Technical Report

NetApp Cloud Backup

Technology Overview

Christopher Wong, NetApp

March 2018 | TR-4427

Abstract

This technology overview provides insight into the technical capabilities of the NetApp® Cloud Backup (formerly AltaVault™) cloud-integrated storage appliances to better assist users in understanding the technologies available in cloud storage appliances on the market today. Cloud Backup appliances provides a simple, efficient, and secure way to move data off site to either public or private cloud storage providers. Using advanced deduplication, compression, and encryption, Cloud Backup enables organizations to eliminate reliance on older, less reliable data protection solutions while improving backup windows and disaster recovery capabilities.

This document is designated for use with Cloud Backup in countries other than China and Russia. China and Russia should use TR-4463 instead.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Overview

1.1 Introduction

This technology overview discusses the technical foundations of NetApp Cloud Backup (formerly AltaVault), to help users understand the technologies available in cloud storage appliances on the market today. This report focuses on core technologies used by the Cloud Backup appliance to keep data protected from end to end, providing the highest level of integrity and recoverability. It provides insight into various deployment scenarios for implementing Cloud Backup and describes how the Cloud Backup appliance works in the context of backup, archive, and disaster recovery. It also discusses how Cloud Backup appliance support can make all the difference in recovering a business environment.

1.2 Audience

The audience for this guide includes NetApp customers, partners, IT architects, technology decision planners, and professional services engineers who are interested in discovering more about the technologies used by the Cloud Backup appliance. NetApp highly recommends that readers have previous experience with backup applications as well as general knowledge about disk systems and storage technologies.

2 Cloud Backup Appliance Overview

2.1 Cloud Backup Introduction

With the never-ending demand to maintain the highest levels of data integrity for increasingly large datasets, companies are increasingly challenged to find effective data protection solutions that balance cost, data protection, and disaster recovery features. Historical approaches such as tape backup and disk-to-disk replication for protecting data and ensuring recoverability in disaster scenarios face enormous constraints. That is because of the amount of human interaction, technical complexity, and costs involved in implementing such solutions to meet recovery requirements.

NetApp Cloud Backup enables customers to securely and efficiently back up data to any cloud at up to 90% lower cost compared with on-premises solutions. Cloud Backup gives customers the power to tap into cloud economics while preserving investments in existing backup infrastructure and meeting backup and recovery SLAs.

The Cloud Backup appliance is a disk-to-disk data storage optimization system with cloud storage integration. It easily integrates with existing backup and archive applications to securely protect critical production data off site without the complexity of using tape management solutions or the cost of using in-house disaster recovery sites and services. The backup server simply connects to the Cloud Backup appliance using SMB, NFS, or OST protocols. In addition, Cloud Backup now supports the ability to directly receive ONTAP® Snapshot™ backups using SnapMirror®, NetApp's snapshot replication engine.

When you back up or archive to an Cloud Backup appliance, it performs inline variable-length deduplication of the backup data and securely replicates data into the cloud. Cloud Backup appliances use the local disk to store enough data for recovery of recent information. For example, with the AVA800 appliance, you can store up to 384TB of deduplicated data per appliance. This mechanism provides LAN performance for the most likely restores. The Cloud Backup appliance then writes the deduplicated, compressed, and encrypted backup data to your public or private cloud storage provider. Cloud Backup appliances also optimize replication restores from the cloud because they move only deduplicated data over the WAN.

Cloud Backup appliances are designed around the need for maintaining the highest data integrity while delivering the performance and costs that companies need in such a backup and disaster recovery solution. Cloud Backup appliances are file-based data deduplication storage devices with SMB, NFS, SnapMirror, and OST connectivity to back up applications to a variety of class-leading cloud storage providers.

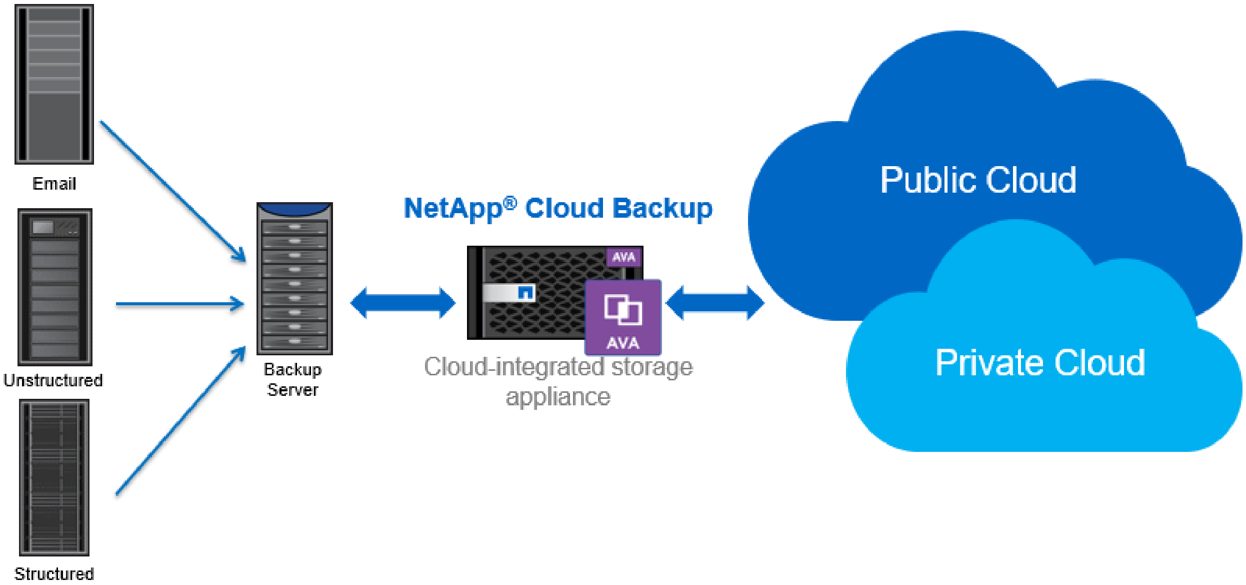
These large, sequential data streams are ingested into Cloud Backup appliances by using multiple 1GbE or 10GbE connections and are inline deduplicated and compressed before they are written to the Cloud Backup local cache and asynchronously replicated through secure TLS connections to cloud storage. Cloud Backup appliances protect data by using class-leading deduplication technologies and scalable and cost-effective cloud storage to provide long-term data storage for backup data. They also implement a local disk cache for immediate restoration.

Cloud Backup appliances are available in a variety of sizes to scale with business requirements and growth. They are also available in virtual format editions for virtualized environments such as VMware vSphere, Microsoft Hyper-V, Linux KVM, and the Amazon EC2 and Microsoft Azure marketplaces for cloud-to-cloud backups. The flexibility of product types offers alternative methods of recovering data in the event of a disaster, when infrastructure and resources might not be available in the same manner as in the lost primary data center.

2.2 Cloud Backup Appliance Deployment Scenarios

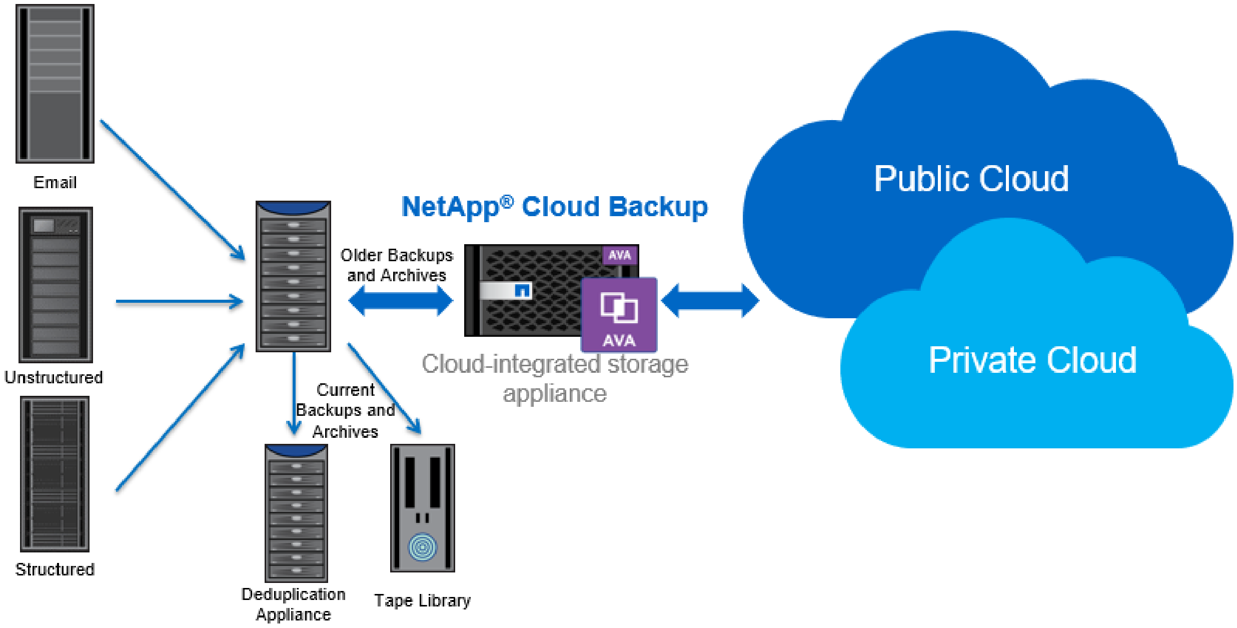
Cloud Backup appliances are designed to be easily integrated into a backup application infrastructure. They can be deployed in a number of scenarios, depending on the scope and size of the environment being protected. For example, a typical scenario places an Cloud Backup appliance directly behind the backup or archive application, protecting that data to the cloud. Data is stored on the appliance cache for quick restore of backups and is maintained in cloud storage for long-term archive, audit, and compliance.

Figure 1) General Cloud Backup deployment example.



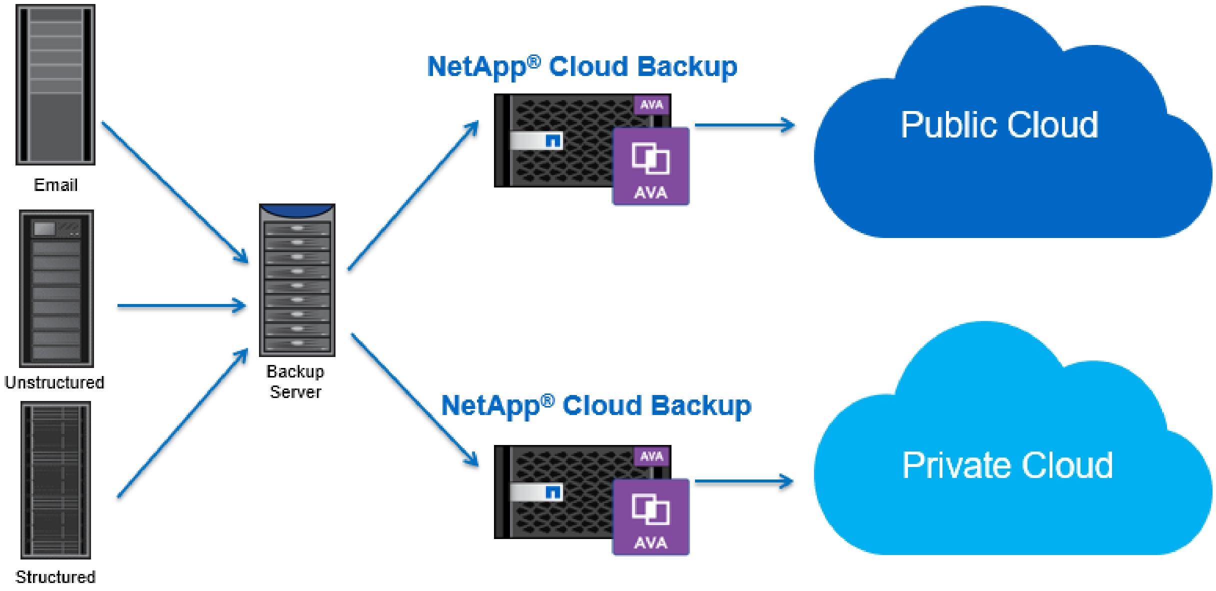
Many organizations, however, have more complex environments. They might have existing disk or deduplicated disk infrastructure configurations for retaining data for short-term requirements and tape for longer-term storage requirements. In these scenarios, companies are typically cost constrained when attempting to add more local infrastructure; or manageability of off-site tape might become a problem as data grows. The Cloud Backup appliance can work seamlessly by inserting itself as a lower tier of storage designed to off-load less critical data from the existing disk storage system so that the disk can be reused for higher-priority data. Cloud Backup can also provide off-site data protection to replace the need to maintain a large tape infrastructure in the data center.

Figure 2) Cloud Backup appliance configured for data tiers.



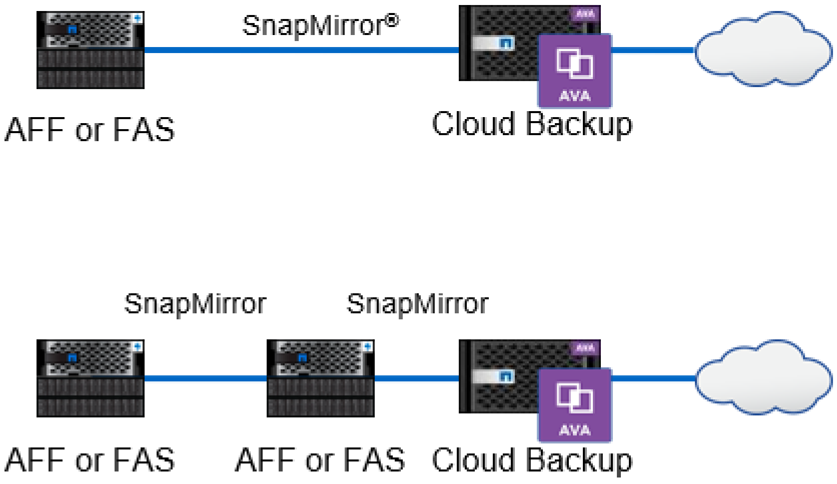
In addition to storage infrastructure requirements, some organizations have specific requirements for different retention rates. In these cases, multiple Cloud Backup appliances might be appropriate to help divide the storage for each priority tier of data. For example, suppose that some data must be protected for long-term audits or government compliance and other data follows normal retention policies. In this scenario, Cloud Backup appliances can be aligned to particular backup or archive policies to keep the data separate, such as one cloud storage target per Cloud Backup appliance.

Figure 3) Cloud Backup appliance configured for multiple data retentions.



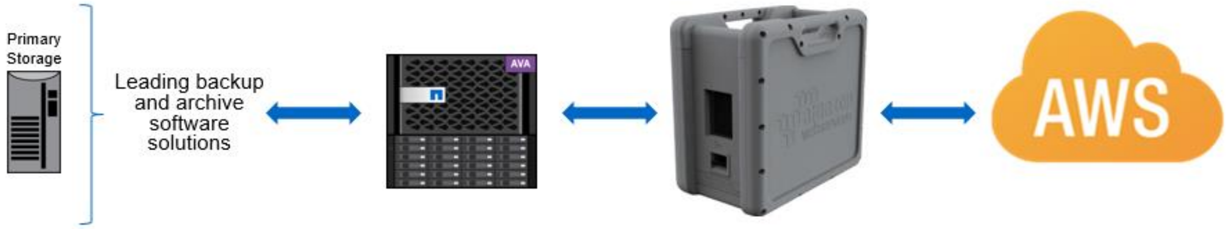
Starting with version 4.3, Cloud Backup can protect data directly from NetApp ONTAP FAS and All Flash FAS (AFF) systems by using Snapshot technology based on SnapMirror replication. The SnapMirror to Cloud Backup feature brings native snapshot capability to Cloud Backup, allowing users to leverage Cloud Backup as an end point for data protection of NAS file services workloads from NetApp FAS or AFF systems. With the ONTAP CLI to manage the data protection workflow, snapshots can be sent through either a direct relationship with a primary NetApp system, or as a secondary mirror of an existing mirrored relationship between two NetApp systems, and protected to the cloud provider of your choice for long-term retention and recovery.

Figure 4) Cloud Backup appliance using SnapMirror relationships.



Cloud Backup version 4.3 also introduces the support for Amazon Snowball, a seeding appliance that can drastically improve initial dataset transfer time to Amazon S3 object storage. As dataset growth continues to increase, pressure on existing internet links to send data to Amazon also increases. Despite Cloud Backup’s industry leading deduplication and compression, organizations can sometimes be challenged by moving very large initial datasets to the cloud. Snowball provides an on-premises storage device that Cloud Backup can use to quickly move large amounts of data to a secured appliance, which is then shipped back to Amazon and loaded into S3 object storage.

Figure 5) Cloud Backup appliance leveraging Snowball.



Regardless of the scenario, Cloud Backup appliances provide a flexible storage point for an organization’s growing data requirements so that users can select the location and retention tier in which the Cloud Backup appliance is utilized.

2.3 Cloud Backup Architecture

Cloud Backup appliances are file-based appliances designed to be easily implemented in an existing backup infrastructure. They offer flexible high-performance storage for backup applications through the following protocols: Windows SMB (also known as Server Message Block); UNIX or Linux NFS (also known as Network File System); NetApp ONTAP SnapMirror; and Veritas OpenStorage (also known as OST). Unlike block-level appliances, Cloud Backup appliances do not require extensive IT architecture redesign, configuration, and implementation to integrate into an existing storage infrastructure. Organizations simply connect the Cloud Backup appliance directly to the network and quickly set up shared storage folders that backup and archive applications can easily point to for subsequent backup operations.

Cloud Backup appliances use proven NetApp enterprise-grade storage chassis engineered to rigorous storage design standards. All Cloud Backup appliances utilize dual power supplies for power redundancy to protect the appliance in the event of an individual power supply failure. Dual boot partitions likewise enable Cloud Backup appliances to power and boot properly, even in the event of an unplanned power outage or failed software upgrade. To provide the required horsepower to efficiently drive the inline deduplication functions to meet the data ingest rate, Cloud Backup appliances utilize dual CPUs, each with multiple cores, and 256GB of low-latency ECC memory.

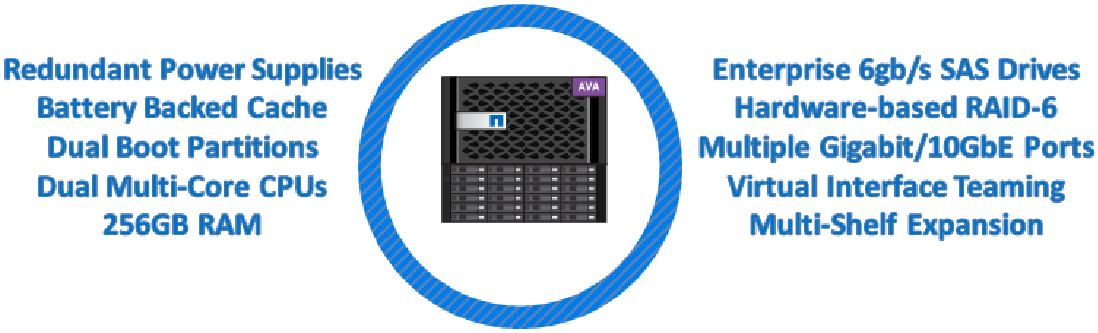
Within the AVA10S disk shelves are enterprise-grade near-line (NL) SAS disks, configured by hardware-based RAID controllers into RAID 6 groups for consistent, reliable data read and write performance as well as data integrity in the event of disk failures. The RAID controllers are protected by a battery-backed unit for uninterrupted operations, even in the event of a power outage. RAID 6 allows up to two drive failures, insuring that the time needed to rebuild the array if one disk fails does not impact data integrity if a second drive fails during the rebuilding process.

Data connectivity is enabled through multiple 1GbE and 10GbE connections, delivering ingest rates of up to 14.1TB per hour from backup applications. In addition, the ports can be aggregated together to form virtual interfaces for easier appliance manageability using the 802.3ad industry standard. The multiple 1- and 10-gigabit connections provide accessibility for Cloud Backup management, as well as for outbound data leaving the Cloud Backup appliance for public cloud storage. Flexibility is provided to allow users to select which interfaces to utilize, depending on the complexity of the user network environment in which the Cloud Backup appliance is placed. VLAN support is offered starting at version 4.2, which allows large organizations to place Cloud Backup within the context of the appropriate backup infrastructure network scope.

Cloud Backup also has significant expansion capability, with the capacity to support up to a total system capacity of 384TB of usable cache, which can manage up to 1.92PB of cloud storage. Assuming deduplication rates of up to 30 times, this means that a single Cloud Backup solution can support over 57PB of logical data in the cloud. By delivering flexible expansion capability, Cloud Backup can grow as business needs expand capacity requirements.

Finally, Cloud Backup appliances include a service processor card to perform platform management. This important feature provides access to Cloud Backup appliances that are having normal run-time problems that prevent regular access through the graphical user interface. It also helps administrators centrally manage and monitor Cloud Backup appliances located in remote sites.

Figure 4) Cloud Backup appliance features.



2.4 Cloud Backup Appliance Data Integrity and Security

One of the most important facets of any reliable data protection appliance is its ability to provide end-to-end data integrity while maintaining a high level of internal security for that data as long as it is owned and managed by the appliance. Cloud Backup appliances offer many layers of integrity checks, including transactional consistency logging and checksum verification at every stage of ingest and recovery.

As backup and archive data is ingested into a Cloud Backup appliance; it is segmented in real time into small chunks and placed into Cloud Backup memory. At the same time, fingerprint labels are created to individually identify these chunks of data. The first checksum is generated to verify that the data written to memory represents the incoming data stream received over TCP/IP from the backup application.

The Cloud Backup appliance then hashes these fingerprint labels and compares these hashes to hashes that were previously written by Cloud Backup. Because Cloud Backup performs variable-byte-length deduplication using one of the finest granularities in the industry, the maximum amount of data deduplication can be achieved. If a match does not occur, then Cloud Backup compresses the data with the LZ4 (Lempel-Ziv) compression algorithm, encrypts the data with 256-bit AES encryption, and flushes the resulting data container called a slab to disk. Because encryption is performed in memory, the data is considered secure at rest after being written to disk. A checksum is generated against the written data to disk, and it is checked when slabs are loaded for future comparisons.

At the same time the new segment is written, an entry called a label map is created in memory. In the event of a restore request, the label map allows the data stored by an Cloud Backup appliance to be decoded back to the source application. This label map entry is similarly flushed to disk and the checksum is verified as well. At any point in time, data that is accepted into an Cloud Backup appliance can be recovered back to its original form.

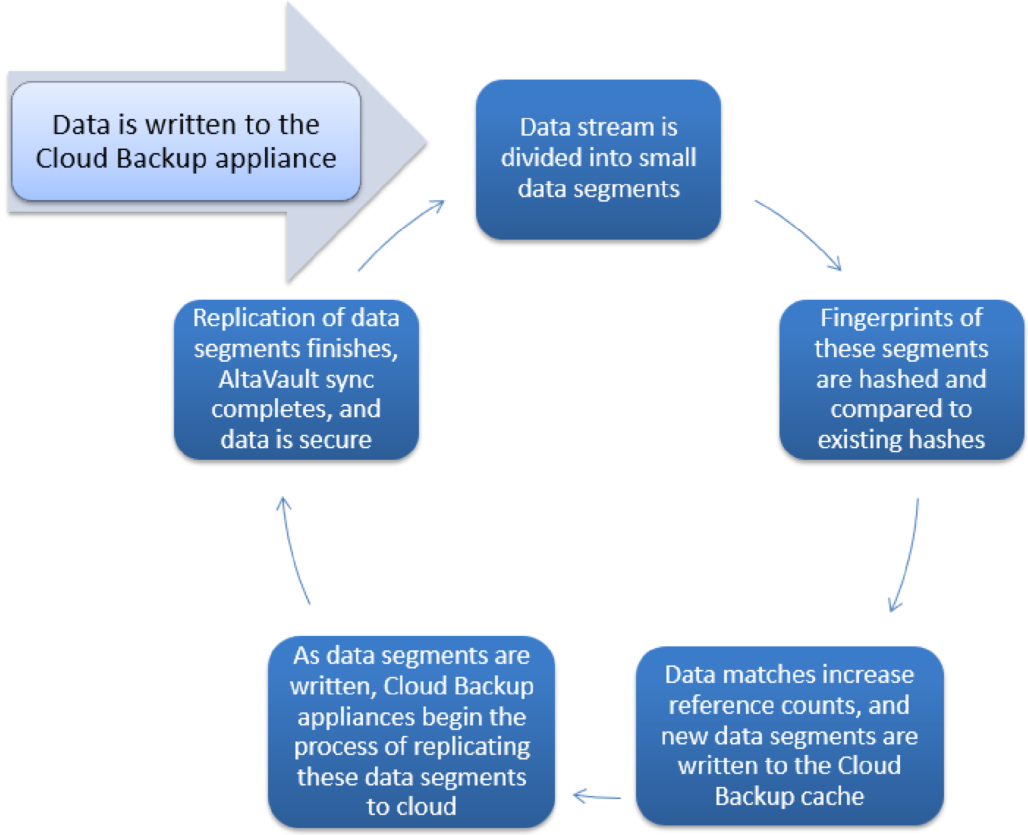
Slabs and associated metadata written to the Cloud Backup appliance are also asynchronously replicated to cloud storage through TLS v1.1 or 1.2 connections for disaster recovery and long-term backup and archive. Replication consistency to a cloud provider is validated by first transferring the slab data, then performing a checksum to verify that the content arrived, then sending the corresponding metadata for the slab, and finally checksum verifying the metadata. Performing replication in this controlled fashion offers crash consistency and rollback mechanisms if the replication process is interrupted during the sending of data. Cloud Backup appliances can examine and confirm unfulfilled transactions and, in the event of partial slab synchronization, delete the problematic data and resend the data.

In addition to data security at rest and in flight, Cloud Backup supports the Key Management Interface Protocol (KMIP) protocol, allowing Key Management Servers (KMS) to offload Cloud Backup's secure information, such as the encryption key and cloud security credentials. This procedure centralizes security and also reduces the risk of key information being compromised or improperly accessed. Supported versions of KMS are documented in IMT.

To protect access to the appliance, Cloud Backup offers role-based permissions to control user access to certain actions, such as exporting the Cloud Backup configuration, changing cloud credentials, and adding NFS shares. In addition, Cloud Backup appliances starting at version 4.3 also support Windows Active Directory domain authentication when logging into Cloud Backup. This seamless integration simplifies administration based on Windows AD permissions for users who manage Cloud Backup appliances.

To provide additional security control for federal organizations, Cloud Backup supports government clouds, such as Azure Government Cloud, and Amazon GovCloud, as secure cloud targets for Federal and higher education data protection workloads. With Amazon, Cloud Backup also supports Amazon's Security Token Service (STS). This web service enables finer control of security credentials used when authenticating with Amazon S3, S3-IA and Glacier storage targets.

Figure 5) Cloud Backup appliance data flow.


Data is written to the Cloud Backup appliance
Data stream is divided into small data segments
Fingerprints of these segments are hashed and compared to existing hashes
Data matches increase reference counts, and new data segments are written to the Cloud Backup cache
As data segments are written, Cloud Backup appliances begin the process of replicating these data segments to cloud
Replication of data segments finishes, AltaVault sync completes, and data is secure

All transactions performed on an Cloud Backup appliance use a transaction log that records the state of data and the actions taken. It asynchronously replays these same changes with cloud storage through separate threads. This allows the Cloud Backup appliance to be consistent in transactions and provides crash-consistent transactional recovery capabilities for the local storage and cloud replicated copy if power or other unexpected hardware outages occur.

In addition to the exhaustive set of practices used to protect data being written on Cloud Backup appliances and to the public cloud, Cloud Backup appliances also provide additional data verification tools to perform manual verifications of data. Those tools are the online filesystem check and verify, which correspond to file system checks for local systems and the cloud, respectively. Online filesystem check diagnoses the integrity of the disk storage file system used by a Cloud Backup appliance. It provides a thorough check of not only the metadata, but also the data content itself. If damaged content is discovered, an option is available to retrieve data from the cloud copy. Verify checks replication consistency and is available to validate that replicated data is indeed in the cloud storage target specified by the Cloud Backup configuration.

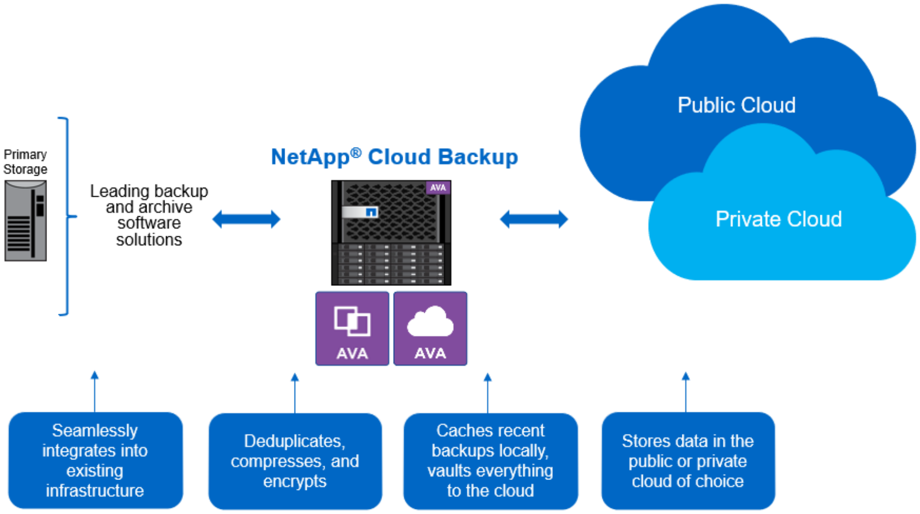
Cloud Backup appliances also offer the capability to encode data that conforms to the Federal Information Processing Standard (FIPS) 140-2 level 1 standard. The NetApp Cryptographic Security Module (NCSM) that Cloud Backup appliances use is validated with the standard to guarantee that data cannot be compromised by insecure cryptographic algorithms. When the appliance is paired with a compliant cloud storage provider, users can rest assured that data will be maintained with the highest level of security. This is important for various business sectors including government, legal, and healthcare.

Further details on security related features are available in the TR-4405: Cloud Backup Security Guide.

2.5 Cloud Backup Appliance Ecosystem Integration

Purpose built specifically for backup and archive markets, Cloud Backup appliances can claim best-in-class backup application integration. With integration for all leading backup, database, and virtual backup solutions as well as with all the leading cloud storage providers, Cloud Backup appliances offer the widest range of coverage in the industry today, covering all leading backup software including NetApp SnapMirror, and cloud storage providers including all the major hyperscaler clouds.

Figure 6) Cloud Backup appliance ecosystem.


Primary Storage
Leading backup and archive software solutions
NetApp® Cloud Backup
AVA
AVA
AVA
Public Cloud
Private Cloud
Seamlessly integrates into existing infrastructure
Deduplicates, compresses, and encrypts
Caches recent backups locally, vaults everything to the cloud
Stores data in the public or private cloud of choice

Cloud Backup appliances have a unique understanding of the nuances of each backup application, and they optimize data transfers based on the methodologies of the backup application. For example, Commvault inserts tape markers into its backup streams. If not handled correctly by the deduplication engine, this can affect data deduplication performance. Similarly, Veeam Backup & Replication uses specific data read routines when performing vPower-based VM recovery operations. Being able to detect and optimize performance for this type of activity is what separates Cloud Backup appliances from others when performing activities with the backup application.

In addition to being fully integrated with backup applications, Cloud Backup appliances are among the few to also be certified with the backup applications. Cloud Backup appliances received Tivoli Ready certification from IBM for the Tivoli Storage Manager backup family of products. They also are on the hardware compatibility list (HCL) for Veritas Backup Exec and Enterprise Vault and are Veeam Verified for Veeam Backup & Replication. These certifications are living partnerships with these vendors, and they guarantee that the highest level of integration and testing has been done to qualify the cross-compatibility of the two products in business environments. As new backup application versions are brought to market, Cloud Backup performs regular release cycle validation to ensure that the new backup versions are supported for use. Cloud Backup identifies the supported versions of backup software in IMT.

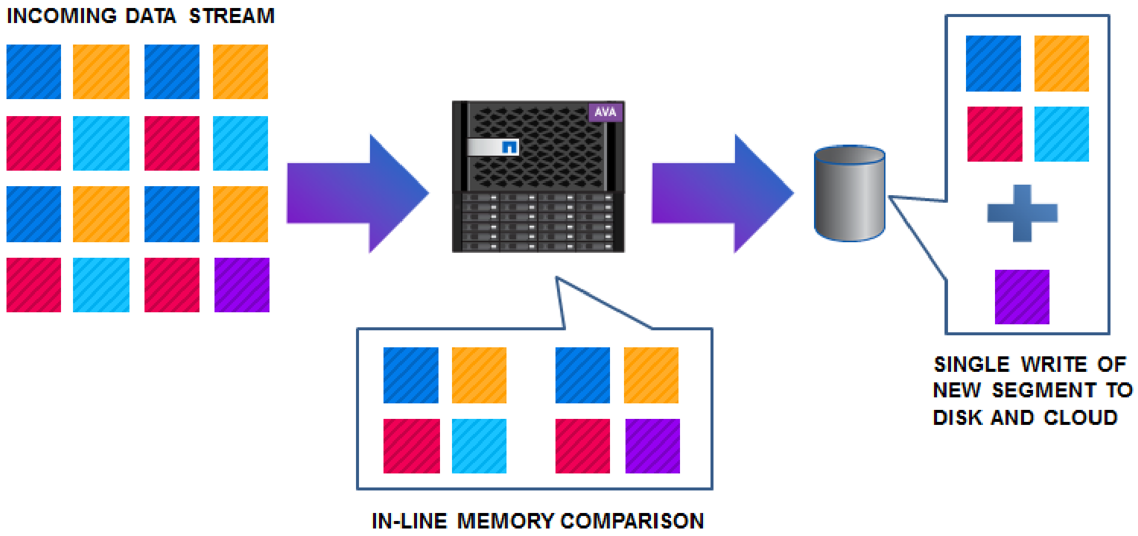
3 Industry Deduplication Overview

Deduplication lies at the heart of any cloud storage appliance or deduplication storage device. When evaluating deduplication storage platforms, it's important to understand the technology design and what it can provide from a storage perspective. Today's deduplication storage products implement one of two major types of deduplication processes and, within each, various methods that achieve the deduplication.

3.1 Inline Deduplication

The first type of deduplication process is inline deduplication, or in-band deduplication. The deduplication engine in this design examines data as it is received in real time by the storage unit. It then deduplicates by using memory to quickly examine the source data stream against previously examined data to determine whether the source data is a duplicate of existing deduplicated data stored on the storage device. If the source data is a duplicate, then only a small reference counter is incremented for that data and the data is discarded. If the source data is uniquely different, then that data is written to the storage unit at the cost of a single write and read.

Figure 7) Cloud Backup appliance inline deduplication.


INCOMING DATA STREAM
AVA
SINGLE WRITE OF NEW SEGMENT TO DISK AND CLOUD
IN-LINE MEMORY COMPARISON

The main benefit of performing data deduplication with this method is that data is stored in its most efficient format on the storage unit, resulting in the largest amount of available space for source data from the user environment. Additional important benefits include the performance cost of getting the data to disk, which in this example requires only one write operation to store the data and one write to validate the data. It also includes the ability to immediately replicate that data, such as to the public cloud, with Cloud Backup appliances, and the absence of a deduplication window, such as with postprocess deduplication storage devices. In the past, storage write performance used to degrade with inline deduplication because of the CPU overhead required to perform deduplication. However, the current generation of powerful hardware systems, which include large amounts of RAM, powerful multicore CPUs, and flash memory, completely resolves this concern.

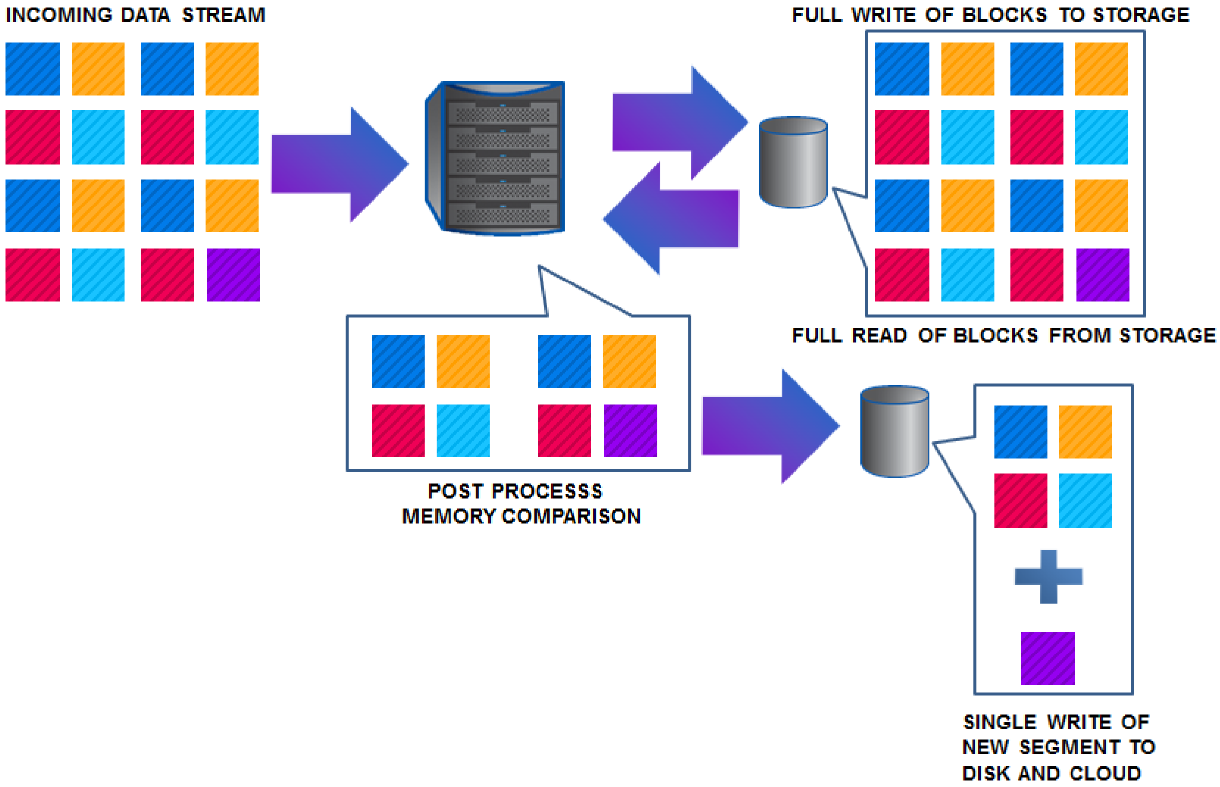
3.2 Postprocess Deduplication (Other Vendors)

The second type of deduplication process that an appliance can use is postprocess deduplication, or out-of-band deduplication. Although Cloud Backup does not support postprocess deduplication, it's important to understand how it differs from inline deduplication. In this type of storage device, the source data must be completely written to the deduplication storage unit's disk storage in its original nondeduplicated form.

Only after the data has been completely written to disk can the deduplication process start (commonly referred to as a postprocess deduplication window). This includes reading back the data from disk into memory, reducing the amount of data by examining the source data to compare it to previously deduplicated data stored, and finally writing new unique deduplicated data back to disk. This can be very time consuming as well as wasteful of space, because data must be written twice and read three times during this process:

- Written once in source form during the initial ingest
- Read from disk storage for validation against the source data
- Read again in source form during postprocess deduplication
- Written back to disk in deduplicated form if the data is unique
- Read from disk storage after the write to validate the deduplicated data

Figure 8) Postprocess deduplication.



Because of the very heavy I/O requirements, postprocess deduplication systems typically require many more CPU and disk resources than inline deduplication storage systems, such as SSDs and additional storage disks, to handle and store the incoming source data stream in its nondeduplicated format. These resources are realized as increased capital costs in equipment, energy, and space expenditures, as well as in performance costs if the system is unable to perform deduplication of the source data during the now required postprocess deduplication window. Additionally, the ability to protect the system through replication can be significantly affected because data remains unprotected while waiting for the postprocess deduplication to complete before replication can begin. This can effectively limit the real capacity of data that the unit is able to store and protect, even if there is sufficient storage for much greater quantities of data. That is because the source data ingest and postprocess window effectively reduce the replication window.

3.3 Additional Deduplication Methodologies

Beyond the deduplication technology type used, another important consideration to evaluate is whether the source data itself is being deduplicated to its most efficient deduplicated size. Deduplication algorithms typically use two methodologies to examine data:

- Fixed-length or variable-length segments
- Size of a segment

3.4 Fixed-Length Versus Variable-Length Segments

In many deduplication strategies, vendors implement algorithms to divide the data for deduplication comparison. For example, consider the blocks of data representing a common English pangram phrase broken into 8-character segments. (A *pangram* is a sentence that contains all the letters of the alphabet.) Fixed-length algorithms examine source data based on fixed block sizes and find matches to deduplicated data. Although this is easier to implement and potentially faster to execute on a data stream, it does not find nearly as much data to reduce as can be found with more robust deduplication strategies. Additionally, changes to the data within those blocks can lead to increased storage consumption, because data shifts within the blocks can greatly increase the amount of unique data in subsequent writes to the deduplication storage device. Changing the first word from “A” to “The” in the following example affects all of the segments because data shifts and none of the segments is the same as previously.

Figure 9) Original data segments.

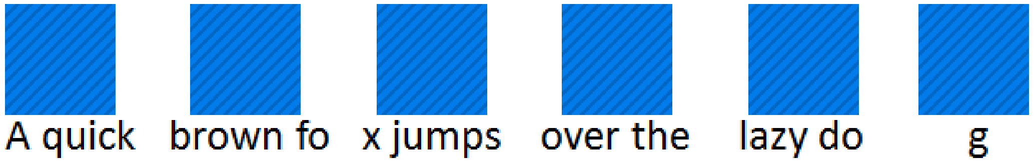
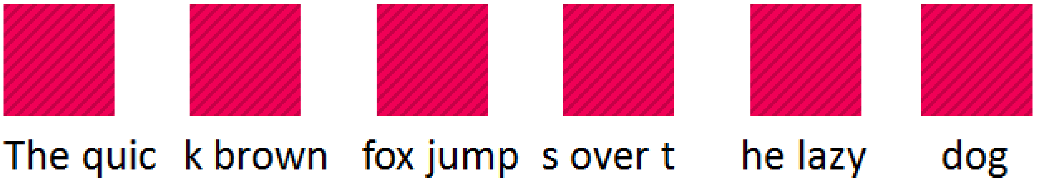
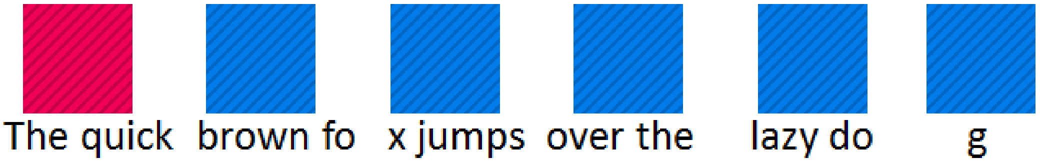


Figure 10) Fixed-length segments after a change in data.



On the other hand, deduplication systems such as Cloud Backup appliances that implement variable-length deduplication algorithms can result in greatly increased efficiency in storage device utilization. This occurs because Cloud Backup appliances can detect the subtle changes in the data stream and account for the shift of data in the blocks that would otherwise reduce duplicate matches. The result of variable length deduplication is increased deduplication matches and overall less storage required to hold the changed data blocks. In this variable length example, the change from “A” to “The” does not affect all of the segments because only the first segment was not previously seen. In this case, it grows in variable size to accommodate the additional characters. All the other segments are still the same, and thus are considered duplicate segments and are not stored as new segments.

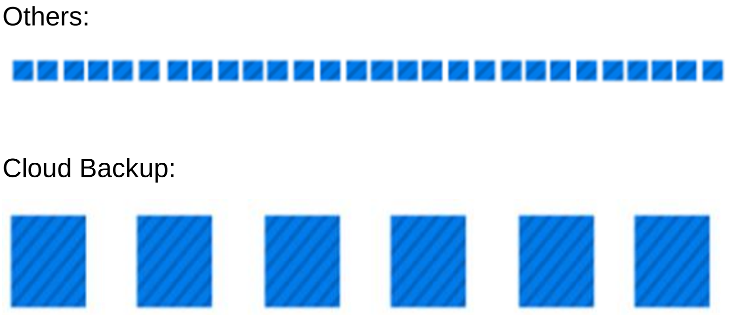
Figure 11) Variable-length segments after a change in data.



3.5 Data Segment Size

Another consideration regarding deduplication functionality is the size of the data segment being examined. In general, the larger the data segment size, the more potential there is for unique data. Conversely, the smaller the data segment size, the more potential there is for duplicate data. Vendors typically deduplicate data by using multikilobyte data segments, ranging from 8KB to 256KB. Cloud Backup appliances use very granular data segments starting at 512 bytes, which, with variable-length segments, improves the ability to deduplicate the data efficiently

Figure 13) Data segment size.



3.6 Deduplication Feature Summary

The overall deduplication performance of a cloud storage gateway solution can be measured as a relationship between the deduplication process and the type of length algorithm, segment size, and comparison methods used. Although good deduplication functionality is important in increasing the amount of source data that a user can store and protect on a deduplication storage appliance, with cloud storage gateways this becomes even more important. That's because it also affects the amount of stored data in the cloud, which in turn affects the overall recurring cost per month of cloud storage consumed. Because they meet this requirement for reducing storage costs, Cloud Backup appliances take the best feature of each category and deliver the best deduplication storage appliance experience for users who need to protect data to cloud storage.

Table 1) Deduplication feature comparison.

Feature Comparison	Best Performing for Backups	Reason
Inline versus postprocess deduplication	Inline deduplication	Fewer reads/writes to disk, fewer resources consumed, allows larger backup windows
Fixed-length versus variable-length segments	Variable-length segments	Fewer changed blocks to write, increased storage utilization
Large versus small data segment size	Small data segment size	Finer granularity for greater deduplication efficiency

4 Additional Cloud Backup Features

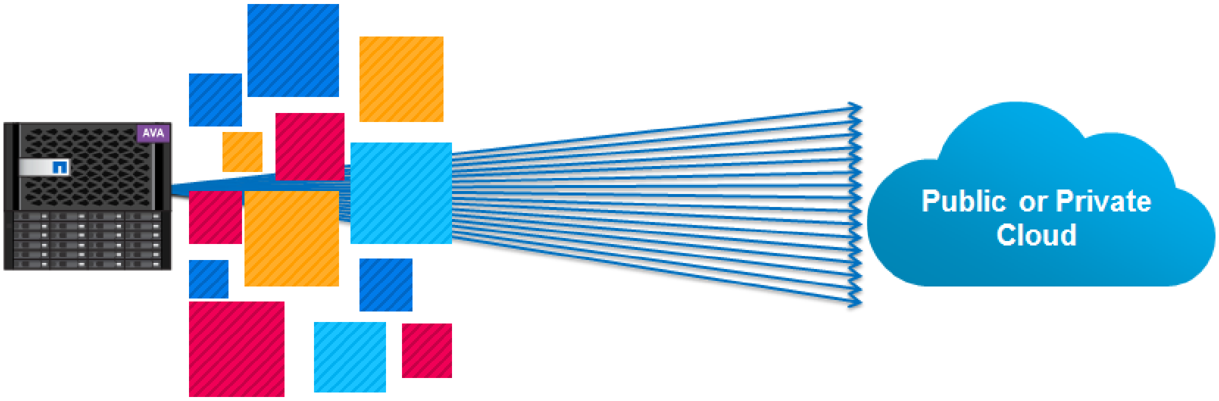
In addition to its class-leading enterprise capabilities, end-to-end data integrity and security, and superior deduplication functionality, Cloud Backup has several other functional capabilities that make it an exceptionally strong fit for today's increasingly large backup and archive workloads.

4.1 Cloud Backup Appliance Replication Capabilities

Replicating to the public cloud storage provider is an important piece of a cloud storage gateway's capability to protect a business environment. Cloud Backup appliances contain advanced WAN optimization capabilities to optimize replication efficiently and effectively to cloud storage while providing users with options to throttle bandwidth usage and to schedule when replication can run.

Cloud Backup appliance replication is done by using cloud provider APIs, which are typically HTTP/REST-based protocols. These protocols are designed for multisession ingests of small data segments that are stored collectively in a single target location, commonly referred to as a cloud bucket. NetApp has worked with each supported cloud provider to carefully tune networking parameters to optimize throughput capabilities to cloud storage. This process also dynamically and intelligently allocates multiple threads to transmit data to and from the cloud provider based on networking performance, the data in the queue for replication, and the cloud provider selected. Transactions are batched together for maximum throughput and efficiency of the replication threads. Objects sent are typically anywhere from several KB up to a few MB in size. If there is a problem replicating an object, replication automatically pauses and alerts the user that replication was unable to maintain connectivity to the cloud storage bucket.

Figure 12) Dynamic replication thread allocation.



Replication itself is automatic and provides an estimated time of completion when running. This information is important if restrictions are placed on bandwidth or if the user is employing scheduling windows. By establishing a replication completion date and time, users know clearly when data is in sync with cloud storage and protected.

4.2 Cloud Backup Appliance Eviction Process

The Cloud Backup appliance cache is designed to serve as a temporary data storage location for the most recent backup and archive data in case an immediate restore is needed. This immediate recovery range can be anywhere from weeks to a month, depending on the needs of the business. Because Cloud Backup is designed to recover more than 90% of the typical use cases, most restores are performed solely from the local cache.

Over time, data that is ingested and protected by a Cloud Backup appliance might not fit completely in the local cache of the appliance. When the appliance storage fills up, eviction might be required to remove the least recently used but previously replicated data segments from the Cloud Backup cache so that new data can be received. Evicted data segments continue to exist as a cloud-only copy in cloud storage, and the backup application continues to see the data segments as if they were local to the Cloud Backup appliance.

Eviction is performed when the Cloud Backup appliance reaches 90% of utilization and removes the least recently used data segments until the cache falls back under 90% utilization. In addition, policy-based eviction can be configured on a given share or export, providing expedited eviction processing on a share ahead of any recently used data segments. This feature is useful for the archive data use case and for data that is not needed for quick recovery and is needed only in the event of a disaster or long-term actions such as an audit. Cloud Backup appliances perform eviction automatically, so no user interaction is required to make sure that the Cloud Backup cache has sufficient space for new data. Typically, eviction occurs several weeks after a Cloud Backup appliance is deployed. Therefore, the least recently used blocks being evicted represent much older backups that are typically not required for most common restore requests.

If the backup application requires the data for recovery, the deduplicated segments of the file that are not in the Cloud Backup cache can be recalled from the cloud to the Cloud Backup appliance. This process is done invisibly, so that all restores for the backup application look like they are coming back from the local disk on the Cloud Backup appliance. It is likely that the data being restored still has many of the segments on the Cloud Backup cache, because those segments are duplicates of other segments that are more frequently referenced. Therefore, the amount of cloud access needed to recover evicted segments is typically very low. This minimizes recovery performance for restores.

To give users further insight into eviction, Cloud Backup appliances provide comprehensive eviction views in the GUI. This enables users to identify the remaining cache available in the Cloud Backup appliance, the amount of data evicted from the Cloud Backup cache over time, and the average data age of the evicted bytes of data. It also enables users to clearly understand which local recovery time frame is now available in the Cloud Backup cache.

4.3 Cloud Backup Appliance Disaster Recovery Capabilities

Cloud Backup appliances can help organizations recover business servers and business processes more quickly than with traditional tape backup solutions. As a disk-based deduplication solution, recovery for the most common data loss scenarios (which are typically within 24 to 48 hours of the data loss event) occur quickly because the data is being read back from the local Cloud Backup appliances. Cloud Backup appliances maintain a local cache size that varies from 2TB all the way up to 384TB of usable data. This range typically allows a localized recovery of data aged anywhere between one day and a couple of months. If not all of the data is in the local cache, the Cloud Backup appliance recalls only the segments of the missing data needed from the cloud provider to complete the recovery. These data segments are typically from 1MB to 4MB in size, and they thus save the company money by not having to recover unnecessary data from cloud storage to complete the restore.

A larger outage might require a significant restore action, which could include recovering the Cloud Backup appliance and the backup infrastructure itself in addition to the production business systems. These are typically true DR scenarios in which the entire working infrastructure is lost, such as in a fire or flood. The effect on the business can be tremendous, including impacts such as lost productivity, lost sales, and the inability to generate products to market. In these scenarios, rapidly meeting a recovery time objective (RTO) and minimizing the recovery point objective (RPO) are essential to business continuity.

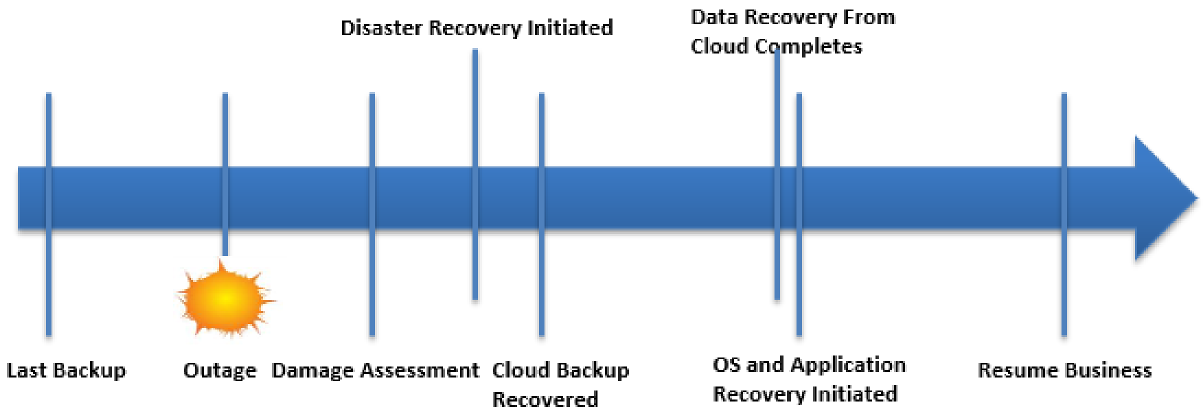
Cloud Backup appliances can be deployed as a virtual instance in the business's secondary data center as a cold standby to perform recovery operations. The appliance model deployed for DR does not need to directly match the original Cloud Backup model, although NetApp recommends that matching if you intend to perform subsequent backups after DR.

In a cold standby DR scenario, the secondary Cloud Backup appliance uses a wizard-driven process to aid in recovering the configuration from the original Cloud Backup appliance. This is followed by steps to recover the backup application namespace and then prepopulating the most recently backed up data from the cloud (typically from the last day or week). Because Cloud Backup appliances maintain the association between data stored on volumes as it pertains to the backup application and the data it needs to recover from the cloud, they use smart prefetch algorithms to efficiently restore the needed data to improve the overall recovery process. If the Cloud Backup appliance is available at the secondary data center site, the process can be started within minutes of a disaster. For example, it is possible to deploy and use the virtual Cloud Backup appliance in evaluation mode for disaster recovery purposes without any license. This appliance can be downloaded from the NetApp Support website.

Figure 13) Cloud Backup appliance DR recovery timeline.

This type of recovery can save enormous amounts of time compared to traditional tape-based recovery, in which tape volumes must be identified, moved to the secondary data center site, mounted, and then read. It also reduces the risks associated with physical volume movement (that is, tape corruption, tape misplacement, tape security, and so on). By combining it with the best practice of securing the backup application catalog or backup database to the Cloud Backup appliance, businesses can further reduce RTO by having the most recent backups available almost immediately. By allowing data recovery to begin in almost real time in response to a major outage event, users can quickly return to normal business operations.

Figure 14) Traditional tape DR recovery timeline.

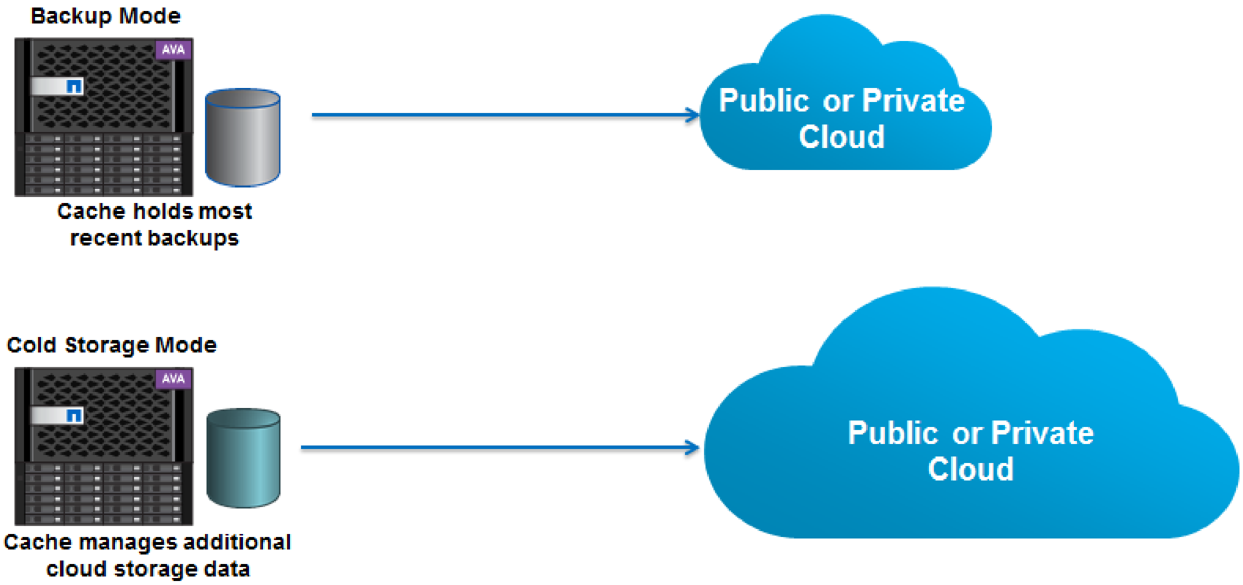

Disaster Recovery Initiated
Data Recovery From Cloud Completes
Last Backup
Outage
Damage Assessment
Cloud Backup Recovered
OS and Application Recovery Initiated
Resume Business

4.4 Cloud Backup Appliance Cold Storage Mode

Government regulations, legal requirements, and audit compliance have vastly changed the data retention requirements that businesses must adhere to. Regulatory statues mandate that companies must regularly archive data for a minimum of 7 years. These regulations cause a huge strain on enterprise data protection budgets to manage cold, or infrequently accessed, data such as audio recordings, medical images, completed projects, and legal documents.

To assist companies in making the best use of the cloud scale and cost efficiencies of elastic storage to reduce management costs for this infrequently accessed long-term data, Cloud Backup can be configured in cold storage mode. Note: Cold storage mode is not intended for use with backup applications, nor with SnapMirror operations. In this mode, the Cloud Backup appliance cache is used to store more metadata and less deduplicated data than in backup mode. This significantly expands the volume of cold data managed in the cloud. Because less deduplicated data can be stored, eviction of data to the cloud occurs more frequently. Eviction also expels the corresponding deduplication information when the deduplicated data is evicted. As a result, most restores of data are done from the cloud instead of from the local cache, and the overall deduplication rate is typically lower than that in backup mode. However, when configured in cold storage mode, Cloud Backup can protect 60 to 250 times more data than in backup mode, or, in terms of object counts, more than 1.3 billion file objects.

Figure 15) Backup and cold storage modes.



4.5 Cloud Backup Share Policy Management

Several policy management options are available to offer users flexibility in how data is maintained on Cloud Backup. Each SMB, NFS, or OST share can be configured with the following options:

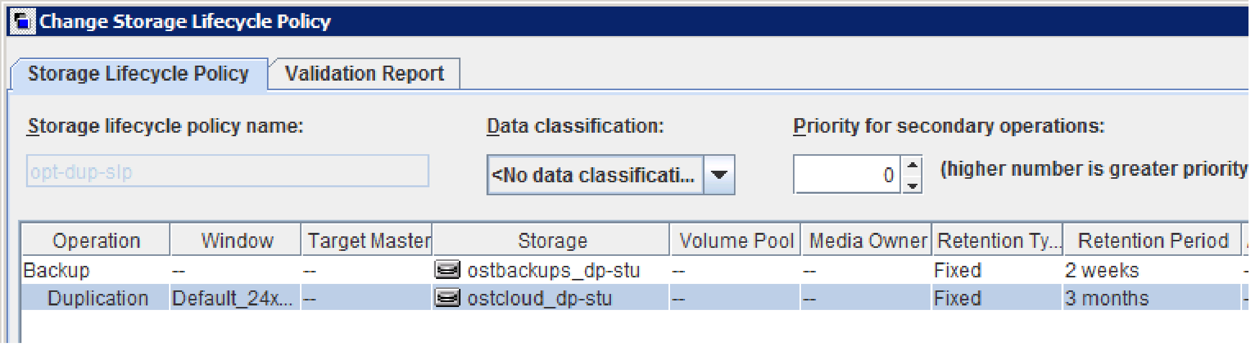
- **Pinning.** With this option, important data that cannot wait for recovery from cloud storage can be made available permanently on the Cloud Backup cache. Data is never eligible for eviction from the Cloud Backup appliance cache.
- **Early Eviction.** When Cloud Backup needs to make additional space available on the cache for new backups, shares with the least important data can be enabled for early eviction. Data in these shares is the first eligible for removal during eviction. Combined with pinning and the normal Cloud Backup eviction mechanism, users effectively have three tiers of data priority on the Cloud Backup appliance – critical, normal, and least important.
- **Disable Deduplication.** Most datasets that Cloud Backup receives are backup workloads, which can be repetitive in nature and very efficient to deduplicate. However, workloads such as archives, which do not see repeating patterns, may not benefit from deduplication. These datasets can be written to a share that is not enabled for deduplication, thereby preserving system resources to use for deduplicating datasets that can benefit from Cloud Backup deduplication.
- **Disable Compression.** Similar to disabling deduplication, disabling compression for shares with data types that are already compression efficient can allow Cloud Backup to better use resources for

datasets that can benefit from Cloud Backup compression. Examples include photo image JPG files, video MP4 files, and archives such as GNU zip files.

4.6 Cloud Backup Integration with OST

Veritas NetBackup can communicate directly with Cloud Backup by using the OST protocol and, through this protocol, allow NetBackup and BackupExec to manage the storage lifecycle policy of backup copies independently in the Cloud Backup cache as well as in the cloud.

Figure 16) Storage lifecycle policy with OST.



Change Storage Lifecycle Policy

Storage lifecycle policy name: opt-dup-slp

Data classification: <No data classificati...

Priority for secondary operations: 0 (higher number is greater priority

Operation	Window	Target Master	Storage	Volume Pool	Media Owner	Retention Ty..	Retention Period
Backup	--	--	ostbackups_dp-stu	--	--	Fixed	2 weeks
Duplication	Default_24x...	--	ostcloud_dp-stu	--	--	Fixed	3 months

4.7 SnapMirror to Cloud Backup

With the SnapMirror to Cloud Backup feature available starting in Cloud Backup 4.3, storage administrators and IT generalists can perform Snapshot copy-based data protection operations for Windows and UNIX file shares to and from public or private cloud object stores. The feature integrates NetApp ONTAP and Cloud Backup cloud-integrated software.

- ONTAP provides storage management, and initiation and management of protection operations.
- Cloud Backup provides storage optimized space in cloud storage providers for snapshots of file share data from ONTAP.

SnapMirror to Cloud Backup enables administrators to back up NFS and SMB NAS file services (NFS v4, SMB3) and home directories to the cloud by using Cloud Backup cloud provider connections.

Advantages of using SnapMirror to Cloud Backup include:

- Storage-efficient incremental forever snapshots with ONTAP
- Storage-efficient backups of snapshots to cloud storage with Cloud Backup
- Data protection flexibility on premises in the cloud
- Optional extended retention through Cloud Backup for 10 years of daily snapshots

ONTAP SnapMirror backup operations to Cloud Backup are performed with an XDP (extended data protection) relationship. The functionality is similar to the former NetApp SnapVault® type relationships where snapshots are retained on the destination separately from the source, and are not a mirror of snapshots. After the relationship between ONTAP and Cloud Backup is created and initialized, a baseline transfer occurs, sending the initial data from the volume snapshot. After the initial baseline transfer of data, subsequent updates send only the changed blocks between the snapshots, which gives block-level incremental updates. Data protection policy management continues to be handled by ONTAP as it is done today. Optionally, it is possible to extend retention through Cloud Backup, which allows up to 10 years of daily snapshots to be stored.

4.8 Cloud Backup Software Upgrades

All Cloud Backup models receive regular software upgrades, allowing users to maintain currency across clouds, backup applications, and KMIP providers as well as to take advantage of enhancements that are continually developed and brought to market. Software upgrades are free to download on current release tracks from the NetApp support website and are easy to install via the upgrade GUI page. Cloud Backup software upgrades can be scheduled to be performed during upgrade change windows to minimize their impact on operations. Because Cloud Backup maintains dual boot partitions, it can roll back to the previous software version if a software upgrade causes an unintended or serious issue.

Cloud Backup appliances include sophisticated technologies to help diagnose and resolve problems quickly. Every Cloud Backup appliance comes with built-in detailed system monitoring to identify whether a serious problem has developed in any of its core functions. Users are alerted to one of the three prominent status values representing the system state: green for healthy, yellow for degraded, and red for critical. Users can also elect to receive notifications by e-mail when the system becomes critical, such as for a bad disk or replication service failure.

User logs and system logs can be easily viewed in the user interface. They are color coordinated to the status values to help users quickly identify key messages about the problem that a Cloud Backup appliance might be experiencing. In addition to taking actions regarding Cloud Backup appliance changes in state, users can also have Cloud Backup appliances capture and automatically submit system diagnosis information to NetApp Support for diagnosis. There is no need to download information and manually send it to Support.

Finally, Cloud Backup appliances have NetApp AutoSupport® capability for improved support. AutoSupport sends daily reports of Cloud Backup health status along with Support logs to help troubleshoot problems. NetApp Support uses this information to help troubleshoot the Cloud Backup appliance more effectively, and to proactively notify customers about such things as available updates and critical notifications. For critical Cloud Backup events, AutoSupport triggers automatic case generation to ensure that problem analysis and resolution can begin immediately, even if the user is not aware of the problem.

Where to Find Additional Information

To learn more about the information described in this document, refer to the following documents and/or websites:

- NetApp Cloud Backup (formerly AltaVault) product page
  https://www.netapp.com/us/products/cloud-storage/cloud-backup.aspx
- NetApp Cloud Backup (formerly AltaVault) Resources page
  http://mysupport.netapp.com/altavault/resources

Version History

Version	Date	Document Version History
Version 1.0	May 2015	Initial version
Version 1.1	August 2015	Updated for 4.0.1 release
Version 1.2	November 2015	Updated for 4.1 release
Version 1.3	April 2016	Updated for 4.2 release
Version 1.4	January 2017	Updated for 4.3 release

Version	Date	Document Version History
Version 1.5	April 2017	Updated for 4.3.1 release
Version 1.6	November 2017	Updated for 4.4 release
Version 1.7	March 2018	Updated for 4.4.1 release, Cloud Backup, SnapMirror update

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Copyright Information

Trademark Information

TR-4427-0318