



AWS Cloud Storage with Veritas NetBackup™

Long-Term Retention Solution

This whitepaper provides a technical overview of AWS cloud storage as a long-term retention storage solution with NetBackup. It highlights the overall solution architecture components, integration flow, and best practices.

VERITAS™

The truth in information.

TABLE OF CONTENTS

INTRODUCTION	5
EXECUTIVE SUMMARY	5
SCOPE	5
TARGET AUDIENCE	5
SOLUTION VALUE	5
SOLUTION KEY FEATURES	6
INSIGHT	6
STORAGE EFFICIENCIES	6
SECURITY	6
PROTECTION	6
MIGRATION OPTIONS	7
SOLUTION ARCHITECTURE OVERVIEW	7
SOLUTION COMPONENTS	8
INFORMATION STUDIO	8
NETBACKUP	11
Deduplication	13
Traditional Duplication (Without Deduplication)	14
AWS	14
S3 Storage Classes	15
Network Connectivity	16
AWS Snowball and AWS Snowball Edge	17
SOLUTION INTEGRATION FLOW	18
IDENTIFICATION OF DATA TO SEND TO CLOUD	18
OPTIMIZED DUPLICATION (DEDUPLICATION) DATA FLOW TO AMAZON S3 STORAGE CLASSES	20
ACCESS APPLIANCE OPTIMIZED DUPLICATION (DEDUPLICATION) DATA FLOW TO AWS CLOUD STORAGE	22
TRADITIONAL DUPLICATION DATA FLOW	23
MIGRATION TO CLOUD	26
DISASTER RECOVERY IN THE CLOUD	27
BEST PRACTICES AND RECOMMENDATIONS	28
PRIVACY LAWS	28
COMPRESSION	28
DEDUPLICATION	29
SNOWBALL/SNOWBALL EDGE	29
NETBACKUP RETRIEVAL ATTRIBUTES	29
CONCLUSION	29
REFERENCES	30
APPENDIX	31
AWS	31
CREATION OF NETBACKUP CLOUD STORAGE – S3 GLACIER STORAGE CLASS	34
DEFINITION OF AN SLP	38

CREATION AND MODIFICATION OF THE BACKUP POLICY TO USE THE SLP 41

VALIDATION 43

VERITAS INFORMATION STUDIO 46

GENERATE CSV REPORT 46

EXTRACT THE FILE PATHS FROM REPORT 49

ENTER THE CLIENT INFORMATION AND FILEPATHS INTO NETBACKUP 50

Revision History

Version	Date	Changes
1.00	11/1/2019	Initial Version

INTRODUCTION

EXECUTIVE SUMMARY

More and more companies are venturing to the public cloud as an option to retain their data for long-term and/or to safeguard their data from on-premises failures, attacks and disasters. Cost, security, and insight into what data can be sent to the cloud has been some of the main requirements. Veritas suite of products alleviates some of these issues and concerns. Veritas NetBackup, for instance, has encryption features for data at rest and in motion for security. It also has storage efficiency technologies such as deduplication and compression to reduce egress and ingress costs to and from the cloud. The utilization of Veritas Information Studio provides valuable insights to identify which data should remain on-premises or which is ideal to be moved to the cloud. NetBackup supports sending and retrieving data to and from the different storage classes offered by AWS with varying costs, availability and performance for long-term retention and preservation of organizations critical digital assets. In addition, NetBackup supports AWS Snowball and Snowball Edge for large-scale data migration or initial seeding of data to public cloud. Veritas products and AWS cloud storage and services work together to optimize data protection, reduce cost and minimize risks and liability.

SCOPE

The purpose of this document is to provide technical details to assist in understanding AWS cloud storage with NetBackup as a solution for long-term retention of backup data. It describes the components of this solution, its value, and some best practices. It is advised to refer to Veritas product documentation or AWS 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 AWS cloud storage and services with NetBackup solution for long-term retention. It provides a technical overview of this solution and highlights some best practices.

SOLUTION VALUE

Veritas portfolio of data management and protection products with AWS cloud storage and services provides some advantages which include:

- **Valuable insights** –Organizations tend to blindly backup data or not remove stale or orphaned data from their primary storage. Veritas Information Studio provide a view on digital assets residing on primary storage and backup storage for assistance in data placement and lifecycle. With new regulatory requirements imposed by numerous countries, it is crucial to get an understanding of data, it's value and liability.
- **Reduced risk** – NetBackup allows for encryption of data prior to being transmitted to AWS cloud storage. Data maintains its encrypted form in AWS cloud storage for data at rest. For data in flight, NetBackup uses SSL (Secure Sockets Layer) protocol for data transfers between NetBackup and cloud storage for enhanced security.
- **Minimize cost** – with NetBackup deduplication and compression features, the amount of data sent or retrieved from cloud is reduced and thus minimizes overall costs. Also, support of sending data to Amazon Simple Storage Service (Amazon S3) including Amazon S3 Glacier and Amazon S3 Glacier Deep Archive storage classes results in more cost savings.

SOLUTION KEY FEATURES

Certain key features that companies look for in an off-premise long-term retention solution product include insights, security, storage efficiency and migration path. AWS cloud storage services with NetBackup provides these features to assist customers in preserving their most valued data.

INSIGHT

One of the challenges often faced by organization is deciding which data can be sent to the public cloud. [Veritas Information Studio](#) has the capability to classify primary sources and/or scan the NetBackup catalog to quickly acquire information about the data on the sources that have been backed up. The ability to identify areas of risk, value, and ROT (Redundant, Obsolete, Trivial) improves operational efficiency, reduces storage cost and minimizes risk and liability.

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 on the AWS cloud storage can be deduplicated using NetBackup Media Server Deduplication Pool (MSDP) technology. Data is sent to Amazon S3 storage classes via CloudCatalyst.

NetBackup also supports compression prior to sending data to the AWS cloud storage. 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. Detailed information on NetBackup compression attributes can be found in the [NetBackup Cloud Administrators Guide](#) and [NetBackup Deduplication Guide](#).

SECURITY

NetBackup has security features that protect all NetBackup components and operations at different security implementation levels such as datacenter and enterprise. Refer to the [NetBackup Security and Encryption Guide](#) for further details on NetBackup's security implementations and levels.

For enhanced security, NetBackup also offers encryption of data. Any encryption done by NetBackup is maintained on the cloud storage target. For more information on how NetBackup conducts encryption specifically for cloud storage servers, refer to [NetBackup Cloud Administrators Guide](#). 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](#). Additional security that is employed for this solution is the requirement to use access keys and credentials when configuring AWS as a cloud storage destination. Data transfers to and from AWS can also be secured by enabling the NetBackup SSL feature.

PROTECTION

In NetBackup 8.2 release as part of the automated disaster recovery in cloud, there is a new feature that extends the capabilities of CloudCatalyst called Image Sharing. With Image Sharing, NetBackup will not only send deduplicated data to a cloud storage target but the image meta-data as well. This new feature allows for recovery of images in the cloud when the on-premises NetBackup catalog is not available due to corruption, power outage, network issues, etc. Data and meta-data can be re-constructed in the cloud or different datacenter by instantiating another CloudCatalyst and attaching it to a new master domain.

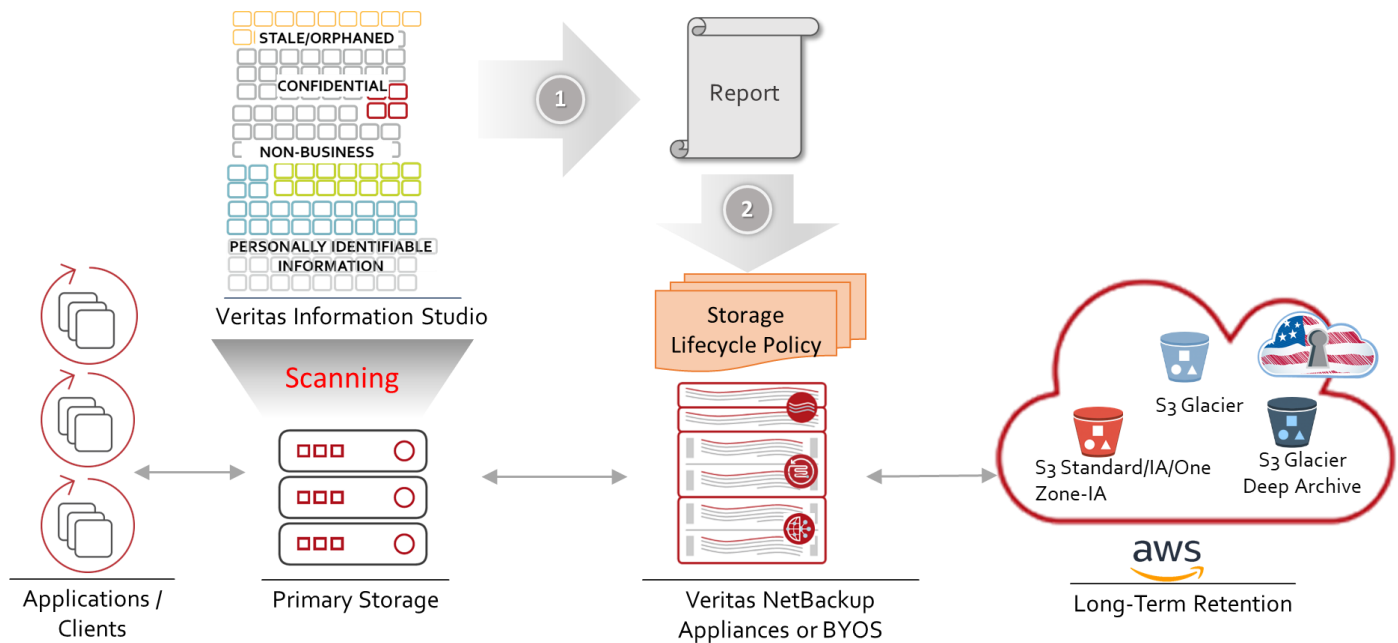
MIGRATION OPTIONS

For customers who are starting to migrate to the cloud and would need to move their “initial data set” or replace their tapes with AWS cloud storage, NetBackup supports AWS Snowball and Snowball Edge. These devices with NetBackup provide an easy path to the cloud when dealing with large datasets.

SOLUTION ARCHITECTURE OVERVIEW

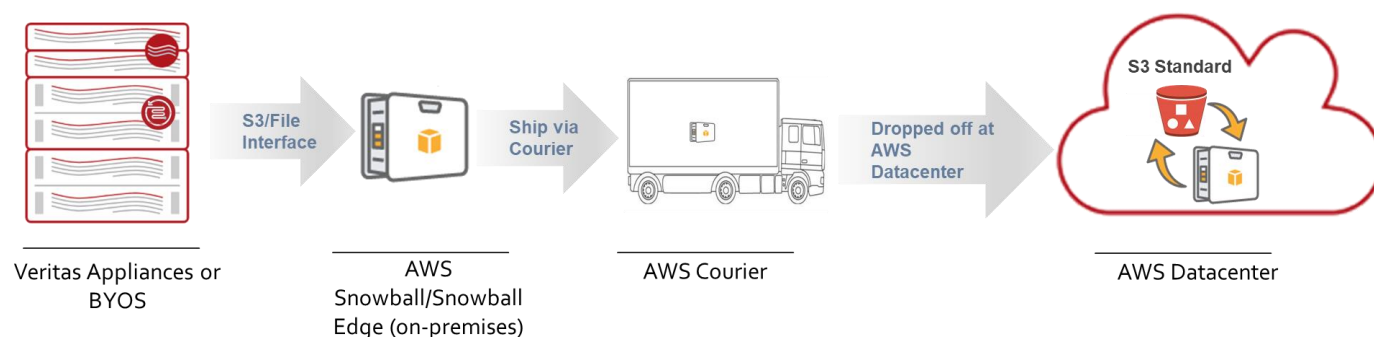
Figure 1 depicts a high-level overview of the Veritas data protection solution with AWS cloud storage for long-term retention. An integral part of this solution is the use of Veritas Information Studio to provide information on the data residing on-premises to help make decisions on the lifecycle of the data. A report generated by Information Studio assists users in making informed decisions to determine which data should be stored on-premises or sent to cloud and/or define the storage lifecycle policies of the data. For instance, the lifecycle of backup data can first reside on-premises like a NetBackup Appliances or BYOS for short-term, then, to Access Appliances for mid-term retention and then to AWS cloud for long-term retention and disaster recovery. NetBackup supports several Amazon S3 storage classes such as S3 Standard, S3 Standard Infrequent Access, S3 Standard One Zone Infrequent Access, S3 Glacier, S3 Glacier Vault and S3 Glacier Deep Archive. These classes of storage differ in sense of cost, usage, restore time, availability and other services.

Figure 1 - Solution Overview of NetBackup with AWS Cloud Storage Classes



For migration of data from tape or initial seeding of cloud with backup data that are on premises, AWS offers shippable physical storage devices, Snowball and Snowball Edge. NetBackup supports these devices to provide customers a migration path to AWS cloud storage. In general, as shown in Figure 2, AWS Snowball and AWS Snowball Edge devices are placed on-premises and communicate with NetBackup either via Amazon S3 or File interface (NFS) protocols depending on which Snowball device is used. Once data is transferred to the AWS storage device, it is then shipped to AWS datacenter via courier. After its arrival, AWS migrates the data to the destined S3 Standard storage class specified by NetBackup during cloud storage server configuration.

Figure 2 - Migration with Veritas NetBackup and AWS Snowball and AWS Snowball Edge



SOLUTION COMPONENTS

In order to get a better understanding of how AWS cloud storage services work with NetBackup for long-term retention, all involved solution components are explained in further detail in the following sections. Main components that are discussed include Information Studio, NetBackup, Amazon S3 storage classes, AWS Snowball and AWS Snowball Edge.

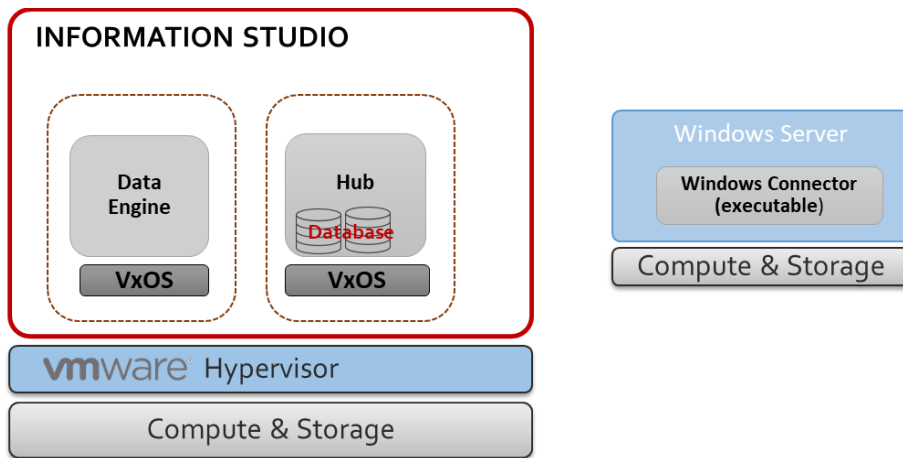
INFORMATION STUDIO

Veritas Information Studio provides visibility by connecting to varied primary data sources on-premises which include: OpenText™, OpenText LiveLink®/Documentum, IBM FileNet, Microsoft® Exchange/SQL Server/SharePoint®, Oracle® database, and SMB shares such as NetApp, Dell EMC Celerra/VNX/Isilon®, Hitachi, and Windows®. It can also scan NetBackup catalogs to harvest information on data that has been backed up. In addition to primary sources, Veritas Information Studio can connect to cloud sources such as Microsoft® OneDrive, SharePoint Online, AWS, Google Cloud platform and Azure. Depending on the data source, it classifies the data into certain categories such as ownership, age, size, activity, stale, non-business, user risk, type, and data patterns. This helps enable administrators to identify data that can be archived or tiered to cheaper storage, enforce security, and perform information lifecycle management and risk analysis. Architecturally, Information Studio consists of three main components:

- **Hub** – responsible for management of jobs, maintenance of configuration information, storage of metadata, logging, and generation of reports.
- **Data engine** – topology discovery of each data sources, scanning and collection of metadata, classification and sending data to the hub.
- **Windows connector** – only required when connecting and scanning of Windows based applications and data sources such as Microsoft SQL, Microsoft SharePoint, SMB/CIFS shares, and Oracle.

Illustrated in Figure 3, the hub and data engine of Information Studio are containerized services running on top of the Veritas Operating System (VxOS), a customized Linux operating system based on Red Hat® Enterprise Linux (RHEL). The containers and VxOS are packaged as an OVA (Open Virtual Appliance) file and deployable on VMware® hypervisor. The windows connector is a separate executable that is run on a Windows host on bare metal or virtual machines and is only required if connecting to a Windows-based data source. Refer to [Veritas Information Studio whitepaper](#) for more details on the architecture.

Figure 3 - Information Studio Components



Depending on the amount of data to be classified, a single or multiple instance of the data engine component can be deployed. For example, if the data source is NetBackup, then the hub container is only required since there is a small-scale data engine running within the hub by default. If the data source is in some remote location or on several SMB shares, separate or multiple remote data engine instances would be deployed. Usually, the data engine is placed in the same geographic region as the data sources being scanned and classified. Minimum virtual machine requirements to run OVA and executable is shown in Table 1.

Table 1 - Minimum Specifications for Veritas Information Studio Components

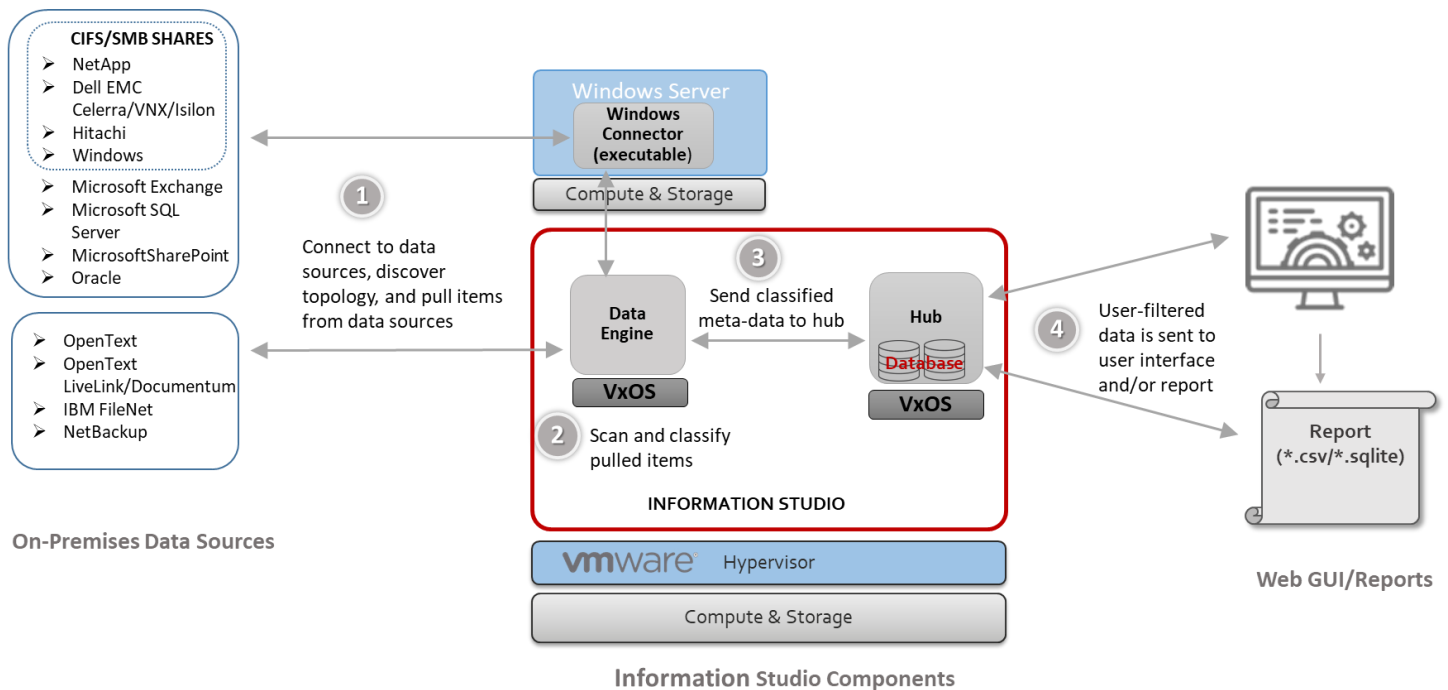
Veritas Information Studio Components	Minimum Specifications
Hub	16 cores 80 GB of RAM Disks: <ul style="list-style-type: none"> • 500 GB for Veritas Operating System • 1 TB for data ESX Server minimum version 6
Remote Data Engine	16 cores 32 GB of Memory Disks: <ul style="list-style-type: none"> • 500 GB for Veritas Operating System • 500 GB disk for data ESX Server minimum version 6
Windows Connector	4 cores 8 GB Memory Disk: 100 GB free disk space Windows 2012 R2 or Windows 2016 (64-bit versions)

In Figure 4, Veritas Information Studio involves the following when processing data sources for classification:

- 1) The Data Engine or Windows connector connects to the data sources and conducts a topology discovery. The data engine then pulls the items from the data sources. As previously mentioned, if the data source is Windows based, then the Windows connector is part of the communication and data path.
- 2) The data engine scans to capture the meta-data and then classifies the data.
- 3) The meta-data captured during scan and classification is sent to the hub for storing and further querying.
- 4) Users can query and filter the data using the APIs or web GUI based on the custom or preconfigured criteria. The results of this filtration are presented within the web graphical user interface and/or a comma-separated variable (CSV) or SQLite file report. The report contains a list of files or items and their associated meta-data that meet the user-specified filters.

Classification can be conducted manually or scheduled in regular intervals (daily, weekly, etc.) specified by user. For details on installation, configuration and deployment, refer to [Veritas Information Studio](#) product documentation.

Figure 4 - Veritas Information Studio Process Flow



Veritas Information Studio is a powerful tool. In addition to custom policy definition where users can find data items based on specific pattern or conditions, it has over 120 preconfigured data classification policies and 700 data patterns to identify common data privacy and regulatory compliance principles. Some examples of preconfigured policies include:

- **Corporate Compliance:** authentication, “company confidential” and IP, ethics and code of conduct, IP addresses, PCI-DSS and proposals/bids.
- **Financial Regulations:** bank account numbers, credit card numbers, GLBA, SOX, SWIFT Codes and U.S. financial forms.
- **Health Regulations:** Australia Individual and Canada Healthcare Identifier, IDC 10 CM diagnosis indexes, medical record numbers, U.S. DEA numbers and U.S. HIPAA (Health Insurance Portability Accountability Act of 1996).

- **International Regulations:** Australia/Canada/U.K./U.S. driver's license and passport numbers, Australia tax, U.K. Unique Tax Reference (UTR), U.S. Social Security number and Taxpayer ID, Canada SIN, France National ID, Italy Codice Fiscale and Switzerland National ID.
- **Personal Identifiable Information (PII)** from over 35 countries.
- **Sensitive data policies** from over 35 countries.
- **U.S. State Regulations:** Criminal history, FCRA, FERPA, FFIECE, FISMA, IRS 1075 or SE.)
- **U.S. Federal Regulations:** California Assembly Bill 1298 (HIPAA), California Financial Information Privacy Act (SB1) and Massachusetts regulation 201 CMR 17.00 (MA 201 CMR 17), and newly evolving policies like the California Consumer Privacy Act (CCPA).

An example of the Veritas Information Studio web graphical user interface (GUI) is shown in Figure 5. The information displays information such as amount of stale data, where your data resides, item extensions, and age. The reports generated by Veritas Information Studio can be manually inspected and be used to define the "Backup Selection List" within NetBackup backup policies. Storage lifecycle policies are used to specify the data that can be kept on-premises and/or duplicated to the AWS cloud. Knowledge of what type of information is stored in data sources allows organizations to make more informed decisions on what to do with the data for storage optimization, security, compliance, archival and long-term retention.

Figure 5 - Example View of Information Studio Web Graphical User Interface.



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 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

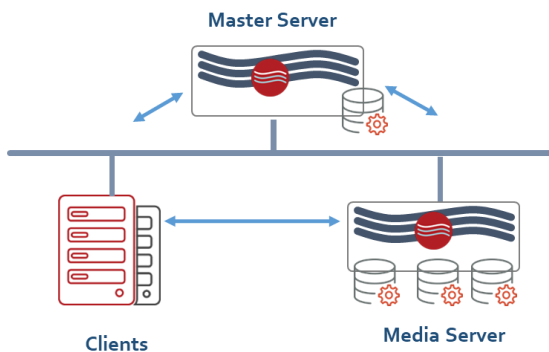
to tape, SAN, NAS, public or private cloud. Schedules, retention periods, and the ability to tier to different types of storage are defined in policies or storage lifecycle policies (SLP).

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 domain is independent but can be centrally managed by NetBackup OpsCenter, a web-based console for managing, monitoring, and reporting on NetBackup operations. Figure 6 illustrates a sample configuration of a set of clients with a dedicated master server and one media server.

Figure 6 – 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 as on-premises and cloud-based virtual machines or in “containers” when using the Flex 5340 Appliance. Highlights of each of these deployment options are as follows:

- **NetBackup Appliance** - purpose-built, highly tuned, scalable, and resilient integrated appliance for NetBackup components. These appliances address the most demanding backup and recovery requirements of enterprises.
 - [NetBackup 5240](#) - meant for moderate workloads and can scale up to 323 TB storage capacity.
 - [NetBackup 5330 Datasheet/NetBackup 5340 Datasheet](#) - the 5330 and 5340 appliances are for demanding workloads and requiring higher usable capacity that can scale up from 1506 TB and 2160TB respectively. The 5330 and 5340 models have high availability configurations that include an additional node to continue operations should the active node fail.

- [NetBackup Flex 5340 Datasheet](#) - The Flex 5340 Appliance supports container technology allowing for multiple containers with different roles of master, media, or CloudCatalyst 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.
- **NetBackup Build Your Own Server (BYOS)** - NetBackup components can be deployed on commodity servers and can run on a Linux or Windows platform. Refer to the [NetBackup Software 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 250 TB per server for systems configured with Red Hat Enterprise Linux (RHEL), Windows Server 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 Software Compatibility List](#) contains more information on the supported hypervisors.
- **NetBackup in Cloud Marketplace** – NetBackup is available for automated deployment in AWS marketplace. For more information, check the “NetBackup in the Cloud – Deployment Templates” section of the [NetBackup Software Compatibility List](#).

The [Access Appliance](#) is part of the Veritas appliance portfolio; however, it mainly acts as an on-premise, mid-term or long-term retention storage target for NetBackup. Hence, for those seeking to extend their on-premises disk-based storage platform for faster recovery times, control and/or simplicity, the Access Appliance is a turn-key storage solution designed for high capacity and cost optimization. The Access Appliance model 3340 is comprised of two clustered nodes and one primary storage shelf and up to three additional expansion storage shelves. It can scale up to 2,800 TB of usable space.

There are two ways to store backup images in public cloud from NetBackup which include:

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

In the next section, these two approaches are described further.

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. Thus, if the data is first placed in an MSDP and then duplicated to another storage platform that does not support MSDP technology, NetBackup would rehydrate deduplicated data prior to

sending data to the storage platform. Rehydration involves putting the backup image back to a non-deduplicated form. 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 target, 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** (spool) - 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. Both compression and encryption (if enabled) are performed after the fingerprint is calculated and prior to sending data to the target storage. For more information on the architecture of NetBackup MSDP deduplication technology, refer to the [NetBackup Deduplication Guide](#).

CloudCatalyst is a crucial component in sending deduplicated data to the AWS cloud without rehydration. For best performance, CloudCatalyst is deployed on a dedicated host, an appliance, as a container in Flex Appliance or BYOS (Build Your Own Server). It can also be deployed as virtual machine running on-premises or cloud-based hypervisors such as [AWS](#). As for the [CloudCatalyst appliance](#), it is essentially the NetBackup 5240 appliance with the G option configured to have more memory (192 GB). There is a separate SKU for the CloudCatalyst appliance when ordering. Multiple CloudCatalyst instances can be implemented, however, each CloudCatalyst instance can only write to one on-premise or public cloud storage platform and only to one bucket (1 PB maximum qualified). If BYOS deployment of CloudCatalyst requires RHEL 7.3 or higher and the minimum amount of memory and disk cache for BYOS CloudCatalyst is configurable and dependent on amount of data being backed up.

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. For these types of data or backup, images are best sent to the public cloud without deduplication. Data is sent to the AWS cloud from NetBackup using the S3 protocol and stored in an Amazon S3 storage classes. Furthermore, NetBackup traditional duplication to AWS cloud supports S3 lifecycle policies such that data can transition from Amazon S3 Standard to Standard Infrequent Access, and/or S3 Glacier.

AWS

AWS has a wide range of cloud storage services for backup, archival, and long-term retention use cases. NetBackup supports several of the Amazon S3 storage classes as indicated by the [NetBackup Hardware and Cloud Storage Compatibility List](#). There are also different connectivity options provided by AWS based on desired transmission performance and cost. For movement of large-

scale data to cloud, AWS offers, AWS Snowball and AWS Snowball Edge, on-premise compute and storage devices which include physical and logical security to securely send large amounts of data to the cloud.

A NetBackup Open Storage Technology (OST) plugin is necessary to send or retrieve data to and from AWS cloud storage target. OST is an Application Programmable Interface (API) developed by Veritas in order that plugins can be developed or created to connect and manage third party vendors storage platform from within NetBackup. OST plugins were developed to interact with S3 compatible cloud providers. The OST cloud plugins are by default installed on NetBackup media, master servers and CloudCatalyst to send data to the AWS cloud storage classes.

S3 Storage Classes

NetBackup can backup and retrieve data from Amazon S3 storage classes and include the following:

- **Amazon S3 Standard (S3 Standard)** – used for frequently accessed data
- **Amazon S3 Standard Infrequent Access (S3 Standard-IA)** – used for data that is infrequently accessed.
- **Amazon S3 Standard One Zone Infrequent Access (S3 One Zone-IA)** – used for data that is infrequently accessed that does not require the availability and resilience of S3 Standard or S3 Standard-IA. Data is stored and available only in one zone.
- **Amazon S3 Glacier (S3 Glacier)**– mainly used for archival or long-term retention of data. For data archival or long-term retention with specific lock policies such as WORM (Write Once Read Many) or legal holds, S3 Glacier Vault is available.
- **Amazon S3 Glacier Deep Archive (S3 Glacier Deep Archive)** - the cheapest of the Amazon S3 storage classes and priced to compete with tape or cheap and deep on-premises storage platforms. Best for long-term retention of data.

The Amazon S3 storage classes above are available with NetBackup in several regions. For up to date region availability, please refer to the [NetBackup Hardware and Cloud Storage Compatibility List](#). Here are some of the regions that are available: **Asia Pacific:** Mumbai, Seoul, Singapore, Sydney, Tokyo, Bahrain; **Canada:** Central; **China:** Beijing, Ningxia, Hong Kong; **Europe:** Frankfurt, Ireland, London, Paris; **South America:** Sao Paulo; **US East:** N. Virginia, Ohio; and, **US West:** N. California, Oregon. For government agencies who require more stringent security and compliance features, NetBackup also supports the above storage classes within the GovCloud for region US West 1 (Gov FIPS 140-2, Gov Non-FIPS). When utilizing these storage classes in GovCloud, SSL is required, and encryption is offered.

Deciding on which AWS cloud storage classes to send backup images are usually based on several factors which include restore time, cost, and NetBackup feature support. Table 2 provides a comparison of each of the AWS cloud storage based on these factors. Refer to AWS website for more information on [Amazon S3 storage classes](#).

Table 2 - Comparison of Amazon S3 Storage Classes

Amazon S3 Storage Classes	Restore Time with NetBackup	Cost**	IPv6 with NetBackup	CloudCatalyst	Accelerator	Minimum Storage Duration
S3 Standard	milliseconds	\$\$\$\$	Supported	Supported	Supported	Not Applicable

S3 Standard Infrequent Access	milliseconds	\$\$\$	Supported	Supported	Supported	30 days
S3 Standard One Zone Infrequent Access	milliseconds	\$\$\$	Supported	Supported	Supported	30 days
S3 Glacier	With CloudCatalyst: <ul style="list-style-type: none"> Expedited: 1-5 minutes Standard: 3-5 hours Bulk: 5-12 hours Without CloudCatalyst: <ul style="list-style-type: none"> Standard: 3-5 hours 	\$\$	Not Supported	Supported	Not Supported	90 days
S3 Glacier Vault	Standard: 3-5 hours	\$\$ (about 2% > Glacier)	Not Supported	Not Supported	Not Supported	90 days
S3 Glacier Deep Archive	With CloudCatalyst: <ul style="list-style-type: none"> Standard: within 12 hours Bulk: within 48 hours Without CloudCatalyst: <ul style="list-style-type: none"> Standard: within 12 hours 	\$	Not Supported	Supported	Not Supported	180 days

** The "\$\$" represents a simple cost view. For exact pricing, please refer to the [AWS Pricing](#).

Network Connectivity

The speed of backups and restores are highly dependent on the type of network connectivity. AWS offers three different types of connectivity to their AWS regions from on-premises datacenter which include:

- **Internet** – Basic internet connectivity and speeds offered by internet service providers.
- **[AWS Direct Connect](#)** – private and dedicated connectivity between on-premises environment to an AWS region. AWS Direct Connect addresses some of the standard internet challenges such as speed (bandwidth and throughput), network congestion and or contention. It supports two bandwidth levels of 1 GbE and 10 GbE fiber-optic connection.
- **[AWS Virtual Private Network \(VPN\)](#)** – secure and private tunnel from on-premises network to the AWS global network. It consists of the following two services:
 - **[AWS Client VPN](#)** – allows secure access to AWS resources via on-premise network. An endpoint is configured to set a secure Transport Layer Security (TLS) VPN session.
 - **AWS Site-to-Site VPN** – enables secure connection to on-premise network to AWS Virtual Private Cloud (VPC).

AWS Snowball and AWS Snowball Edge

If large amounts of data are required to be sent to the Amazon S3 storage classes regularly, to initially seed data and/or for data migration, AWS developed physical storage devices, Snowball and Snowball Edge. These devices are deployed on-premises and are required to be in the same region as the destination bucket in the AWS cloud. NetBackup connects to these devices and configured as a cloud storage server. Regular backup data (live) or data residing on tapes (old) or secondary storage can be duplicated to these devices using NetBackup storage lifecycle policies. The data are transferred to these devices on-premises using Snowball and Snowball Edge tools. Once the transfer is complete, devices are shipped by courier to the AWS data center where it will be uploaded to the destined Amazon S3 bucket. If duplicating live regular backup data, then the policies are suspended during the physical transport and resumed once data is available in the cloud or another AWS storage device is available on-premises. **NOTE:** *NetBackup currently only supports sending data to the Amazon S3 Standard storage class with these devices.* Information on how to configure NetBackup with Snowball and Snowball Edge devices, refer to [NetBackup Cloud Administrator's Guide](#) and AWS website. Highlights of the differences between these devices are shown in Table 3. Refer to the AWS website for more information or the latest on the specifications of [Snowball](#) and [Snowball Edge](#).

Table 3 - Main Differences between Snowball and Snowball Edge

Features	AWS Snowball	AWS Snowball Edge
Raw Capacity	50 TB (US Region Only) 80 TB	100 TB – Storage optimized 42 TB plus 7.68 TB dedicated NVMe SSD – Compute optimized
Usable Capacity	42 TB (US Region Only) 72 TB	80TB – Storage optimized 39.5 TB– Compute optimized NOTE: <i>NetBackup does not support Clustered Snowball Edge.</i>
Network Connector	RJ45, SFP+, SFP+ (with optic connector)	RJ45, SFP+, SFP+ (with optic connector), QSFP
Tools	Snowball client – transfers data and encrypts to or from Snowball. Needs to be downloaded from Amazon resources site and installed on a separate server. Amazon S3 Adapter - transfers data to and from Snowball using Amazon S3 Rest API. Needs to be downloaded and installed on a separate server. Minimum server specifications dedicated to run tools: 16 core CPU and 16 GB of Memory. NOTE: <i>If running</i>	Snowball Client – used to unlock the Snowball Edge. Needs to be downloaded from Amazon resources site and installed on separate server. Amazon S3 Adapter – transfers and encrypts data to and from Snowball Edge using Amazon S3 REST API. Installed on device by default. File Interface (NFS v3/v4/v4.1) – allows transfers and encryption of data into a bucket on Snowball using an NFS mount point. Installed on device by default. NOTE: <i>Not supported with CloudCatalyst.</i>

	<i>multiple instances of S3 Adapter or client, 7 GB of memory is required.</i>	
--	--	--

SOLUTION INTEGRATION FLOW

All data goes through a lifecycle from being created, read, modified, moved to other tiers of storage and eventually expired or deleted once it is no longer of use. When the data is actively used, it resides in primary storage and for data protection backed up to secondary storage. As data or backup data becomes infrequently accessed, it gets moved to cheaper storage on-premises and/or moved off-premises. Understanding the data in terms of usage, age, type, contains personal identifiable data, non-business, subject to regulatory compliance, etc. is crucial in determining where that data can move across storage platforms or different tiers of storage on-premises or off-premises.

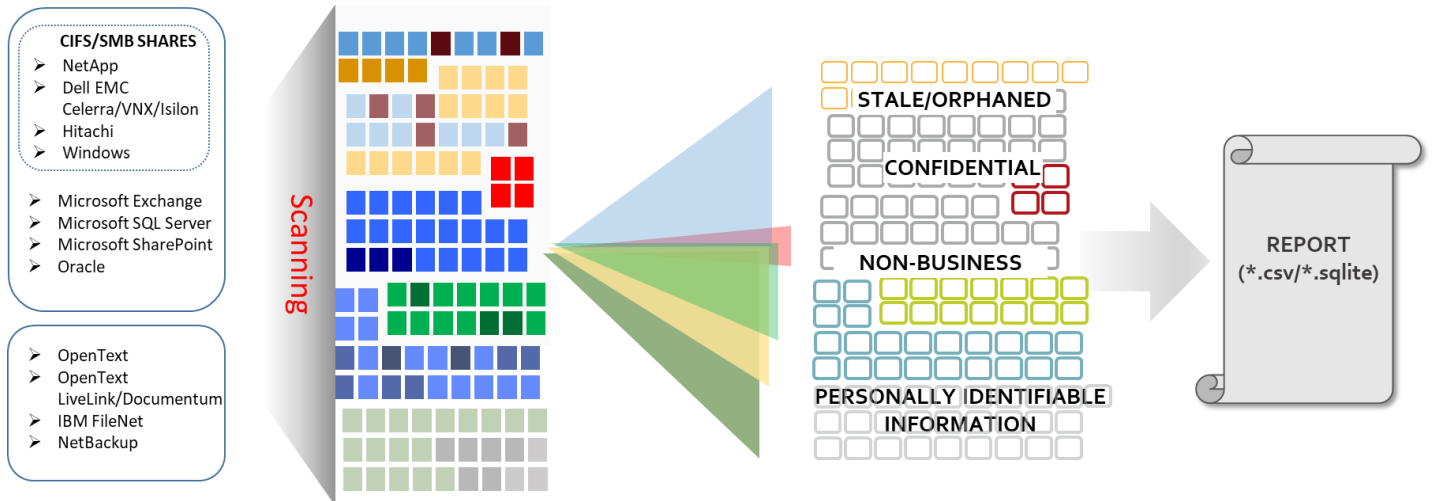
NetBackup manages the lifecycle of the data using storage lifecycle policies (SLP). NetBackup backup policies and/or SLP define the path or flow of data. Backup policies can be defined to send data either to a single target or to an SLP. An SLP defines the lifecycle objectives of the data from backup to duplication to varying storage types and/or replication to different domains. As previously mentioned, data is sent to an Amazon S3 storage classes utilizing the S3 protocol with the S3 OST cloud plugin installed on CloudCatalyst or media server. The solution integration flows described in this section includes:

- Identification of data to send to the public cloud using Veritas Information Studio
- Optimized duplication (deduplication) to cloud using CloudCatalyst
- Optimized duplication (deduplication) to cloud from Access Appliance
- Traditional duplication to cloud
- Migration to cloud

IDENTIFICATION OF DATA TO SEND TO CLOUD

Identification of what data to send to the cloud can be done via visual inspection or if applicable, Veritas Information Studio can be utilized to do a more granular and user-defined filtered search to identify data that can be sent to the cloud. As shown in Figure 7, items within data sources such as SMB shares, Microsoft SharePoint, etc. are scanned and classified based on certain pre-configured filters or user-defined criteria such as personal identifiable information, non-business, and activity. A report is generated either via the Veritas Information Studio web GUI or through APIs. **NOTE:** *If the data source is NetBackup catalog, no classification is done by Veritas Information Studio however, certain attributes is harvested from the catalog such as age, and file type.*

Figure 7 – Report Generation Flow of Veritas Information Studio



If the report was generated using the GUI, then a comma-separated variable (CSV) file is created and if using an API, the file generated can either be CSV or SQLite format. Some customers may opt to generate reports of SQLite format to further conduct more advanced SQL queries and generate a list of items to send to the cloud based on these queries. Sample snippet output in CSV format generated is shown below in Figure 8. The contents of this report include a header that describes the type of data in the report such as name, owner, extension, size, count, classification tags, etc. The extracted data is in the subsequent lines.

Figure 8 - Sample snippet of *.csv report.

```
$ cat nbu.csv | more
```

```
Name,Owner,Extension,Size,Count,NetBackupPolicies,MasterServer,CreatedTime,ModifiedTime,AccessedTime,Repository,ContentSource,DataStore,Path,Location,ClassificationTags,ClassifiedTime,PeopleTags,PlacesTags,OrganizationTags,LastNerScanTime,LastContentMatchedTime
```

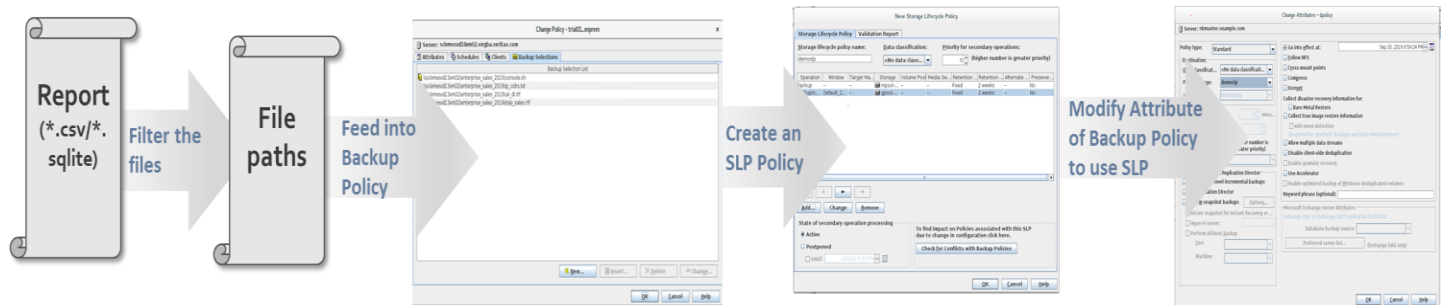
```
ArchiveMigration.pdb,user,pdb,65866752,1,VM_NBU_PROD_M,"xxxxx.xxxxx.veritas.com",2013-09-16T07:42:24Z,2013-09-16T07:40:09Z,2013-09-16T07:40:08Z,F,VM_NBU_PROD_M,NBU,"/F/DFS1/FromHROUS/Products/BE/COLUMBUS/1375D/exe.o/i386/ArchiveMigration.pdb",Pune,,,,,
```

```
"int8-exp-three-digits.out",uid-26,out,28578,1,VM_SDIO_CICD_Q,"xxxxx.xxxxx.veritas.com",2016-12-16T07:16:59Z,2012-02-23T22:59:21Z,2012-02-23T22:59:21Z,usr,VM_SDIO_CICD_Q,NBU,"/usr/lib64/pgsql/test/regress/expected/int8-exp-three-digits.out",Pune,,,,,
```

As depicted in Figure 9, this report would need to be further filtered to pull the file paths and then be manually entered into the NetBackup backup policy "Backup Selection List" attribute using the NetBackup administration console. Alternatively, an "inclusion" script or "exclusion" list file can also be created and fed into the NetBackup commands using [bpplinclude](#) and [bpsetconfig](#) respectively. Example scripts that incorporate these commands to populate the "Backup Selection List" in the backup policy are discussed in the Appendix. Once the backup policy is defined, it can be targeted to a storage lifecycle policy to first do a backup to a local storage target followed by a duplication to a cloud target. Parameters in the SLP determine at what point the

second copy is made as well as the retention policies for all the copies in the SLP. The backup policy attribute “Policy Storage” is modified to utilize the desired storage lifecycle policy.

Figure 9 -Veritas Information Studio Integration Flow with NetBackup

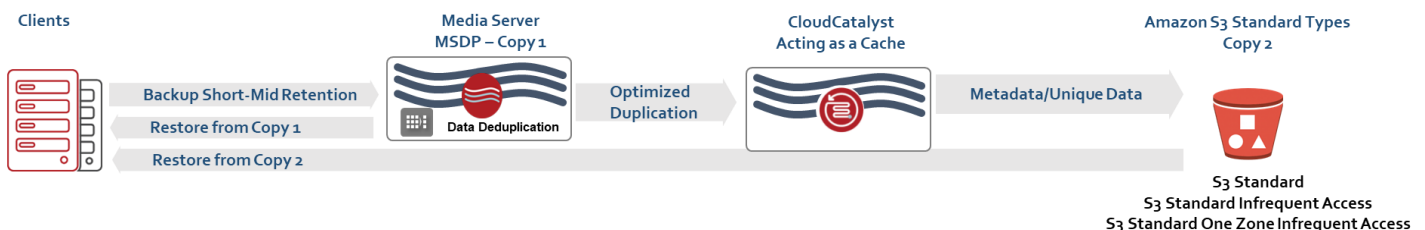


OPTIMIZED DUPLICATION (DEDUPLICATION) DATA FLOW TO AMAZON S3 STORAGE CLASSES

CloudCatalyst is required when sending deduplicated data to any of the Amazon S3 storage classes supported by NetBackup. Depending on the target Amazon S3 storage class the data paths and restore time differs. Below are example paths for each of the Amazon S3 storage classes supported by NetBackup with CloudCatalyst. The most common path to cloud storage is when data from clients are initially backed up and deduplicated to an MSDP storage target residing on a media server for short-term retention and duplicated to cloud for long-term. For restores, the data can be restored from any of the copies that reside on-premises or public cloud. Data is sent to an Amazon S3 storage class utilizing the S3 protocol with the S3 OST cloud plugin installed on CloudCatalyst as previously discussed.

Figure 10 illustrates an example data path from NetBackup client to an Amazon S3 Standard, Standard-IA, and Standard One Zone-IA storage classes. In this example, the client data are backed up and deduplicated to an MSDP storage target residing on a media server for short or mid-term retention (copy 1). This deduplicated data is sent to CloudCatalyst which caches the data for performance and then uploads the unique data to the Amazon S3 Standard, Standard-IA, or Standard One Zone-IA storage class for long-term retention (copy 2). It is best to have the deduplication be done on a media server instead of on CloudCatalyst. Having a separate media server to do the backup and deduplication allows the CloudCatalyst to use all its resources mainly for caching and transferring the data to the AWS cloud. For restores, data goes through a similar path but in reverse direction. By default, if the data is still in the media server, then it is restored from the media server. However, if restoring data from Amazon S3 Standard/S3 Standard-IA/S3 Standard One Zone-IA storage class, it will first check the CloudCatalyst cache and if it exists in the cache, it is sent back from the cache. If data is not present in the CloudCatalyst cache, the data is retrieved from Amazon S3 Standard/Standard-IA/Standard One Zone-IA storage class and passed to the CloudCatalyst to send up to client.

Figure 10 - Data Flow of NetBackup to and from Amazon S3 Standard, Standard IA, Standard One Zone IA Storage Classes

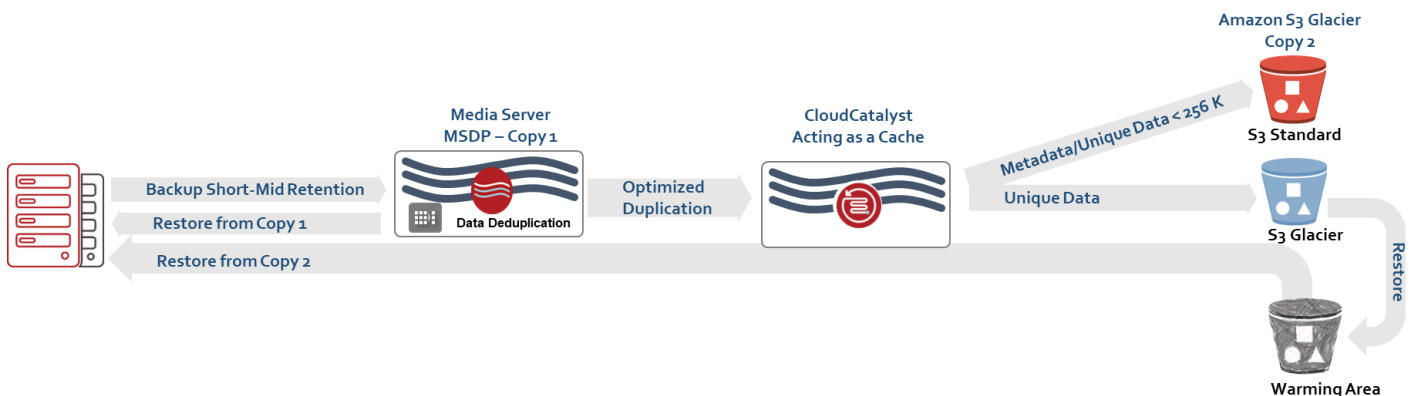


When writing to an Amazon S3 Glacier storage class as shown in Figure 11, the meta-data and data less than 256K are written to an S3 Standard storage class and regular data is written to an S3 Glacier storage class. In this example, client data is first backed up and deduplicated to a MSDP (copy 1) on a media server and then duplicated to S3 Glacier (copy 2) via CloudCatalyst using S3 protocol. Restoration of data from S3 Glacier requires a “warming” step where the data is recalled in a temporary area (Reduced Redundancy Storage) within AWS cloud prior for the data to be retrieved. The media server would poll the status of the warming operation as part of the restore job. Once all fragments of the data are warmed for recovery, then CloudCatalyst would restore the data and stream it out to the client. AWS offers 3 types of retrieval operations for S3 Glacier storage class:

- Expedited – completes within 1-5 minutes
- Standard – completes within 3 – 5 hours
- Bulk - completes within 5-12 hours

With NetBackup, the default retrieval type is “bulk”. If the other types are desired, then a GLACIER_RETRIEVAL file containing one of the strings: “expedited” or “standard” is placed in the /usr/opensv/netbackup/bin directory of the master server. **NOTE:** The restore in the warming area is done in a serial fashion and thus if image consists of multiple fragments, then restore time may be longer.

Figure 11 – Data Flow of NetBackup to and from Amazon S3 Glacier Storage Class

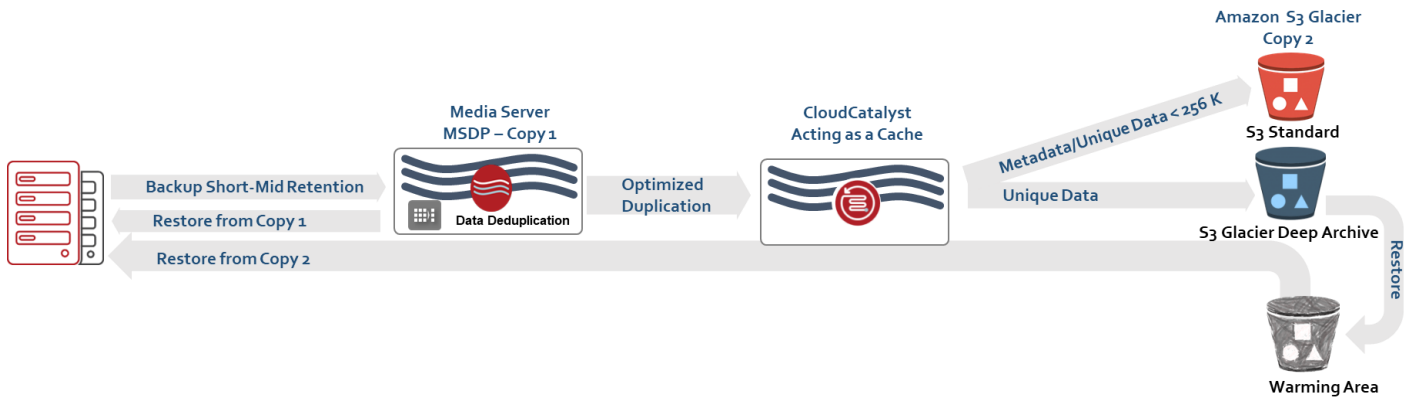


As in the previous example, when sending data to Amazon S3 Glacier Deep Archive storage class pictured in Figure 12, the client data is backed up and deduplicated on a media server and placed on an MSDP and then duplicated to the AWS cloud via CloudCatalyst. Metadata and unique data less than 256 K are sent to an S3 Standard storage class and larger sized unique data is sent to S3 Glacier Deep Archive storage class. A warming area within AWS is in the path during restores. When using S3 Glacier Deep Archive storage class, supported retrieval types are:

- Standard – completes within 12 hours
- Bulk - completes within 48 hours

Bulk is the default type unless the GLACIER_RETRIEVAL file in /usr/opensv/netbackup/bin directory of the master server contains the string “standard”. NetBackup requests for restore from S3 Deep Archive storage class and checks or polls the full availability of the data prior to retrieval of the data.

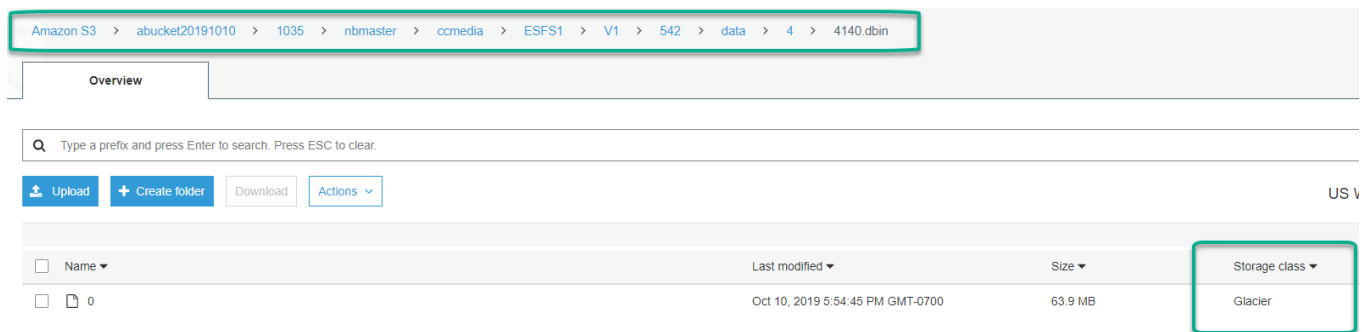
Figure 12 – Data Flow of NetBackup to and from Amazon S3 Glacier Deep Archive



MSDP separates a backup image into 64 MB 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 of size 64 MB and header file for each container of size 55KB. Once it passes through CloudCatalyst, there will also be a block map file and image properties file of sizes less than 55 bytes for both the data container and header file. If encryption is enabled, there is additional information relating to the keys and header for the keys. A sample view from the AWS management console of the 64 MB container file in an S3 Glacier is shown in Figure 13. As can be observed, the storage class is Glacier and the sample path from bucket name is of the form:

`/<hash code>/master server/media server/<ESFS INSTANCE NAME>/<OBJECT LAYOUT VERSION>/<inode>/path/to/file/`

Figure 13 - Sample View of Deduplicated Data in Amazon S3 Glacier Storage Class

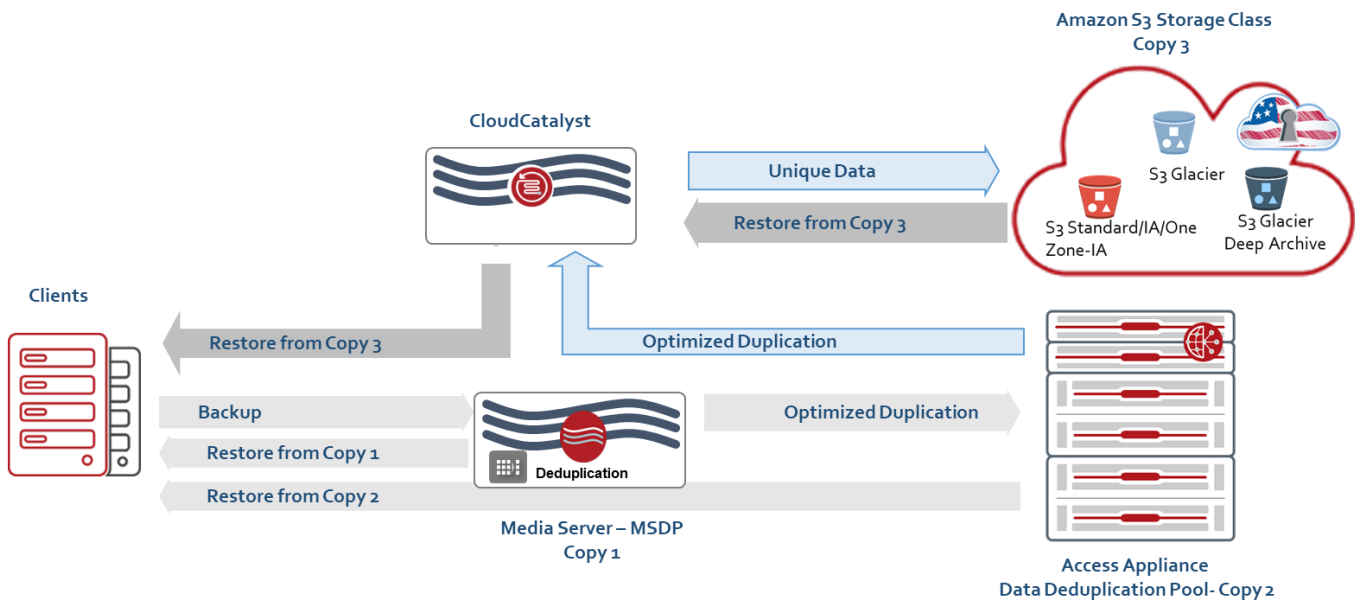


ACCESS APPLIANCE OPTIMIZED DUPLICATION (DEDUPLICATION) DATA FLOW TO AWS CLOUD STORAGE

In some scenarios, the Access Appliance is in the data path for on-premises mid-term or long-term retention prior to sending off-premises. When using Access data deduplication with NetBackup, one can duplicate data to the cloud utilizing NetBackup SLP policies and CloudCatalyst. The SLP would specify to duplicate the data from Access Appliance to CloudCatalyst to send the data to the cloud. Figure 14 provides a view of how this approach to send data to the cloud from Access data deduplication. In this example, deduplicated data is sent to the Access Appliance from the media server and then it does an optimized duplication to the cloud via CloudCatalyst. The role of the media server during the optimized duplication to cloud is to control and orchestrate the

transfer between Access Appliance and CloudCatalyst. The actual I/O is between the Access Appliance and CloudCatalyst. A restore can be done either from the Access Appliance (copy 2) or from the AWS cloud (copy 3). The data path from CloudCatalyst to the varying Amazon S3 storage classes is the same as described in the previous sections. By default, copy 1 is used for restores unless specifically specified to restore from the different copies.

Figure 14 - Access Appliance Data Deduplication Path to Amazon S3 Storage Classes via CloudCatalyst

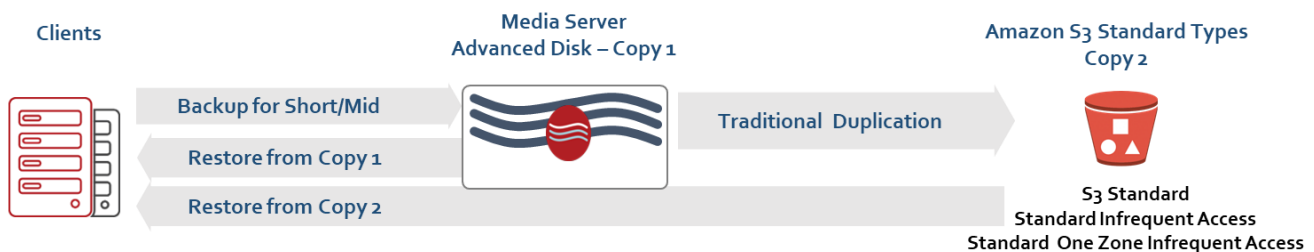


TRADITIONAL DUPLICATION DATA FLOW

For data that does not benefit from deduplication, backup data is duplicated from media server to the Amazon S3 storage classes using the NetBackup S3 OST plugin installed by default on the media server. Pictured below in Figure 15-17 are examples of the data flow from NetBackup to and from for the differing Amazon S3 storage classes. The traditional duplication flow is similar to the optimized duplication (deduplication) without the employment of CloudCatalyst.

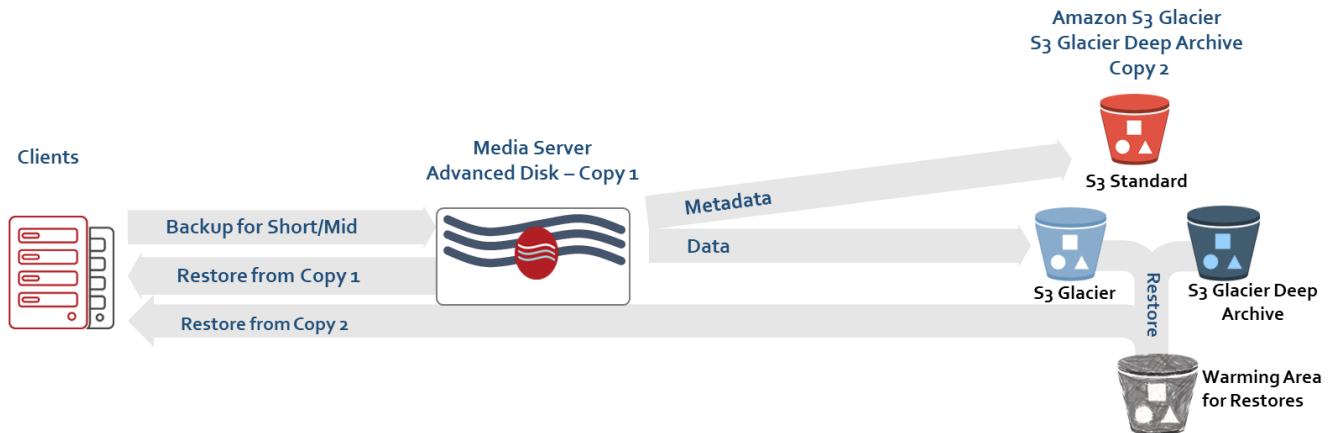
The traditional flow for sending data to Amazon S3 Standard, Standard-IA, and Standard One Zone IA storage classes involves the data to be initially stored in an advanced disk (copy 1) on a media server for short to mid-term retention and then copied to the Amazon S3 storage class (copy 2) for long-term retention. Restores can be done from either the advanced disk (copy 1 - default) or from Amazon S3 storage classes (copy 2) as depicted in Figure 15.

Figure 15 – Data Flow of NetBackup Traditional Duplication to and from Amazon S3 Standard storage classes.



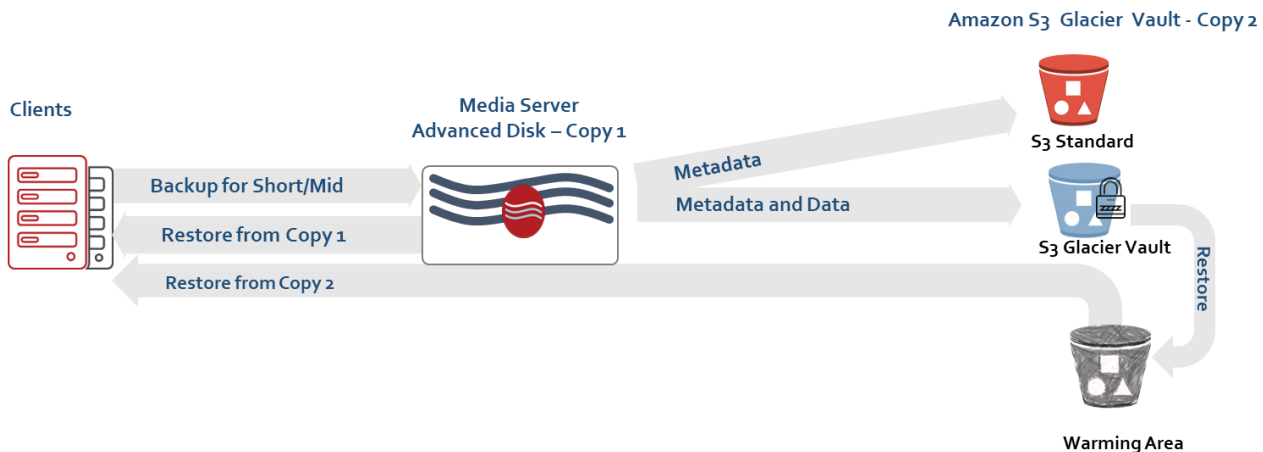
When sending data to S3 Glacier or S3 Glacier Deep Archive, metadata are stored in an S3 Standard storage class and data are stored in S3 Glacier or S3 Glacier Deep Archive storage class as pictured in Figure 16. Restoration for any of the Glacier types of S3 storage classes require pulling the data first to an Amazon warming area prior to retrieval from NetBackup. For traditional duplication, retrieval type supported for both is standard, however, S3 Glacier storage class completes within 3-5 hours whereas Glacier Deep Archive storage classes completes within 12 hours.

Figure 16 - Data Flow of NetBackup Traditional Duplication to and from Amazon S3 Glacier and Glacier Deep Archive Storage Classes



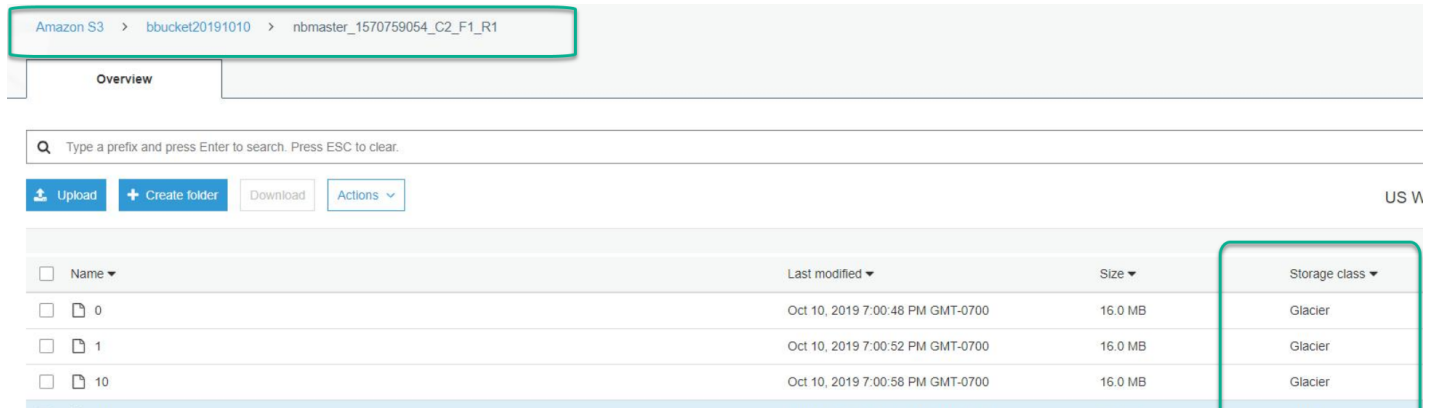
If there is a requirement to store data with certain immutable or lock policies such as WORM or retention for compliance or regulatory rules, Amazon S3 Glacier Vault is used. The vault lock policies are defined and managed within AWS after the S3 Glacier vault is created as a cloud storage target within NetBackup. For S3 Glacier Vault shown in Figure 17, both the metadata and data are stored in the S3 Glacier Vault with vault lock policies applied. In addition, metadata is stored in an S3 Standard storage class. As with S3 Glacier storage class, restores from S3 Glacier Vault would require placing the data in a warming area prior to restoring from NetBackup. However, unlike S3 Glacier, the retrieval type supported by NetBackup with S3 Glacier Vault is standard and thus restoration takes a minimum of 3 to 5 hours. NetBackup media server checks the status of the restore in the warming area and pulls the data once the data is fully available. **NOTE: The data is available for retrieval within AWS warming location for a maximum of 24 hours.**

Figure 17 - Data Flow of NetBackup Traditional Duplication to and from Amazon S3 Glacier Vault



A sample view of data in an S3 Glacier storage class for traditional duplication is shown in Figure 18. Backup images and its associated header information are stored in a directory structure. Each directory contains the image properties, block map file, and the actual backup image. The header directory contains the header information, properties of header information, and block map file for the header. The S3 OST Cloud plugin breaks the backup image up into fixed object sizes of 16 MB. With encryption, there is one key per OST image so there will be additional directories and objects related to the keys.

Figure 18 - Sample View of Duplicated Data in S3 Glacier Storage Class



The screenshot shows the Amazon S3 console interface. At the top, the breadcrumb navigation is 'Amazon S3 > bbucket20191010 > nbmaster_1570759054_C2_F1_R1'. Below this is a search bar and a toolbar with buttons for 'Upload', 'Create folder', 'Download', and 'Actions'. The main area displays a table of objects:

Name	Last modified	Size	Storage class
0	Oct 10, 2019 7:00:48 PM GMT-0700	16.0 MB	Glacier
1	Oct 10, 2019 7:00:52 PM GMT-0700	16.0 MB	Glacier
10	Oct 10, 2019 7:00:58 PM GMT-0700	16.0 MB	Glacier

Additionally, with traditional duplication, NetBackup supports tiering data (not metadata) across the different levels of Amazon S3 object storage classes using Amazon S3 storage lifecycle policies. For instance, tiering data from S3 Standard to Glacier or S3 Standard to S3 Standard Infrequent Access is available and supported by NetBackup. However, Amazon storage lifecycle policies are only supported for instances where Accelerator (feature that provides full backups with reduction of backup window and compute and storage resources of an incremental backup) and CloudCatalyst are not utilized to send data directly to the AWS cloud. So, for data that does not require deduplication, NetBackup has parameters such as `UPLOAD_CLASS`, `TRANSITION_TO_STANDARD_IA_AFTER` and `TRANSITION_TO_GLACIER_AFTER` to integrate with the AWS's lifecycle cloud tiering. For example, backup data can first reside in an S3 standard storage class and then after 30 days transition to S3 Standard Infrequent Access and then after 60 days transition to Glacier as illustrated in Figure 19. **NOTE:** If transitioning to Glacier, check to make sure that Glacier is supported for the region which the bucket belongs to. Also, Amazon S3 lifecycle is not supported with NetBackup Accelerator.

Figure 19 - NetBackup with Amazon S3 Lifecycle Policy



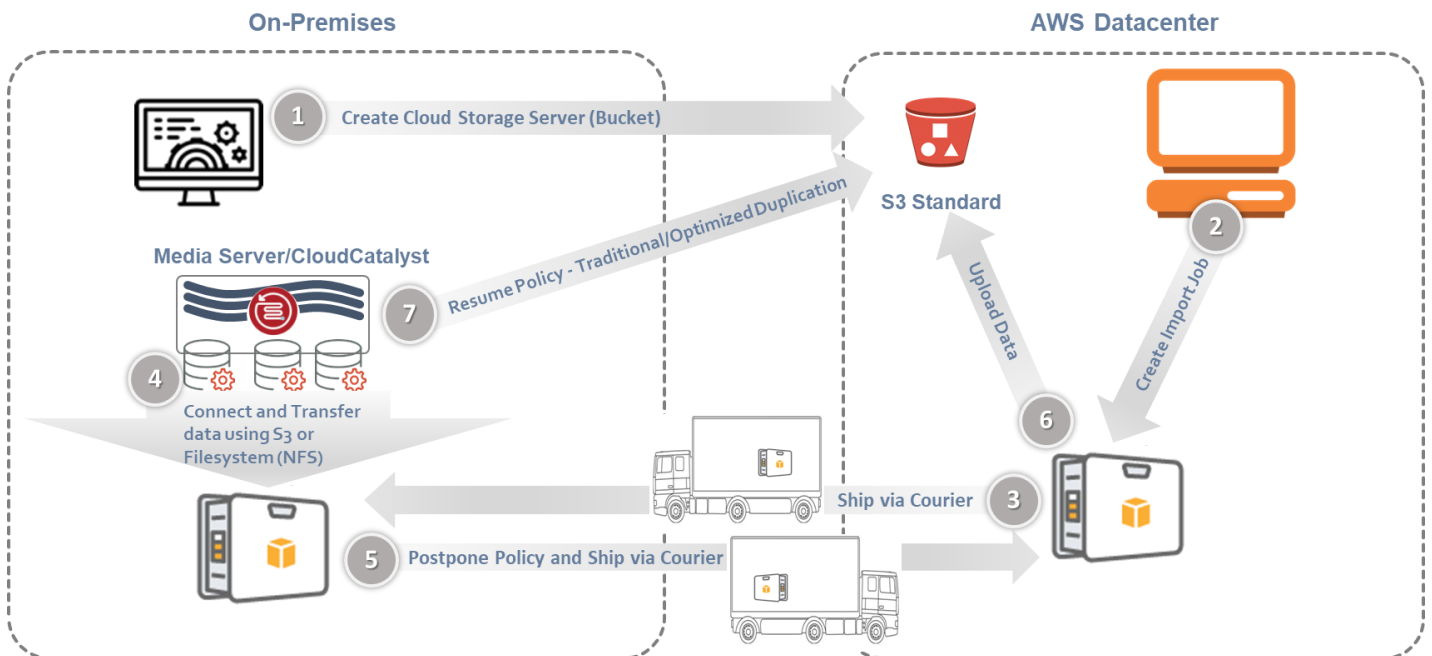
*NOTE: Amazon Storage Lifecycle Policies not supported with CloudCatalyst and Accelerator

MIGRATION TO CLOUD

Migration to cloud or initial seeding of data into cloud is simple when using NetBackup with AWS Snowball and AWS Snowball Edge. These devices are placed on-premises in order to handle large-scale data transfers quickly without overloading the corporate network. As illustrated in Figure 20 this involves the following:

1. From NetBackup administration console, configure a cloud storage server and create Amazon S3 Standard class as target storage. The bucket created should be in the same region as where the Snowball is available to be ordered.
2. On AWS management console, create an import job specifying the bucket created in the previous step and indicate region and shipping information.
3. The Snowball or Snowball Edge is shipped by courier to the datacenter where the NetBackup domain resides.
4. The device is connected to the same network as the NetBackup domain. As previously discussed, for Snowball the tools required for transfers are downloaded from AWS resources website onto a server and for Snowball Edge the tools are already on the device. Once the devices are configured, a cloud storage server is created in NetBackup with the device's properties (e.g. IP address, type of transfer (e.g. S3 or Filesystem (NFS), etc.). A backup policy and/or storage lifecycle policy is defined to transfer data to Snowball or Snowball Edge. **NOTE: File interface transfer is not supported with CloudCatalyst, only S3 adapter is supported.**
5. After all data is transferred, the policy is postponed or deactivated in NetBackup, the device is disconnected and shipped back to AWS datacenter via courier.
6. Once device is received at the AWS datacenter, data is transferred to the destined S3 Standard class (bucket). The progress of transfer can be viewed or monitored on the AWS console.
7. After all data has been uploaded, the backup policy and/or SLP can be resumed on NetBackup.

Figure 20 - Migration to Cloud Data Flow



An example estimated timeline based on an actual test of this whole process is outlined in Table 4 for 72 TB of data for the US region using NetBackup CloudCatalyst and S3 Adapter. The total time is dependent on the network connection, the type of AWS device used, the region, the transfer protocol (e.g. S3 adapter or File interface), etc.

Table 4 - Estimated Timeline for Migration of 72 TB of Data to AWS Cloud in US Region (Based on Actual Testing)

Activity	Time
Order and AWS prepare to ship	3 days
Shipped to datacenter	2 days
Transfer 72TB to Snowball from NetBackup with CloudCatalyst using S3 Adapter at 200 MB/second	4.2 days
Shipped to AWS	2 days
Import to Bucket in AWS	2 days

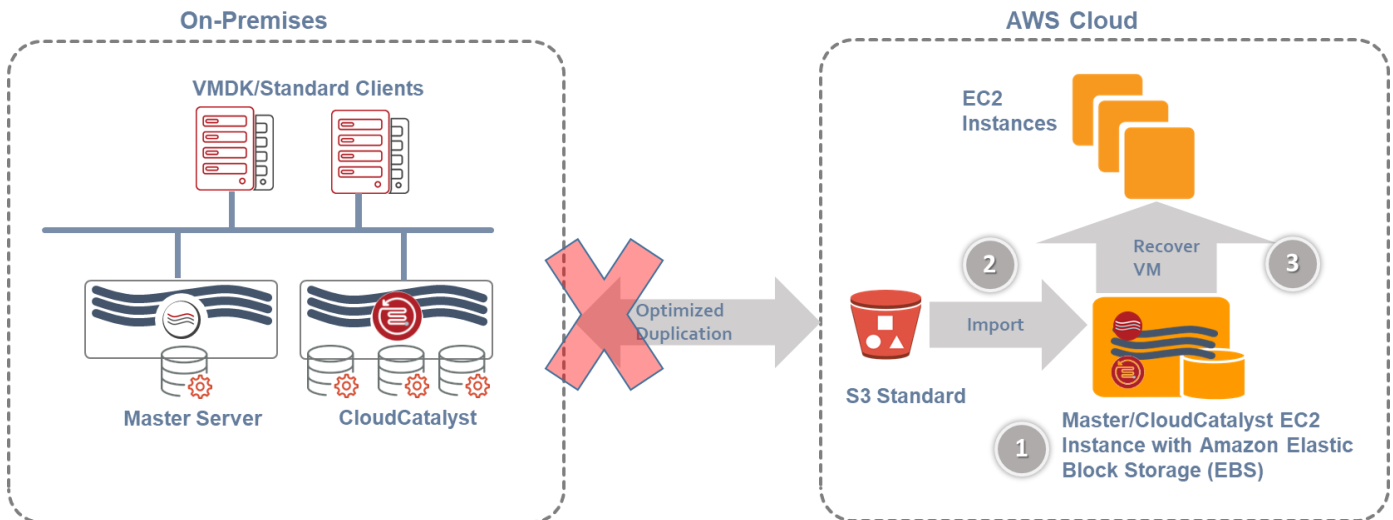
DISASTER RECOVERY IN THE CLOUD

Introduced in NetBackup 8.2 is a feature called "[automated disaster recovery](#)" which allows recovery of the backup images in the cloud using the CloudCatalyst Image Sharing functionality in scenarios where the on-premises NetBackup environment is lost due to power outage, natural disaster or catastrophic events. Utilizing CloudCatalyst with this feature allows for creation of "self-describing" images where both the data and metadata of the backup image are sent to the AWS cloud. With this feature demonstrated in Figure 21, when the on-premise NetBackup environment becomes unavailable the flow involves:

1. A [NetBackup master and CloudCatalyst instance](#) is instantiated on AWS.
2. The images read from the Amazon S3 Standard cloud storage and imported to recreate the NetBackup catalog
3. The image(s) is recovered within the cloud.

The NetBackup master and CloudCatalyst instance deployed on AWS should be based on RHEL 7.3 or later with a recommended 64 GB of memory, 8 CPUS and IPv4 network. A new command "[nbimageshare](#)" is available to pull the data and metadata using the Amazon S3 REST APIs. **NOTE:** *This feature is only qualified with VMware and Standard policy types and data stored only in an Amazon S3 Standard storage class (bucket). It is not supported with S3 Glacier or S3 Glacier Deep Archive. Also, the S3 Standard bucket is "read-only" from the NetBackup instances in the cloud.*

Figure 21 – Automated Disaster Recovery in AWS Cloud



BEST PRACTICES AND RECOMMENDATIONS

Following best practices is important in creating an optimum deployment. This section covers highlights some best practices relating to the AWS cloud storage as a long-term retention solution for NetBackup. Some of these best practices are covered more in-depth in the following documentations:

- [NetBackup Cloud Administrator's Guide](#)
- AWS [Snowball](#) and [Snowball Edge](#) Best Practices

PRIVACY LAWS

Certain countries have employed General Data Protection Regulation (GDPR) and privacy laws specifically related to data stored in the public cloud. Although NetBackup stores data in a backup image format, the data can still be perused or scanned for personal identifiable information, confidential and business critical information. Use tools such as Veritas Information Studio to decide what information is best on-premises versus in the public cloud to reduce liability and violation of these laws. Also be aware of restoring the data in the cloud when using NetBackup automated disaster recovery, "self-describing images". For example, if data is backed up in Europe region, the NetBackup cloud instances and data restored in the cloud should be in the same region or adhere to your companies GDPR policies. Use of NetBackup encryption would also assist in addressing some of the concerns relating to privacy.

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.

DEDUPLICATION

When using CloudCatalyst, it is recommended that CloudCatalyst be used solely for caching and transferring data to the AWS cloud storage and a separate media server be utilized to conduct the backup and deduplication, especially when handling large amounts of data. If more than 250 TB of MSDP is required, use the NetBackup and CloudCatalyst 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 323 TB or 1056 TB depending on the model. The maximum MSDP capacity depends on the NetBackup appliance used as a media server.

When the first copy resides within an MSDP and then duplicated to the cloud without CloudCatalyst, the data is rehydrated prior to being sent to the AWS cloud. The rehydration would increase the time and network resources to send data to cloud. As a best practice, it is recommended to use CloudCatalyst to send deduplicated data to the cloud or if deduplicated data is not required, do not place data within an MSDP but instead in other media types such as advanced disk.

SNOWBALL/SNOWBALL EDGE

Some of the best practices for when using Snowball and/or Snowball Edge include:

- For Snowball Edge, use the S3 Adapter to transfer data as oppose to the File Interface. It has been observed that the performance using the S3 Adapter is significantly better when compared to the File Interface.
- Keep the copies on-premises until data has been fully imported into the Amazon S3 and it has been verified.
- Create the bucket within the NetBackup administration console to conform to the bucket naming convention.

NETBACKUP RETRIEVAL ATTRIBUTES

There are several NetBackup attributes that are best to use when sending or restoring data from when using Glacier, Glacier Vault and/or Glacier Deep Archive. For instance, it is best to use the option [True Image Recovery](#) whenever possible. It has been observed that enabling this option significantly reduces the retrieval or restore of data from hours to minutes in some cases. With True Image Recovery, it keeps track of information such that it only restores the files that were present at the time of the last backup otherwise it will restore all files.

Another useful attribute is RETRIEVAL RETENTION PERIOD to implement with Glacier and Glacier Deep Archive only. Set this value within the NetBackup storage server to a minimum of 3 days. By setting the value to days allows the data to be kept within the warming area for scenarios where it might take several retries to pull data due to network issues, contention and other failures.

CONCLUSION

Veritas products along with AWS cloud storage services provide a long-term retention solution for data protection and management of customer's data. AWS offers varying storage classes at different levels of availability, cost, and restore to meet the various needs of customers. Veritas Information Studio is instrumental to assist in identifying data that can be safely moved to the cloud to minimize risk and liability. Utilizing AWS cloud storage with NetBackup provides a compelling option for protecting and preserving data from on-premise challenges such as failures, attacks and disasters as well as an off-premise long-term retention solution.

REFERENCES

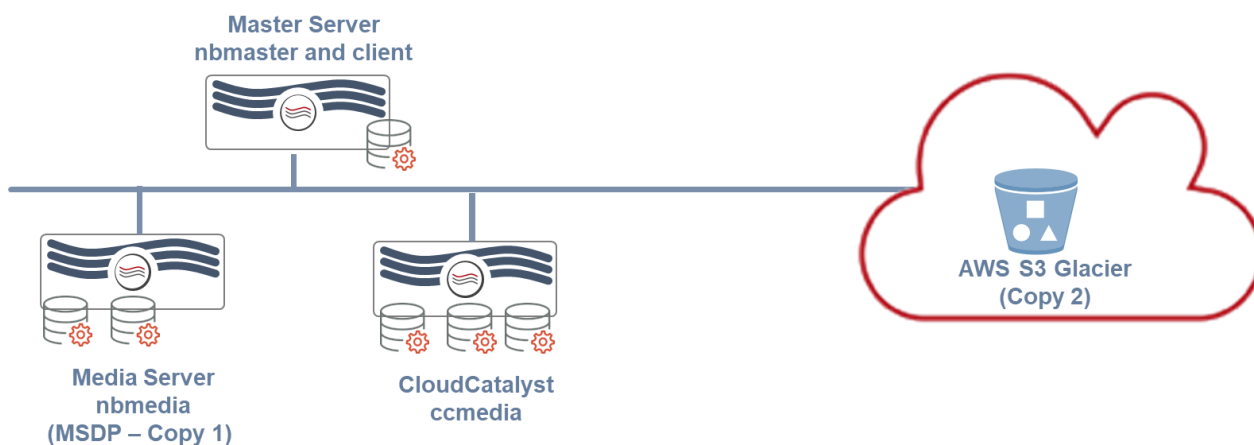
- NetBackup
 - Product Documentation
 - https://www.veritas.com/support/en_US/article.DOC5332
 - NetBackup Deduplication Guide
 - https://www.veritas.com/content/support/en_US/doc/25074086-136046435-0/index
 - NetBackup Cloud Administrator's Guide
 - https://www.veritas.com/content/support/en_US/doc/58500769-135186602-0/v58383369-135186602
 - Disaster Recovery
 - Veritas NetBackup in Highly Available Environments Administrator's Guide
 - https://www.veritas.com/support/en_US/doc/39129704-133515633-0
 - Disaster Recovery procedure for CloudCatalyst
 - https://www.veritas.com/content/support/en_US/doc/25074086-127355784-0/v127468421-127355784
- Access Appliance
 - Product Documentation
 - https://www.veritas.com/content/support/en_US/dpp.Access.html
 - Whitepapers/Datasheets
 - <https://www.veritas.com/protection/access-appliance/resources>
- AWS Cloud Storage
 - AWS S3 Storage Classes
 - <https://aws.AWS.com/s3/storage-classes/>
 - Snowball
 - <https://docs.aws.AWS.com/snowball/latest/ug/whatissnowball.html>
 - Snowball Edge
 - <https://docs.aws.AWS.com/snowball/latest/developer-guide/whatisedge.html>
- Information Studio
 - Product Documentation
 - https://sort.veritas.com/documents/doc_details/INFOSTUDIO/1.1/General/Documentation/
 - Whitepaper
 - https://www.veritas.com/content/dam/Veritas/docs/white-papers/Vo956_WP_Information-Studio.pdf

APPENDIX

This section provides an example of how to configure NetBackup to send data to cloud as well as how to utilize the report from Veritas Information Studio within example scripts to generate a backup policy. This is only a sample deployment and readers are expected to refer to the Veritas product documentation and AWS for definitive and specific installation, administration and configuration details.

This example utilizes BYOS NetBackup environment which consists of master server, media server and CloudCatalyst server and configuration is as pictured in Figure 22.

Figure 22 - Example Configuration Used in this Example



It is assumed that the Veritas Information Studio and NetBackup components are configured and installed. It also assumes that the media server deduplication pool (MSDP) has already been created. The example consists of the following main steps:

- 1) Create a user, access and secret keys on AWS management console.
- 2) Creation of NetBackup cloud storage server, enabling CloudCatalyst, and using S3 Glacier storage class as a target
- 3) Creation of SLP definition where backup and deduplication are done on a media server and placed in an MSDP and then duplicated to cloud storage server (S3 Glacier) done in previous step.
- 4) Create a backup policy and modify the backup policy to utilize the SLP.
- 5) Run a manual backup.
- 6) Verification of backup and duplication.
- 7) Generate the *.csv report using Information Studio GUI based on certain pre-configured filters.
- 8) Extract file path from report and create file paths that can be fed into NetBackup backup policies.
- 9) Use a script to create a backup policy defined above to include the file paths from the report.

AWS

On AWS management web console, create a user with an associated Access Key ID and Secret Access Key. These keys are used when configuring the AWS cloud storage server within NetBackup administration console.

Step 1) Log on to the AWS Web console GUI. Open the IAM console and click on **Manage Users** under the Create Individual IAM users.

The screenshot shows the AWS IAM console interface. On the left is a navigation menu with 'Identity and Access Management (IAM)' selected. Under 'AWS Account (215277274125)', 'Dashboard' is highlighted. Other options include Groups, Users, Roles, Policies, Identity providers, Account settings, and Credential report. A search bar is present. Under 'AWS Organizations', options for Organization activity and Service control policies (SCPs) are listed. The main content area is titled 'Welcome to Identity and Access Management'. It includes a sign-in link, IAM Resources summary (Users: 3, Roles: 8, Groups: 2, Identity Providers: 1, Customer Managed Policies: 1), and a Security Status section with a progress bar (3 out of 4 complete). The security tasks listed are: 'Activate MFA on your root account' (warning icon), 'Create individual IAM users' (checkmark icon, with a 'Manage Users' button), 'Use groups to assign permissions' (checkmark icon), and 'Apply an IAM password policy' (checkmark icon). A descriptive text for the 'Create individual IAM users' task advises against using the root account for day-to-day interaction.

Step 2) Click on **Add user**.

The screenshot shows the 'Users' page in the AWS IAM console. The left navigation menu is the same as in the previous screenshot, but 'Users' is now selected. The main content area has 'Add user' and 'Delete user' buttons at the top. Below them is a search bar and a table of users. The table has columns for 'User name', 'Groups', 'Access key age', 'Password age', 'Last activity', and 'MFA'. There are three users listed: 'duser', 'test', and 'test2'. All three users have 'Not enabled' MFA. The 'test2' user has a warning icon next to their 'Access key age' (157 days). The table indicates 'Showing 3 results'.

User name	Groups	Access key age	Password age	Last activity	MFA
duser	s3amazon	14 days	None	Today	Not enabled
test	s3amazon	160 days	None	159 days	Not enabled
test2	s3amazon and s3group	157 days	None	114 days	Not enabled

Step 3) Enter **user name** and place check on access type to be **"Programmatic access"**.

The screenshot shows the AWS IAM 'Add user' console. The top navigation bar includes the AWS logo, 'Services', 'Resource Groups', and a user profile 'Admin/Agnes.Jacob@veritas.co'. The page title is 'Add user' with a progress indicator showing 5 steps, with step 1 highlighted. The section 'Set user details' includes a text input for 'User name*' containing 'auser' and a link 'Add another user'. Below this, the 'Select AWS access type' section shows two options: 'Programmatic access' (selected with a checked radio button) and 'AWS Management Console access' (unchecked). Descriptions for each option are provided. At the bottom, there is a 'Cancel' button and a 'Next: Permissions' button. A note '* Required' is visible on the left.

Step 4) Click **Next Tags** and just click **Next: Review**. In this example, the s3AWS group selected has AWSS3FullAccess. For details on minimum permissions required, refer to [Veritas NetBackup Cloud Administrator's Guide](#).

The screenshot shows the AWS IAM 'Add user' console at Step 4: Set permissions. The progress indicator shows 5 steps, with step 2 highlighted. Under the 'Set permissions' section, three options are available: 'Add user to group' (selected), 'Copy permissions from existing user', and 'Attach existing policies directly'. Below this, the 'Add user to group' section shows a search bar and a table of groups. The table has two columns: 'Group' and 'Attached policies'. Two groups are listed: 's3amazon' (selected with a checked checkbox) and 's3group' (unchecked). The 'Attached policies' for 's3amazon' is 'AmazonS3FullAccess' and for 's3group' is 'AdministratorAccess'. At the bottom, there are 'Cancel', 'Previous', and 'Next: Tags' buttons.

Group	Attached policies
<input checked="" type="checkbox"/> s3amazon	AmazonS3FullAccess
<input type="checkbox"/> s3group	AdministratorAccess

Step 5) Click on **Create user**.

Add user 1 2 3 4 5

Review

Review your choices. After you create the user, you can view and download the autogenerated password and access key.

User details

User name	auser
AWS access type	Programmatic access - with an access key
Permissions boundary	Permissions boundary is not set

Permissions summary

The user shown above will be added to the following groups.

Type	Name
Group	s3amazon

Tags

No tags were added.

Cancel Previous **Create user**

Step 6) Download the *.csv file with **access key and secret key** or save information in a file.

Add user 1 2 3 4 5

Success

You successfully created the users shown below. You can view and download user security credentials. You can also email users instructions for signing in to the AWS Management Console. This is the last time these credentials will be available to download. However, you can create new credentials at any time.

Users with AWS Management Console access can sign-in at: <https://215277274125.signin.aws.amazon.com/console>

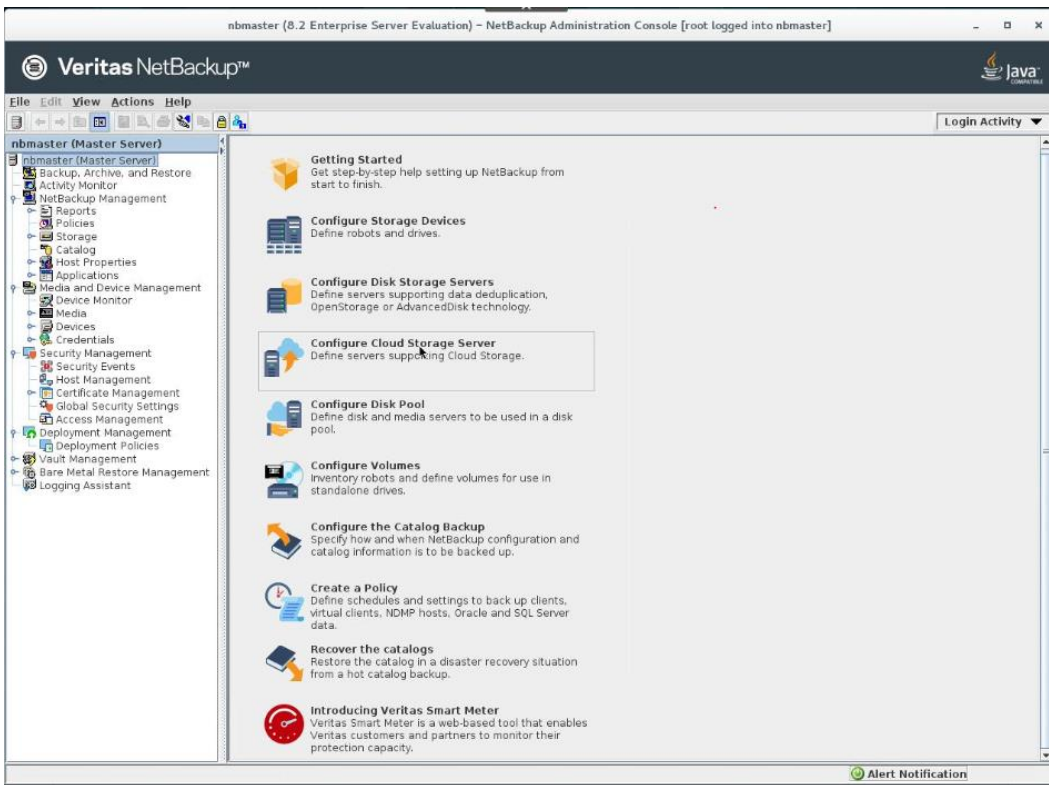
Download .csv

User	Access key ID	Secret access key
auser	AKIA7EH4G	***** Show

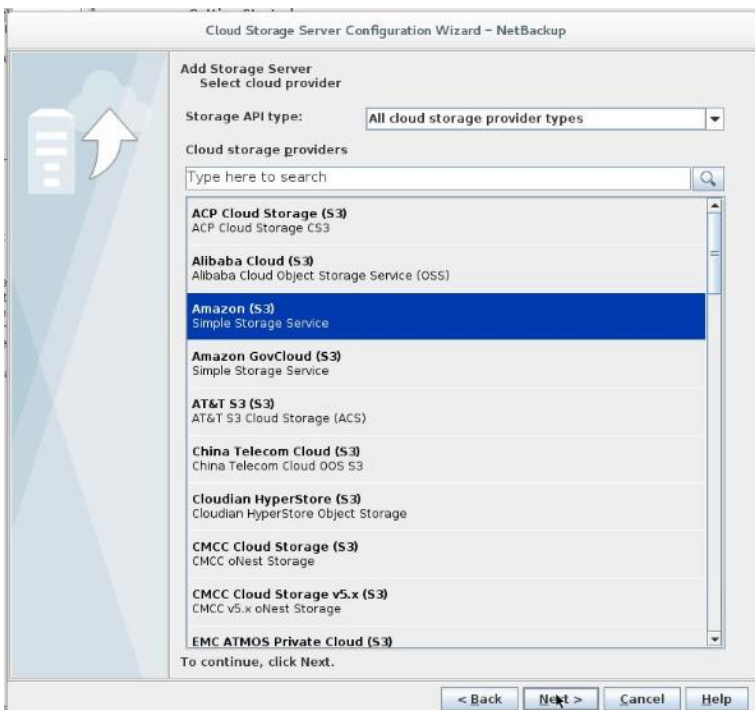
CREATION OF NETBACKUP CLOUD STORAGE – S3 GLACIER STORAGE CLASS

In this example, NetBackup is configured to send data to S3 Glacier storage class via CloudCatalyst. The other AWS S3 Storage classes would follow a similar configuration but would require different information such as region, etc.

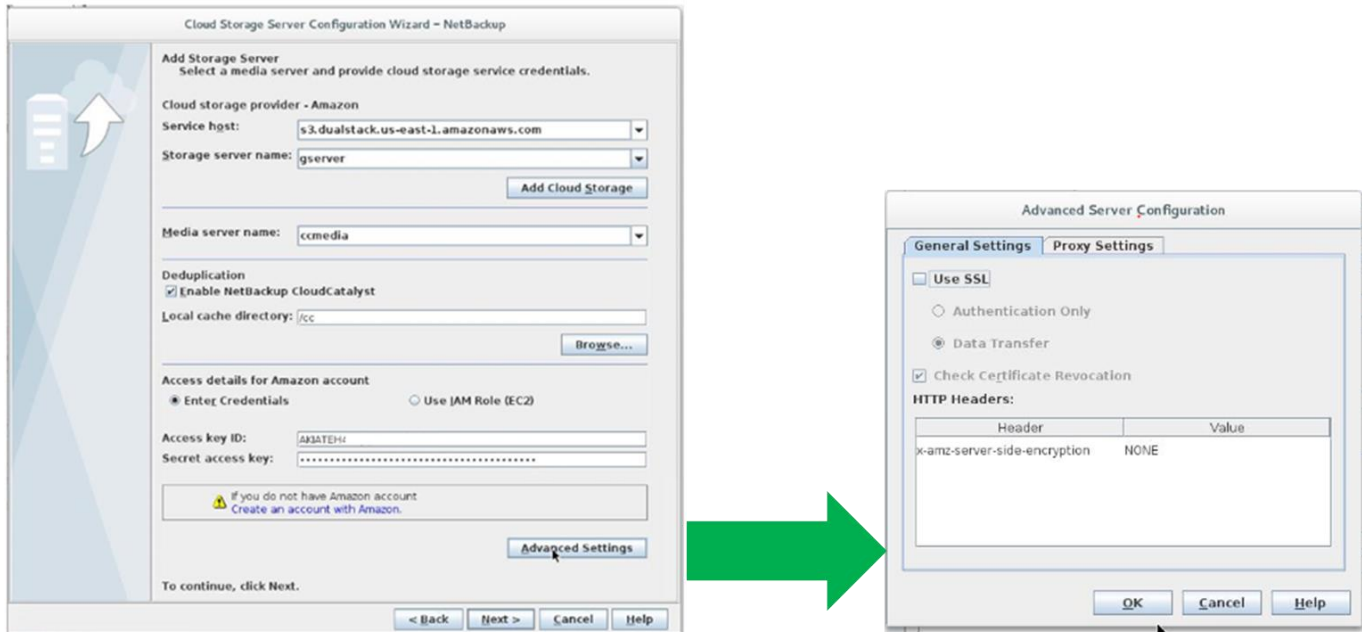
Step 1) Logon to NetBackup administration console. Click on **Configure Cloud Storage Server** and then click **Next**.



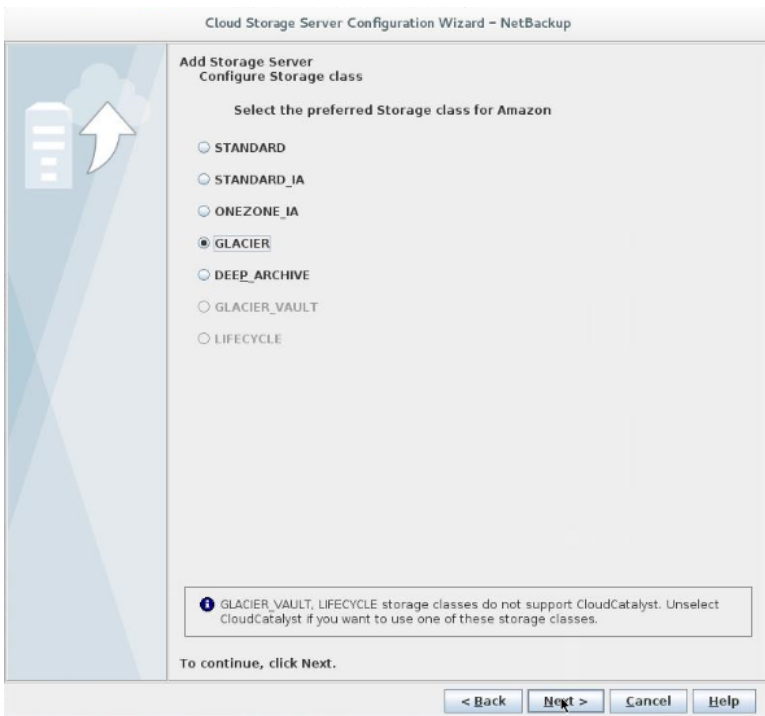
Step 2) Select the **AWS S3** as the Cloud Storage provider.



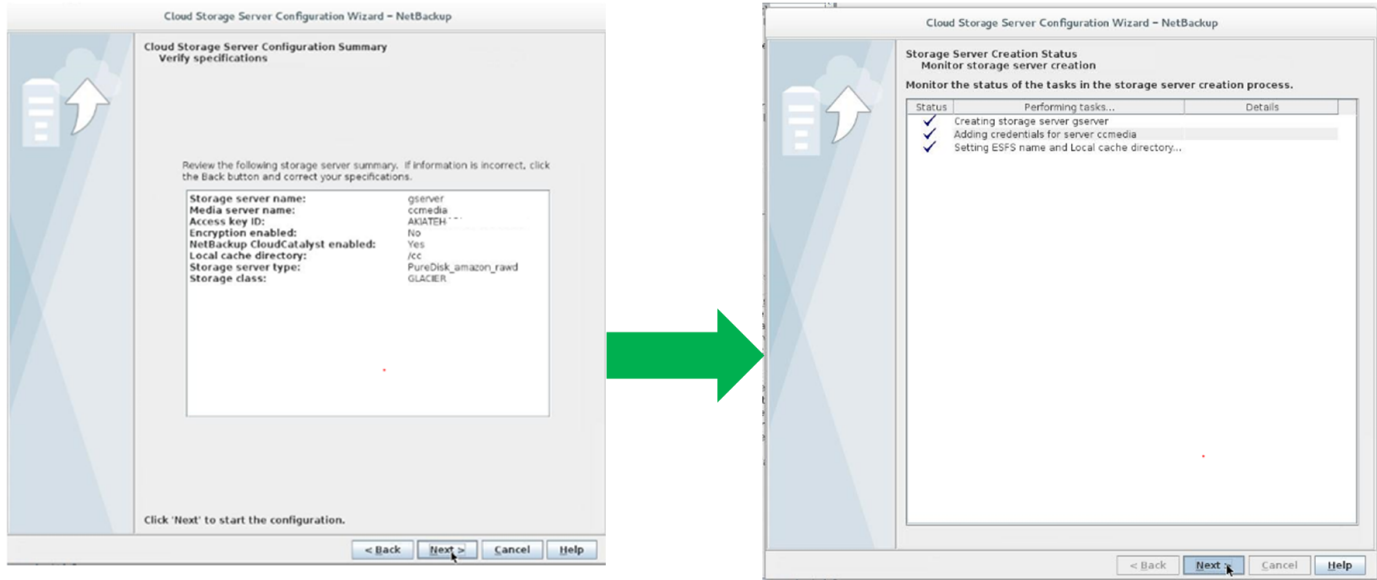
Step 3) Enter **storage server name**, the **CloudCatalyst server name**, click on **Enable NetBackup CloudCatalyst**, and enter **access key and secret key**. In this example, SSL is disabled. So, Under Advanced settings, disable **SSL** by unchecking the box.



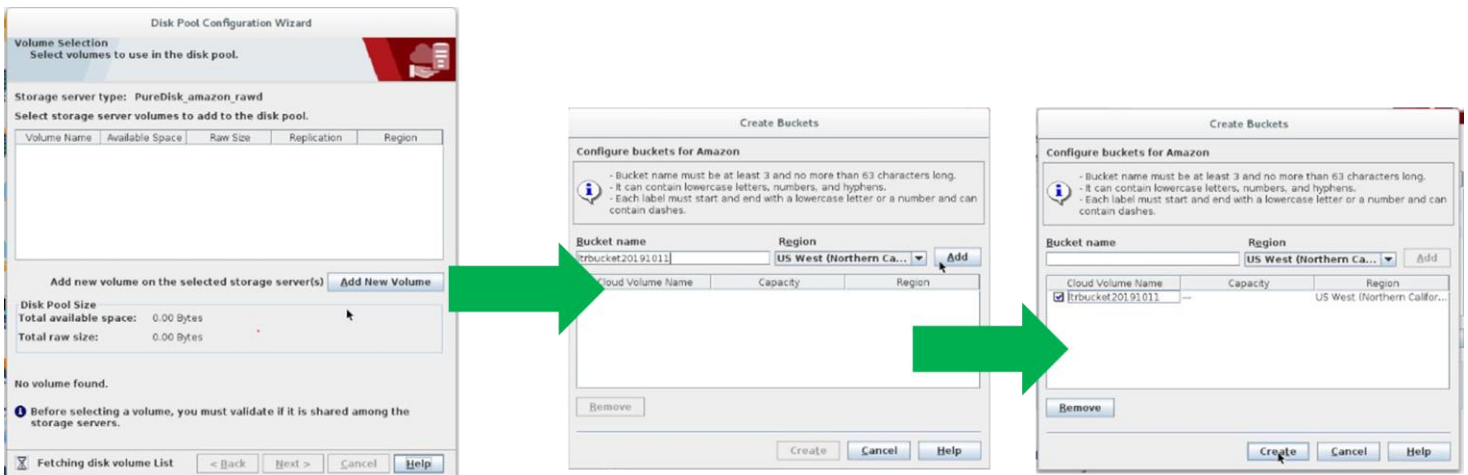
Step 4) Select **Glacier** as the preferred storage class.



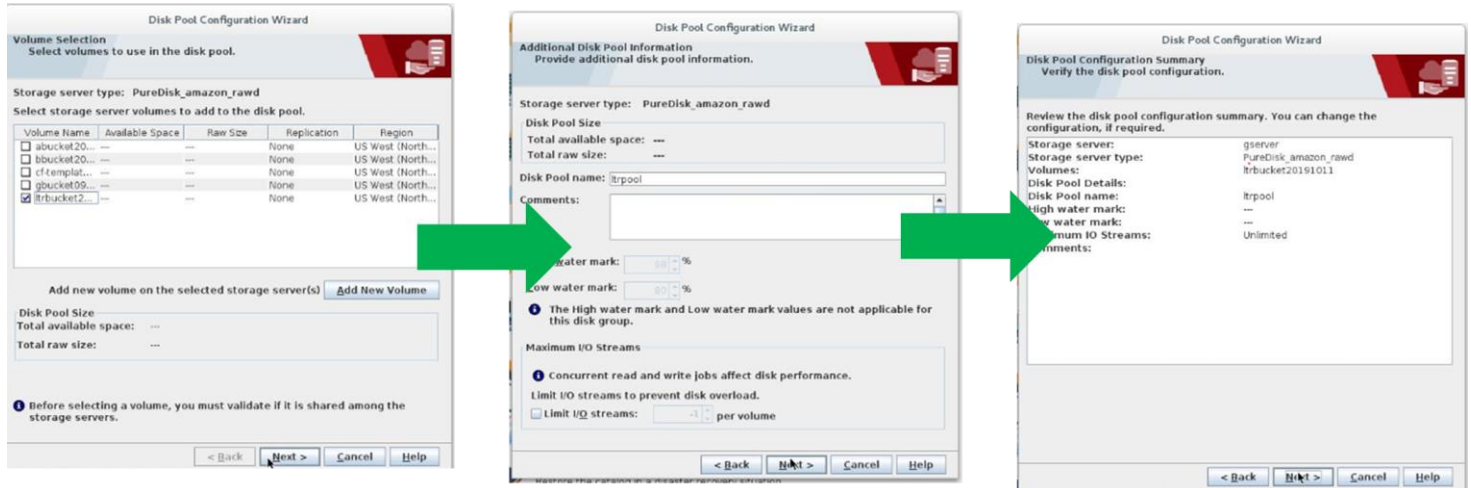
Step 5) Review the configuration and click **Next** and **Next**.



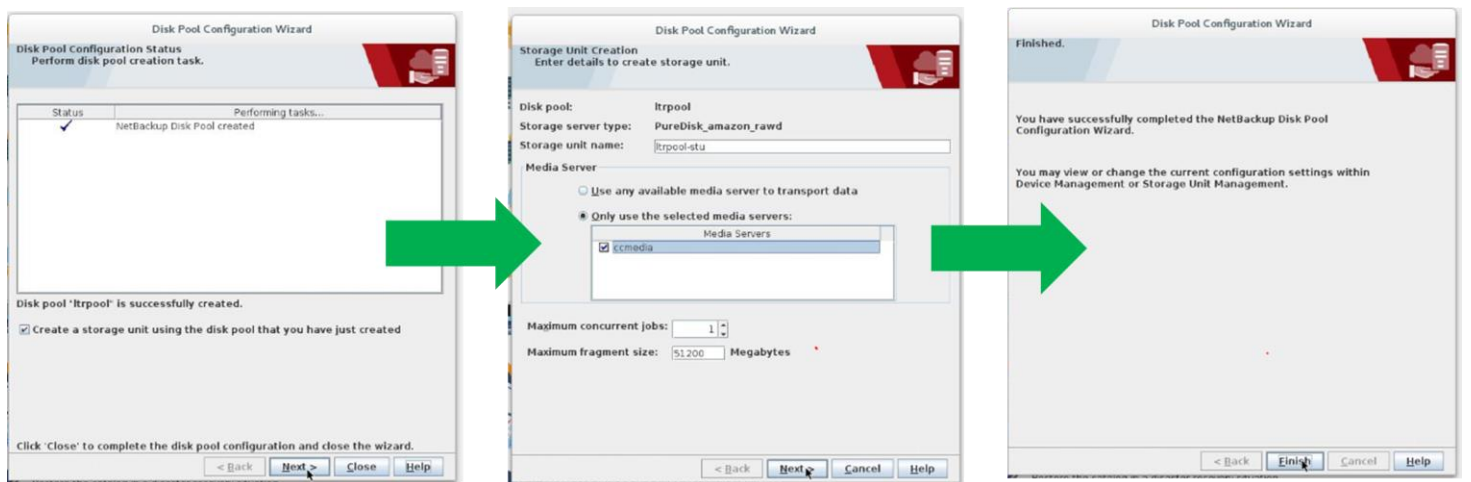
Step 6) Click on **Add New Volume**. On next screen, enter a **bucket name** and click on **Add**. **Select** the new bucket added and click **Create**.



Step 7) Select the newly created bucket and click **Next**. Enter **name** of disk pool, click **Next**, review the disk pool configuration and click **Next**.



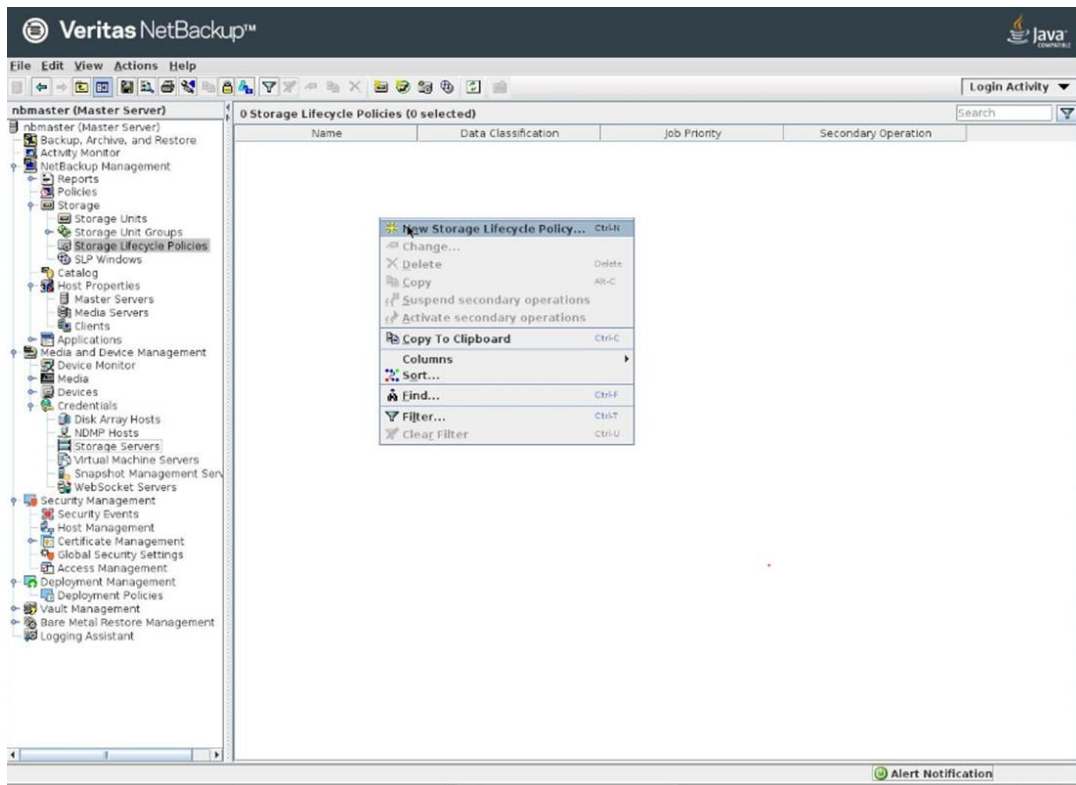
Step 8) After disk pool gets created, continue with the storage unit creation clicking **Next**, **Next** and then **Finish**.



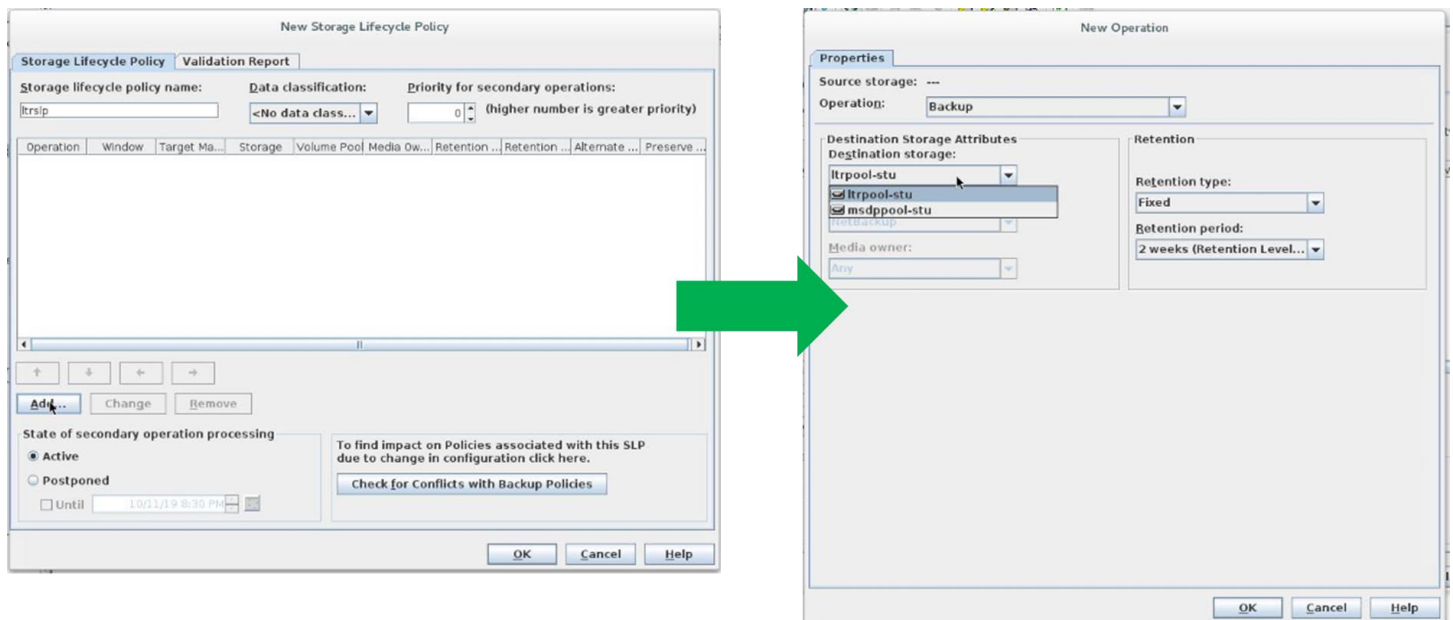
DEFINITION OF AN SLP

In this example, an SLP is created to first backup into an MSDP pool and then to the cloud storage server created in previous section. It is assumed that an MSDP has already been defined.

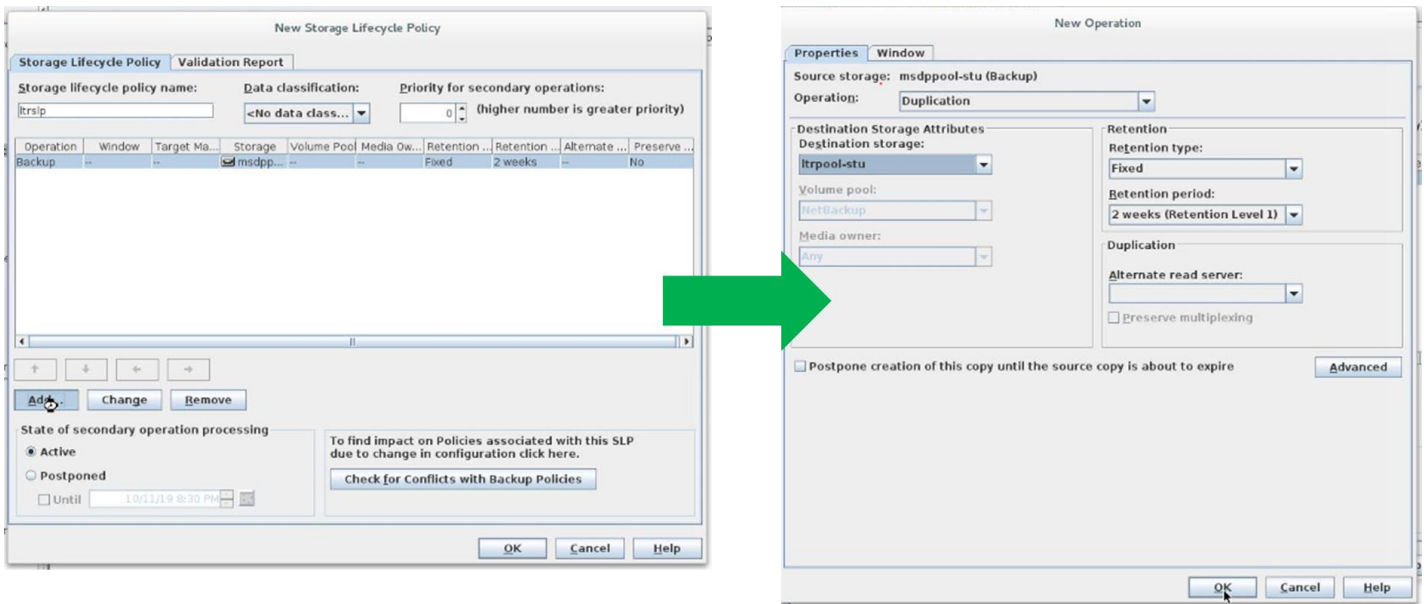
Step 1) Right click right pane and select **New Storage Lifecycle Policy**.



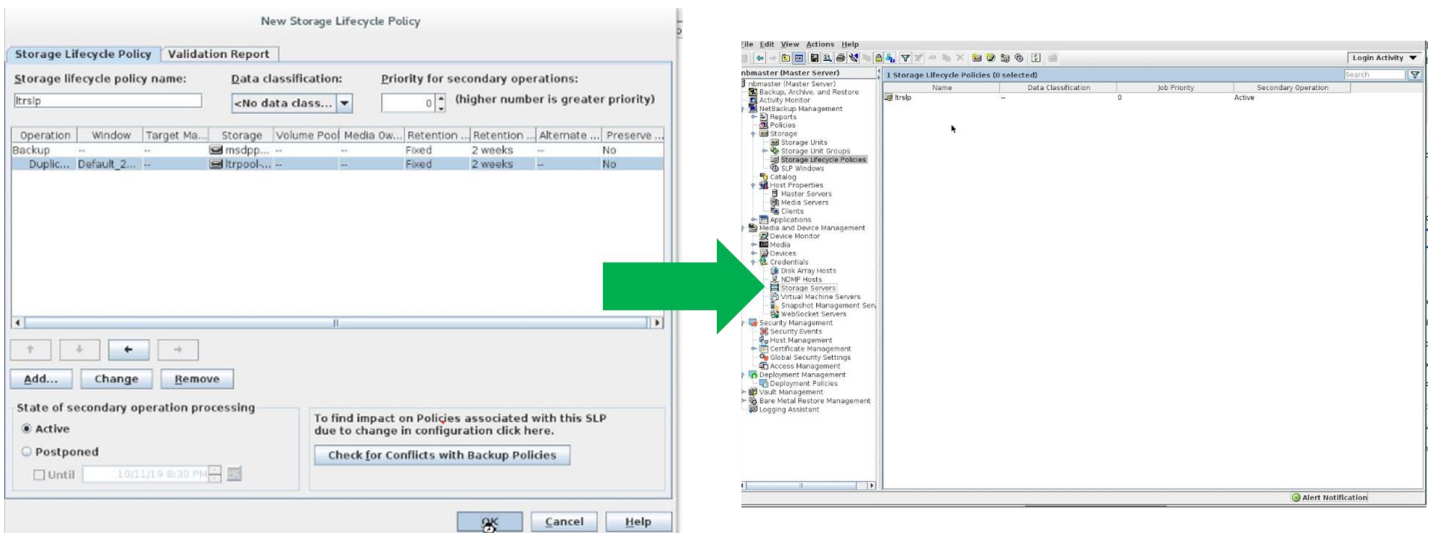
Step 2) Enter storage lifecycle policy **name**. Click **Add** and select Operation to be **Backup** and destination storage to be the MSDP (e.g. msdppool-stu).



Step 3) Click **Add** again. In next pane, select operation to be *Duplication* and destination storage to be cloud storage pool (e.g. *ltrpool-stu*).



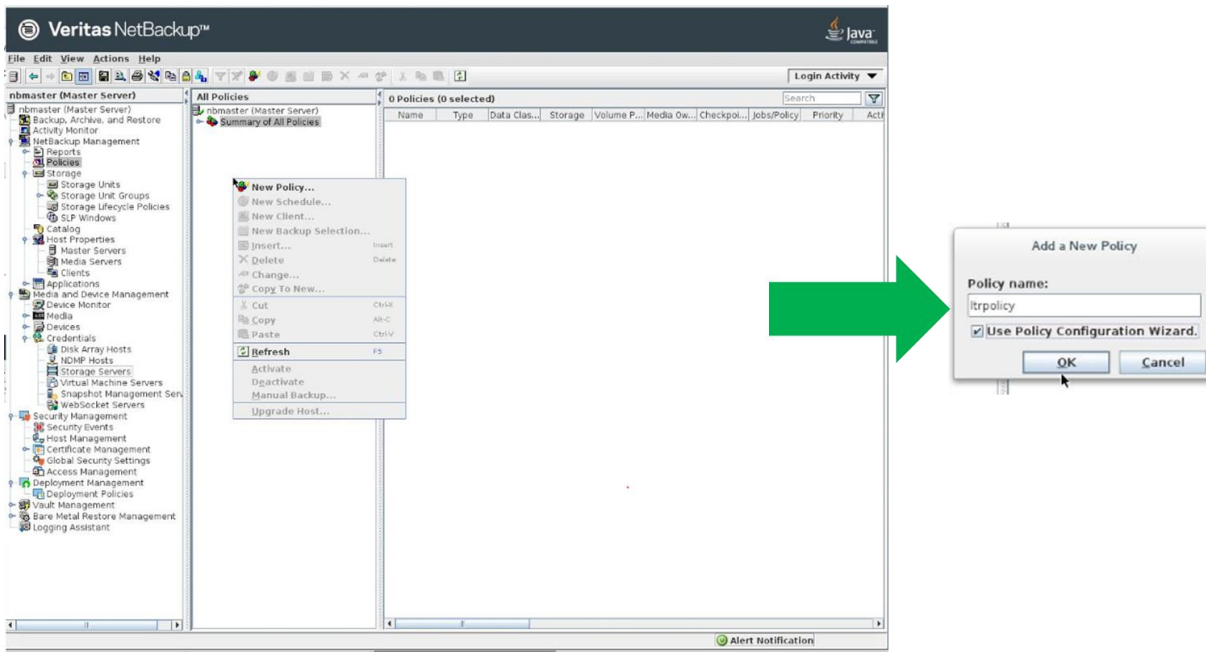
Step 4) Click Ok and the defined SLP is on the right-side pane.



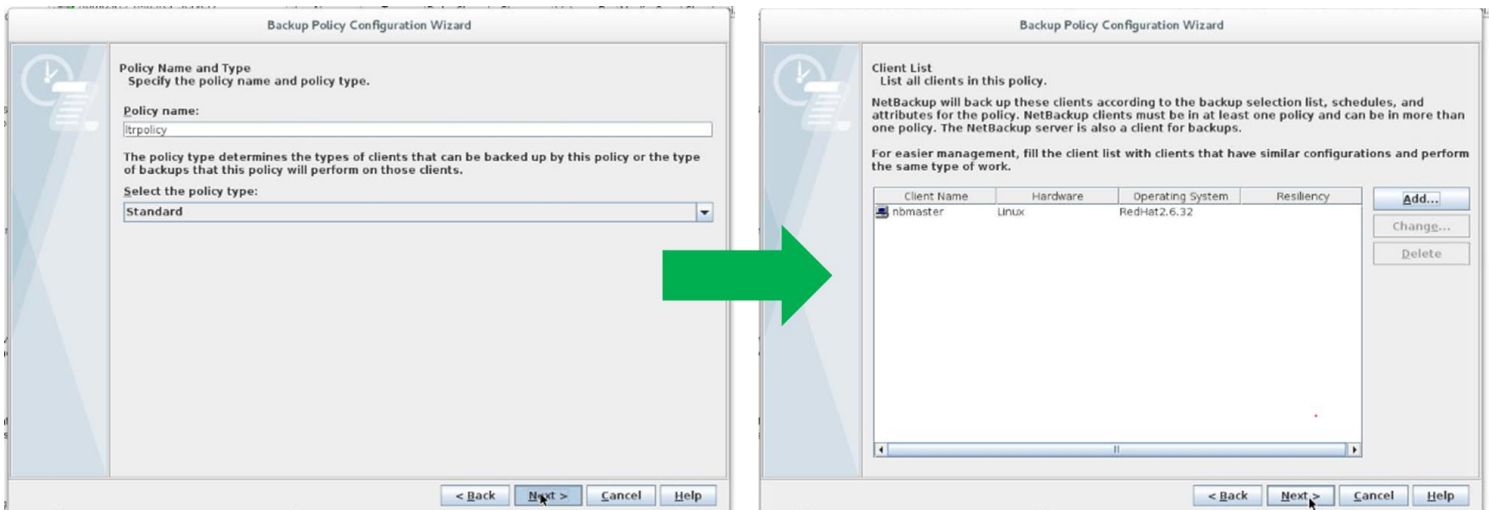
CREATION AND MODIFICATION OF THE BACKUP POLICY TO USE THE SLP

This section describes the next steps which is to create a backup policy and then modify the backup policy attribute to use the SLP defined in previous step.

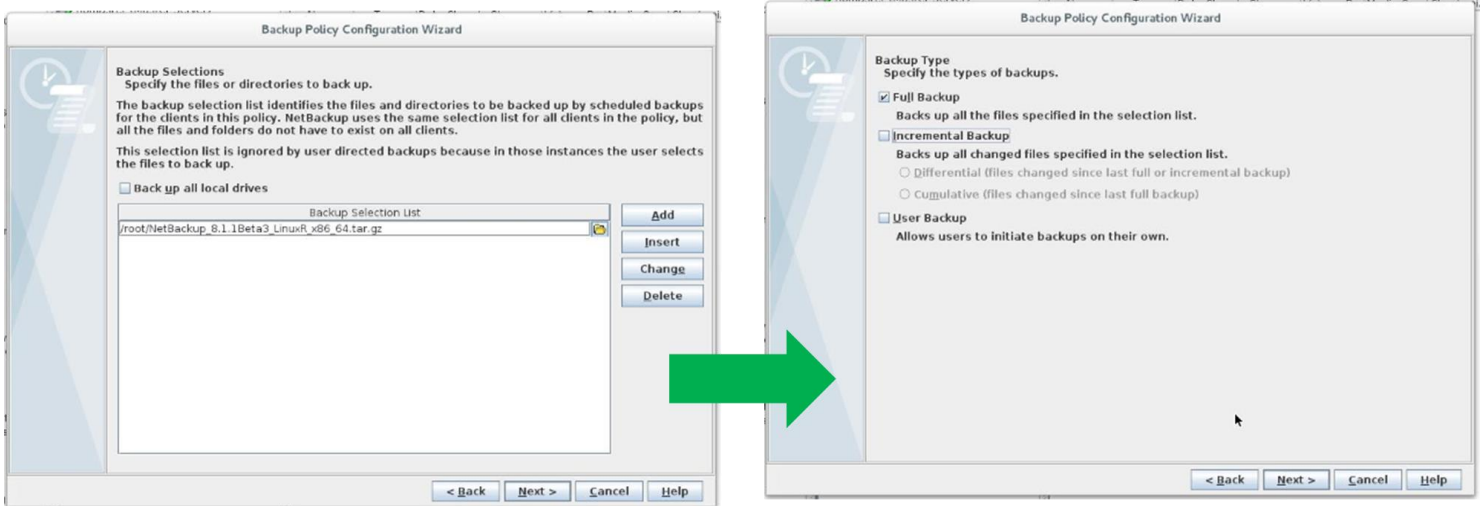
Step 1) Click on **Policies** on left pane and then right click on right pane and select **New Policy**. Enter **name** and place check on **Use Policy Configuration Wizard**. Follow the wizard.



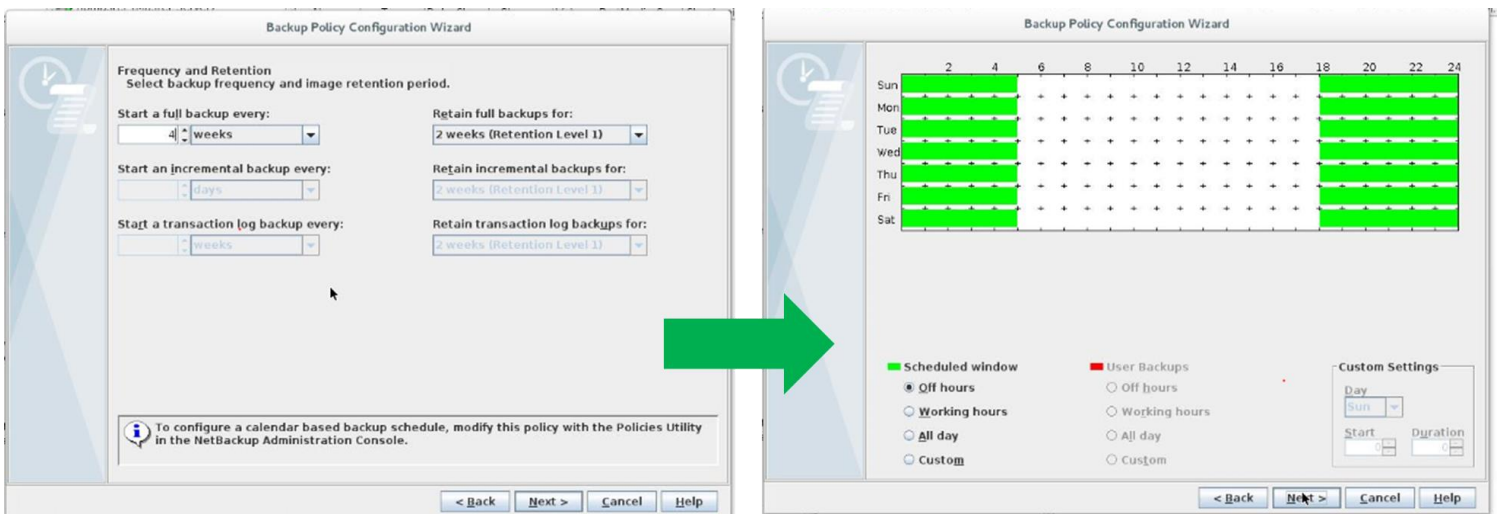
Step 2) Select the policy type. In this example, selecting **Standard**. Enter the **client** information in next screen.



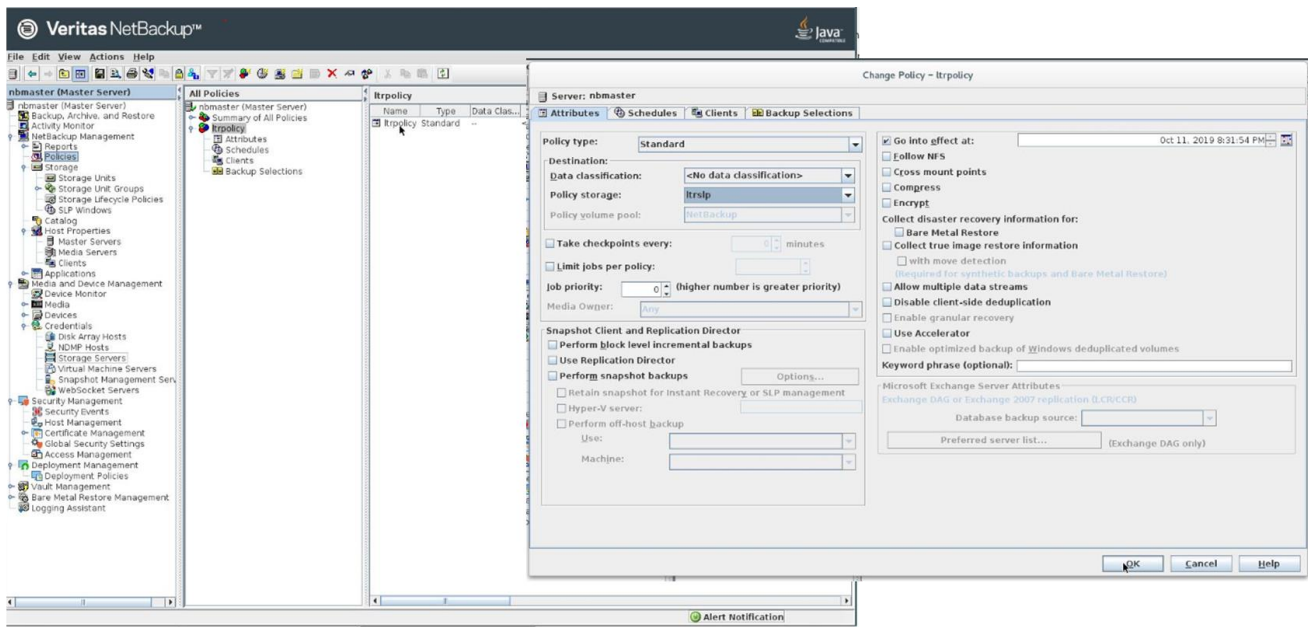
Step 3) Select the **Backup Selection List** and click **Next**. In the next screen, select the **Backup Type**.



Step 4) Enter **Frequency and Retention**. Click **Next** and **Next**.



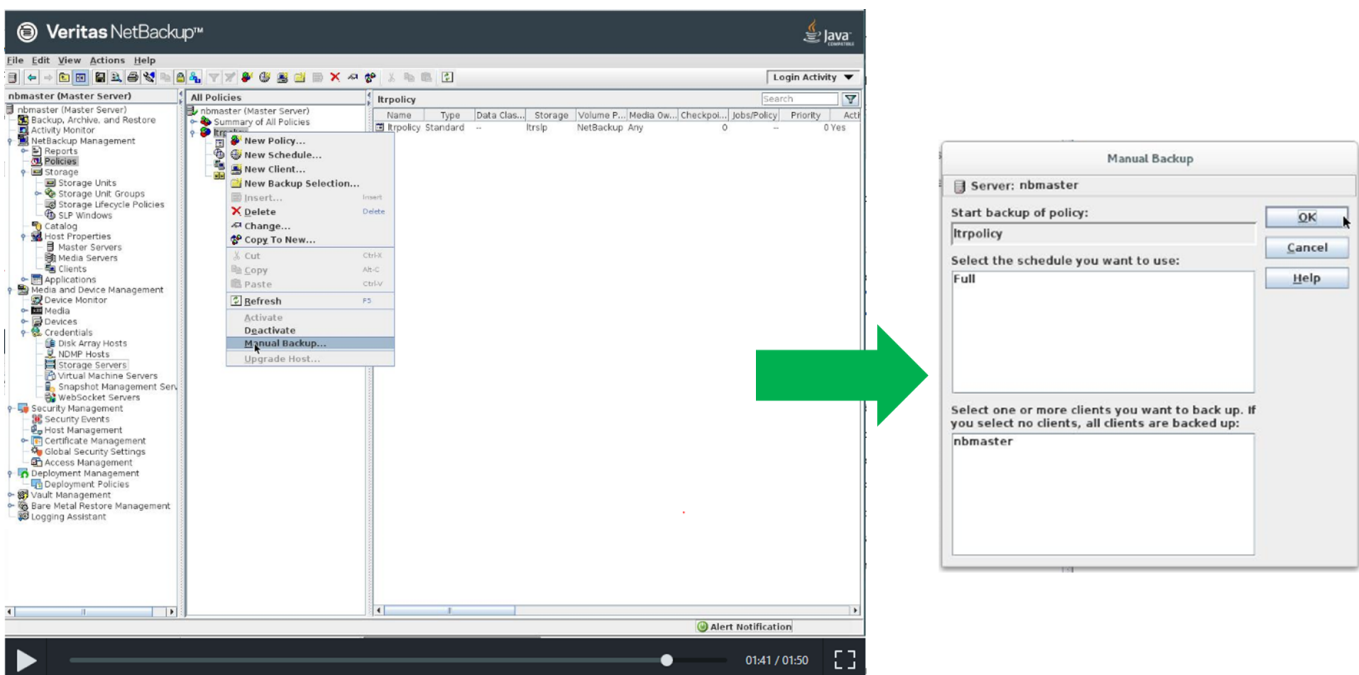
Step 5) Double click on the new created policy and modify the **Attribute Policy Storage** to use the SLP created in previous section.



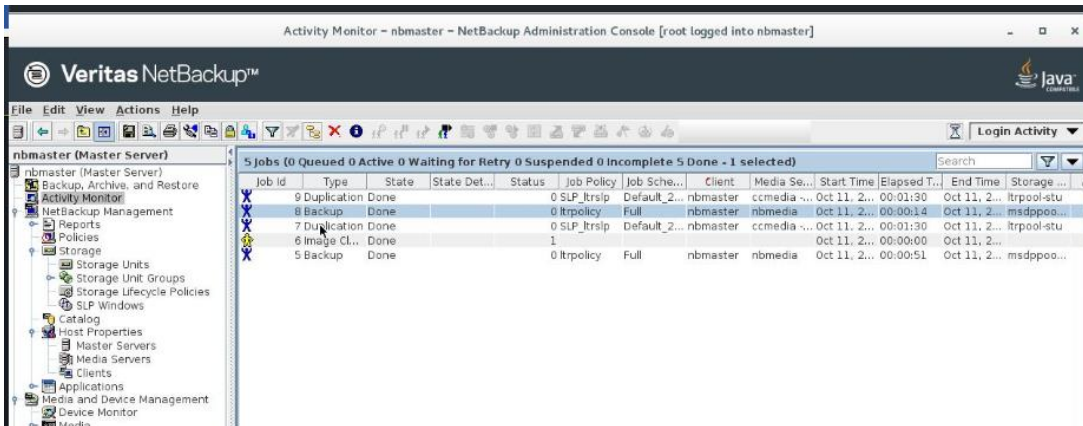
VALIDATION

Validation of the configuration involves running a manual backup which would follow the defined SLP in previous section. Based on the policy, it will first do a backup and place the deduplicated data on MSDP. Afterwards it will duplicate to cloud.

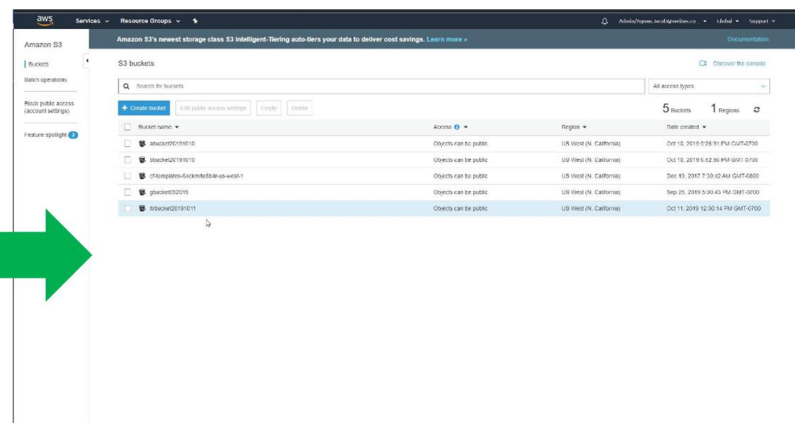
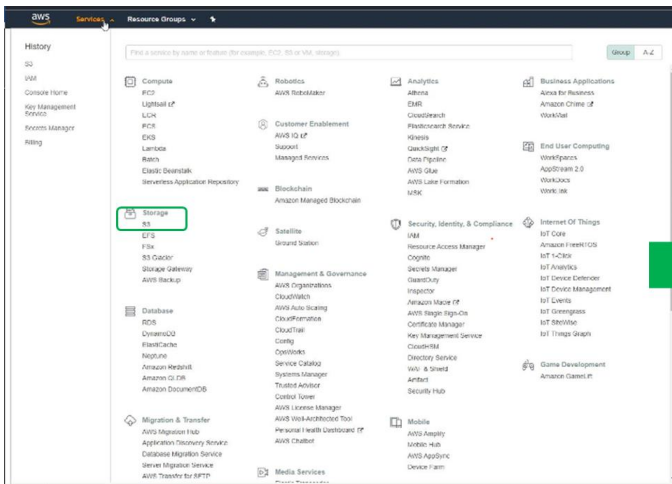
Step 1) Right click on the backup policy and select **Manual Backup**. Click **Ok**.



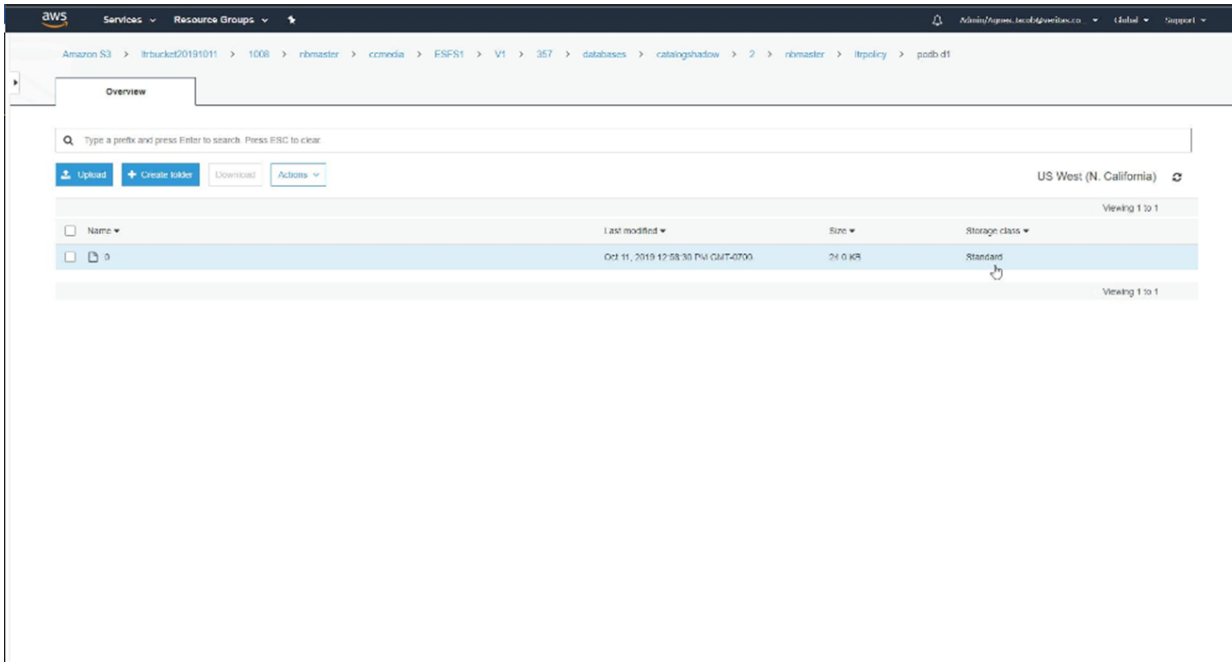
Step 2) Click on **Activity Monitor** on left-side pane. Wait for the backup and duplication job to be done.



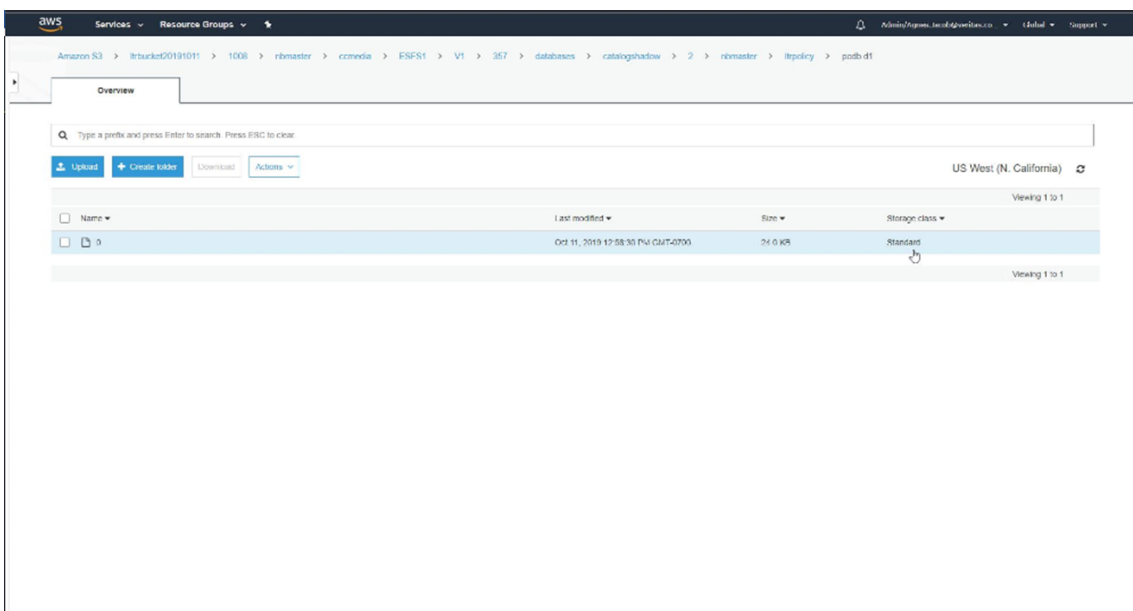
Step 3) Log on to the AWS management console and select **S3** to bring up the bucket views. Traverse the directory structure.



Step 4) As can be seen below, data less than 256K is placed in an AWS S3 **Standard storage class**. As discussed previously, when sending data to Glacier, metadata and data less than 256K is placed in a Standard storage class and the data in Glacier storage class.



Step 5) Look at another location in the bucket and one can observe that 64 MB containers with unique data are placed in **Glacier storage class**.



Step 3) Click on **Classifications** and select the pre-defined classification filters **IT-Codice-Fiscale** and **Intellectual Property**. Then a set of results will be presented on the right pane as shown below. **NOTE:** The preview of items in the GUI is limited to 1000 lines, however, when report is generated it will contain all filtered data.

The screenshot shows the Veritas Information Studio interface. On the left, the 'Classification (2)' panel displays two selected filters: 'IT-Codice-Fiscale' (4.0 KB) and 'Intellectual-Property' (2.9 KB). A green arrow points from this panel to the main content area on the right. The main area shows a table of items with columns: Path, Name, Extension, Size, Last Modified, Content Source, Data Store, and Location. Two items are visible:

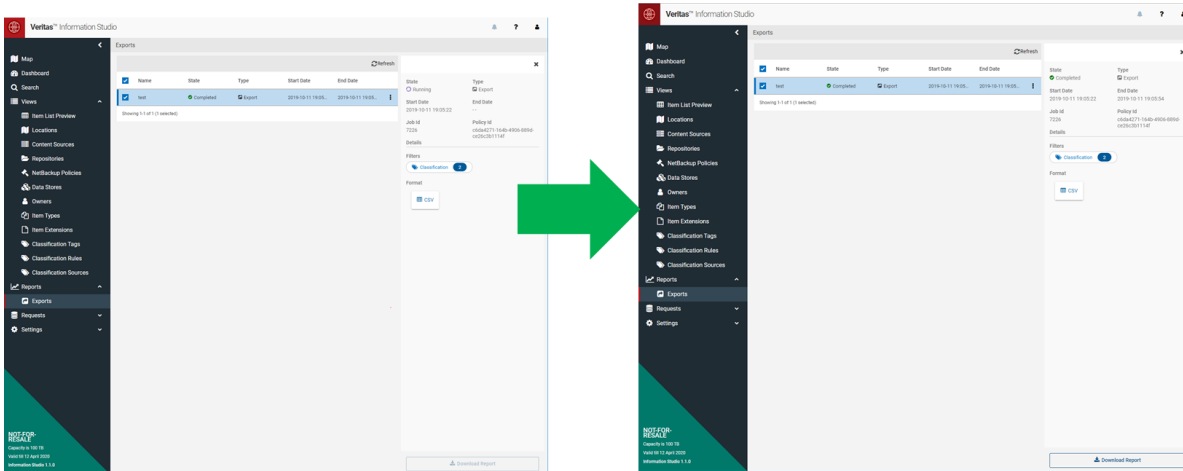
Path	Name	Extension	Size	Last Modified	Content Source	Data Store	Location
\\schmesd13...19\console.sh	console.sh	sh	2.9 KB	2019-09-09 12	schmesd13vm02...	NativeFileSe...	Sant...
\\schmesd13v...italy_sales.rtf	italy_sales.rtf	rtf	4.0 KB	2019-09-09 12	schmesd13vm02...	NativeFileSe...	Sant...

At the top of the main area, there are summary statistics: Total ID (6.84 KB, 2 items), Orphaned (0 items), State (100%, 6.84 KB, 2 items), and Non Business (0 items). A green arrow points from the Classification panel to the main content area.

Step 4) Generate report (*.csv) by clicking on **Actions** at the top and selecting **Export**. Then enter **name** of report and description. Select the scheduler type to be **"Run Now"**.

The screenshot shows the 'Create Job & Schedule' dialog in Veritas Information Studio. The dialog has four steps: Step 1 of 4 (Filters applied), Step 2 of 4 (Name This Report), Step 3 of 4 (Choose Report Format), and Step 4 of 4 (Job Schedule). In Step 2, the 'Report Name' is 'test' and the 'Report Description' is 'test'. In Step 3, the 'Choose Report Format' is 'CSV'. A green arrow points from the main content area to the dialog. At the bottom right, there is a 'Save Report' button.

Step 5) Click on Reports on left-side pane to view the progress. Once done, click on **Download Report**. A *.zip file will be downloaded and extract the file.



Step 6) Extract the file from the zip. A sample view of report is shown below. The file paths are highlighted.

```
[root@nbmaster]# cat testout_o.csv
Name,Owner,Extension,Size,Count,NetBackupPolicies,MasterServer,CreatedTime,ModifiedTime,AccessedTime,Repository,
ContentSource,DataStore,Path,Location,ClassificationTags,ClassifiedTime,PeopleTags,PlacesTags,OrganizationTags,LastNe
rScanTime,LastContentMatchedTime
italy_sales.rtf,Administrators,rtf,4046,1,,,2019-09-09T19:13:23Z,2019-09-09T19:13:39Z,2019-09-
09T19:13:23Z,enterprise_sales_2019,"nbuclient.com",NativeFileServer,"\\nbuclient.com\enterprise_sales_2019\italy_sales.rtf"
,"Santa Clara CA","IT-Codice-Fiscale:'Date of Birth','Italy Codice Fiscale'
",2019-09-11T21:53:23Z,,,,
ip_cidrs.txt,Administrators,txt,321,1,,,2019-09-09T19:14:06Z,2019-09-09T19:14:25Z,2019-09-
09T19:14:06Z,enterprise_sales_2019,"nbuclient.com",NativeFileServer,"\\nbuclient.com\enterprise_sales_2019\ip_cidrs.txt",
Santa Clara CA,"IP-Address:'IPv6 Addresses'
",2019-09-11T21:53:23Z,,,,
console.sh,Administrators,sh,2962,1,,,2019-09-09T19:16:39Z,2019-09-09T19:16:39Z,2019-09-
09T19:16:39Z,enterprise_sales_2019,"nbuclient.com",NativeFileServer,"\\nbuclient.com\enterprise_sales_2019\console.sh",
Santa Clara CA,"Intellectual-Property:'Confidential - Generic','Programming Language Source Code'
",2019-09-11T21:53:23Z,,,,
uk_dl.rtf,Administrators,rtf,1406,1,,,2019-09-09T19:12:38Z,2019-09-09T19:12:52Z,2019-09-
09T19:12:38Z,enterprise_sales_2019,"nbuclient.com",NativeFileServer,"\\nbuclient.com\enterprise_sales_2019\uk_dl.rtf",
Santa Clara CA,"UK-Drivers-License:'U.K. Drivers License Number'
",2019-09-11T21:53:23Z,,,,
_sales.rtf","","o","nbuclient.com","4046","italy_sales.rtf","20190909T191323+0000","b3e73199-d348-11e9-boab-
215a5907467d","20190909T191339+0000","fbc1cc3e-odb7-4ecc-acbf-1676dbcf089c","false"
```


EXTRACT THE FILE PATHS FROM REPORT

Scripts, awk/sed or other tools can be developed to extract the file paths from the report generated in the previous section. Below is a sample script that would extract paths and create file paths that NetBackup command tools can use as inputs.

Step 1) From the *.csv report downloaded in the previous section, extract the file paths. The sample script below will extract the files from this output and put it in a form that it can be entered manually into the "Backup Selection List" of the policy or fed into a script that would create or modify an existing script. **Usage:** `./getfiles.py [csv_file]`

```
[root@nbumaster]# cat getfiles.py

#!/usr/bin/env python
from csv import DictReader
import sys
def read_command_line_arguments():
    if len(sys.argv) != 2:
        print_usage()
        exit()
def print_usage():
    print ("Usage: %s" % (sys.argv[0]) + " [csv_file]" )
    sys.exit()
read_command_line_arguments()
csv_inputfile = (sys.argv[1])

#Extract filenames from Information Studio Report
with open(csv_inputfile) as file:
    csv_reader = DictReader(file)
    for row in csv_reader:

        ##Format the list as per Netbackup exclusion list syntax:

        print("Exclude = " + row['Path'])
        # Use the below line if do not want to add "Exclude" in output
        # print(row['Path'])
```

Step 2) When the script is run, it will extract the file paths as shown below.

```
[root@nbumaster]# ./getfile2.py testout_o.csv
Exclude = \\nbuclient.com\enterprise_sales_2019\console.sh
Exclude = \\nbuclient.com\enterprise_sales_2019\ip_cidrs.txt
Exclude = \\nbuclient.com\enterprise_sales_2019\uk_dl.rtf
Exclude = \\nbuclient.com\enterprise_sales_2019\italy_sales.rtf
```

ENTER THE CLIENT INFORMATION AND FILEPATHS INTO NETBACKUP

The client information and file paths extracted can be manually entered into the backup policy. In this example, the client is a Windows box and the file paths are fed to an exclusion list such that these set of files are not sent to the cloud. An example script to modify an existing policy for inclusion is also provided.

Step 2) Redirect the output of the shell script to a *.txt file to feed into the command “bpsetconfig” to exclude. An example is shown below.

```
[root@nbumaster]# cat win_exclude.txt
Exclude = \\nbuclient.com\enterprise_sales_2019\console.sh
Exclude = \\nbuclient.com\enterprise_sales_2019\ip_cidrs.txt
Exclude = \\nbuclient.com\enterprise_sales_2019\uk_dl.rtf
Exclude = \\nbuclient.com\enterprise_sales_2019\italy_sales.rtf
```

Step 3) Run bpsetconfig on master to set the exclusion and then bpgetconfig to confirm the exclusion has been configured.

```
root@nbumaster]#
/usr/opensv/netbackup/bin/admincmd/bpsetconfig -h nbuclient.com win_exclude.txt

root@nbumaster]# /usr/opensv/netbackup/bin/admincmd/bpgetconfig -M nbuclient.com | grep '^Exclude'
Exclude = \\nbuclient.com\enterprise_sales_2019\console.sh
Exclude = \\nbuclient.com\enterprise_sales_2019\ip_cidrs.txt
Exclude = \\nbuclient.com\enterprise_sales_2019\uk_dl.rtf
Exclude = \\nbuclient.com\enterprise_sales_2019\italy_sales.rtf
```

Step 4) The example script shows how a NetBackup policy can be created to use the report generated from Veritas Information Studio and conduct a backup. The getfiles.py script used in the previous step was modified to NOT add the word “Exclude” and was used to generate an output TEMPFILE (e.g. /root/demo/tmpfile.out) of files to be added to “Backup Selection List” of policy to conduct a backup. The policy created is for one client and a single SMB share and utilizes the NetBackup commands such as bppolicynew, bppllist, bpplsched, bpplinclude, and bpplclients.

```
[root@nbumaster]# cat create_nbu_policy.sh
echo -e "Creating NBU Policy for classified Files using on-prem storage unit(tier)\n"
CIFS_SHARE=enterprise_sales_2019
NEW_POLICY_NAME=demotest1
SCHEDULE_NAME=myschedo1a
NBU_MASTER=nbumaster.com
NBU_CLIENT_HOST=nbuclient
ON_PREM_STU=stu001
CSV_REPORT=/root/demo/testout_o.csv
TEMPFILE=/root/demo/tmpfile.out
export PATH=$PATH:/usr/opensv/netbackup/bin/admincmd

bppolicynew $NEW_POLICY_NAME -M $NBU_MASTER
bppllist $NEW_POLICY_NAME -U
bpplsched $NEW_POLICY_NAME -add $SCHEDULE_NAME -st FULL -residence $ON_PREM_STU -window 82800 3600
bpplsched $NEW_POLICY_NAME -U
bpplinclude $NEW_POLICY_NAME -delete -M $NBU_MASTER -f "C:\enterprise_sales_2019"

/root/demo/getfiles.py $CSV_REPORT > $TEMPFILE

for file in `cat $TEMPFILE`
do
    bpplinclude $NEW_POLICY_NAME -add $file
done

bppllist $NEW_POLICY_NAME -U
bpplclients
bpplclients $NEW_POLICY_NAME -M $NBU_MASTER -add $NBU_CLIENT_HOST Windows-x64 MS-Windows
bppllist $NEW_POLICY_NAME -U |grep HW

echo "Done!!!"
```

Step 5) An output of running this script is shown below.

```
[root@nbumaster]# sh -x create_nbu_policy.sh
+ echo -e 'Creating NBU Policy for classified Files using on-prem storage unit(tier)\n'
Creating NBU Policy for classified Files using on-prem storage unit(tier)

+ CIFS_SHARE=enterprise_sales_2019
+ NEW_POLICY_NAME=demotest1
+ SCHEDULE_NAME=myschedo1a
+ NBU_MASTER=nbumaster.com
+ NBU_CLIENT_HOST=nbuclient.com
+ ON_PREM_STU=stu001
+ CSV_REPORT=/root/demo/testout_o.csv
+ TEMPFILE=/root/demo/tmpfile.out
+ export PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/root/bin:/usr/opensv/netbackup/bin/admincmd
+ PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/root/bin:/usr/opensv/netbackup/bin/admincmd
+ bppolicynew demotest1 -M nbumaster.com
+ bpspllist demotest1 -U
-----

Policy Name:    demotest1

Policy Type:    Standard
Active:         yes
Effective date: 10/15/2019 16:33:33
Client Compress: no
Follow NFS Mounts: no
Cross Mount Points: no
Collect TIR info: no
Block Incremental: no
Mult. Data Streams: no
Client Encrypt: no
Checkpoint:     no
Policy Priority: 0
Max Jobs/Policy: Unlimited
Disaster Recovery: 0
Collect BMR info: no
Residence:      (specific storage unit not required)
Volume Pool:    NetBackup
Server Group:   *ANY*
```

Keyword: (none specified)
Data Classification: -
Residence is Storage Lifecycle Policy: no
Application Discovery: no
Discovery Lifetime: 0 seconds
ASC Application and attributes: (none defined)

Granular Restore Info: no
Ignore Client Direct: no
Use Accelerator: no

Clients: (none defined)

Include: (none defined)

Schedule: (none defined)
+ bpplsched demotest1 -add mysched01a -st FULL -residence stu001 -window 82800 3600
+ bpplsched demotest1 -U

Schedule: mysched01a
Type: Full Backup
Frequency: every 7 days
Excluded Dates-----
No specific exclude dates entered
No exclude days of week entered
Synthetic: 0
Checksum Change Detection: 0
PFI Recovery: 0
Maximum MPX: 1
Retention Level: 1 (2 weeks)
Number Copies: 1
Fail on Error: 0
Residence: stu001
Volume Pool: (same as policy volume pool)
Server Group: (same as specified for policy)
Residence is Storage Lifecycle Policy: 0
Daily Windows:

```

Sunday 23:00:00 --> Sunday 24:00:00
Monday 23:00:00 --> Monday 24:00:00
Tuesday 23:00:00 --> Tuesday 24:00:00
Wednesday 23:00:00 --> Wednesday 24:00:00
Thursday 23:00:00 --> Thursday 24:00:00
Friday 23:00:00 --> Friday 24:00:00
Saturday 23:00:00 --> Saturday 24:00:00
+ bpplinclude demotest1 -delete -M nbumaster.com -f 'C:\enterprise_sales_2019'
+ /root/demo/getfiles.py /root/demo/testout_o.csv
++ cat /root/demo/tmpfile.out
+ for file in `cat $TEMPFILE`
+ bpplinclude demotest1 -add '\\nbuclient.com\enterprise_sales_2019\italy_sales.rtf'
+ for file in `cat $TEMPFILE`
+ bpplinclude demotest1 -add '\\nbuclient.com\enterprise_sales_2019\ip_cidrs.txt'
+ for file in `cat $TEMPFILE`
+ bpplinclude demotest1 -add '\\nbuclient.com\enterprise_sales_2019\console.sh'
+ for file in `cat $TEMPFILE`
+ bpplinclude demotest1 -add '\\nbuclient.com\enterprise_sales_2019\uk_dl.rtf'
+ bppllist demotest1 -U
-----

```

Policy Name: demotest1

Policy Type: Standard

Active: yes

Effective date: 10/15/2019 16:33:33

Client Compress: no

Follow NFS Mounts: no

Cross Mount Points: no

Collect TIR info: no

Block Incremental: no

Mult. Data Streams: no

Client Encrypt: no

Checkpoint: no

Policy Priority: 0

Max Jobs/Policy: Unlimited

Disaster Recovery: 0

Collect BMR info: no

Residence: (specific storage unit not required)
 Volume Pool: NetBackup
 Server Group: *ANY*
 Keyword: (none specified)
 Data Classification: -
 Residence is Storage Lifecycle Policy: no
 Application Discovery: no
 Discovery Lifetime: 0 seconds
 ASC Application and attributes: (none defined)

Granular Restore Info: no
 Ignore Client Direct: no
 Use Accelerator: no

Clients: (none defined)

Include: \\nbmaster.com\enterprise_sales_2019\italy_sales.rtf
 \\nbmaster.com\enterprise_sales_2019\ip_cidrs.txt
 \\nbmaster.com\enterprise_sales_2019\console.sh
 \\nbmaster.com\enterprise_sales_2019\uk_dl.rtf

Schedule: myschedo1a
 Type: Full Backup
 Frequency: every 7 days
 Excluded Dates-----
 No specific exclude dates entered
 No exclude days of week entered

Synthetic: 0
 Checksum Change Detection: 0
 PFI Recovery: 0
 Maximum MPX: 1
 Retention Level: 1 (2 weeks)
 Number Copies: 1
 Fail on Error: 0
 Residence: stu001
 Volume Pool: (same as policy volume pool)
 Server Group: (same as specified for policy)
 Residence is Storage Lifecycle Policy: 0

Daily Windows:

```
Sunday 23:00:00 --> Sunday 24:00:00
Monday 23:00:00 --> Monday 24:00:00
Tuesday 23:00:00 --> Tuesday 24:00:00
Wednesday 23:00:00 --> Wednesday 24:00:00
Thursday 23:00:00 --> Thursday 24:00:00
Friday 23:00:00 --> Friday 24:00:00
Saturday 23:00:00 --> Saturday 24:00:00
```

+ bplclients

Hardware OS Client

Windows-x64 Windows nbuclient.com

Linux RedHat2.6.32 nbumaster.com

+ bplclients demotest1 -M nbumaster.com -add nbuclient Windows-x64 MS-Windows

+ bpllist demotest1 -U

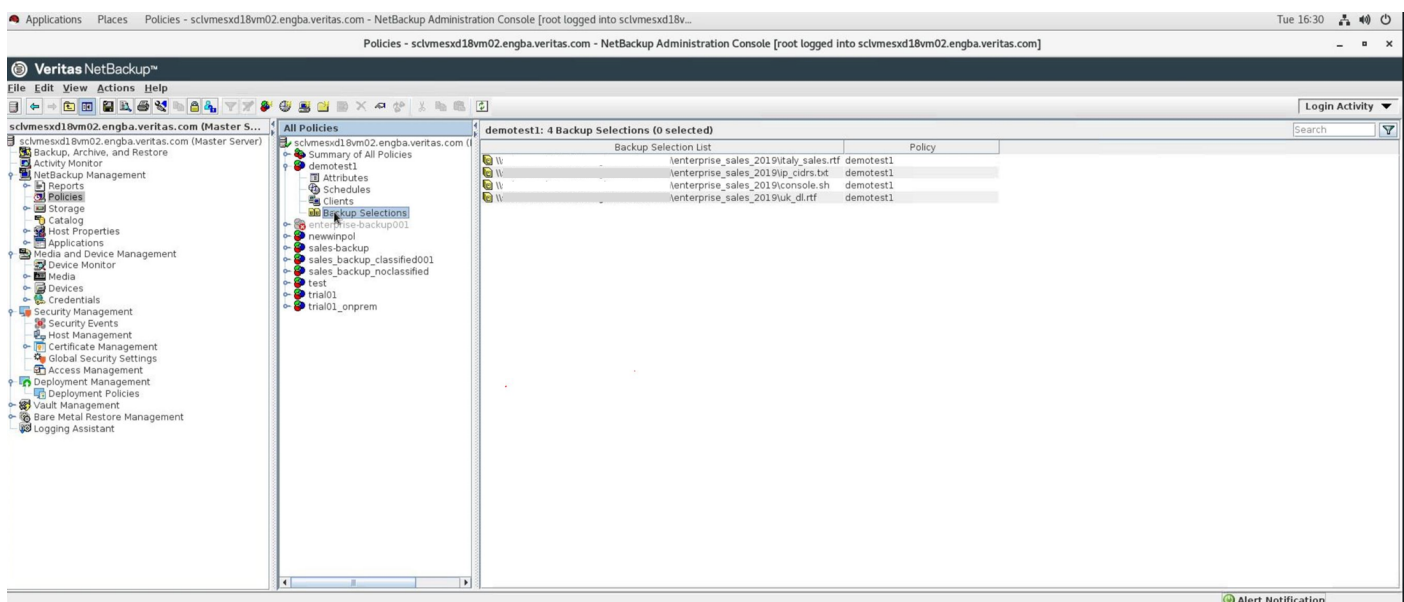
+ grep HW

HW/OS/Client: Windows-x64 MS-Windows nbuclient

+ echo 'Done!!!'

Done!!

A screenshot of what the created policy looks like in NetBackup administration console.



DISCLAIMER

THIS PUBLICATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID. VERITAS TECHNOLOGIES LLC SHALL NOT BE LIABLE FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH THE FURNISHING, PERFORMANCE, OR USE OF THIS PUBLICATION. THE INFORMATION CONTAINED HEREIN IS SUBJECT TO CHANGE WITHOUT NOTICE.

No part of the contents of this book may be reproduced or transmitted in any form or by any means without the written permission of the publisher.

ABOUT VERITAS TECHNOLOGIES LLC

Veritas Technologies empowers businesses of all sizes to discover the truth in information—their most important digital asset. Using the Veritas platform, customers can accelerate their digital transformation and solve pressing IT and business challenges including multi-cloud data management, data protection, storage optimization, compliance readiness and workload portability—with no cloud vendor lock-in. Eighty-six percent of Fortune 500 companies rely on Veritas today to reveal data insights that drive competitive advantage. Learn more at <http://www.veritas.com/> or follow us on Twitter at [@veritastechllc](https://twitter.com/veritastechllc).

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

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

VERITAS[™]
The truth in information.