Technical Report

NetApp AltaVault Cloud-Integrated Storage Appliances

Best Practices for Disaster Recovery

Christopher Wong, NetApp
November 2017 | TR-4420

Abstract

This guide outlines the considerations and best practices for using NetApp® AltaVault™ cloud-integrated storage appliance appliances to prepare for and perform disaster recovery. AltaVault appliances provides a simple, efficient, and secure way to off-site data to either public or private cloud storage providers. Using advanced deduplication, compression, and encryption, AltaVault enables organizations to eliminate reliance on older, less reliable data protection solutions while improving backup windows and disaster recovery capabilities.

TABLE OF CONTENTS

LIST OF FIGURES

1 AltaVault Overview and Disaster Recovery Introduction

1.1 Executive Overview

NetApp AltaVault storage enables customers to securely back up data to any cloud at up to 90% lower cost compared with on-premises solutions. AltaVault gives customers the power to tap into cloud economics while preserving investments in existing backup infrastructure and meeting backup and recovery SLAs. AltaVault appliances simply act as a network-attached storage target within a backup infrastructure, enabling organizations to eliminate their reliance on tape infrastructure and all of its associated capital and operational costs, while improving backup windows and disaster recovery capabilities.

It is simple to set up the AltaVault appliance and start moving data to the cloud in as quickly as 30 minutes, compared to setting up tape or other disk replication infrastructures, which can take days. Leveraging industry-leading deduplication, compression, and WAN optimization technologies, AltaVault appliances shrink dataset sizes by 10x to 30x, substantially reducing cloud storage costs, accelerating data transfers, and storing more data within the local cache, speeding recovery.

Security is provided by encrypting data on site, in flight, as well as in the cloud using 256-bit AES encryption and TLS v1.1/1.2. AltaVault appliances provide a dual layer of encryption that makes sure that any data moved into the cloud is not compromised, and it creates a complete end-to-end security solution for cloud storage.

Because an AltaVault appliance is an asymmetric, stateless appliance, no hardware is needed in the cloud, and you can recover the last known good state of a broken or destroyed AltaVault appliance to a new AltaVault appliance. AltaVault appliances provide flexibility to scale cloud storage as the business requirements change. All capital expenditure planning required with tape and disk replication-based solutions is avoided, saving organizations up to 90%.

1.2 Disaster Recovery Planning

Disaster recovery planning is a key component of the larger business continuity planning process, which focuses on establishing procedures and protocols for recovering business processes and systems in the event of large, unplanned outages or disasters. Preparing for unplanned outages or disasters requires a prepared and consistently reviewed approach to how to identify tiers of data/process criticality (risk analysis), services and processes that need to be implemented to react to drastic changes in availability (disaster planning), and evaluating and addressing weaknesses that could inhibit the ability to identify or respond to disasters.

1.3 Disaster Recovery Planning Terms and Definitions

Business Continuity

Business continuity (BC) is the set of processes and procedures an organization implements to make sure that essential business functions can continue during and after a disaster.

Business Continuity Planning

Business continuity planning (BCP) attempts to address and prevent interruption of mission-critical services and to reestablish full functioning as swiftly and smoothly as possible.

Risk Analysis

Risk analysis identifies key functions and assets that are critical to an organization’s operations and the probabilities of disruption to those functions and assets in the event of a disaster. Risk analysis is useful to understand what objectives and strategies must be employed to reduce avoidable risks and minimize impacts of unavoidable risks.

Disaster Recovery Plan

A disaster recovery plan (DRP) is a plan that is designed to help an organization's IT infrastructure team restore service and operational abilities to one or more target systems, applications, or facilities in the event of a disaster at a primary facility.

Disaster Recovery Site

A disaster recovery site (DRS) is a location that is separate from the primary processing facility for an organization that can house hardware, communications interfaces, and environmentally controlled space capable of providing backup data processing support:

- A DR hot site can typically deploy its resources to restore services within a very short time because production resources are replicated in almost immediate time to this type of site.
- A DR warm site can bring up services within a reasonably short time (but longer than a hot site can restore services because replication of services may not be performed as regularly).
- A DR cold site is typically a preestablished space that may or may not have the necessary equipment on site, but can be set up in the event of a disaster.

High Availability

High availability (HA) describes the ability for a service of system to continue servicing functioning for a certain period of time, normally a very high percentage of time, for example, 99.99%. High availability can include redundant resources that can be implemented to eliminate single points of failure or clustering services or processes across two or more systems to provide distributed workload availability.

Recovery Time Objective

Recovery time objective (RTO) is the duration of time that a business process or service must be restored after a disaster. RTO is typically established for services and processes within the scope of the BCP.

Recovery Point Objective

Recovery point objective (RPO) describes the acceptable amount of data loss and is measured in time such as hours. Typically, RPO is used to describe the point in time to which an organization must recover data. BCP helps to establish guidelines for backups or replication of systems such that RPO can be met for systems.

1.4 Tiers of Disaster Recovery

In 1992 the SHARE user group established seven tiers of disaster recovery, which describe methodologies for recovering mission-critical computer systems as required to support business continuity. Commonly used today by the disaster recovery industry, the tiers are described next.

Tier 0: Do Nothing

No backups are taken, and no business continuity plan exists. This tier features the highest risk with a strong possibly of no ability to recover systems, data, or processes.

Tier 1: Off-Site Vaulting

Describes the method of transporting backups to a secure off-site location and typically describes a tape-based backup environment. This tier lacks systems on which to restore data and focuses on the transport of data at an off-site storage facility. This process requires minor operator involvement to generate and transport tapes off a production site.

Tier 2: Off-Site Vaulting with a Hotsite

This tier is similar to tier 1 in that backups are transported off site, but tier 2 includes an off-site facility and resources in which to recover data in the event of a disaster. These resources may or may not be enabled, but can be activated in the event of DR. This process requires minor operator involvement to generate and transport tapes off a production site, but also additional involvement in preparing and maintaining the DR facility in the event that it needs to be activated in an emergency.

Tier 3: Electronic Vaulting

This tier improves upon tier 2 capabilities by providing an electronic vault of a subset of backup data, such that some recovery processes can be implemented without the need to wait for backups to be prepared. A tier 3 environment may consist of VTL disk libraries, for example.

Tier 4: Electronic Vaulting to DR Hotsite

Resources at the DR site are on and available, and backup copies are deployed typically to a disk subsystem that represents a point in time of the production dataset. Backups are also typically taken more frequently, because the medium on which they are written is disk based.

Tier 5: Two-Site Two-Phase Commit

Tier 5 requires that both the primary and secondary platforms' data be updated before the update request is considered successful. This satisfies the need for businesses that must have data consistency between production and DR sites.

Tier 6: Zero Data Loss

Tier 6 implements the highest level of data currency across production and DR facilities and typically needs to be implemented without dependence on the application or application staff to provide consistency. Examples include disk mirroring in either asynchronous or synchronous form, depending on the RPOs and RTOs.

1.5 Data Classification

Data classification is an important component of DRP, storage management planning, and backup application planning. Within every operating environment, various types of data exist and can be classified into four tiers as shown in Table 1.

Table 1) Data classification.

Data Class	Description
Critical	Application data critical for business processes that provide minimum acceptable levels of service in the event of a disaster or data that must be available for regulatory audits (for example, customer orders and financial data).
Important	Application data for standard business processes, which is impossible or extremely expensive to recreate, or data that has significant operating value (for example, classified data).
Semi-important	Application data for normal operational procedures, but can be cost effective in recreating from original data sources at minimal to moderate costs (for example, support documentation).

Data Class	Description
Nonimportant	General data that can easily be recreated from original source data (for example, reports).

By classifying business processes around their associated data, restore procedures (as documented in the DRP and implemented by a backup policy) can be ordered to recover mission-critical servers, applications, and data first. Doing recovery of business resources based on priority helps maximize the use of the limited computing and storage resources that may be available to do disaster recovery.

1.6 Costs Related to Disaster Recovery Solutions

When disaster recovery objectives move toward the higher tiers, costs associated with providing hardware, staff, and maintenance grow exponentially. Examples of costs relating to a DR solution:

- A secondary site with operational equipment, software, software licenses, and standby IT resources
- Bandwidth connections over long distances between primary and recovery sites
- Additional backup software to support advanced features (such as add-ons for database applications)
- High availability or clustering equipment or software
- Hardware supporting replication and/or point-in-time Snapshot® copies

1.7 Infrastructure Preparation

Having highly available infrastructure and associated resources at both primary and disaster recovery sites is necessary to make sure that when a disaster strikes, an organization can bring back the critical systems necessary to restore and reliably run business services and processes. Preparing infrastructure can be broken down into the following areas:

- **Space.** Production and disaster recovery sites must be capable of physically containing the necessary infrastructure related to the business processes implemented or to be recovered. Both should take into consideration growth of infrastructure, technological density (virtualized systems vs. physical systems), cooling, power, and weight requirements.
- **Power.** Power infrastructure must provide redundancy and scalability without disruption. Every power management device (transformers, systems, UPSs, and so on) must be built with redundancy in mind, just like high-availability systems architecture.
- **Security.** Controlling access to a data center is extremely important to help fortify operations against malicious behavior, while allowing access to the key personnel that are responsible for managing the infrastructure resources.
- **Hardware capacity planning.** Plan for systems that can include redundant power supplies, redundant cooling devices, and hot-swappable internal disks. If virtualizing a physical production environment at a disaster recovery site, carefully consider the resources and capabilities of the hardware deployed to make sure that it meets the expectations of the workload placed on it after the production system is recovered.
- **WAN bandwidth.** Sizing and establishing reliable access to the Internet are critical in making sure that replication of off-site backups can be performed and that those backups can be brought back to a disaster recovery site during a disaster.
- **Software.** Make sure that production operating system, virtualization, application software, upgrades, and patches are available at the off-site disaster recovery site. Software should be documented and cataloged for easy access.

After data priority is established, a backup policy for the organization can be built by the backup administrator to meet the objectives resulting from the planning phases.

2 Comparing Disaster Recovery Processes

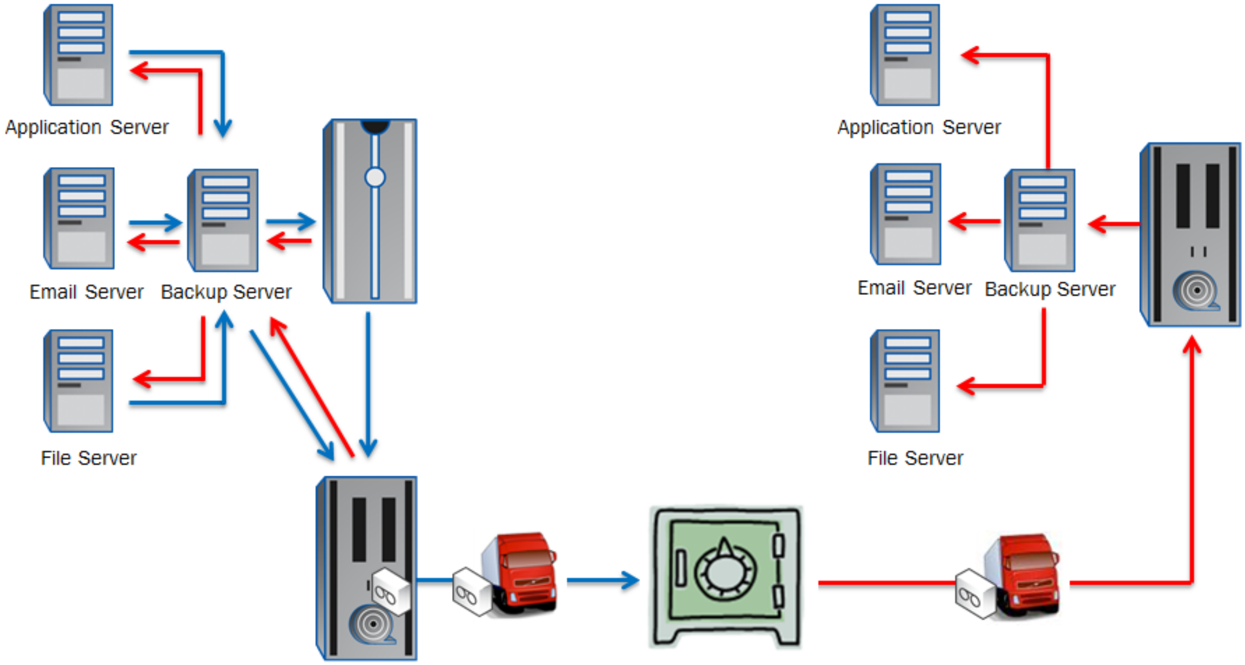
A DR outage might require a significant restore action to be undertaken, which can include recovering the backup infrastructure, in addition to the production business systems. These are typically true DR scenarios, in which the entire working infrastructure is lost, such as in a fire or flood. The effect to the business is significant and includes impacts such as lost productivity, lost sales, and 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 successful business continuity.

Efficiently restoring an enterprise environment after a disaster requires planning, which includes classifying systems, data, and resources. Typically, DRP and storage management planning occur as separate activities in many environments. Business continuity planning generally provides information about critical systems, their supporting systems, the value of each system to the business, risk analysis, and the recovery time objective for each system. These concepts can be associated to DRP and then translated into requirements for storage management planning.

2.1 Traditional Disaster Recovery Process

Typical backup applications consist of a client component installed on each production server, which processes a copy of the active files or blocks of a server and sends them to a central backup server target that can service many client backups simultaneously. Backups are initially written to a disk storage target, which can handle the workload of several backups streaming through the backup server. After backups are completed to the backup server, migration of the data usually occurs to move the backup data from expensive disk to less costly tape storage on site. Additionally, a second copy of the backups is created to another set of tapes to be sent off site to a vault for disaster recovery purposes. When a disaster occurs, a request is made to the vaulting location to have the second copy of backups returned to a disaster recovery facility, upon which a second set of resources is turned on to facilitate the disaster recovery process.

Figure 1) Traditional disaster recovery data flow.



Modern backup applications all involve the use of policy management to help define data prioritization into storage and management planning. Backup applications are carefully configured through the use of backup schedules, versioning, and storage device tiering to effectively organize data availability relative to the storage resources available. Because prioritization of critical and important data is first, it is desired to maintain the data for these systems on storage mediums that are the quickest in recovering the data, usually disk. But because disk systems are a high-cost storage medium, less costly but more labor-intensive tape system solutions are typically involved in order to hold old data or versions of critical or important data that has aged sufficiently.

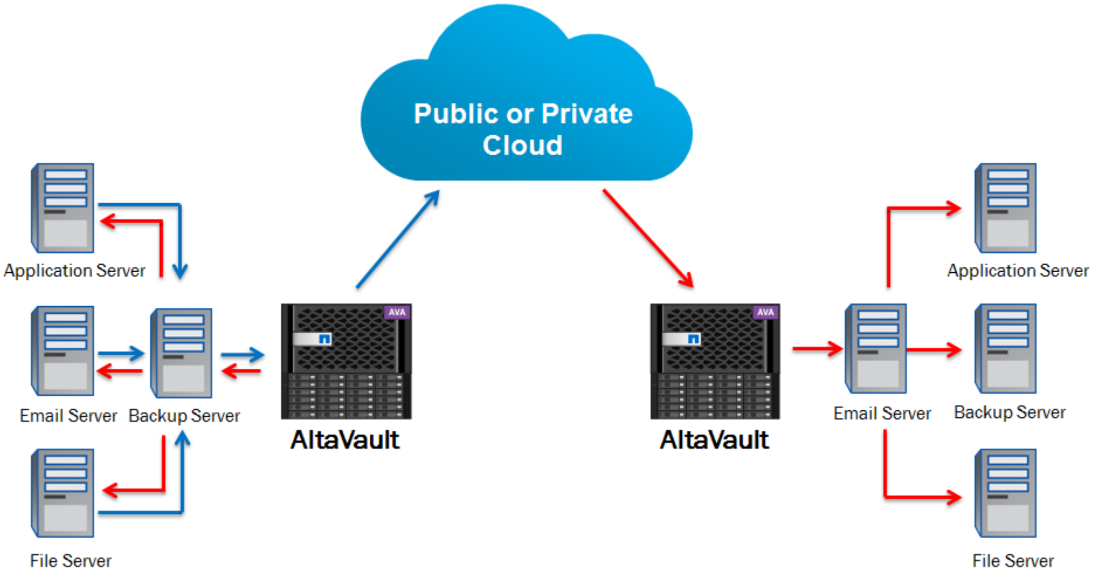
Defining the policy point in which data must be moved from one medium to another is often a difficult decision, because financial, resource, and/or physical system or site limitations restrict solutions that can best implement the business continuity and disaster recovery plans. Successfully implementing a storage management policy within these bounds is further restricted by requirements to the data lifecycle, in which older data that has reached a sufficient age must be removed from the backup server and all subsequent copy locations. System administrators in typical disk-to-tape backup environments must manage an increasing amount of operational overhead related to creating, storing, shipping, vaulting, and reclaiming tape volumes to and from the production and disaster recovery site facilities.

2.2 Simplifying the Disaster Recovery Process with AltaVault Appliances

AltaVault appliances simplify the complex nature of traditional backup strategies. As a disk-based deduplication solution, AltaVault appliances accept backup streams from the backup, archive, or database server into an AltaVault appliance, which serves as both the local disk storage target for local restores and as the cloud storage gateway for the off-site DR copy of data to cloud storage.

By leveraging highly efficient compression, deduplication, and encryption technologies on the incoming data stream, AltaVault can replace on-site disk and tape storage systems for holding the most recent local backups for immediate restore operations. AltaVault appliances maintains a local cache size varying from 8TB all the way up to 192TB of deduplicated compressed data, typically allowing a localized recovery to occur for data aged anywhere between one day to a couple of months. In addition, AltaVault replicates the backup data through encrypted TLS v1 to a cloud storage target, providing a cost-effective, secure, and fully automated process for disaster recovery copies of backups.

Figure 2) AltaVault disaster recovery data flow.



In the event that not all of the data is within the local cache when a restore is requested, an AltaVault appliance recalls just the necessary segments of the missing data needed from the cloud provider to complete the recovery. Typically, these data segments are from 1MB to 4MB in size and thus save the company money by not having to recover unnecessary data from cloud storage to complete the restore.

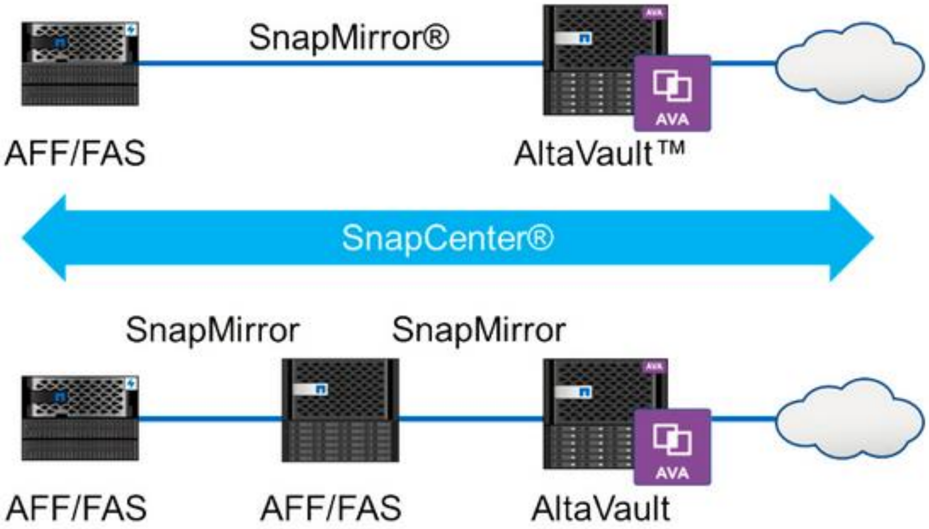
Because cloud storage providers build their sites on economy of scale and offer high durability protection, replicated copies of backup data stored at a cloud storage provider are even more protected from data corruption than if they were still on local disk at the production facility and are protected much more so than tape volumes, which are a highly exposed point of failure. AltaVault replication also eliminates the various system administration overhead and expense required to generate, collect, and ship volumes to and from the production and disaster recovery sites relative to the data lifecycle of storage management.

At the disaster recovery site, AltaVault appliances can quickly be brought up and connected to the cloud storage provider to begin restores. There is no downtime required as in the case when waiting for tape volumes to be found, shipped, and loaded into a tape system. AltaVault virtual appliances can also be spun up in the Amazon EC2 and Microsoft Azure compute clouds to provide organizations the ability to perform DR recovery of data, processes, or applications to cloud compute instances, leveraging pay-as-you-go mechanics of cloud compute resources to deliver a cheaper alternative to costly DR sites, which can be heavily underutilized a majority of the time.

With the addition of NetApp Snapshot™ support starting in AltaVault version 4.3, AltaVault can now be leveraged as an additional data protection tier for NetApp ONTAP Fabric-Attached Storage (FAS) and All-Flash FAS (AFF) systems. This reduces the complexity of protecting data by eliminating the requirement for a traditional backup application to perform data backup, archive, and disaster recovery. The NetApp SnapMirror protocol is used by AltaVault to receive incremental snapshot backups of volumes, and AltaVault securely stores the data in the cloud for long term protection and disaster recovery.

AltaVault appliance solutions cover tiers 1 through 4 of the seven disaster recovery tiers noted previously, replacing or consolidating multiple types of traditional data protection solutions, including tape libraries, vaulting, and disk-replicated solutions. When combined with an existing NetApp FAS or AFF infrastructure to create the Data Fabric Solution for Cloud Backup, users can seamlessly gain advantages of all 6 disater recovery tiers.

Figure 3) Data Fabric Solution for Cloud Backup.



2.3 Benefits of Using AltaVault Appliances for Disaster Recovery

Deploying AltaVault appliances in a production environment can significantly reduce the amount of resources and costs associated with protecting business processes and services, because it provides several capabilities found in higher disk-based tiered solutions but at 30% to 50% of the costs as would be required with implementing and supporting these tiers of solutions. Organizations that typically could only afford a tier 1 or 2 tape-based solution can now afford a tier 3 or 4 disk replication-based solution with improved recoverability across all their services and processes. The additional capabilities provided can help an organization achieve much higher levels of business continuity and recoverability that they previously would be unable to achieve due to limits in capital or operational funding, physical and network resources, and operational staff to service such a solution. Key benefits of an AltaVault appliance:

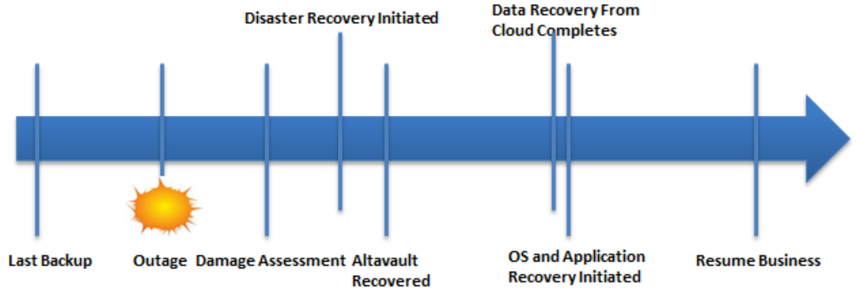
- **Ease of use.** Management of the appliance is achieved with a simple GUI interface accessed directly from the appliance or through the network. Multiple appliances can be managed remotely.
- **Reduction in administration.** AltaVault appliances free up IT users from traditional backup management time sinks such as tape vaulting and tape management. IT can now use that time to focus on higher priority projects.
- **Interoperability.** The appliance is designed to drop into an organization's existing backup and archive environment seamlessly, as a standard network-attached storage target. It supports all of the major backup applications currently available and in use by the top companies of the world and can also serve as an archive target for long-term datasets.
- **Storage optimization.** Leveraging industry-leading compression and deduplication technologies that are the cornerstone of current NetApp solutions, AltaVault appliances provide performance gains when replicating data to the cloud. By reducing the footprint of storage requirements significantly (up to 30x reduction), storage and access costs associated with protecting data are significantly reduced, while improving utilization of existing WAN connections to cloud storage.
- **Advanced data management policy.** Because data likely has different business priorities for different parts of the business, AltaVault appliances provide deep data management policies, including local data pinning for guaranteed availability, options to disable deduplication and compression for optimized data types, and the ability to prioritize data movement off the AltaVault appliance for the least important data.
- **Security.** Using dual-level encryption standards, data is protected both at rest using AES-256 bit encryption and in transit using TLSv1 encryption. Data is protected using encryption keys that can be standardized with an organization's encryption key policies.
- **Data integrity.** Delivering unmatched data integrity, cloud storage providers can provide up to 12 9s of data durability because of the high degree of data replication and redundancy checking performed across storage disks, arrays, local site buildings, and geographically distributed data centers.
- **Stateless appliance.** The appliance can store and can rebuild the most recent backups locally, and for all older backups the appliance can restore from the cloud as necessary.
- **DR flexibility.** AltaVault appliances can be deployed to a DR facility as a cold or warm physical or virtual appliance or as an AltaVault virtual machine appliance within the Amazon EC2 or Microsoft Azure cloud compute infrastructure. AltaVault Virtual and AMI appliances are free to use for performing DR operations.
- **Near-infinite scalability.** With AltaVault appliances being able to address up to over 14PB of backup capacity, even the largest enterprises can achieve off-site data protection for all their data with just a few appliances. And because cloud storage is elastic, capacity can be grown or scaled back instantaneously.

2.4 Disaster Recovery Timelines

AltaVault appliances can be deployed as a virtual or physical instance within the business's secondary data center to perform recovery operations.

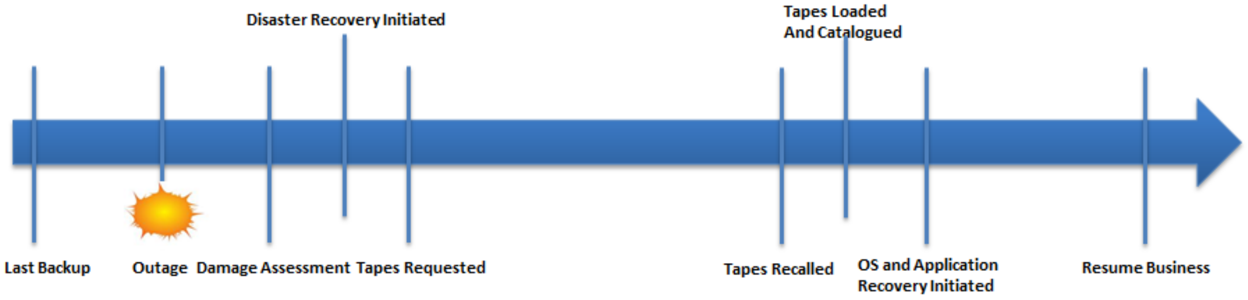
In a DR scenario, the secondary AltaVault appliance uses a wizard-driven process to aid in the recovery of the configuration from the original AltaVault appliance, followed by steps to recover the backup application namespace and then prepopulating the most recently backed up data from the cloud (typically the last day or week). Because AltaVault 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, AltaVault appliances uses smart prefetch algorithms to efficiently restore the needed data to improve the overall recovery process. The process can be started within minutes of the disaster if the AltaVault appliance is available at the secondary data center site. For example, Virtual AltaVault software is free to download and use, providing this immediate need if the business does not want to invest in a secondary AltaVault appliance initially.

Figure 4) AltaVault disaster recovery timeline.


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

This can save enormous amounts of time over traditional tape-based recovery, in which tape volumes must be identified, moved to the secondary data center site, mounted, and then read, as well as reduce the risks associated with physical volume movement (for example, tape corruption, tape misplacement, tape security, and so on). And when combined with the best practice of securing the backup application catalog or backup database to the AltaVault appliance, businesses can further reduce RTO by having those 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 business operational capabilities by utilizing AltaVault appliances.

Figure 5) Traditional tape disaster recovery timeline.


Disaster Recovery Initiated
Tapes Loaded And Catalogued
Last Backup
Outage
Damage Assessment
Tapes Requested
Tapes Recalled
OS and Application Recovery Initiated
Resume Business

The following table summarizes the time lines and benefits of AltaVault to improving DR.

Table 2) Benefits comparison.

Type of Disaster Recovery	Recovery Speed	Administration Overhead	Data Loss Risk	Accessibility
Traditional DR	Days	Very high	Medium	Regional
Improved DR using AltaVault appliances	Several hours	High	Very low	Worldwide

3 Disaster Recovery with AltaVault Appliances

Disasters can occur without warning, knocking out key business processes or entire facilities. As companies grow globally and become increasingly connected in supply chains, purchases, and trade, quickly rebuilding business activity in these events is critical.

3.1 Guidelines for Deploying AltaVault Appliances to Prepare for DR

The AltaVault appliance has several best practices and considerations before being implemented in a production and disaster recovery environment. Key considerations and recommendations are as follows:

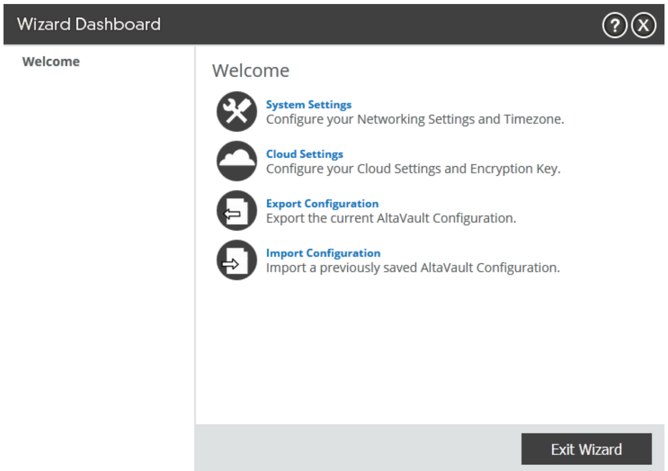
- AltaVault folder shares can be configured to help describe a policy target. For example, critical system backups may be directed by a backup application to point to a critical folder on one AltaVault data connection, while noncritical backups may be directed by a backup application to point to a noncritical folder on the remaining AltaVault data connections. This methodology can help balance priorities of data over the network, as well as organize data for recovery in case of a disaster.
- When possible, attempt to organize backup policies such that the most similar backup datasets arrive to the same AltaVault appliance. For example, if backing up a Windows Server farm to multiple AltaVault appliances, operating system backups likely have the best deduplication rates when grouped together to the same AltaVault appliance. File and application server backups may see better deduplication when grouped together, due to the likelihood that similar data is stored in each location.
- The AltaVault appliance can export its configuration to a file. An example file name may look like altavault_config_(HOSTNAME)_(DATETIME).tgz. It is recommended to store this file in different physical locations. You should also keep the exported configuration file within the DR site. The configuration file contains information about the configuration, including the encryption key. Alternatively, just the encryption key can be exported if it is managed by an encryption key vault or key management system.
- AltaVault appliances deployed at a disaster recovery site for a test DR scenario must not perform cloud access activities while AltaVault appliances deployed at the production site are engaged in replication activity. An AltaVault appliance at the production site must complete replication activity and then disable replication during the period that an AltaVault appliance at the disaster recovery site recovers data for disaster test warming preparation.
- AltaVault appliances can be deployed to only one cloud storage provider at a time. If a need arises where an AltaVault appliance must back up to a different cloud storage provider, the AltaVault cache must be cleared in preparation for reconfiguration to the new cloud storage provider credentials. All existing data with the previous cloud storage provider remains and can be recovered through a virtual AltaVault appliance if necessary.
- If the AltaVault storage capacity is less than the space used during DR, you can still initiate the recovery process. However, in this case the AltaVault appliance recovers only as much data as can be accommodated by the cache storage, and will evict older data from cache. This can potentially lengthen restores because restores need to be tiered by priority.

3.2 Production AltaVault Appliance Preparation

After an AltaVault appliance has been configured and deployed, protecting it is essential in order to recover any data that has been stored. The AltaVault appliance protects all information related to its configuration: CIFS and NFS shares, cloud credentials, user accounts, network configuration, security and reporting settings, and most importantly, the encryption key. All of these settings can be saved to a single file through the AltaVault GUI as follows:

1. From the AltaVault GUI, browse to Configure > Setup Wizard and select Export Configuration to export the configuration file. An example file name may look like altavault_config_(HOSTNAME)_(DATETIME).tgz.

Figure 6) Setup wizard to export configuration.



Wizard Dashboard

Welcome

Welcome

System Settings
Configure your Networking Settings and Timezone.

Cloud Settings
Configure your Cloud Settings and Encryption Key.

Export Configuration
Export the current AltaVault Configuration.

Import Configuration
Import a previously saved AltaVault Configuration.

Note: NetApp recommends that you store the exported configuration file in different physical locations. You should also keep the configuration file within the DR site. If an encryption key passphrase is present at the time of the configuration export, this passphrase will need to be used to import that configuration on a new AltaVault appliance.

3.3 Best Practices for Implementing Disaster Recovery with AltaVault

Implementing a DR strategy in the event of a real disaster can be harmful to the business if the respondents are unable to successfully and efficiently perform their duties to recover the business environment. Because DR is particularly impactful for all respondents such as system and backup administrators, having a clear and succinct DR execution strategy in place can help reduce the amount of user error, system error, and time involved in preparing and bringing up the environment at a DR site. Exercising DR preparedness by testing various DR scenarios regularly can help to reduce the amount of potential roadblocks faced when actually implementing a DR in a live situation. While testing DR can be somewhat impactful to production time, the costs associated with this are typically much smaller versus performing a live DR without being prepared. Key considerations and recommendations are as follows:

- Site and resource activation plans should be readily available, with clear instructions built from the business continuity and disaster recovery plan on how to perform the needed disaster recovery activities relative to the type of outage that has occurred. This should include information and alignment about all physical systems, software, patches, networking, power, and other related assets needed to recover the business environment.
- If deploying multiple AltaVault appliances for disaster recovery, it is recommended to configure them based on the criticality of the data in which they need to recover. Assuming that AltaVault appliances were deployed using the preceding guidelines for AltaVault deployment, then the DR plan should clearly indicate which data needs to be recovered first.
- If AltaVault deployment guidelines were used when aligning backup data criticality to AltaVault storage folders, populate data back onto the AltaVault appliance using the Prepopulation page of the GUI, where the folder name specified represents those backups that are most critical to recover. This maximizes the WAN usage relative to the data that needs to be restored first, effectively providing tiers to data recovery from cloud storage.
- Do not populate an AltaVault appliance at a DR site with recovered data that does not need to be restored, relative to the priorities established in the DR plan. This can lead to reduced performance in restore capability by the backup application, due to potential eviction activity on the AltaVault cache storage as segments are recovered from cloud storage.
- If it is anticipated that the recovery needs for restore will extend beyond the local cache storage capabilities of the AltaVault appliance deployed, recovery performance may suffer because many

segments of data might have to be evicted in order to make space for other segments of data required for restore. The approximate time frame needed to account for possible declined performance can be roughly estimated by identifying the overall amount of data needed to recover and dividing this by the deduplication rate achieved with the original AltaVault appliance. From this subtract the amount that will be stored locally on the DR AltaVault appliance. This new result will be the amount that will be subject to potentially slower recovery. Divide this result by the throughput of the WAN connection to identify the approximate additional time needed.

3.4 Disaster Recovery Scenarios with AltaVault Appliances

There are two DR scenarios that can occur, but both result in the same DR procedures.

Disaster Recovery Scenario 1: Production Site Down, Cold Disaster Recovery

This scenario assumes that the entire production site, including application and database servers, backup servers, and AltaVault appliance, is unavailable and must be recovered. In this true disaster recovery scenario, all replacement devices must be newly deployed at a DR site, with the backup servers and AltaVault appliance being deployed first to restore the production server environment (DR tier 2). An AltaVault appliance, whether deployed as a physical, virtual or, Amazon AMI, must be deployed as a new replacement device and requires initialization time to download the backups from the cloud so that the backup application can fulfill the recovery tasks. However, this can be done in parallel with the tasks required to rebuild the backup server, and in general recovery can begin as soon as the backup server is made ready. This scenario typically has a longer recovery time objective due to the requirement of having to first receive new equipment and set it up, prior to restoring backups from the cloud to the AltaVault appliance.

Pros: No upfront capital costs required until time of disaster; can leverage most current technologies available at time of DR.

Cons: Costs more to acquire equipment in response to a disaster; longest recovery time, which includes equipment procurement and setup time followed by reseed data time from cloud storage back to the AltaVault appliance.

Disaster Recovery Scenario 2: Production Site Down, Warm Disaster Recovery

This scenario is similar to the preceding second scenario, but rather than having to deploy the backup servers and AltaVault appliance as new replacement devices at the time of a disaster, these servers and AltaVault appliance have already been provisioned and made available and are running in standby mode at the DR site. The available AltaVault appliance in this scenario can immediately request access to the cloud storage backups and receive the most recent replicated backup data sent from the production AltaVault appliance. Thus when a disaster occurs at the production site, the resources can almost immediately be put to use to begin recovering production systems. Recovery time is reduced in this scenario because there is no replacement and configuration time taken into account because the environment is already staged.

Pros: Equipment costs are significantly lower due to upfront cost to acquire the hardware; minimizes downtime to time required to reseed data to appliance from cloud storage; full ability to test restores and DR procedures in advance of true DR event.

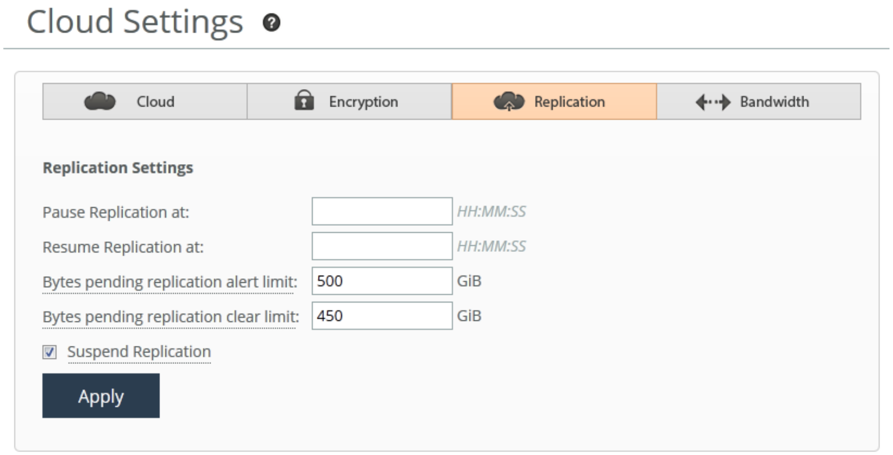
Cons: Capital costs are paid in advance to guarantee DR readiness; Additional facility and management costs associated with maintaining a physical or virtual AltaVault appliance in cold standby mode.

3.5 AltaVault Appliance Considerations for DR Testing

While testing DR is not much different from performing actual DR, there are considerations to take to make sure that AltaVault is properly utilized and does not interfere with production operations:

- AltaVault appliances deployed for a test DR scenario must not perform recovery activities while AltaVault appliances deployed at the production site are engaged in replication activity. AltaVault appliances at the production site must complete replication activity and then disable replication during the period that AltaVault appliances at the disaster recovery site recover data for disaster testing. Replication can be suspended on a production AltaVault appliance by accessing the AltaVault GUI interface and going to the Configure → Cloud Settings page, clicking the Suspend Replication checkbox, and clicking Apply.

Figure 7) Suspending replication.



Pause Replication at: HH:MM:SS

Resume Replication at: HH:MM:SS

Bytes pending replication alert limit: 500 GiB

Bytes pending replication clear limit: 450 GiB

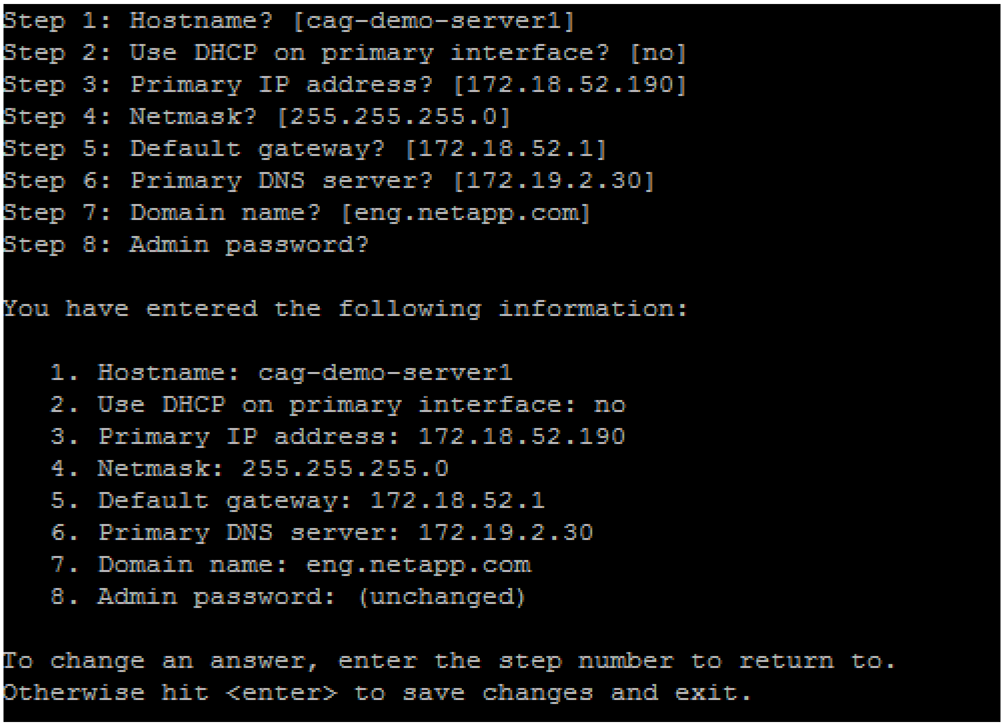
☑ Suspend Replication

- Because typical DR restore tests only need to recover a subset of the production backup data, it is suggested to use the deployment guidelines to align data that is restored to a specific AltaVault folder share, so as to maximize WAN usage during the datastore prepop phase of the recovery to the data that must be restored.
- If available, have WAN burst capabilities enabled by the Internet provider to assist in speeding up recovery of data from cloud storage. Because this would typically be the case in true DR, this can help establish a better understanding of recovery times pulling data back from cloud storage.

3.6 Performing Disaster Recovery with AltaVault Appliances

The first step to recover from a catastrophic failure of a production site is to install and configure for disaster recovery a new physical or virtual AltaVault appliance. It is recommended to use virtual AltaVault for the initial recovery, which can be downloaded from the support website and quickly deployed within a VMware environment at the DR site. While not required, the AltaVault appliance at the DR site is suggested to have the same or greater local storage capacity as the original AltaVault appliance at the lost production site, should you decide to make these resources at the DR site your production resources after the DR is complete. The following describes the steps to fully recover and restore the backup data from the cloud to the new AltaVault appliance.

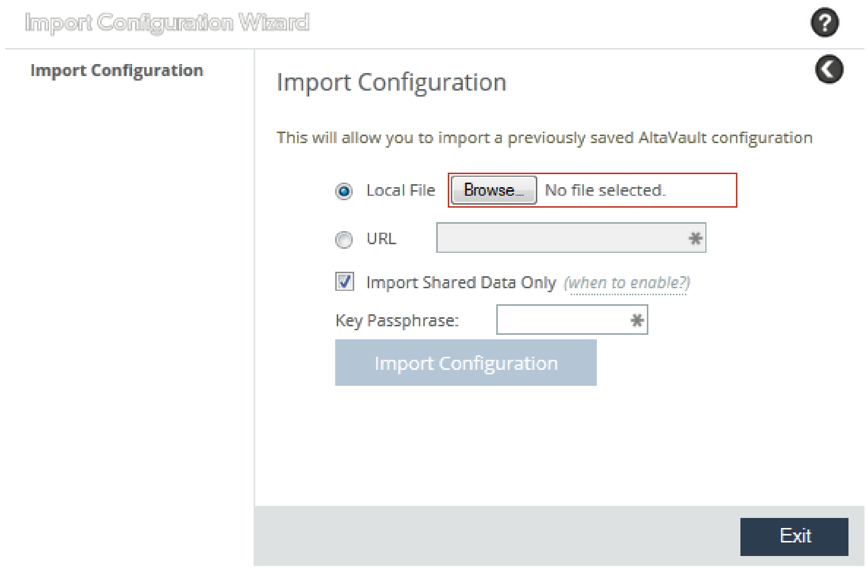
1. Configure the AltaVault appliance to the new network environment at the DR site.
   a. Plug a serial cable into the console port and a terminal, or in the case of the virtual AltaVault appliance use the hypervisor console.
   b. Log in to the AltaVault CLI using the default login admin and default password password.
   c. Configure the AltaVault network information. For details, see the AltaVault Administration Guide.



```
Step 1: Hostname? [cag-demo-server1]
Step 2: Use DHCP on primary interface? [no]
Step 3: Primary IP address? [172.18.52.190]
Step 4: Netmask? [255.255.255.0]
Step 5: Default gateway? [172.18.52.1]
Step 6: Primary DNS server? [172.19.2.30]
Step 7: Domain name? [eng.netapp.com]
Step 8: Admin password?

You have entered the following information:

   1. Hostname: cag-demo-server1
   2. Use DHCP on primary interface: no
   3. Primary IP address: 172.18.52.190
   4. Netmask: 255.255.255.0
   5. Default gateway: 172.18.52.1
   6. Primary DNS server: 172.19.2.30
   7. Domain name: eng.netapp.com
   8. Admin password: (unchanged)

To change an answer, enter the step number to return to.
Otherwise hit <enter> to save changes and exit.
```

2. Recover the original configuration of the AltaVault appliance to the new AltaVault appliance at the DR site. Browse to the menu Configure → Setup Wizard and import the previously saved configuration file. Make sure you leave the default "Import Shared Data Only" checkbox selected, which imports only the needed configuration settings to this AltaVault appliance. Networking and other appliance-specific settings are not recovered from the original AltaVault appliance to the new AltaVault appliance.

Figure 8) Setup wizard to import configuration.



Import Configuration Wizard

Import Configuration

Import Configuration

This will allow you to import a previously saved AltaVault configuration

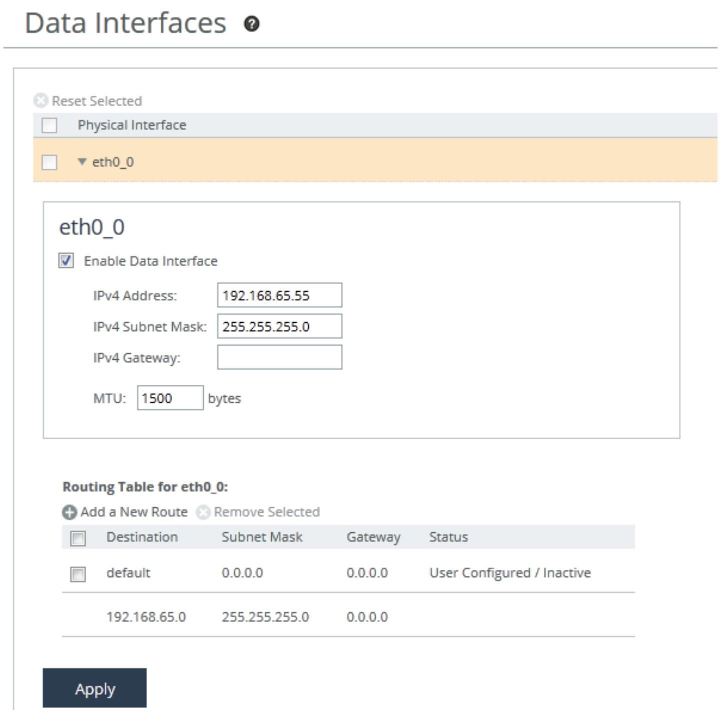
● Local File Browse... No file selected.

○ URL

☑ Import Shared Data Only (when to enable?)

Key Passphrase:

3. Configure AltaVault data interfaces to the new network environment at the DR site. Browse to the menu Configure > Data Interfaces and configure data interfaces network information.



Data Interfaces

Reset Selected

□ Physical Interface

□ ▾ eth0_0

eth0_0

☑ Enable Data Interface

IPv4 Address: 192.168.65.55

IPv4 Subnet Mask: 255.255.255.0

IPv4 Gateway:

MTU: 1500 bytes

Routing Table for eth0_0:

Add a New Route   Remove Selected

□	Destination	Subnet Mask	Gateway	Status
□	default	0.0.0.0	0.0.0.0	User Configured / Inactive
	192.168.65.0	255.255.255.0	0.0.0.0	

4. After the configuration is complete, connect to the AltaVault CLI using SSH and initiate the DR recovery procedure. If performing a true DR, issue the following commands:

```
AltaVault > enable
AltaVault # configure terminal
AltaVault  (config) # no service enable
AltaVault  (config) # datastore format local
AltaVault  (config) # replication recovery enable
AltaVault  (config) # service restart
```

Note: The replication recovery enable command fails to execute if the optimization service is enabled or if the AltaVault appliance detects existing data in the new AltaVault cache. Assuming this is a new, empty AltaVault appliance, you do not receive any failures, and the commands are executed without error. This process can take a few seconds to a few hours depending on the backups being restored. During the recovery process, the system communicates with the cloud provider and recovers all the namespace files that existed before the failure.

5. If you are performing DR testing and the production AltaVault appliance should continue to own the cloud storage contents following the DR test, then use the following commands on the secondary AltaVault appliance instead:

```
AltaVault > enable
AltaVault # configure terminal
AltaVault  (config) # no service enable
AltaVault  (config) # datastore format local
AltaVault  (config) # replication dr-test enable
AltaVault  (config) # service restart
```

Note: The production AltaVault appliance must have replication paused while performing the preceding DR test commands. The replication dr-test enable command fails to execute if the optimization service is enabled or if AltaVault appliance detects existing data in the new AltaVault cache. Assuming this is a new, empty AltaVault appliance, the commands all execute without error. This process can take a few seconds to a few hours depending on the backup metadata being restored.

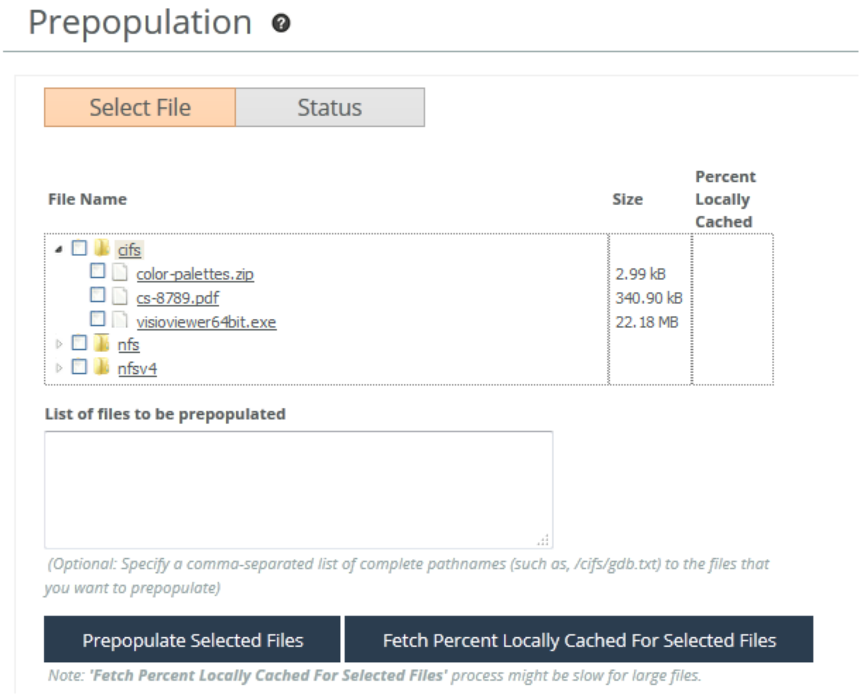
6. If the secondary AltaVault appliance is not correctly configured using the preceding steps, the replication recovery enable or replication dr-test enable command may fail, indicating the AltaVault cache is not empty. To correct this condition, issue the command datastore format local to delete the local cache contents and then rerun the replication command. This command deletes any existing contents of the secondary AltaVault cache, so use this command with caution. Do NOT run the datastore format command on the primary AltaVault appliance.
7. After the appliance has recovered the file system metadata, you may optionally perform prepopulation of the data contents back to the AltaVault cache from cloud storage. This requires additional time and WAN resources, but can significantly improve restore time due to data being retrieved in an optimal fashion relative to the volumes that are required for recovery. It is highly recommended to perform prepopulation of the required volumes for DR, as outlined in the following sections.

3.7 Prepopulation Using the AltaVault GUI

Because the recovery process downloads only the namespace metadata to the AltaVault cache, subsequent initial file access might be slow because the AltaVault appliance downloads all of the data from cloud storage. It is highly recommended to also prepopulate the data from the cloud back onto the new AltaVault appliance in order to accelerate the recovery process. Doing so makes sure that WAN utilization is maximized to transfer the needed backup data to the AltaVault appliance. If using Amazon Glacier, refer to the AltaVault Cloud Integrated Storage Administration Guide for further recommendations about performing prepopulation for backup applications.

To prepopulate data using the AltaVault GUI:

1. Browse to Configure > Prepopulation.
2. When the prepopulation page appears, select the items to restore from the file explorer window presented.
3. Click the Prepopulate Selected Files button to initiate the prepopulation job.


Prepopulation
Select File
Status
Prepopulate Selected Files
Fetch Percent Locally Cached For Selected Files

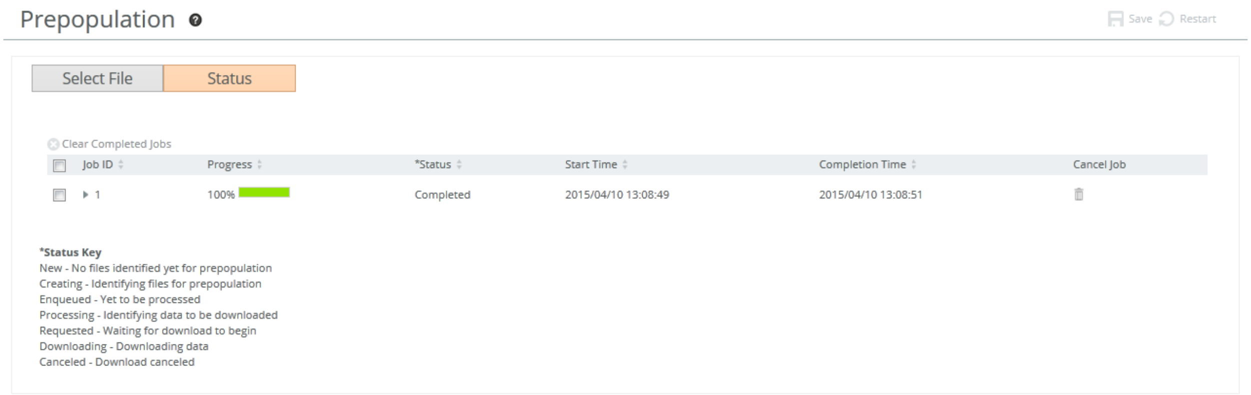
File Name	Size	Percent Locally Cached
◢ □ cifs		
□ color-palettes.zip	2.99 kB	
□ cs-8789.pdf	340.90 kB	
□ visioviewer64bit.exe	22.18 MB	
▷ □ nfs		
▷ □ nfsv4		

List of files to be prepopulated

(Optional: Specify a comma-separated list of complete pathnames (such as, /cifs/gdb.txt) to the files that you want to prepopulate)

*Note: **'Fetch Percent Locally Cached For Selected Files'** process might be slow for large files.*

4. Monitor the status of the prepopulation operation using the Status tab.



□	Job ID	Progress	*Status	Start Time	Completion Time	Cancel Job
□	1	100%	Completed	2015/04/10 13:08:49	2015/04/10 13:08:51	🗑

***Status Key**
New - No files identified yet for prepopulation
Creating - Identifying files for prepopulation
Enqueued - Yet to be processed
Processing - Identifying data to be downloaded
Requested - Waiting for download to begin
Downloading - Downloading data
Canceled - Download canceled

Note: If the AltaVault appliance storage capacity is less than the space used in the cloud, you can still initiate the recovery process. However, in this case the AltaVault appliance recovers only as much actual data as the size of its storage. If the recovery process attempts to bring back more data than the disaster recovery AltaVault appliance can handle, then the recovery process might fail. Virtual AltaVault AVA-v8, for example, can store up to 8TB of cloud data. For more details on AltaVault appliance sizes, see the AltaVault Installation Guide for Virtual Appliances.

3.8 Prepopulation Using the AltaVault Command Line Interface

Prepopulation is also provided using a command-line interface (CLI) tool that further enhances the ability of the AltaVault appliance to recover specific data from cloud storage. The prepop CLI tool provides powerful capabilities to help speed up data recovery from cloud storage, when paired together with a properly planned backup and storage management configuration. To use the prepop tool:

```
Connect to the AltaVault command line using a SSH client and enter the following commands:
AltaVault > enable
AltaVault # configure terminal
AltaVault  (config) # datastore prepop {[num-days <number of days>] | [pattern <pattern>] |
[recursive]} dryrun
```

Where the parameters are provided as shown in the following table.

Table 3) Datastore prepop command parameters.

Parameter	Description
num-days <number of days>	Specify the number of last-modified days to start data retrieval (from the present date to the number of days you specify).
start-date <start date>	Specify the date from which the data retrieval should start. The system prepopulates the files modified on or before this date
end-date <end date>	Specify the date on which the data retrieval should end. Stop prepopulating files on or after this date.
pattern <pattern>	Filters the data retrieved by the pattern you specify. The pattern specified contains a required internal share name created on the AltaVault appliance, one or more optional subfolder names from the external share name visible to the user, and finally a required regular expression describing the file or files to be prepopulated. The * symbol with the regular expression matches all characters.

Parameter	Description
dryrun	AltaVault calculates the estimated amount of cloud data to be recovered by the operation, and the amount of actual data to be recovered by the operation. No data is restored in this case.

5. To view the state of a running prepopulation operation, enter the following command:

```
AltaVault > show datastore prepop
```

> **Note:** If the AltaVault appliance storage capacity is less than the space used in the cloud, you can still initiate the recovery process. However, in this case the AltaVault appliance recovers only as much actual data as the size of its storage. If the recovery process attempts to bring back more data than the disaster recovery AltaVault appliance can handle, then the recovery process might fail. Virtual AltaVault, for example, can store up to 8TB of cloud data. For more details on AltaVault appliance sizes, see the AltaVault Cloud Integrated Storage Installation Guide for Virtual Appliances.

DR Example 1

Backups of a critical server are performed with NetBackup where:

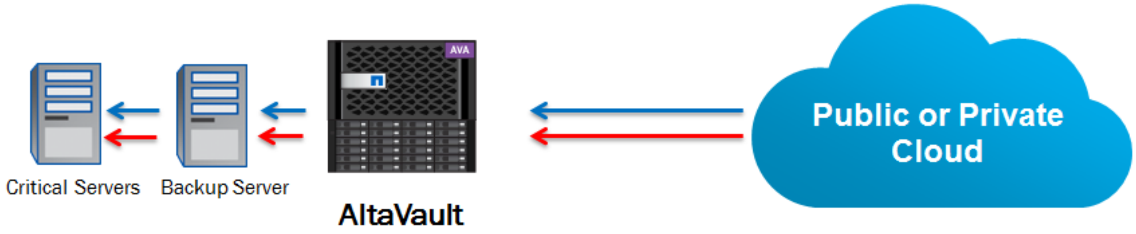
- Weekly full and daily incremental backups of the operating system drive are being sent to AltaVault CIFS share backupOS, which links to folder /backupOS in the AltaVault appliance.
- Weekly full and daily incremental backups of the application drives are being sent to AltaVault CIFS share backupApp, which links to folder /backupApp in the AltaVault appliance.

To populate an AltaVault appliance at a DR site with the most recent backup data, issue the following commands:

```
datastore prepop num-days 7 pattern backupOS/*
datastore prepop num-days 7 pattern backupApp/*
```

This recovers the most recent segments of data pertaining to those last seven days of backups, which would include the baseline full and subsequent incremental backups needed by NetBackup to bring the system to its most current state.

Figure 9) DR example 1.



DR Example 2

Several production systems are being backed up with IBM Tivoli Storage Manager where:

- Weekly full and daily incremental backups of critical servers (OS + app drives) are being sent to AltaVault CIFS share tier1bkups, which links to folder /tier1bkups in the AltaVault appliance.
- Bimonthly full and daily incremental backups of important servers (OS + app drives) are being sent to AltaVault CIFS share tier2bkups, which links to folder /tier2bkups in the AltaVault appliance.
- Quarterly full and daily incremental backups of workstations and laptops (OS + app drives) are being sent to AltaVault CIFS share tier3bkups, which links to folder /tier3bkups in the AltaVault appliance.

To populate an AltaVault appliance at a DR site respective to the preceding system tiers, issue the following commands:

```
datastore prepop num-days 7 pattern tier1bkups/*
```

This recovers the most recent segments of data pertaining to those last seven days of backups, which would include the baseline full and subsequent incremental backups needed by TSM to bring the critical systems to the most current state. After those systems are recovered by TSM, perform the datastore prepop command again, this time as:

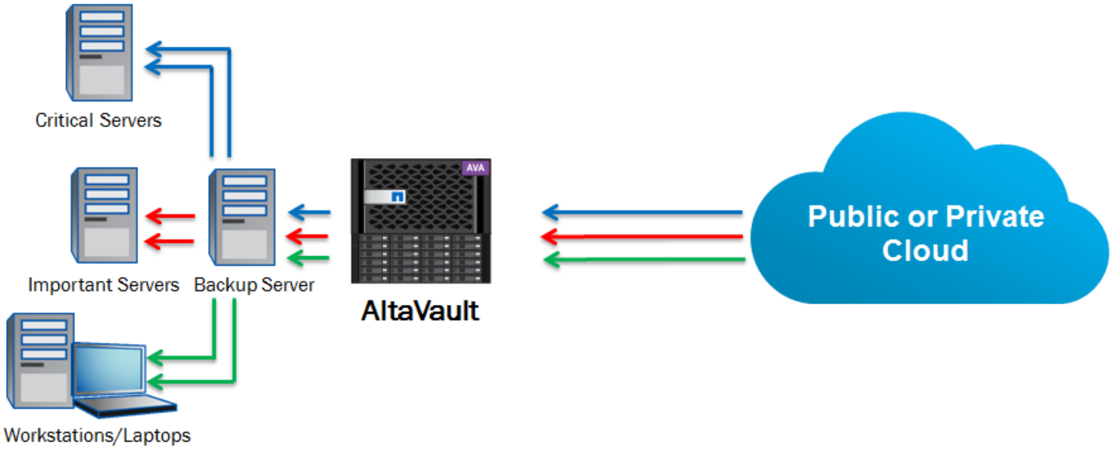
```
datastore prepop num-days 14 pattern tier2bkups/*
```

This recovers the most recent segments of data pertaining to those last 14 days of backups, which would include the baseline full and subsequent incremental backups needed by TSM to bring the important systems to the most current state. After those systems are recovered by TSM, perform the datastore prepop command one last time, this time as:

```
datastore prepop num-days 90 pattern tier3bkups/*
```

This recovers the most recent segments of data pertaining to those last 90 days of backups, which would include the baseline full and subsequent incremental backups needed by TSM to bring the important systems to the most current state.

Figure 10) DR example 2.



DR Example 3

This scenario is the similar to scenario 1 and 2 combined. The environment is much larger than the previous scenario and thus requires multiple AltaVault appliances at the DR site to restore the scope of systems in the DR plan:

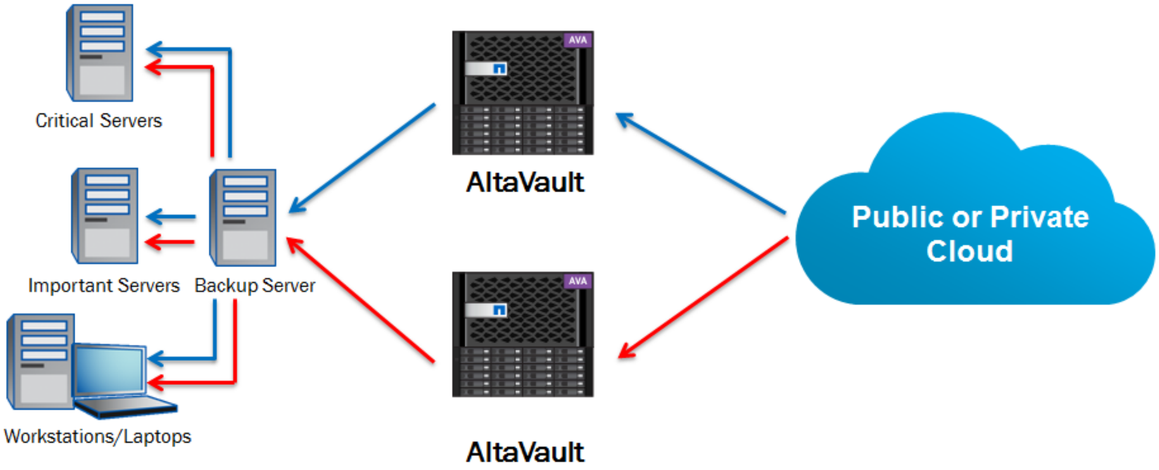
- Weekly full and daily incremental backups of critical servers (OS drives) are being sent to AltaVault CIFS share tier1bkupsOS, which links to folder /tier1bkupsOS in the AltaVault appliance.
- Weekly full and daily incremental backups of critical servers (app drives) are being sent to AltaVault CIFS share tier1bkupsApp, which links to folder /tier1bkupsApp in the AltaVault appliance.
- Bimonthly full and daily incremental backups of important servers (OS drives) are being sent to AltaVault CIFS share tier2bkupsOS, which links to folder /tier2bkups in the AltaVault appliance.
- Bimonthly full and daily incremental backups of important servers (app drives) are being sent to AltaVault CIFS share tier2bkupsApp, which links to folder /tier2bkupsApp in the AltaVault appliance.
- Quarterly full and daily incremental backups of workstations and laptops (OS drives) are being sent to AltaVault CIFS share tier3bkupsOS, which links to folder /tier3bkupsOS in the AltaVault appliance.

- Quarterly full and daily incremental backups of workstations and laptops (app drives) are being sent to AltaVault CIFS share tier3bkupsApp, which links to folder /tier3bkupsApp in the AltaVault appliance.

To populate the multiple AltaVault appliances at a DR site respective to the preceding system tiers, one of two methods can be performed, depending on requirements:

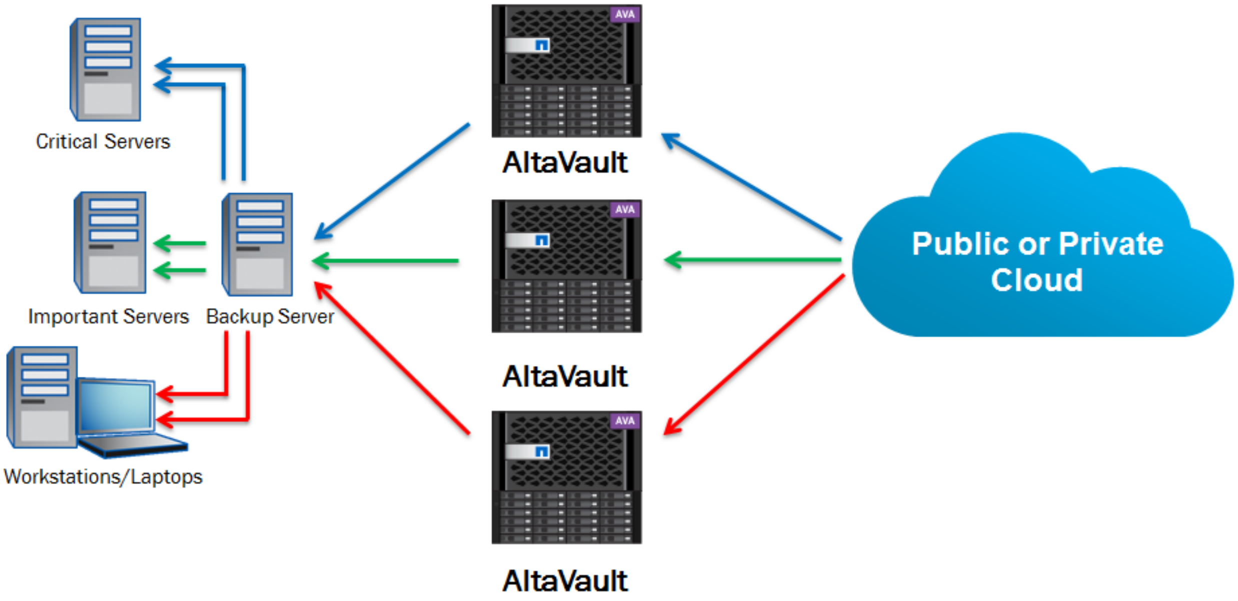
- **Method A.** Use one AltaVault appliance to perform datastore prepop of the tier1bkupsOS folder and a second AltaVault appliance to perform datastore prepop of the tier1bkupsApp folder, similar to scenario 1 but parallelized because of the access to multiple AltaVault appliances. This method allows for recovery of critical server OS and data volumes simultaneously. Recover the critical systems, and when complete reuse the AltaVault appliances to datastore prepop the OS and app drives for the next tier (important servers), perform restores, and again reuse the AltaVault appliances for the last tier of workstations and laptops.

Figure 11) DR example 3a.



- **Method B.** Use one AltaVault appliance to perform datastore prepop of the tier1bkupsOS folder and tier1bkupsApp folder, a second AltaVault appliance to perform datastore prepop of the tier2bkupsOS folder and tier2bkupsApp folder, and a third AltaVault appliance to perform datastore prepop of the tier3bkupsOS folder and tier3bkupsApp folder. In this fashion, the restores are parallelized differently such that you can recover multiple tiers simultaneously. Because it is likely that there is less data that must be prepopulated at higher tiers due to fewer servers at that tier, recovery can begin sooner for those systems without delaying the ability for recovery of additional systems at lower tiers.

Figure 12) DR example 3b.



3.9 Post-DR Considerations

After successfully deploying an AltaVault appliance and using a backup application to recover the necessary operating systems and critical data to resume business operations, the AltaVault appliance can subsequently be used for new backups from those recovered systems, should this DR site become a temporary production site. To make sure that the AltaVault appliance operates effectively, consider the following:

- If the AltaVault appliance is used for new backups, it is highly recommended that the AltaVault appliance type deployed at the DR site be of equivalent make and size as the lost production AltaVault appliance used.
- Most new backups following a full recovery are full rather than incremental backups. Phase in full backups appropriately to prevent overwhelming the backup application and AltaVault appliances.
- Because an AltaVault appliance may only have a subset of the recovered data within localized disk cache, expect deduplication performance to be lower while new backups are taken. If possible, consider performing additional data recovery using the datastore prepop command to increase potential deduplication rates. This results in higher recovery costs, but saves in future cloud storage costs per month.

After a true production site is available, the DR site may no longer be required. Migrating systems from a DR environment back to a production site can be done either by physically moving the resources from the DR site back to the production site or by performing another DR process at the new production site to recover the resources back on new production hardware. In the latter case, the activities are similar to those in a true disaster recovery scenario, and similar practices should be implemented accordingly to restore the operating systems and business data.

Where to Find Additional Information

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

- AltaVault Cloud-Integrated Storage product page
  http://www.netapp.com/us/products/cloud-storage/altavault-cloud-backup.aspx

- 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	April 2016	Updated for 4.2 release
Version 1.2	January 2017	Updated for 4.3 release
Version 1.3	April 2017	Updated for 4.3.1 release
Version 1.4	November 2017	Updated for 4.4 release

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

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

Trademark Information

TR-4420-1117