Access 3340 Appliance with NetBackup

Long-Term Retention Solution

Contents

Revision History

Rev 1.0 05 Mar 2018	Initial version
Rev 1.01 08 March 2018	Minor corrections on support
Rev 1.02 09 Apr 2018	Minor updates
Rev 2.0 01 Oct 2018	Updates based on 7.4.2 release
Rev 2.1 05 Mar 2019	Updates to level of support Veritas Data Deduplication
Rev 2.2 01 July 2019	Updates to memory requirements for Veritas Data Deduplication
Rev 3.0 10 Feb 2021	Updates based on 7.4.2.301 release and NBU 8.3
Rev 4.0 3 May 2021	Updates based on 7.4.3 release

Introduction

Executive Summary

Veritas Technologies is a leader in developing data resiliency solutions that focuses on protection and management of companies' digital assets critical for their success and business continuity. One of Veritas flagship products is NetBackup that is designed to protect datacenters, hybrid, and multi-cloud environments. Adding to Veritas portfolio and legacy of creating stable solutions that customers have trusted and relied on is the Access Appliance, a turn-key storage appliance created to address the long-term retention needs of organizations. The Access Appliance acts as an on-premises storage target for data that has been backed up using NetBackup. The Access Appliance has been optimized and designed to work seamlessly with Veritas NetBackup. The integration of these solutions provides a compelling offering for the long-term retention use case.

Scope

The purpose of this document is to provide technical details to assist in understanding the Access Appliance with NetBackup as a solution for long-term retention of backup data. It describes the components of this solution, its value, sizing guidance, and some best practices. It is advised to refer to Veritas product documentation for installation, configuration and administration of each of the products discussed in this whitepaper. **NOTE**: This document gets updated periodically and if you downloaded a local copy of this document, please get the latest from this link.

Target Audience

This document is targeted for customers, partners, and Veritas field personnel interested in learning more about the Veritas Access Appliance with the NetBackup solution for long-term retention. It provides a technical overview of this solution, guidance in sizing, and highlights some best practices.

Solution Value

Companies usually have a strategy to protect data in case of a failure, disaster, or crisis and NetBackup is an industry leader in this area. However, as data increases at an accelerating pace, companies are striving to determine the best strategy in the management, preservation, and retention of their valued data for long-term. There are several challenges that come to mind when talking about a long-term solution which include cost, complexity, control, and visibility. Traditionally, the solution for long-term retention has been tape because of its low cost. However, the complexity in tape management in addition to the time to restore has been an issue. Recently companies have looked to the public cloud for a possible solution, however, issues in total cost of ownership and control become a concern.

To address all these challenges, Veritas has designed the Access Appliance as a purpose-built, on-premises storage appliance for long-term retention use cases. Together with NetBackup, the Access Appliance provides a resilient and cost-effective solution for the preservation of data backups that companies want to retain and have readily available for further use.

As this document unfolds the architecture and features of this solution, it will showcase the following key values:

- **Minimize cost** – Access Appliance provides a low-cost, disk-based solution that is easy to manage. With NetBackup deduplication feature, the amount of storage space is reduced by saving only one copy of the data blocks and having the duplicates point to that one copy, thus providing a more storage efficient solution and reducing overall costs. When using Veritas Data Deduplication (VDD) feature introduced in version 7.4.2, further reduction can be observed in a multi-domain NetBackup environment as Access supports global deduplication of data across the domains or across multiple media servers.

- **Simplify Management** – Access Appliance with NetBackup has deep integration features that simplify the configuration and administration, such as invoking policy-based storage management and intelligent data movement between tiers.
- **Increase visibility and control** – more and more companies would like to leverage the data that has been archived and retained for IT or business analysis and investigations so having the data on-premises under the company's control and visibility allows for quick restores to conduct these studies.

Solution Key Features

There are certain key features that companies look for in a long-term retention solution product: flexibility, storage efficiency, and ease of management. The Access Appliance with NetBackup provides these features to assist customers in preserving their most valued data.

Seamless Integration with NetBackup

The Access Appliance has been integrated as a cloud provider in NetBackup 8.1 (with updates) and as a target open storage server for the Veritas Data Deduplication feature. For instance, during configuration of a cloud storage server on the NetBackup master server's administration graphical user interface (GUI), Access is listed as one of the cloud storage providers that can be selected.

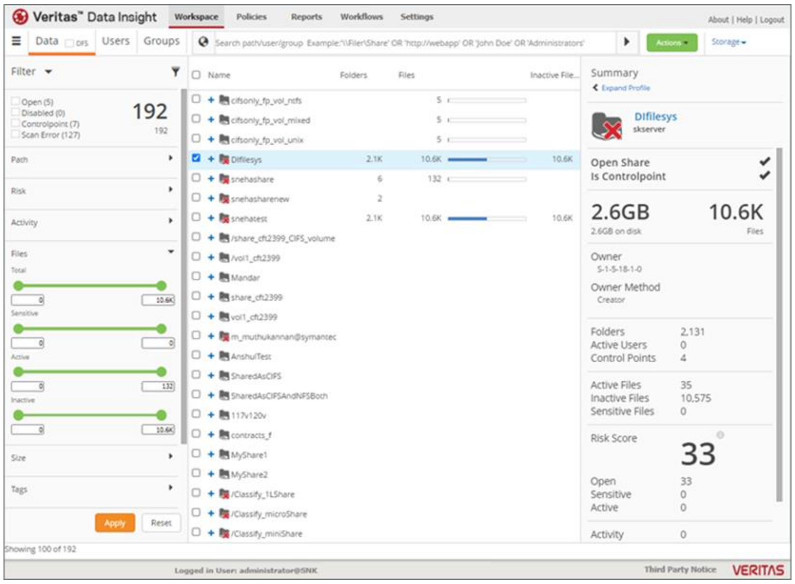
In addition, the Access Appliance GUI provides a configuration wizard and policies to make it easier to configure an S3 bucket for NetBackup or Veritas Data Deduplication pool. Provisioning can also easily be done using "Quick Actions" to provision another S3 bucket or configure Veritas Data Deduplication.

A container-based NetBackup client add-on package is also available to be installed on the Access Appliance. A good use case for having the NetBackup client integrated into the Access Appliance is the protection of the deduplication catalog when using Veritas Data Deduplication pool as a target. It can be backed up to another location for added protection.

Use with Other Veritas Products for Better Visibility

Veritas Data Insight has the capability to scan and classify primary file system sources such as filers, SharePoint, Documentum repositories and cloud storage. It classifies the data into certain categories such as ownership, age, size, activity, and access patterns in order that administrators can identify data that can be archived or tiered to cheaper storage such as the Access Appliance, enforce security, perform information lifecycle management and risk analysis. The ability to identify areas of risk, value, and ROT (Redundant, Obsolete, Trivial) improves operational efficiency, reduces storage cost and minimizes risk and liability. For instance, reports generated from Data Insight can be inspected to determine which files can be backed up for long-term retention storage on-premises or safe to place in the cloud. An example of Data Insight web graphical user interface (GUI) is shown in Figure 1. The information displays information such as number of inactive files, where your file resides, file extensions, and age.

Figure 1 - Sample View of Data Insight



□ Open (5)
□ Disabled (0)
□ Controlpoint (7)
□ Scan Error (127)

192
192

Files
Total: 0 – 10.6K
Sensitive: 0 – 0
Active: 0 – 132
Inactive: 0 – 10.6K

Name	Folders	Files	Inactive File...
□ cifsonly_fp_vol_ntfs		5	
□ cifsonly_fp_vol_mixed		5	
□ cifsonly_fp_vol_unix		5	
☑ Difilesys	2.1K	10.6K	10.6K
□ snehashare	6	132	
□ snehasharenew	2		
□ snehatest	2.1K	10.6K	10.6K
□ /share_cft2399_CIFS_volume			
□ /vol1_cft2399			
□ Mandar			
□ share_cft2399			
□ vol1_cft2399			
□ m_muthukannan@symantec			
□ AnshulTest			
□ SharedAsCIFS			
□ SharedAsCIFSAndNFSBoth			
□ 117v120v			
□ contracts_f			
□ MyShare1			
□ MyShare2			
□ /Classify_1LShare			
□ /Classify_microShare			
□ /Classify_miniShare			

Showing 100 of 192

Summary

DIfilesys
skserver

Open Share ✓
Is Controlpoint ✓

2.6GB
2.6GB on disk

10.6K
Files

Owner
S-1-5-18-1-0

Owner Method
Creator

Folders	2,131
Active Users	0
Control Points	4
Active Files	35
Inactive Files	10,575
Sensitive Files	0
Risk Score	33
Open	33
Sensitive	0
Active	0
Activity	0

Logged in User: administrator@SNK

Storage Efficiencies

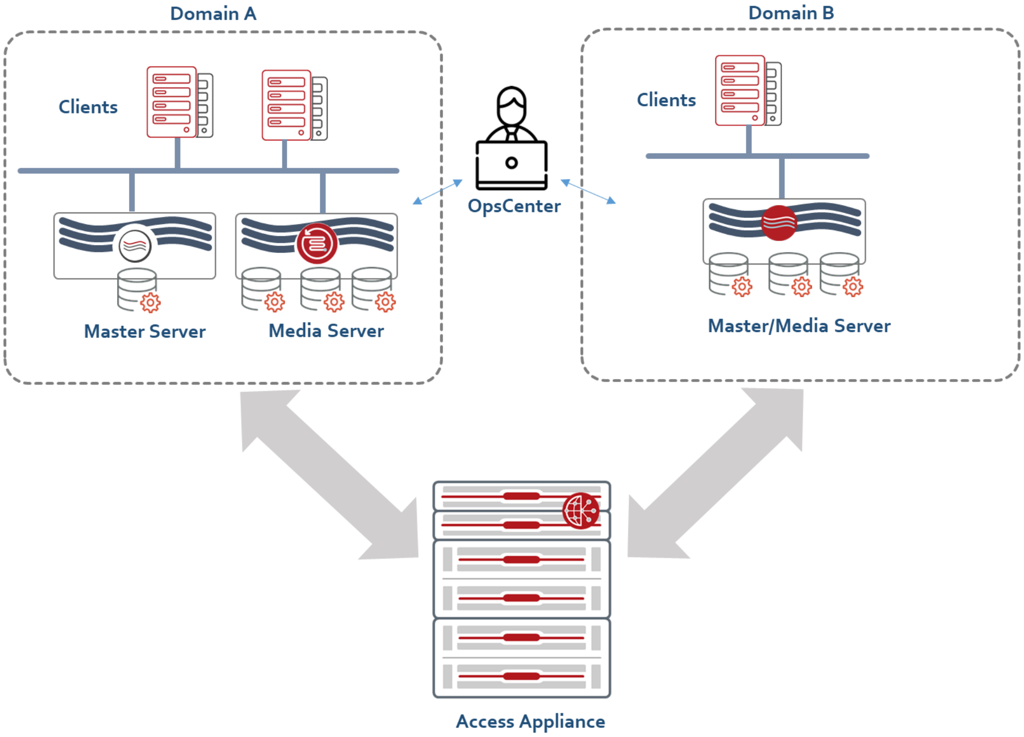
Support for storage efficiency is one of the main factors when choosing and purchasing a long-term retention storage platform solution. The ability to maximize storage space assists in reducing overall cost. Backup images stored in the Access Appliance can be deduplicated using NetBackup Media Server Deduplication Pool (MSDP) technology. Data is sent to Access either via MSDP Cloud Tiering (MSDP-C) or directly to Veritas Data Deduplication.

NetBackup also supports compression prior to sending data to the Access Appliance. So, Access stores the compressed format of the data. Compression improves storage utilization by reducing the number of bits required to represent data. The type of data defines the degree a file can be compressed. Data types that compresses well include text files or unstripped binaries. Data that is already compressed and stripped binaries are not good candidates for compression.

Support for Multiple NetBackup Domains

The Access Appliance can support multiple NetBackup domains. For example, multiple instances of NetBackup protecting multiple clients from different sites, locations, or data centers can use a single Access Appliance for long-term retention as shown in Figure 2. The Access Appliance must be reachable from either site and data is sent to Access from the NetBackup instances using the S3 protocol on HTTP or HTTPS transport when using MSDP-C or with a proprietary protocol when using Veritas Data Deduplication. When using Veritas Data Deduplication in a multi-domain NetBackup environment, more storage efficiency is observed since only one copy is saved if a duplicate block is encountered from both domains.

Figure 2- Example of Multi-domain NetBackup with the Access Appliance.



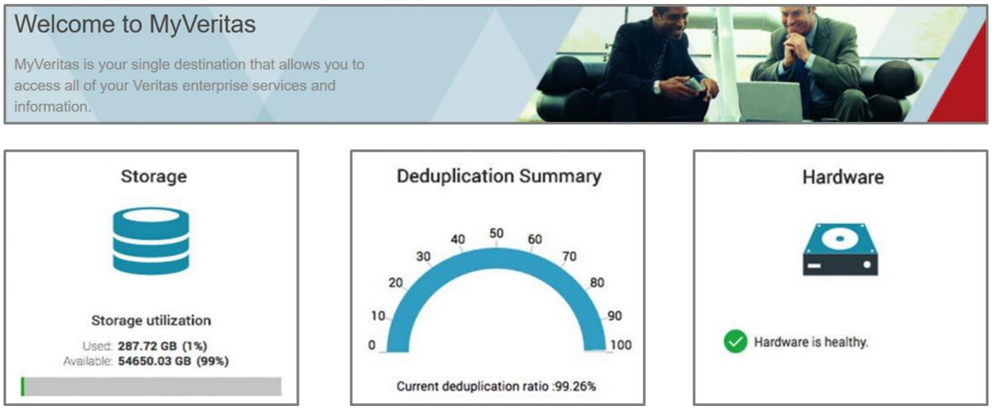
Security

NetBackup has security features that protect all NetBackup components and operations at different security implementation levels such as datacenter, world, and enterprise. For enhanced security, NetBackup offers encryption of data. Any encryption done by NetBackup is maintained on the Access Appliance. NetBackup sends data over dedicated and secure network ports to Access. Additional security that is employed for this solution is the requirement to use Access user keys and credentials when configuring Access as a cloud storage destination. When using MSDP-C, data sent to Access can also be secured by enabling the SSL feature on Access. When SSL is enabled, certificates are generated on Access and placed in NetBackup master and media servers and data is sent via HTTPS. Refer to the NetBackup Security and Encryption Guide for further details on how NetBackup conducts encryption. When using NetBackup deduplication technology, there is encryption for deduplicated data which is separate and different from the NetBackup policy-based encryption. For more information on the implementation, refer to the NetBackup Deduplication Guide.

AutoSupport Feature

Veritas Appliances such as the Access Appliance and NetBackup Appliance can call home if their health monitoring services observe hardware or software issues. The advantages of using Veritas appliances for the entire solution are the ability to automate support case management and leverage guided workflows for faster resolutions of issues and mitigation of risks. Veritas AutoSupport service provides proactive monitoring and alerting 24x7 on the health of the appliances. This feature alerts customers and/or service engineers to quickly handle the issue and reduce further risks. Enabling this feature can be done simply by registering the appliance(s) at the Veritas MyAppliance portal as shown in Figure 3 and enabling the call-home functionality. A single vendor provides end-to-end support for quicker resolution and response as opposed to having to contact multiple vendors to handle issues related to varying products and/or hardware implemented in the solution.

Figure 3- MyAppliance Portal View



Welcome to MyVeritas

MyVeritas is your single destination that allows you to access all of your Veritas enterprise services and information.

Storage

Storage utilization

Used: 287.72 GB (1%)
Available: 54650.03 GB (99%)

Deduplication Summary

Current deduplication ratio :99.26%

Hardware

Hardware is healthy.

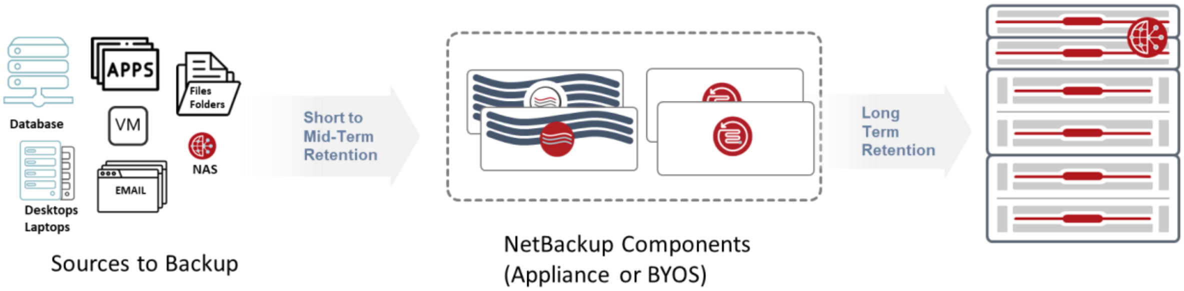
Monitoring and Detection

Available on the Access Appliance is Symantec Data Center Security (SDCS), an intrusion detection system. SDCS is a real-time monitoring and auditing software. It performs host intrusion detection, file integrity monitoring, configuration monitoring, user access tracking and monitoring, and produces logs and event reports. SDCS adds security hardening and monitoring for the Access Appliance to reduce security risks and attacks. For more information on the Access Appliance intrusion detection system, refer to the Access Appliance Initial Configuration Guide.

Solution Architecture

The high-level architecture of this solution mainly consists of the sources to backup (i.e. database, applications, virtual machines, files, and email), NetBackup components (appliance or build your own server (BYOS) and Access Appliance. As pictured in Figure 4, data is backed up to NetBackup for short-to- mid-term retention and then moved to Access Appliance for long-term retention.

Figure 4 - Solution Overview



Two ways to store backup images on Access from NetBackup include:

- **Deduplication** (Optimized Duplication) – deduplication using NetBackup MSDP deduplication technology.
- **Without Deduplication** (Traditional Duplication) - backup images are duplicated to the Access Appliance from NetBackup.

Components

In order to get a better understanding of how the Access Appliance integrates with NetBackup, all involved solution components are explained in further detail in the following sections. Information on the varying ways to deploy NetBackup in addition to the deduplication options is also discussed.

NetBackup

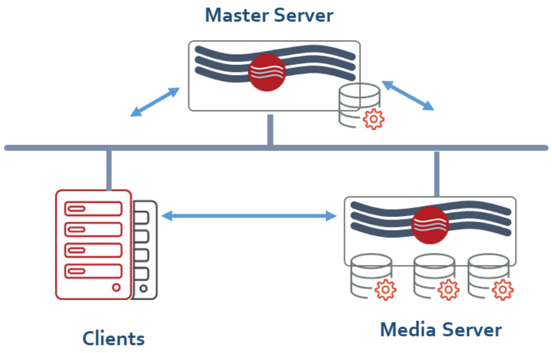
Veritas NetBackup provides protection for a wide variety of data and platforms such as operating systems, virtual systems, databases and applications, files, and all kinds of content. It has many add-on features to speed up backups, snapshot management, backup automation, and provide insights on where the active and inactive backups are located. It has the capability to backup data to tape, SAN, NAS, public or private cloud. Schedules, retention periods, and the ability to tier to different types of storage are defined in policies or storage lifecycle polices (SLP).

A typical NetBackup environment consists of three components:

- **Master Server** – manages and controls the backup and recovery activities and hosts the catalog that contains policies and schedules, metadata about the backup jobs, and media, device, and image metadata information.
- **Media Server** – writes client data as backup images to varying types of storage such as local disks, tape, network-attached storage (NAS), storage array network (SAN), cloud, etc. and later restores the data to the client as instructed by the master server.
- **Clients** – NetBackup client components are installed on hosts that have the data to be backed up and responsible for sending and receiving data to and from media server for backup and recovery.

The master and media server components can be in one system or distributed to several servers depending on the number of clients and backup workload. Typically, for a small environment, the master and media server can exist in one server and for a large environment there is one dedicated master server and several media servers. NetBackup can also be configured in a multi-domain where there are separate instances of master servers in different locations. In a multi-domain environment, each NetBackup instance is independent but can be centrally managed by NetBackup OpsCenter, a web-based console for managing, monitoring, and reporting on NetBackup operations. Figure 5 illustrates a sample configuration of a set of clients with a dedicated master server and one media server.

Figure 5 – An Example of a Dedicated Master Server and One Media Server Protecting Several Clients



NetBackup is very flexible in its deployment. It can be deployed on an all appliance solution offered by Veritas, on commodity servers (also known as Build Your Own Servers (BYOS)), or a mixture of both. It also can be run on virtual machines or in “containers” when using the Flex 5340 Appliance. Highlights of each of these deployment options are described below.

NetBackup Appliance

Veritas offers a purpose-built, highly tuned, scalable, and resilient integrated appliance for NetBackup components. These appliances address the most demanding backup and recovery requirements of enterprises. There are four different models available for varying types of workload, capacity, and performance needs. Depending on the model, usable capacity can range from 4 TB to 2160 TB and the number of expansion shelves can vary between 4 to 6 shelves. The NetBackup 5250 appliance is meant for moderate workloads and can scale up to 442 TB storage capacity, while the 5340 appliances are for demanding workloads and requiring higher usable capacity that can scale up from 1506 TB and 2160TB respectively as shown in Figure 6. The 5340 models have high availability configurations that include an additional node to continue operations should the active node fail. The NetBackup Flex Appliance(s) supports container technology allowing for multiple containers with different roles of master and media servers to be created in one appliance. It also has support to create multiple domains in one appliance. The appeal of the Flex Appliance is the "multi-tenant" capability and the ease of deploying a full NetBackup environment with multiple independent versions of NetBackup quickly. The Flex Appliance is available in two models, the 5150 appliance for small and moderate workloads and the 5340 appliance for larger and more demanding workloads.

A high-level comparison of each of the models is shown in Table 1. These enterprise -class appliances have been engineered for resiliency and versatility. They provide inline deduplication and depending on the model has a media server deduplication pool (MSDP) for intelligent deduplication to reduce the size of backups and network bandwidth to other tiers of storage such as the Access Appliance.

Table 1 - Summary of NetBackup Appliances Specifications

Model	**CPU Processor**	**RAM**	**Ports**			**Capacity**		**HA**	**Rack Units**
			1 GbE	**10/25GbE**	**8/16 Gb FC**	**Usable Capacity**	**MSDP Capacity (Max)**		
5250	2 x Xeon® 4214 (2.4 GHz) Total 24 Cores	64-512 GB (DDR4)	Up to 4	Up to 6 (10/25 GbE)	Up to 8 (8/ 16 Gb)	10 -442 TB (9.1-402 TiB)	442 TB (442 TiB)	No	Server: 2U Storage per Shelf: 2U
5340	2 x Xeon® 6138 (2 GHz) Total 40 Cores	786GB-1.6 TB (DDR4)	4	2-10 (10 GbE)	Up to 8 (8 Gb)	132-2160 TB (120-1920TiB)	1056 TB (960 TiB)	Yes	Server: 2U Storage per Shelf: 5U
Flex 5150	1 x Xeon Bronze 3106 (1.7 GHz) Total 8 cores	64 GB (DDR4)	4	2 (10 GbE/25 GbE) (optional)		16 TB (14.5 (TiB)	16 TB (14.5 (TiB)	No	Server: 1U

Flex 5340	2 x Xeon® 6138 (2 GHz) Total 40 Cores	786GB-1.6 TB (DDR4)	4	2-10	Up to 8	264-2160 TB (240-1920TiB)	1056 TB (960 TiB)	Yes	Server: 2U Storage per Shelf: 5U

NOTE: TB - Capacity values are calculated using Base 10; TiB - Capacity values are calculated using Base 2.

For more information on the specific details of each of these models such as power requirements, number of network interfaces, refer to the following datasheets:

- NetBackup 5250 Datasheet
- NetBackup 5340 Datasheet
- NetBackup Flex 5150 Datasheet
- NetBackup Flex 5340 Datasheet

NetBackup Build Your Own Server (BYOS)

NetBackup components can be deployed on commodity servers referred to as BYOS (Build Your Own Servers). NetBackup can run on a Linux or Windows platform. Refer to the NetBackup Master Compatibility List for a full list of platform versions supported. In production, the minimum requirement for a master server is 4 cores and 16 GB of memory. For each media server, there is a 4 GB minimum memory requirement and for clients a minimum of 512 MB is required.

Other than the hardware and platform differences, there is a difference in the maximum MSDP capacity that can be set up on a single server in BYOS. The MSDP capacity for BYOS is limited to 250TB per server for systems configured with Red Hat Enterprise Linux (RHEL) and SUSE Linux and 64 TB for others.

NetBackup Virtual Appliances

NetBackup can also be run on virtual appliances; however, this deployment is appropriate mostly for remote offices. Implementing NetBackup on virtual machines provides a simple deployment and minimizes capital expenditures. The NetBackup Hardware and Cloud Storage Compatibility List contains more information on the supported hypervisors. Depending on which hypervisor is chosen to deploy NetBackup and the role of either master or media servers, minimum requirements may differ.

NetBackup MSDP Deduplication

Backup images are generally ideal for deduplication since the probability of encountering duplicated blocks of data are higher when compared to other forms of data types such as encrypted data. The deduplication ratio defines how well data can be deduplicated. The higher the ratio, the more space is saved. Deciding on whether to deduplicate your data or not depends on several factors: data type, data change rate, retention period, and backup policy. For instance, encrypted data is inherently unique and will not benefit from any deduplication savings. Data that has a high change rate will not take advantage of the savings long enough to justify the overhead imposed by deduplication. In the context of backup images, daily full backups will have higher deduplication ratios when compared with incremental or differential backups.

NetBackup MSDP is Veritas proprietary deduplication technology. Without support for NetBackup MSDP in the storage target, deduplicated data from NetBackup MSDP is rehydrated prior to sending data to the storage platform. NetBackup allows for inline deduplication of backup images on either the client or media server. The difference between the client side or media server deduplication is where the deduplication occurs. For client-side deduplication, the backup data is first deduplicated on the client before being sent to target storage. Client-side deduplication uses available resources on clients and reduces the network traffic since deduplicated data is sent over the network. In either scenario, the backup images are placed in a media server deduplication pool (MSDP).

Architecturally NetBackup MSDP deduplication is composed of the following main components:

- **Deduplication Plugin** - Separate the data into segments or chunks. Use a hash algorithm to calculate fingerprints to identify each unique segment. Compare incoming data fingerprints with the fingerprints of existing data.
- **Deduplication Engine** (spoold) **-** manage and store the fingerprint database and metadata, store unique segments or use a reference or pointer to the data already stored, and conducts integrity checks.
- **Deduplication Manager** (spad) – maintains the configuration, controls and dispatches the internal processes, security and events handling.

NetBackup MSDP utilizes SHA-2 (SHA256) for the hash algorithm. The chunk segment size unit used to compute fingerprints is by default a fixed length of 128 KB or configurable to variable-length size based on chunk boundary. NetBackup MSDP also compresses deduplicated data for further storage efficiency. Furthermore, there is an option to encrypt deduplicated data. For more information on the architecture of NetBackup MSDP deduplication technology, refer to the NetBackup Deduplication Guide.

There are two methods to send deduplicated data to Access without rehydration:

- **MSDP Cloud Tiering (MSDP-C) –** deduplication is done by media server and deduplicated data is sent to Access using MSDP cloud tiering. S3 protocol is used in this scenario.
- **Veritas Data Deduplication (VDD)** – deduplication done by NetBackup media server or client prior to sending to Veritas Data Deduplication pool. MSDP proprietary protocol is used to send data to Access.

MSDP Cloud Tiering (MSDP-C)

MSDP has support to send deduplicated data without rehydration directly to Access Appliance referred to as MSDP Cloud (MSDP-C) tiering. Configuring a NetBackup appliance or BYOS as a MSDP-C storage server allows data to be sent via S3 protocol to one local storage target and one or more cloud storage targets. However, there is a combined capacity support of 1.2 PB for both block and object.

Access Appliance Data Deduplication (MSDP Support)

To support NetBackup MSDP in Access, the NetBackup MSDP technology components ported to Access Appliance include the deduplication engine (spoold) and deduplication manager (spad). This feature was introduced in Access Appliance version 7.4.2. The primary functions of these components are to manage and store unique data, fingerprints, metadata, and the associated logs and journals. Both nodes share the storage used as a deduplication pool, but only one runs the management processes at a time, in an active/passive configuration. If the active node fails or is unreachable, Access cluster management will automatically start the necessary components on the passive node and resume the processing and storing of deduplicated data.

NOTE: A NetBackup media server or client with the deduplication plug-in is still required to do the actual deduplication of data which involves segmentation of data, calculating fingerprints and comparison with existing fingerprints. During configuration of the Access Appliance as a storage server in NetBackup, a media server is required and needed during restores.

An advantage of VDD is in a multi-domain NetBackup configuration or when utilizing multiple media servers in one domain, data is “globally” deduplicated in Access such that only one copy of the data is stored between the domains.

Comparison Highlights of MSDP-C with Access and VDD

Highlights of the basic differences between the two options to support NetBackup MSDP with the Access Appliance are in Table 3.

Table 2 - Feature Differences of Utilizing MSDP-C and VDD

Features	**MSDP-C**	**VDD**
Transfer Protocol	S3 (HTTP/HTTPs)	Proprietary (TCP)
Ports	8143	10082 and 10102
Destination Type	Bucket	Data Deduplication Pool
File system Type (default) on Access	Clustered File system and simple or stripe data layout based on Access version.	Clustered File system and striped data layout
Cloud Support	Via NetBackup MSDP-C	Via NetBackup MSDP-C
Data Stored in Access	Metadata, unique data segments, and encryption key if enabled.	Metadata, fingerprints database, unique data segments, and encryption key (if enabled), journals and logs.
Target MSDP Size	Maximum combined (local and object) is 1.2 PB per MSDP-C storage server instance.	Maximum 1.4 PB qualified
Global Deduplication in Multi-Domain	No	Yes

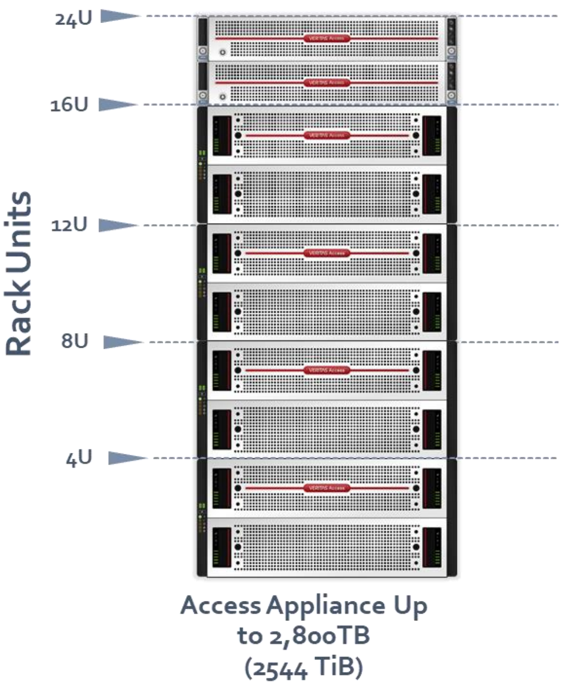
Traditional Duplication (Without Deduplication)

In some cases, deduplication of backup images is not ideal. Backups that have a strict time limit for restores, have a high rate of change, or encrypted are not good candidates for deduplication. Another case is when incremental backups are done instead of full. For these types of data or backup, images are best sent to the Access Appliance without deduplication. Data is sent to Access Appliance using the S3 protocol and stored in an S3 bucket.

Access Appliance

NetBackup can send backup images to various storage types (disk, tape, cloud, etc.) for long-term retention. For those seeking an on-premises disk-based solution for faster recovery times, control and/or simplicity when compared to tape or cloud, Veritas has developed the Access Appliance for ease of acquisition, management, and support. Access Appliance is a turn-key storage solution designed for high capacity and cost optimization, making it well suited for long-term retention. The Access Appliance model 3340 is comprised of two clustered nodes and one primary storage shelf and up to three additional expansion storage shelves. The appliance can scale up to 2,800 TB of usable space as can be seen in Figure 6.

Figure 6 - Access Appliance Rack Units



Highlights of Access Appliance specifications are shown in Table 4. Refer to the Access 3340 Appliance datasheet for more detailed information.

Table 3 - Highlights of Access Appliance Specifications

Model	CPU Processor	RAM	Ports		Capacity	Rack Units
			1 GbE	**10 GbE**		
3340 (2 nodes)	2 x Xeon® 4108 (1.8 GHz) per node Total: 16 core per node	384 GB per node	4 per node	2 per node	280 TB – 2800 TB (254 TiB – 2544 TiB)	Server: 2U Storage per Shelf: 5U

Note: TB - Capacity values are calculated using Base 10; TiB - Capacity values are calculated using Base 2.

The two nodes are clustered in active/active configuration such that each node can handle I/O requests. Storage shelves are connected to each node and configured with dynamic multipathing so I/O can be sent to either node for performance and availability. The redundant hardware RAID controller in the primary storage shelf configures and presents the shelves' physical disks into disk groups (volumes) protected by a RAID 6 storage layout. With a RAID 6 configuration, data with dual parity is striped across the configured volumes (5 volumes per storage shelf with each volume containing 16 disks). Each data volume can remain operational despite two concurrent disk failures.

The nodes run RHEL 7.7 or later as the operating system platform and Access software version 7.4.2.301 or later. The Access Appliance is a scale-up Network-Attached Storage platform that supports multiple protocols, including NFS, CIFS, FTP HTTP and S3. In addition, there is also a proprietary protocol used to communicate between NetBackup MSDP and VDD engine and manager. With NetBackup, the Access Appliance is seen as an s3 target or a data deduplication pool target. In both, NetBackup MSDP is the deduplication technology implemented.

When using Access as an S3 target, deduplicated data written to Access from NetBackup is placed in an S3 bucket. A bucket maps to a single file system of type cluster file system (cfs). The Access Appliance supports a maximum usable capacity of 2.8 PB and thus the maximum size of an S3 bucket in an appliance is 2.8 PB. When using the S3 protocol to backup images from NetBackup, the S3 object URL presented to clients is of the form *s3.<clustername>:8143*. Clients such as NetBackup utilize this URL as the S3 endpoint for reading and writing to the Access bucket. Clients simply map this S3 object URL to one of the Access virtual IPs. A dedicated S3 communication port, 8143, is required for both HTTP and HTTPS and thus firewalls must keep this port open.

Introduced in Access 7.4.2 is VDD. When deploying VDD, the fingerprints, metadata, and deduplicated data are placed in a cluster file system with a stripe layout. Data deduplication requires 5 logical volumes for the storage pool or one full shelf to be allocated. However, it is not required that the entire storage pool be dedicated for the data deduplication filesystems created. When adding to the storage pool, add the 5 volumes on the entire shelf to maintain the stripe. One or more filesystems will be created based on the size of the deduplication pool defined. Data is distributed equally among the file systems; however, the fingerprint database and metadata will be held in a separate filesystem. A proprietary protocol is implemented to communicate between NetBackup MSDP and VDD engine and manager. Ports 10082 for Access deduplication engine and 10102 for the Access deduplication manager must be opened on the corporate firewall for this communication. As previously stated, VDD runs in active/passive mode on the appliance. The active node is determined when specifying the virtual IP during configuration and provisioning of VDD. Access will start deduplication services on the node that owns the physical interface associated with the virtual IP (VIP) specified during configuration. For instance, a VIP of 10.182.81.87 is associated with physical interface "eth4" of node two and this VIP is used during the data deduplication configuration, then node two becomes the active node and the other node is passive. **NOTE**: Qualified maximum data deduplication pool size is 1.4 PB and a minimum of 384 GB of memory per node is required for this feature. By default, the fingerprint cache utilizes 50% of the memory. Also, maximum supported concurrent jobs or streams is 80.

For management, the appliance can be managed by the command-line shell referred to as the CLISH and a web-based graphical user interface (GUI) where one can create and provision Access as an S3 target and/or data deduplication pool. A configuration wizard is also available on the GUI for quick provisioning of an Access S3 bucket and VDD for NetBackup. The wizard walks user through creation of a storage pool of disks, activation of the appropriate policies and provisioning of an S3 bucket or data deduplication pool for NetBackup.

NOTE: For examples of how to deploy and configure the Access Appliance with NetBackup with MSDP-C refer to the Quick Start section of Veritas Deduplication Guide and NetBackup with VDD, refer to the Veritas Access Solutions Guide for NetBackup.

Solution Data Flow

This section explores how all these components integrate and how data flows through each component. The flow of data depends on whether deduplication using MSDP-C, VDD or no deduplication is employed. Also, the definitions within the NetBackup policies and/or Storage Lifecycle policies (SLP) also affect the flow.

In all data flow scenarios, an Open Storage Technology (OST) plugin is necessary to send data to Access Appliance. OST is an API developed by Veritas in order to allow third-party vendors to develop a software plugin module that tightly integrates their products with NetBackup software. This plugin is installed on the media servers to communicate with the vendor's storage device. OST plugins were developed to send data to S3 compatible cloud providers and VDD pool. Both plugins are by default installed on NetBackup media and master servers to send data to Access.

Deduplication Data Flow with MSDP-C

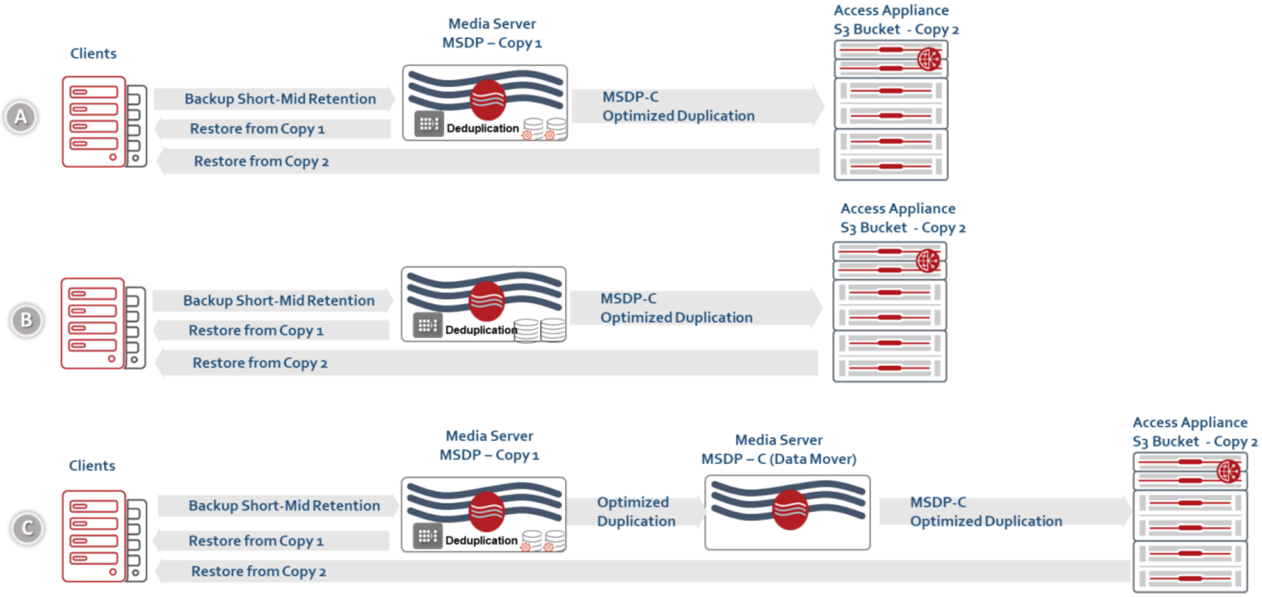
NetBackup polices and/or SLP define the path or flow of data. Regular policies can be defined if the data needs to be sent to a single target and SLP can be set up to backup, duplicate, and/or replicate the data in different storage types or destinations. For instance, an SLP can send backup first to a media server with faster disks and then duplicate to another media server with slower disks or to an MSDP and/or to secondary storage for long-term retention. Data is sent to an Access S3 bucket utilizing the S3 protocol with the S3 OST cloud plugin installed on a media server. Examples of NetBackup MSDP deduplication technology with MSDP-C and Access data paths are explained below and illustrated in Figure 7:

A. Data from clients are initially backed up and deduplicated to an MSDP local storage disk pool residing on a media server for short or mid-term retention (Copy 1). This deduplicated data is sent to Access Appliance via MSDP-C for long-term retention (Copy 2). For restores, a copy of the data can be restored from either the first copy on MSDP or the second copy residing on Access S3 bucket. However, by default, restores are retrieved from Copy 1. The media server is required to rehydrate the deduplicated data from the MSDP copy.

B. Data is backed up on an advanced disk on the media server for short or mid-term retention and then data is deduplicated and unique data is sent to the Access Appliance S3 bucket via MSDP-C. In case of restores, the data by default will be restored from the advanced disk (Copy 1) unless specified to restore from the Access Appliance copy 2. The media server will rehydrate the deduplicated data before sending to the client.

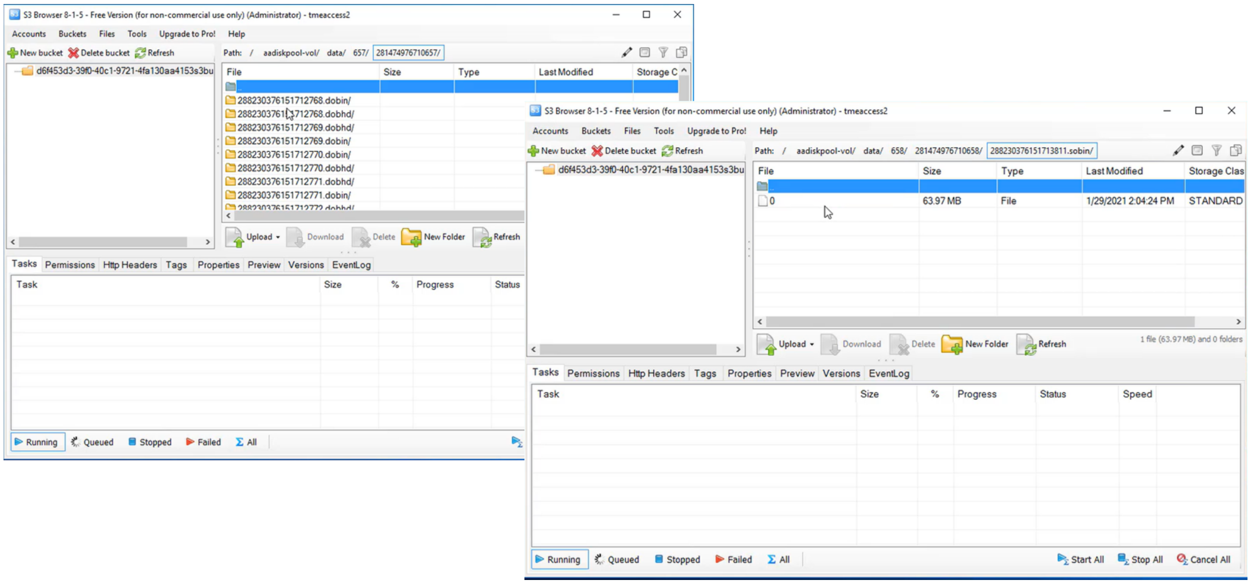
C. The maximum combined MSDP size for local and object is 1.2 PB per MSDP-C storage server. In order to utilize a 1 PB Access S3 bucket and adhere to this requirement, another media server can be deployed to act mainly as a data mover. In example C, data from clients are initially backed up and deduplicated to an MSDP local storage disk pool residing on a media server for short or mid-term retention (Copy 1). Another media server is deployed to be used as a data mover to send deduplicated data to Access Appliance via MSDP-C for long-term retention (Copy 2). As in the other examples, to restore data, a copy of the data can be restored from either the first copy on MSDP local storage or the second copy residing on Access S3 bucket.

Figure 7 – Examples of MSDP Deduplication with MSDP-C to Access Appliance S3 Bucket Data Flows



In addition to the MSDP meta-data, MSDP separates a backup image into container files and adds a header for each container. Thus, there are two files for each MSDP container consisting of a data container file holding unique data with fingerprint and header file for each container. If encryption is enabled, there is additional information relating to the keys and header for the keys. A sample view using the S3Browser application shows the data in an Access S3 bucket after being sent from MSDP-C as shown in Figure 8.

Figure 8 - Sample View of the Data in an S3 Bucket Sent from MSDP-C.


S3 Browser 8-1-5 - Free Version (for non-commercial use only) (Administrator) - tmeaccess2

Path: / aadiskpool-vol/ data/ 657/ 281474976710657/

d6f453d3-39f0-40c1-9721-4fa130aa4153s3bu

File	Size	Type	Last Modified	Storage C
..				
288230376151712768.dobin/				
288230376151712768.dobhd/				
288230376151712769.dobhd/				
288230376151712769.dobin/				
288230376151712770.dobin/				
288230376151712770.dobhd/				
288230376151712771.dobhd/				
288230376151712771.dobin/				
288230376151712772.dobhd/				

S3 Browser 8-1-5 - Free Version (for non-commercial use only) (Administrator) - tmeaccess2

Path: / aadiskpool-vol/ data/ 658/ 281474976710658/ 288230376151713811.sobin/

d6f453d3-39f0-40c1-9721-4fa130aa4153s3bu

File	Size	Type	Last Modified	Storage Clas
..				
0	63.97 MB	File	1/29/2021 2:04:24 PM	STANDARD

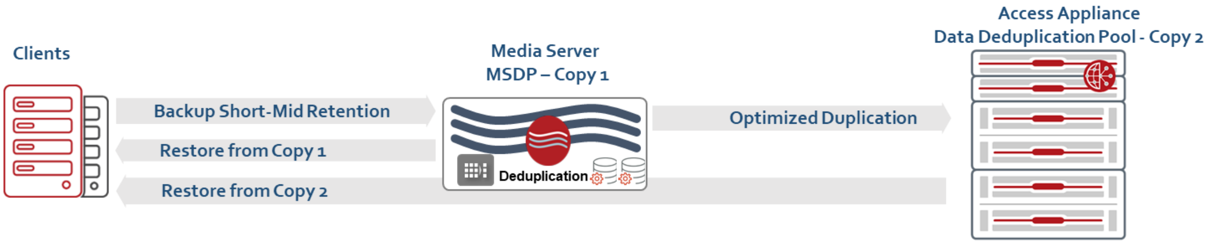
1 file (63.97 MB) and 0 folders

Veritas Data Deduplication (VDD) Flow

As discussed, policies and SLP define how backup images are stored and tiered to other storage platforms. Hence, the data path is dependent on the defined policies and SLP. When utilizing VDD, a media server is required to do the deduplication of the data. Access main role in the path is to store the unique data, fingerprints database, metadata in addition to supporting files such as journals and logs. As shown in the example in Figure 9, data from clients are initially backed up to an MSDP on a media server where it is deduplicated and stored for short to mid-term retention and is the primary copy. An SLP is defined to duplicate the unique data blocks to the Access Appliance for the second copy. Restores can be done from Copy 1 (default) residing on an MSDP in media server or from copy 2 residing in Access. When restoring from copy 2, a media server is required to retrieve the data from Access, rehydrate and send to client.

As previously discussed, data is sent to Access using NetBackup MSDP proprietary protocol based on the OST framework.

Figure 9 - Example of NetBackup Data Flows to VDD Pool



One advantage of VDD is the ability to do global deduplication in a multi-domain environment or multiple media servers in one domain. A single Access Appliance can be the target storage for multiple media servers from same or different domains. Figure 10 illustrates the flow of how global deduplication is achieved with NetBackup and VDD in a multi-domain environment. A description of flow is as follows in either domain:

1. The local media server is responsible for segmentation and fingerprinting of the backup data into data blocks. It will check the media server fingerprint cache for duplicates and stores only unique blocks in local media server media server deduplication pool.
2. A duplication is conducted to create a second copy on the Access Appliance. The local media server fetches and compares the fingerprints from previous backup of image and then on the fingerprint cache on the Access Appliance. If data block does not exist, then the unique data block is sent to the Access Appliance.

It is assumed that the Access Appliance is reachable from both domains. **NOTE**: Conditions in which deduplication and global deduplication is not achievable include a) fingerprint is not in local media server or Access fingerprint cache (size is based on physical memory) during comparison and b) fragmentation where the blocks are in different containers that are far apart such that it affects performance of restores. Frequency of backups and/or duplication jobs can also affect the differences in the deduplication rate observed on media server and the Access Appliance.

Figure 10 - Global Deduplication in Multi-Domain Environment

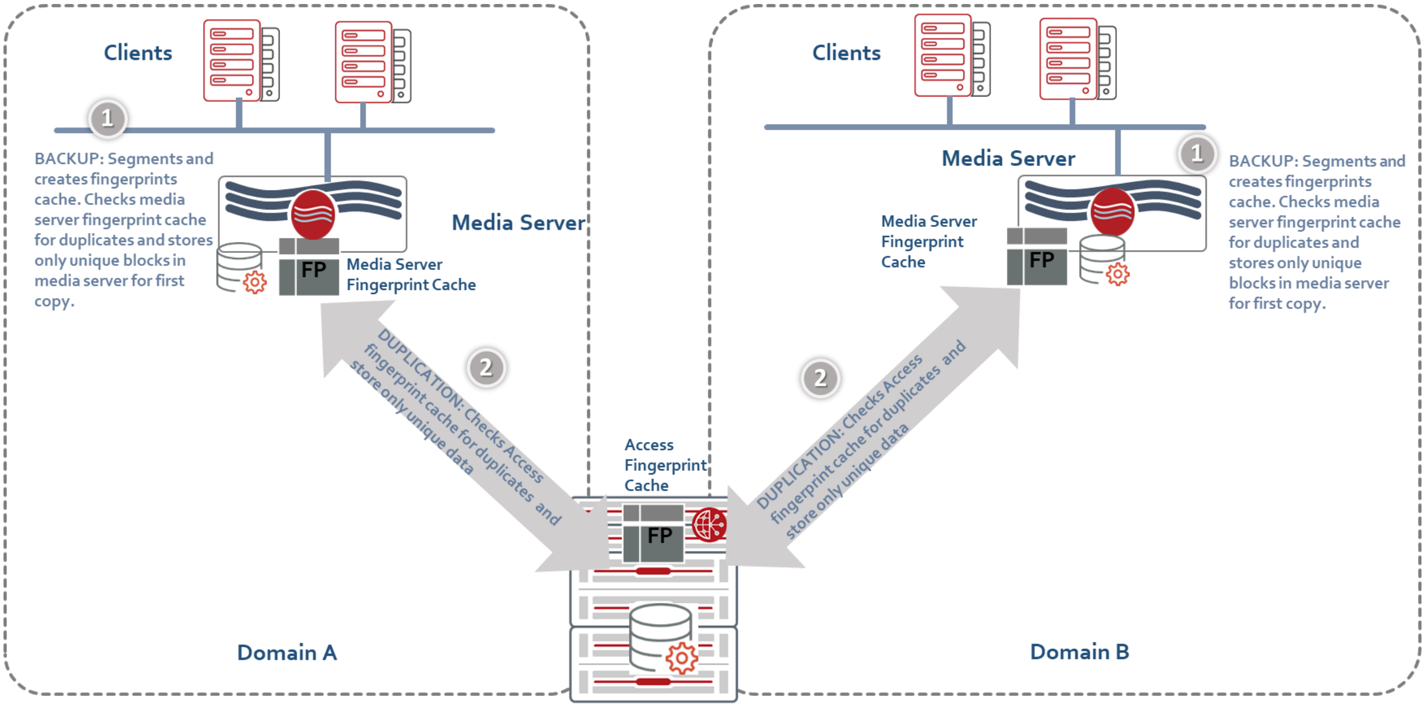
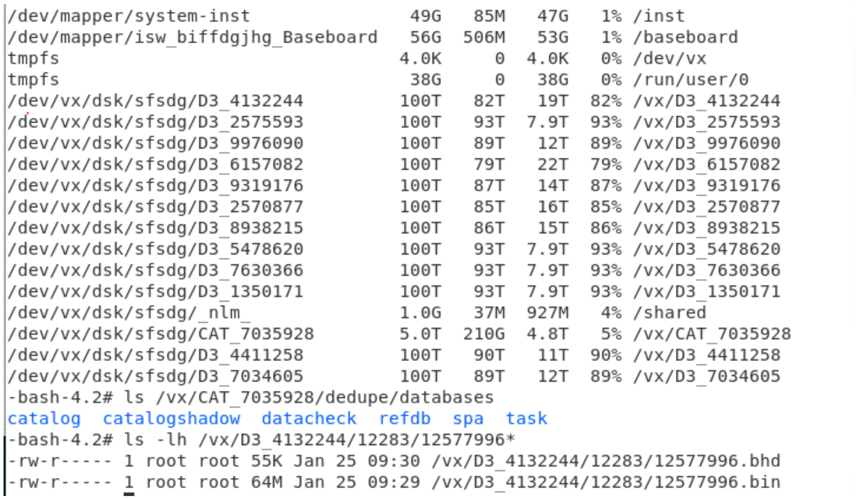

Clients
1
BACKUP: Segments and creates fingerprints cache. Checks media server fingerprint cache for duplicates and stores only unique blocks in media server for first copy.
Media Server
FP
Media Server Fingerprint Cache
2
DUPLICATION: Checks Access fingerprint cache for duplicates and store only unique data
Access Fingerprint Cache
FP
Domain A
Clients
Media Server
1
BACKUP: Segments and creates fingerprints cache. Checks media server fingerprint cache for duplicates and stores only unique blocks in media server for first copy.
Media Server Fingerprint Cache
FP
2
DUPLICATION: Checks Access fingerprint cache for duplicates and store only unique data
Domain B

Figure 11 provides a view of how the data is stored on Access. In this example, there are multiple 100 TB filesystems for the deduplicated data (i.e. /vx/D3_*) and another filesystem for the MSDP catalog (i.e./vx/CAT_*). Listing the contents of the MSDP catalog directory one can see the fingerprint database, journals, logs and other associated directories. The data directory holds the actual the backup image container (i.e. *64.bin) and the header information (i.e. *64.bhd) associated with the backup image container. The image containers hold multiple segmented data objects and by default, the maximum size of the containers is 64 MB. If encryption is enabled on NetBackup, then there are also files for the encryption key and header of key.

Figure 11 - Example View of the Data Stored on VDD Pool



```
/dev/mapper/system-inst             49G   85M   47G   1% /inst
/dev/mapper/isw_biffdgjhg_Baseboard  56G  506M   53G   1% /baseboard
tmpfs                              4.0K     0  4.0K   0% /dev/vx
tmpfs                               38G     0   38G   0% /run/user/0
/dev/vx/dsk/sfsdg/D3_4132244       100T   82T   19T  82% /vx/D3_4132244
/dev/vx/dsk/sfsdg/D3_2575593       100T   93T  7.9T  93% /vx/D3_2575593
/dev/vx/dsk/sfsdg/D3_9976090       100T   89T   12T  89% /vx/D3_9976090
/dev/vx/dsk/sfsdg/D3_6157082       100T   79T   22T  79% /vx/D3_6157082
/dev/vx/dsk/sfsdg/D3_9319176       100T   87T   14T  87% /vx/D3_9319176
/dev/vx/dsk/sfsdg/D3_2570877       100T   85T   16T  85% /vx/D3_2570877
/dev/vx/dsk/sfsdg/D3_8938215       100T   86T   15T  86% /vx/D3_8938215
/dev/vx/dsk/sfsdg/D3_5478620       100T   93T  7.9T  93% /vx/D3_5478620
/dev/vx/dsk/sfsdg/D3_7630366       100T   93T  7.9T  93% /vx/D3_7630366
/dev/vx/dsk/sfsdg/D3_1350171       100T   93T  7.9T  93% /vx/D3_1350171
/dev/vx/dsk/sfsdg/_nlm_            1.0G   37M  927M   4% /shared
/dev/vx/dsk/sfsdg/CAT_7035928      5.0T  210G  4.8T   5% /vx/CAT_7035928
/dev/vx/dsk/sfsdg/D3_4411258       100T   90T   11T  90% /vx/D3_4411258
/dev/vx/dsk/sfsdg/D3_7034605       100T   89T   12T  89% /vx/D3_7034605
-bash-4.2# ls /vx/CAT_7035928/dedupe/databases
catalog  catalogshadow  datacheck  refdb  spa  task
-bash-4.2# ls -lh /vx/D3_4132244/12283/12577996*
-rw-r----- 1 root root 55K Jan 25 09:30 /vx/D3_4132244/12283/12577996.bhd
-rw-r----- 1 root root 64M Jan 25 09:29 /vx/D3_4132244/12283/12577996.bin
```

Traditional Duplication Data Flow

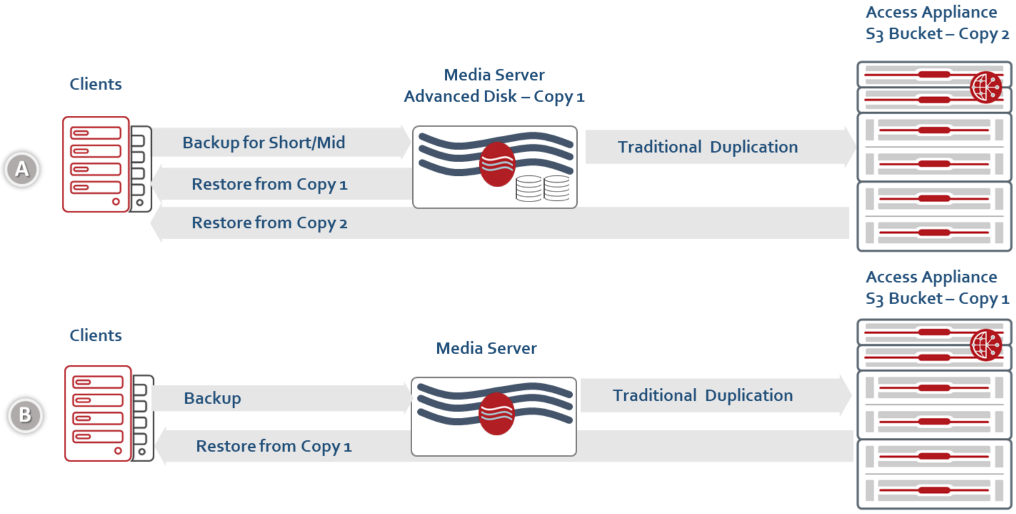
For data that are not deduplicated, a copy of the backup images is sent to Access appliance. Depending on policies and/or SLP defined, examples of traditional duplication to Access appliance are as follows and pictured in Figure 12:

A. Data are first stored in an advanced disk on a media server for short to mid-term retention and later copied to the Access Appliance. Restores can be done from either the advanced disk (Copy 1 - default) or from Access S3 bucket (Copy 2).

B. Data are sent from clients to media server to create a backup images and directly copied to the Access Appliance S3 bucket. Restores from Access (Copy 1) goes through media server to assemble and then sent to client.

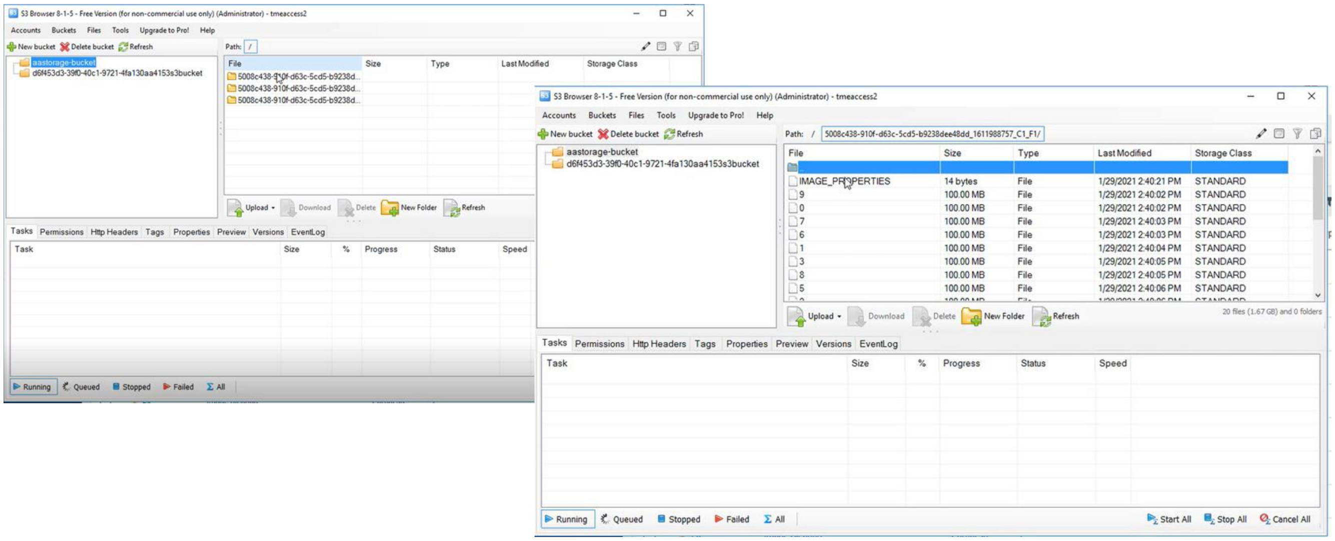
Data is sent to Access Appliance using the S3 OST Cloud plugin to send data to Access S3 bucket.

Figure 12 - Traditional Duplication Data Flow from NetBackup to Access Appliance



A sample view of data on an Access S3 bucket after traditional duplication is shown in Figure 13. Backup images and its associated header information are stored in a directory structure. Inside each directory contains the image properties, block map file, and the actual image. And the header directory contains the header information, properties of header information, and block map file for the header. The S3 OST Cloud plugin will send data to Access S3 bucket into configurable fixed object sizes.

Figure 13 - Example View of Data on Access Appliance After Traditional Duplication.



S3 Browser 8-1-5 - Free Version (for non-commercial use only) (Administrator) - tmeaccess2

aastorage-bucket
d6f453d3-39f0-40c1-9721-4fa130aa4153s3bucket

Path: /

File	Size	Type	Last Modified	Storage Class
5008c438-910f-d63c-5cd5-b9238d...				
5008c438-910f-d63c-5cd5-b9238d...				
5008c438-910f-d63c-5cd5-b9238d...				

S3 Browser 8-1-5 - Free Version (for non-commercial use only) (Administrator) - tmeaccess2

aastorage-bucket
d6f453d3-39f0-40c1-9721-4fa130aa4153s3bucket

Path: / 5008c438-910f-d63c-5cd5-b9238dee48dd_1611988757_C1_F1/

File	Size	Type	Last Modified	Storage Class
..				
IMAGE_PROPERTIES	14 bytes	File	1/29/2021 2:40:21 PM	STANDARD
9	100.00 MB	File	1/29/2021 2:40:02 PM	STANDARD
0	100.00 MB	File	1/29/2021 2:40:02 PM	STANDARD
7	100.00 MB	File	1/29/2021 2:40:03 PM	STANDARD
6	100.00 MB	File	1/29/2021 2:40:03 PM	STANDARD
1	100.00 MB	File	1/29/2021 2:40:04 PM	STANDARD
3	100.00 MB	File	1/29/2021 2:40:05 PM	STANDARD
8	100.00 MB	File	1/29/2021 2:40:05 PM	STANDARD
5	100.00 MB	File	1/29/2021 2:40:06 PM	STANDARD

20 files (1.67 GB) and 0 folders

Task	Size	%	Progress	Status	Speed

Disaster Recovery

Having a disaster protection plan is imperative for business continuity. The NetBackup catalogs and data are key components to protect. NetBackup Auto Image Replication (AIR) is a mechanism in which NetBackup replicates the necessary components from NetBackup domain to another NetBackup domain in one or numerous geographical sites for disaster recovery. At the replicated site, the data is first replicated to an MSDP on media server and then optimized duplicated to the Access Appliance. Each domain should contain an Access Appliance with a bucket or VDD pool defined. The size of the target bucket and VDD pool should be the same size or greater than the source. **NOTE**: Replication to another site is done asynchronously.

AIR is only one approach to protect against storage failure and site loss in a NetBackup environment with Access Appliance and allows for rapid recovery. Protection of the NetBackup catalogs is critical, and it is recommended to do regular backups of the catalogs to protect against corruption or accidental deletion. More information on other ways on how to protect NetBackup environments from disaster can be found in the Veritas NetBackup in Highly Available Environment's Administrator's Guide. Please also refer to the References section for other links relating to this topic.

The following sections illustrate sample flows of NetBackup AIR configuration where Access Appliance is used for long term retention in the deployment. **NOTE**: A single Access Appliance cannot store or act as both the source and destination of AIR. It is expected that the destination of the replication is a different Access Appliance.

AIR and Optimized Duplication (MSDP-C) to Access S3 Bucket

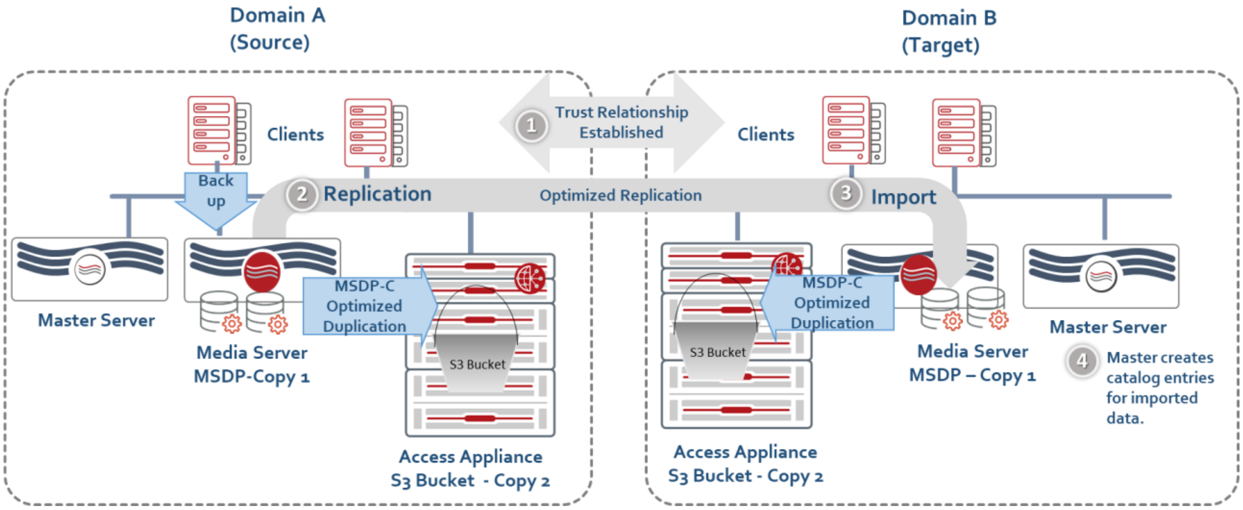
An a MSDP on a media server needs to be configured on both source and target for the replication. Storage lifecycle polices (SLP) is where a replication can be specified after a backup and/or duplication. A sample flow of NetBackup AIR where MSDP-C and Access S3 bucket is utilized involves (also illustrated in Figure 14):

1. A trust relationship is established between the NetBackup servers in the domains where credentials and certificates are required for authentication.
2. A sample SLP on the ***source*** is defined as follows:
    a. Backup to an MSDP local disk storage pool (Copy 1)

   b. Duplicate to Access S3 bucket via MSDP-C (Copy 2)
   c. Replicate to MSDP on target (Domain B). There is no rehydration of optimized data.
3. An SLP on the ***target*** is defined as follows:
   a. Import to an MSDP local disk storage pool (Copy 1)
   b. Duplicate to Access S3 bucket via MSDP-C (Copy 2)
4. The master server at the target domain automatically creates the entries in the NetBackup catalog as the data is being imported.

NOTE: NetBackup AIR is only occurring between the MSDP on the media servers on source and target. The media server is responsible for the control and orchestration of the entire lifecycle policies in each domain. Each site maintains its own metadata and fingerprint database and thus the fingerprints are compared at the target site and only unique data is sent to target Access Appliance.

Figure 14 - Sample Flow of NetBackup AIR with MDSP-C and Access S3 Bucket in the Environment


Domain A (Source)
Domain B (Target)
1 Trust Relationship Established
Clients
Clients
Back up
2 Replication
Optimized Replication
3 Import
Master Server
Media Server MSDP-Copy 1
MSDP-C Optimized Duplication
S3 Bucket
Access Appliance S3 Bucket - Copy 2
S3 Bucket
MSDP-C Optimized Duplication
Media Server MSDP – Copy 1
Master Server
4 Master creates catalog entries for imported data.
Access Appliance S3 Bucket - Copy 2

AIR for Veritas Data Deduplication (VDD)

When using VDD with AIR, the Access Appliance acts as a duplication target for MSDP on target domain. After the replication and import processes are done by the media servers MSDP, the data is duplicated to Access Appliance data deduplication pool. A sample flow of NetBackup AIR with VDD illustrated in Figure 15 is as follows:

1. A trust relationship is established between the NetBackup servers in the domains where credentials and certificates are required for authentication.
2. A sample SLP on the ***source*** is defined as follows:
   a. Backup to an MSDP local disk storage pool (Copy 1 on source)
   b. Duplicate from MSDP to Access Appliance deduplication pool (Copy 2).
   c. Replicate to target MSDP local disk storage pool (Copy 1) to MSDP local disk storage pool (Copy 1) in Domain B.

   The media server will initiate the backup and replication processes on source. There is no rehydration of optimized data.
3. A sample SLP on the ***target*** is defined as follows:
   a. Import to an MSDP local disk storage pool (Copy 1)

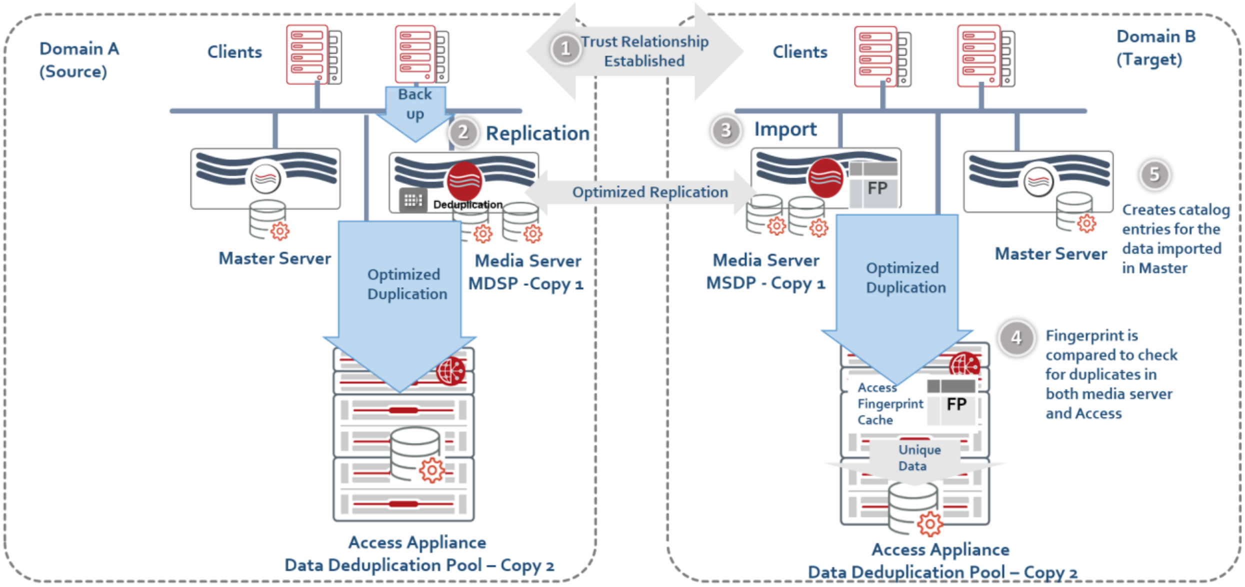
   b. Duplicate from MSDP local disk storage pool to VDD pool (Copy 2 on target)

   The media server will initiate the import process on target.

4. The media server receives optimized data sent, compares the fingerprints in its cache, and stores only unique data on disk and then it is duplicated to Access Appliance where fingerprints are also checked, and only unique data is stored. The fingerprint and metadata (catalogs) on the target side is updated as data is being written.
5. The master server at the target domain automatically creates the entries in the NetBackup catalog as the data is being imported.

NOTE: Each site maintains its own metadata and NetBackup catalogs. Metadata and catalogs are updated as data is imported.

Figure 15 - NetBackup AIR with VDD



Another example of AIR with VDD is shown in Figure 16. In this example, there is only an Access Appliance in the target domain. In this example, the sample flow is as follows:

1. A trust relationship is established between the NetBackup servers in the domains where credentials and certificates are required for authentication.
2. A sample SLP on the ***source*** is defined as follows:
   a. Backup to an MSDP local disk storage pool (Copy 1 on source)
   b. Replicate from source MSDP local disk storage pool (Copy 1) in Domain A to target MSDP Pool (Copy 1) in Domain B.

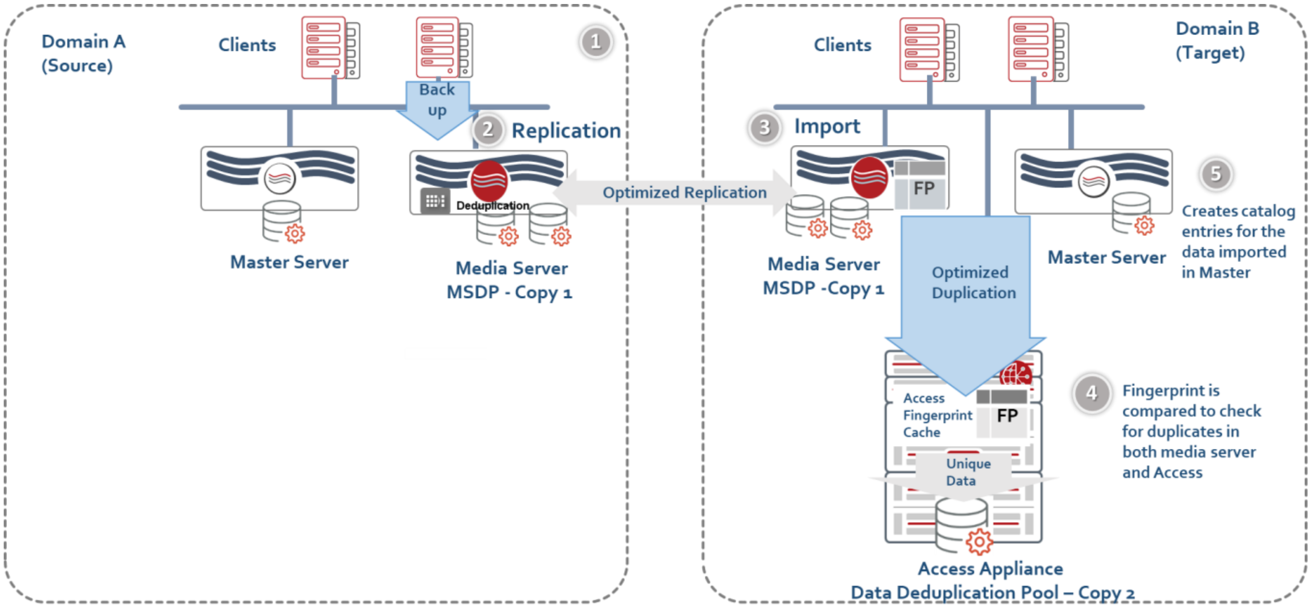
   The media server will initiate the backup and replication processes on source. There is no rehydration of optimized data.
3. A sample SLP on the ***target*** is defined as follows:
   a. Import to an MSDP local disk storage pool (Copy 1) in Domain B
   b. Duplicate from MSDP local disk storage pool to VDD pool (Copy 2 on target)

   The media server will initiate the import process on target.

4. The media server receives optimized data sent, compares the fingerprints in its cache, and stores only unique data on disk and then it is duplicated to Access Appliance where fingerprints are also checked, and only unique data is stored. The fingerprint and metadata (catalogs) on the target side is updated as data is being written.
5. The master server at the target domain automatically creates the entries in the NetBackup catalog as the data is being imported.

   NOTE: As in previous example, each site maintains its own metadata and NetBackup catalogs. Metadata and catalogs are updated as data is imported.

Figure 16 - Another Example of AIR with VDD



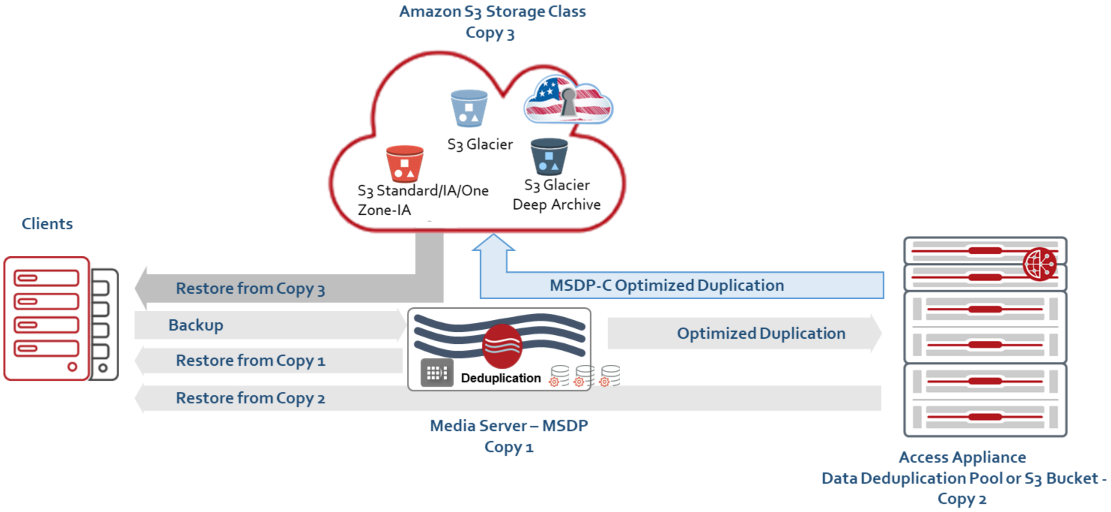
AIR With Access S3 Bucket (Without Deduplication)

Using NetBackup AIR is not supported for the S3 connector/plugin and for advanced disk as a replication target. One can protect the data on Access Appliance by having several copies of the data and NetBackup catalog on-premises or one can duplicate the data and NetBackup catalogs to a remote site. **NOTE**: When duplicating non-deduplicated data to a remote site within the same domain, keep in mind that transferring large amounts of data can consume the network bandwidth and may take a long time.

Cloud Support

When sending MSDP data from Access Appliance to the cloud, one can duplicate data to the cloud utilizing NetBackup SLP policies and MSDP-C. The SLP would specify to duplicate the data from Access Appliance using MSDP-C to duplicate the data to the cloud. Figure 17 provides a view of how data is sent to the cloud from VDD or S3 bucket. In this example, an optimized duplication is sent to the Access Appliance and then it does an optimized duplication to the cloud via MSDP-C. The role of the media server during the optimized duplication is to control and orchestrate the transfer between Access Appliance and MSDP-C. The actual I/O is between the Access Appliance and MSDP-C. A restore can be done either from the Access Appliance (Copy 2) or from public cloud (Copy 3). By default, copy 1 is used for restores unless specified to use different copies.

Figure 17 - Example of Sending Data to Cloud from VDD Pool or S3 Bucket via MSDP-C



Best Practices and Recommendations

Following best practices is important in creating an optimum deployment. This section covers some best practices relating to the Access Appliance as a long-term retention solution for NetBackup.

Data Layout on Access Appliance

The appliance contains hardware RAID 6 controllers and inherently does striping with dual parity across disks on the storage shelves for high performance and data durability. Selecting other layouts such as mirrored or erasure coding layout for data protection is not necessary. As a best practice, for long-term retention use case, it is recommended to configure the Access Appliance for NetBackup using the defaults:

- S3 bucket – cluster file system, simple or stripe layout depending on Access version, and block size of 8 KB
- Data Deduplication Pool – clustered file system, striped layout, and block size of 8 KB

For VDD, the layout is by default stripe for additional performance.   As previously discussed, the Access Appliance creates 5 logical volumes on each storage shelf with each volume containing 16 disks. The number of logical volumes that data deduplication storage pool requires is 5 logical volumes, allowing for the data to be striped across the 5 logical volumes. When growing the storage pool and using VDD, the volumes must be added in multiples of 5 (the entire shelf), and thus, it is advisable to plan or size the system appropriately.

Deduplication

If more than 250 TB of MSDP is required, use the NetBackup appliances as opposed to the BYOS version. NetBackup BYOS has a limitation of 250 TB for the size of MSDP, whereas the size of MSDP on an appliance can be up to 442 TB or 1 PB depending on the model. The maximum MSDP capacity depends on the NetBackup appliance used as a media server.

For MSDP-C, multiple buckets are supported per storage server. By default, the storage server defines cache properties in contentrouter.cfg file with a total of 1 TB. Hence, as a rule of thumb, when configuring the storage server, configure 1 TB of disk storage per cloud target. Also, it is best practice to assign a bucket per MSDP-C storage server.

As previously stated, using VDD requires 384 GB of memory per node and only one node runs the management processes at a time, in an active/passive configuration. If using the Access Appliance for more than the VDD workload, then it is recommended that one of the nodes be dedicated to VDD and the other node for the other workloads. The Access Appliance should be sized appropriately to handle multiple workloads and single-node failure scenarios. Also, it is the not recommended to modify the default setting of MaxCacheSize beyond the default of 50%. The MaxCacheSize in contentrouter.cfg file determines the number of fingerprint indices that can be cached in memory. Allocating more memory to the fingerprint cache can cause memory starvation issues. When specifying the size or growing of the VDD pool, it is also best practice to allocate in multiples of 100 TB. By default, file systems are resized in multiples of 100 TB if grown. For example, if one specifies the initial size of Access deduplication pool to be 80 TB and then increases it to 800 TB, the 80 TB will be resized to 100 TB and seven subsequent filesystems of 100 TB will be created to make 800 TB. **NOTE:** The size of the VDD pool can be increased, however, it cannot be decreased or shrunken.

Another consideration when utilizing VDD is not to send daily incremental backups but instead just duplicating weekly full and/or monthly backups to VDD. Expiration of numerous daily incremental backups from the Access Appliance can overload the garbage collection process and cause performance degradation.

Compression

For better storage utilization, using NetBackup compression might be an option when deduplication is not ideal, and the data type being backed up is compressible. Although compression can reduce the size of a backup, it can consume server resources. As a best practice, the media server should be sized appropriately for compression. For detailed information on NetBackup compression attributes and considerations, refer to the NetBackup Administration Guide, Volume I, and compression for cloud storage targets and deduplication, refer to NetBackup Cloud Administrators Guide and NetBackup Deduplication Guide respectively. **NOTE**: For deduplicated data, compression is enabled by default and different from the NetBackup compression policy attribute.

Encryption

NetBackup encryption can be enabled at the NetBackup attribute policy level or at the MSDP level via the pd.conf file. By default, encryption is off in both levels. If enabling encryption at the NetBackup attributes policy, data will be encrypted prior to the deduplication which would reduce the deduplication rate. However, enabling encryption at the MSDP level will not affect overall deduplication rates since the data is encrypted after deduplication. **NOTE**: NetBackup KMS is not supported with the Access Appliance.

Network Connectivity

The Access Appliance has two 10 GbE uplinks per node. Each physical port maps to a virtual IP. Thus, there are four virtual IP addresses. As a best practice present the fully qualified domain name mapping to the virtual IP so it will automatically transition to the other node if one node fails or the physical links on one node fails or is unreachable. Also utilizing the fully qualified domain name as oppose to the virtual IP is beneficial if the virtual IP changes for scenarios like migration. For instance, map one of the virtual IPs to the S3 object URL s3.<clustername> when using the S3 protocol. Also, if using VDD, specify the fully qualified domain name mapping to Access Appliance virtual IP during NetBackup configuration.

Bonding is an option on Access Appliance. Joining or bonding multiple network interfaces on the Access appliances into a single interface improves the bandwidth and network throughput through the combined single interface. Bonding is only configurable via the Access command-line interface. As a best practice, the switch that the uplinks of the Access Appliance are connected to is configured appropriately for the link aggregation.

Multiple NetBackup Domains

As previously mentioned, the Access Appliance can store data from multiple NetBackup domains. When using MSDP-C with Access, a single bucket can be used to store data for both domains. However, it is advisable when using MSDP-C with Access, to have each NetBackup domain with their own bucket to avoid name collisions in the different domains and for better categorization and identification. However, this best practice is not to be applied when using VDD in which a single data deduplication pool is configurable and global deduplication is observed between the domains.

As previously mentioned, when using AIR with VDD, the Access Appliance cannot be both the source and destination. A separate Access Appliance is needed in each domain.

Load Balancing

There are two nodes on the Access Appliance configured as active/active. As a best practice, balancing the load across nodes on Access is recommended. Load balancing can be achieved using any of the following techniques:

- **External load balancing** – using an external load balancer such as HAProxy or F5, allows for more algorithms to distribute load across nodes such as least connections or weights. It also frees the Access nodes from the proxy handling and balances the network traffic between the nodes.
- **Manual load balancing** – virtual IP addresses of the nodes can be manually assigned to applications in a distributed manner. The disadvantage of this approach is that even distribution might be difficult to gauge since applications are not all equal in sense of workload.
- **DNS load balancing** – the S3 object URL name for an Access S3 bucket, s3.<clustername> is created in DNS and includes all the virtual IP addresses of the nodes. DNS round-robins through the virtual IP addresses. The disadvantage of using DNS is in case of connectivity issues, the virtual IP is still in rotation until it is manually removed.

NOTE: Load balancing does not apply to VDD.

Monitoring

It is important to monitor or be aware of the alerts especially storage utilization warnings and hardware critical alerts. The AutoSupport features assists in this manner, but as a best practice, it is advisable to be pro-active instead of re-active. For instance, once the capacity reaches 60%, it might be a good time to revisit the storage utilization or plan for growth.

Sizing Guidance

In planning for data protection, two considerations come to mind: recovery point objectives (RPO) and recovery time objectives (RTO). From a backup and recovery standpoint, the RPO and RTO determine which policies are implemented, and therefore the resources required by a NetBackup deployment in terms of the necessary amount of systems, appliances, and storage. Other considerations include the number of users and applications, amount of data that is backed up, the frequency, and how long to keep the data. When planning for a long-term retention solution for backup images there are two factors:

- Capacity - how many backup images can be stored
- Performance – how much workload (backup streams and bandwidth) the storage platform can handle.

The Veritas account team will assist in the sizing of the appliance based on your requirements using these factors. Some parameters that might enter in the equation when estimating long-term storage requirements include:

1. Volume of source data.
2. Daily data change ratio
3. Annual storage growth
4. Data retention for daily incremental.
5. Retention for weekly, monthly and yearly full backups.
6. Estimated deduplication ratio for initial backup and daily incremental.
7. Estimated deduplication ratio for weekly, monthly, and yearly full backups.
8. Performance and/or service level requirements.

Samples of Capacity Sizing for Access Appliance

Depending on whether data has been deduplicated or not affects the actual capacity used. There is overhead incurred when using deduplication technologies such as the creation of an encapsulating directory structure, storing of the metadata, and/or hash table. This overhead needs to be accounted for when calculating the capacity available. **NOTE**: The sample calculations in this section only discusses the overhead relating to capacity, however, it doesn't consider the other parameters previously discussed such as annual growth rate, daily change rate, performance, etc. It is best to work with the Veritas Account teams who have the tools and expertise to determine the optimum configuration based on your requirements.

Sizing Deduplication with MSDP-C

NetBackup MSDP deduplication technology places the backup images into 64 MB containers. Using the default block size of the Access Appliance for the cluster file system is 8 KB. The overhead of Access, NetBackup header, and keys (if encryption is enabled) for unique data includes:

- Without encryption enabled – 0.18%
- With encryption enabled – 0.26%

Example without encryption for 800 TB of data with an 8:1 deduplication ratio is as follows:

- Logical Data Stored on Disk = 800 TB
- Unique Data Stored on disk 100 TB
- Access and Header Overhead = 0.18 % x Unique Data
  - $0.18\% \; x \; 100 \; TB \; = \; 0.18 \; TB$
- Total Volume Storage Requirements = Unique Data + Access and NetBackup Header Overhead
  - $100 \; TB \; + \; 0.18 \; TB \; = \; 100.18 \; TB$

Example with encryption for the same size:

- Access, Header and Encryption Overhead
  - $0.26\% \; x \; 100 \; TB \; = \; 0.26 \; TB$
- Total Volume Storage Requirements = Unique Data + Access and NetBackup Header and Encryption Overhead
  - $100 \; TB \; + \; 0.6 \; TB \; = \; 100.26 \; TB$

NOTE: The above calculation does not include the MSDP meta-data on Access Appliance.

Sizing Deduplication with VDD

For VDD, the fingerprint database, the metadata, unique blocks, journals and logs are stored on Access. If defining an 800 TB data deduplication pool, eight 100 TB file systems is created. Each file system created incurs a 0.1% file system overhead, thus, with eight file systems, the maximum would be 0.8%. In addition, the database, the metadata, logs and journals consume up to 4% of the storage. Example for an 800 TB of data to backup with 8:1 deduplication ratio is as follows:

- Fingerprint database, metadata, journal, and logs overhead: 4%
- File system overhead: 0.8 % for 800 TB file system
- Logical Data Stored on Disk = 800TB
- Unique Data Stored on disk is 100TB
- Fingerprint database, metadata, journal, logs overhead = 4% x Logical Data Stored on Disk
  - $4\% \ x\ 800\ TB = 32\ TB$
- File system overhead for one file system = 0.8% x Unique Data
  - $0.8\% \ x\ 100TB \ = \ 0.8\ TB$
- Total Volume Storage Requirements = Unique Data + Fingerprint database, metadata, journal, logs overhead + File system overhead
  - $100\ TB \ + \ 32\ TB \ + \ 0.8\ TB \ = \ 132.8\ TB$

NOTE: The above calculation does not include the 5 TB filesystem created for the MSDP catalog on Access Appliance.

Sizing for Traditional Duplication of Data

For traditional duplication, as previously explained, the S3 OST cloud plugin shards the backup image into 16 MB fixed object and sends it to Access. There is one header information per 50 GB of backup image and keys if encryption is enabled. The overhead values are miniscule and thus, the size of backup image for traditional duplication is about the same size of backup image being stored. Take for example a 150 GB, Access and NetBackup header and keys (if encryption is enabled) overhead are:

- Without encryption enabled – 0.000082%
- With encryption enabled – 0.00014%

Example without encryption for 150 GB backup image size:

- Data Stored on disk 150 GB
- Access and Header Overhead = 0.000082% x Data
  - $0.000082\% \ x\ 150\ GB \ = \ 0.000012\ GB$
- Total Volume Storage Requirements = Data + Access and NetBackup Header Overhead

  - $150\ GB\ +\ 0.000012\ GB\ =\ 150\ GB$

Example with encryption for the same size:

- Access, Header and Encryption Overhead = 0.00014%
  - $0.00014\%\ x\ 150\ GB\ =\ 0.00021\ GB$
- Total Volume Storage Requirements = Data + Access and NetBackup Header and Encryption Overhead
  - $150\ GB + 0.00021\ GB\ =\ 150\ GB$

Conclusion

The addition of the Access Appliance in the Veritas appliance portfolio provides a competitive disk-based solution for long-term retention of data. Integrated with NetBackup, the Access Appliance becomes a more compelling option in data protection, and disaster planning and recovery. Implementing the Access Appliance with NetBackup as a long-term retention solution simplifies management and support, minimizes costs, and improves control and visibility.

References

- NetBackup
  - Product Documentation
    - https://www.veritas.com/support/en_US/article.DOC5332
  - NetBackup Deduplication Guide
    - https://www.veritas.com/content/support/en_US/doc/25074086-131900563-0/index
  - NetBackup Cloud Administrator's Guide
    - https://www.veritas.com/content/support/en_US/doc/58500769-145530841-0/v121751550-145530841
  - Disaster Recovery
    - Veritas NetBackup in Highly Available Environments Administrator's Guide
      - https://www.veritas.com/content/support/en_US/doc/39129704-144650364-0/v38488996-144650364
    - Auto Image Replication (AIR): How To move a copy of an image back to the source master server
      - https://www.veritas.com/support/en_US/article.TECH205923
- Access Appliance
  - Product Documentation
    - https://sort.veritas.com/documents/doc_details/AAPP/7.4.2/Appliance%203340/ProductGuides/
    - https://sort.veritas.com/documents/doc_details/AAPP/7.4.3/Veritas%203340/Documentation/
  - Veritas Access Solutions Guide for NetBackup - Sample configuration of Access Appliance with NetBackup
    - https://www.veritas.com/support/en_US/doc/125197328-133637736-0/index

Table of Figures

ABOUT VERITAS

2625 Augustine Drive, Santa Clara, CA 95054
+1 (866) 837 4827
veritas.com

For specific country offices
and contact numbers,
please visit our website.