the production environment, the VM measures the latency on the source system. These latency measurements determine the optimal number of connections to maintain, balancing transfer speed with the load on the production system. Focusing only on the changed data reduces the time and network bandwidth required for the copy operation.

STEP 05: COMMITTING METADATA

Once the data is successfully written to the target storage, the metadata is committed to the Index DB. This involves acknowledging the completion of the copy job and updating the index to reflect the current state of the data. This step ensures that future differential comparisons are accurate and up-to-date, facilitating efficient subsequent copy operation.

All these steps occur in parallel, with data processed in batches to ensure faster copying of voluminous unstructured data and optimized speed by leveraging advanced techniques such as multi-threaded transfers and latency-based connection management.

API INTEGRATION

NAS-CD delivers universal interoperability with various NAS platforms and public-cloud providers, eliminating the need for vendor-specific disk-to-disk (D2D) data-protection solutions, capacity and licensing costs, and the resulting vendor lock-in. As a result, a single NAS-CD instance provides data protection for heterogeneous storage environments. NAS CD integrates with the following vendors via API:

- Dell PowerScale (Isilon)
- NetApp 7 Mode
- NetApp Cluster Mode
- Pure FlashBlade
- Qumulo
- AWS FSx for NetApp

With its API integration capabilities, Rubrik NAS Cloud Direct (NAS CD) offers powerful features tailored for modern data management needs.

- Firstly, the platform enables automated discovery of multi-tenants and network IPs, streamlining the onboarding process and facilitating seamless integration into diverse environments.
- Additionally, the platform automates permissions provisioning and simplifies access control management by creating an admin account with minimum privileges.
- Furthermore, NAS CD's automatic discovery of exports enables effortless identification and management of data shares, promoting operational efficiency and agility.

Through these API-driven capabilities, NAS CD empowers organizations to streamline their data management workflows, optimize resource utilization, and quickly adapt to evolving business requirements.

IMMUTABILITY

Immutability in data storage refers to the property of data being unable to be altered, modified, or deleted after it has been written. Once data is stored immutably, it remains unchanged and cannot be tampered with or overwritten. When customers use cloud storage providers to store Rubrik NAS CD backups, they can leverage the storage providers' object-level immutability support to protect the NAS CD objects from being modified or deleted for the configured retention period.

Immutability support for NAS CD would leverage the Write-Once Read Many (WORM) mechanisms internally provided by the cloud providers we support. NAS CD will support object-level immutability, and objects will be immutable until the retention period expires. Object-level retention is a policy applied to every object by the client writing the object. In this case, NAS CD will use object-level immutability to manage and track the status of its written objects to ensure it can designate between objects that should be extended and objects that are set to expire.

NAS CD supports immutability on the following cloud providers:

1. AWS
2. Azure
3. Rubrik Cloud Vault (RCV)
4. NetApp StorageGrid

Each cloud provider has its own way of supporting immutability, such as AWS using an S3 object lock and Azure using Immutable storage for its Azure blob.

For AWS, NAS CD leverages the native Amazon S3 Object Lock feature to provide immutable backups. AWS allows the S3 object lock to be enabled at the time of bucket creation and on existing buckets. Versioning should be enabled for the bucket as a prerequisite for Immutability.

For Azure, NAS CD leverages the native version-level immutability feature to provide the immutable backup. Azure allows version-level immutability to be set for both old and new containers. Version-level immutability has to be enabled on storage accounts or containers, and Azure blob versioning is required to be enabled for the immutability

Immutability can be configured on NAS CD by providing the credentials (access keys, secrets) with the cloud accounts that have immutability enabled and then using those locations for backup data using the backup policy. Customers can define the retention period in the backup policy, which will also serve as the immutability period. The immutability window is never greater than the retention/expiration window.

There are a couple of modes for configuring the Immutability:

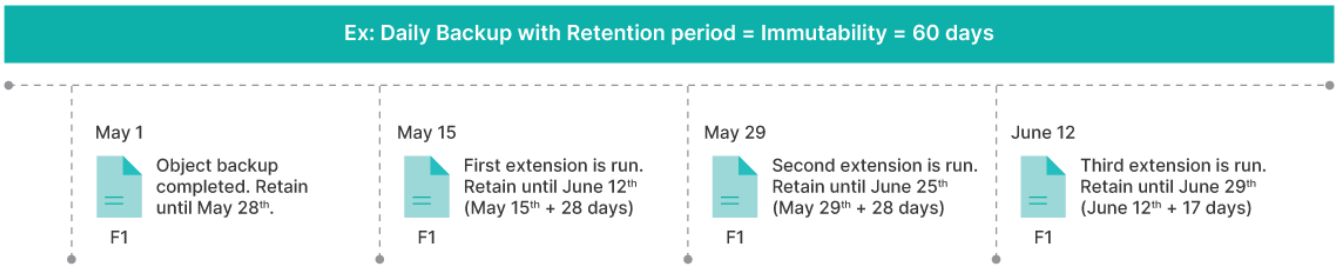
- **Governance mode:** In this mode, a user account with special permissions can remove the immutability setting for an object or delete the immutable object before the retention period ends.
- **Compliance mode:** In this mode, no one can change an object's immutability setting or delete it until the retention period ends.

Rubrik NAS CD will use the compliance or governance mode for immutability. It can set the retention period for an object version in the bucket, and a specific object will be retained until the compliance mode is active.

HIGH-LEVEL ARCHITECTURE

NAS CD employs the concept of Rolling Immutability, where the immutability lock period for NAS CD objects is not set at the beginning but incremented at regular intervals until the retention period is reached. By default, NAS CD extends the immutability lock period by 28 days every 14 days until the retention period is reached. This setup means that the minimum configurable retention period for immutable objects within NAS CD is 28 days.

Let's illustrate this with an example.



In the given example, the NAS CD policy sets a retention period of 60 days. Here's how it unfolds:

- **May 1st:** Rubrik performs the initial backup of an object and sets the retention to 28 days, expiring on May 28th.
- **May 15th:** After 14 days from the backup, NAS CD extends the retention period by another 28 days. This extends the retention until June 12th (May 15th + 28 days).
- **May 29th:** Following the same pattern, NAS CD extends the retention period by another 28 days, setting the new expiration to June 25th (May 29th + 28 days).
- **June 12th:** The final extension occurs, aligning with the 60-day retention configured in the SLA policy. This extends the retention until June 29th (June 12th + 17 days, completing the 60-day period from May 1st).
- The final extension will run on June 12th and will extend until the retention period configured on the SLA policy, which means it will extend the retention until June 29th (June 12th + 17 days, 60 days from May 1st).

GARBAGE COLLECTION

The Garbage Collection (GC) process operates through a series of well-defined steps to ensure efficient data management and storage optimization. Here's a detailed breakdown of the process:

INDEX TRAVERSAL AND CHUNK IDENTIFICATION

- During the retention cycle, a background job systematically traverses the index.
- This traversal involves scanning the index to identify chunks (data blocks) that are older than the specified retention period.
- Any chunks that meet this criterion are marked for deletion.

DELETE STORE AND DELETE PLAN FORMULATION

- After identifying the chunks to be deleted, the background job proceeds to the delete store.
- The delete store formulates a delete plan. This plan outlines which chunks are to be removed based on the retention criteria and marks them accordingly.
- This delete plan is created after midnight.

EXECUTION OF THE DELETE PLAN

- 24 hours after the delete plan is created, the GC background job then swings into action to execute the delete plan.
- It removes the indices associated with the chunks marked for deletion, ensuring that the metadata referencing these chunks is updated to reflect the deletion.

DATA DELETION ON TARGET STORAGE

- Data deletion occurs on the target storage system once the GC process verifies that 100% of the data referenced within a chunk is garbage (i.e., no longer needed or referenced).
- If a chunk is found to contain no valid data references, the GC background job deletes the entire chunk.
- This ensures that storage space is reclaimed and that only necessary data is retained.
- The storage space is typically reclaimed within 48 hours after the backup is deleted or an object expires.

DATA DISCOVER

Data Discover is engineered to provide visibility into an organization's unstructured data. It utilizes a scan engine that identifies new and updated data across multi-petabyte environments with a substantial number of files.

Once enabled, Data Discover scans NFS and SMB exports across all integrated systems. It systematically discovers and indexes metadata from files and directories on NAS to compile a complete representation of an organization's unstructured data.

The performance of the scanning process is dependent on the capabilities of the NAS and the network. The engine has the capacity to scan a large number of objects per second, allowing for the generation of reports within a reasonable timeframe, even in sizable data environments.

With Data Discover, organizations gain an overview of their unstructured data, providing insights into:

- The quantity of data held
- The storage location of the data
- The last access or modification date of any given object

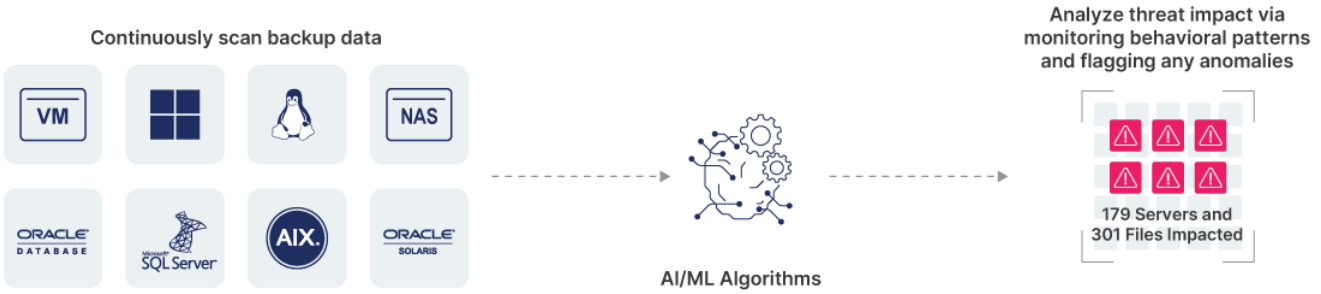
Having access to this information equips organizations to make decisions based on the current state and usage of their unstructured data.

CYBER RESILIENCE

In a cyber attack, swiftly identifying impacted NAS files and objects is crucial for organizations to initiate effective remediation measures and minimize disruption. With Rubrik's comprehensive cybersecurity solutions, including Anomaly Detection and Sensitive Data Monitoring, organizations can quickly pinpoint which NAS files and objects are compromised and where they are located within their infrastructure. Let's discuss each of them in detail.

ANOMALY DETECTION

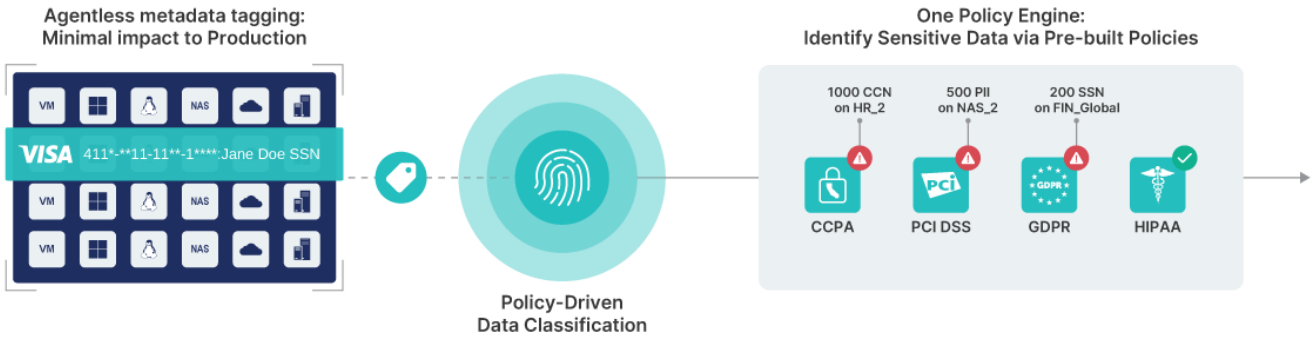
Anomaly Detection by Rubrik offers a proactive solution to combat cyber threats. It leverages advanced machine learning algorithms to identify and respond to anomalies swiftly. Detecting additions, deletions, modifications, and encryptions (anomalies) allows organizations to track data changes over time, simplifying the recovery process with just a few clicks and minimizing business disruption. Anomaly Detection applies machine learning algorithms to establish normal baseline behavior for each system, monitoring behavioral patterns and flagging any deviations from the baseline. Analyzing various file properties such as change rates and system sizes, it detects anomalies and alerts users through RSC, email, or integration with SOAR and SIEM applications. Anomaly Detection remains adaptive to emerging threats by continuously refining its detection model.



Anomaly Detection's streamlined user experience, integrated within the Rubrik global management interface, simplifies recovery. Users can easily select impacted applications and files and restore them to the most recent clean versions with minimal effort. With Rubrik automating the restoration process, organizations can recover quickly and confidently, safeguarding against potential data loss and disruption to business operations.

SENSITIVE DATA MONITORING

Sensitive Data Monitoring offers a proactive approach to reducing sensitive data exposure and managing exfiltration risk by providing comprehensive insights into the types and locations of sensitive data within your organization's data. Sensitive Data Monitoring scans and classifies sensitive data without agents or impacting the production environment. It leverages pre-built policy templates and analyzers to identify common data types from regulations and standards such as GDPR, PCI DSS, HIPAA, GLBA, etc.

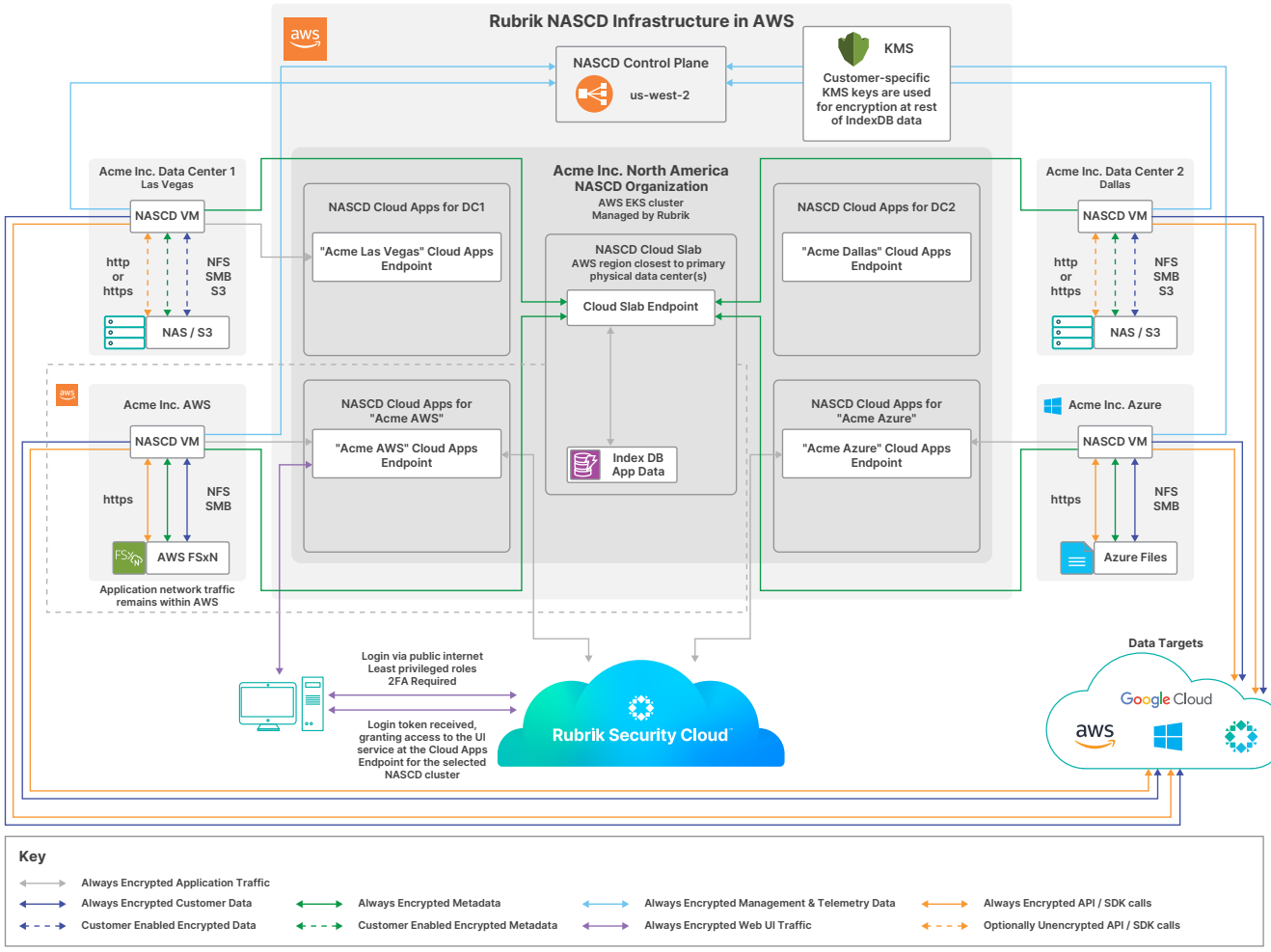


Sensitive Data Monitoring's streamlined user experience, integrated within the Rubrik global management interface, simplifies the organization's ability to assess data exposure and identify sensitive data that may be at risk of unauthorized access or exfiltration, enabling proactive risk mitigation measures. By providing the location of sensitive data, Rubrik facilitates compliance efforts, ensuring organizations maintain regulatory requirements and protect sensitive information effectively.

SECURE INFRASTRUCTURE

NAS CD is designed with security as its crux, emphasizing the safeguarding of data and network operations. As discussed above the backup and recovery task data are meticulously managed by bundling them into groups, compressing them, and writing them to the data target as blobs. These data blobs are designed to include sufficient information to restore customer data even if the metadata is lost, ensuring that data recovery remains possible in adverse situations. However, there are distinctions in how data is encrypted based on the protocol used at the various stages.

Rubrik NAS Cloud Direct (NASCD) Architecture



UNSTRUCTURED DATA SOURCE TO VM

When the data is moving from the customer unstructured data source to the stateless VM, encryption is based on the protocol configured in your datacenter environment. While native NFS data travels unencrypted, the architecture allows for encryption under specific conditions such as:

- For S3 Source NAS to VM: Encryption is used if the S3 is configured as an HTTPS endpoint.
- For SMB Source NAS to VM: Encryption is used if enforced by the source NAS.
- For NFSv4.1 Source NAS to VM: Encryption is used if Kerberos is configured.

It's also important to note that data only passes through the NAS Cloud Direct (NASCD) VM and is not actually stored on the VM, minimizing the risk of data exposure on intermediary systems.

VM TO NAS CD CONTROL PLANE OR TARGET

The network security framework ensures that all internet traffic originating from the stateless VM to the NAS CD control plane or the target is outbound only on port 443 and is secured using TLS 1.3 encryption, providing robust in-flight data protection against eavesdropping and tampering. AWS ECM (Encryption Key Management) is responsible for managing encryption keys, adding an extra layer of security by ensuring that keys are handled securely and efficiently.

NFS/SMB/S3 CLIENT SECURITY

Rubrik has engineered NFS, SMB, and S3 clients from the ground up with a focus on high security and performance. The NFS and SMB clients utilize Remote Procedure Calls (RPC) for accessing data, ensuring a structured and efficient method for data operations. In contrast, the S3 client leverages the AWS S3 SDK for data access, aligning with AWS standards for security and performance. Both NFSv3 and NFSv4 are supported, including the more secure NFSv4.1 with krbp5 (Kerberos) authentication. The SMB protocol is supported in both SMB2 and the more secure SMB3 with AES-CCM Wire Encryption. Additionally, NTLMv2 authentication method is supported, providing secure option for verifying identities.

VM SECURITY

The NAS Cloud Direct VM is securely sourced directly from Rubrik Security Cloud or the NASCD UI, ensuring the integrity of the VM software. Each downloaded bundle contains a unique embedded hardware ID and an order key specific to each Cloud Apps site, providing a means for secure identification and management. When the VM is first started, it undergoes an enrollment process to become a member of a Cloud Apps site, using the embedded order key and further identification by the unique hardware ID. This process ensures that each VM is securely linked to its designated Cloud Apps site. Additionally, all network communication from the VM to internal services is encrypted in-flight with TLS 1.3, ensuring that data remains secure and protected throughout its transmission.

The customer configuration is stored in a database which is encrypted at rest.

These measures ensure that all communications involving the VM are protected, maintaining the confidentiality and integrity of the data throughout its lifecycle.

SUMMARY

Rubrik NAS Cloud Direct delivers a comprehensive solution that maximizes efficiency and security in unstructured data environments. It offers the capability to protect data across cloud and on-premises targets at a petabyte scale, enabling organizations to seamlessly scale their data protection operations while driving operational efficiency with incremental forever backups. Rubrik's parallel data streams optimize network throughput, ensuring rapid and efficient data transfer. Additionally, Rubrik maximizes performance on unstructured data, providing up to 10 times better performance compared to legacy NDMP solutions. Effortlessly scanning, indexing, and moving billions of files and objects, Rubrik streamlines petabyte-scale NAS data protection. With robust security features, including data encryption at rest and in transit, logically gapped backups, and continuous monitoring for cyber threats, organizations can ensure the integrity and confidentiality of their data assets. Rubrik also empowers organizations to identify sensitive data, allowing them to gain insights into their security posture and effectively mitigate the risk of exposure.

For additional information, please visit https://www.rubrik.com or contact your Rubrik Account Team.

VERSION HISTORY

Version	Date	Summary of Changes
1.0	July 2024	Initial Release
2.0	August 2024	Added Security and Data Discover sections
2.1	September 2024	Fixed typos

Global HQ
3495 Deer Creek Road
Palo Alto, CA 94304
United States

1-844-4RUBRIK
inquiries@rubrik.com
www.rubrik.com

rwp-hiw-rubrik-nas-cloud-direct / 20240924