BACKUP TO THE CLOUD WITH DELL EMC DATA PROTECTION SUITE & CLOUDBOOST

Best Practices

ABSTRACT

This white paper discusses features, planning considerations, options, and best practice recommendations for installing and using Dell EMC CloudBoost as part of an overall data protection solution.

November, 2016

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Data protection is evolving as customers increasingly look to move data to public, private or hybrid cloud. Dell EMC® CloudBoost® enables secure, efficient long-term retention of backups to the cloud to eliminate the risk of tape, reduce TCO, and increase IT and business agility. Customers routinely use tape for long-term retention (LTR) of backups. Tape automation, transport, and storage are expensive and ongoing, while the medium itself is prone to data loss or corruption.

CloudBoost cloud-enables Dell EMC Data Protection Suite® and Symantec/Veritas NetBackup. Available in both virtual and physical editions with several local data cache options to suit any environment, CloudBoost requires no in-cloud infrastructure to extend protection software to the customer's chosen public or private cloud. The result: secure, cost-effective, high-performance LTR backed by inherently durable object storage.

This white paper will discuss the features, planning considerations and setup options and provide some best practice recommendations for installing and using Dell EMC CloudBoost for Data Protection.

AUDIENCE

This white paper is intended for customers, prospective customers, Dell EMC partners, Dell EMC System Engineers, and other technical planners or installers who are interested in understanding the features, options, and best practices when considering or deploying Dell EMC CloudBoost for data protection.

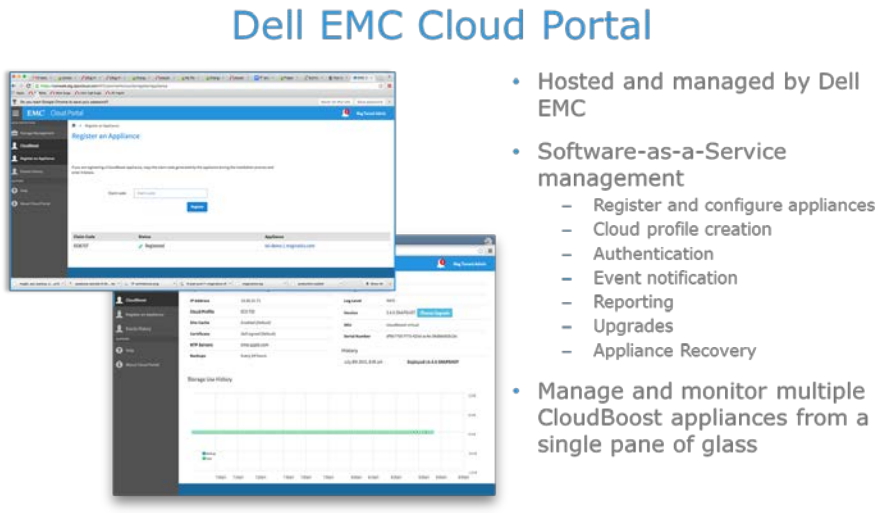
SOLUTION OVERVIEW

Most IT organizations must retain periodic backups for several years. Though long-term retention (LTR) backup copies are infrequently accessed, IT must be able to retrieve them on demand as either full or incremental restores. At the same time, IT is being asked to eliminate the myriad risks associated with tape-based solutions and to control the substantial operating costs imposed by tape LTR.

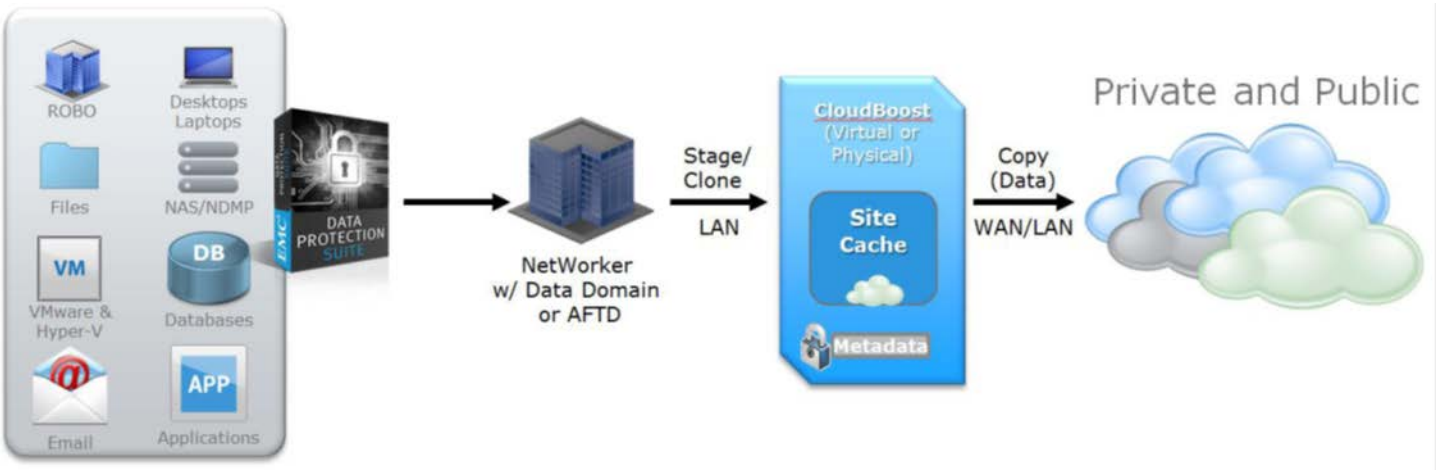
CloudBoost enables users of Dell EMC Data Protection Suite to eliminate the risk and overhead of tape and leverage the economics and agility of cloud by replacing tape with public, private, or hybrid cloud storage for long-term retention of backups. The combination of Dell EMC NetWorker®, Dell CloudBoost, Dell EMC Data Domain®, and an enterprise-grade object store such as Dell EMC Elastic Cloud Storage (ECS)® delivers an integrated, industry-leading solution for all of your operational and LTR backup requirements. With CloudBoost, long-term retention of backups is secure, efficient, high performance, and cost-effective.

Environments that rely on Wide Area Network connectivity often suffer from bandwidth, latency, or reliability limitations that can make cloud tiering seem unworkable, whether the target is a public cloud or centrally-located private cloud. CloudBoost's source-side deduplication, compression, and WAN optimization boost performance and throughput while reducing the consumption and cost of network bandwidth and cloud capacity. A local data cache of up 32 terabytes further speeds backup and restore LTR operations while enabling LAN access to cache-resident data even when the WAN is down.

CloudBoost is managed through the Web-based Dell EMC Cloud Portal, a secure, single-pane-of-glass console. With Cloud Portal, administrators can effortlessly manage and monitor one or many CloudBoost instances. Dell EMC Cloud Portal simplifies and centralizes all administrative tasks, including initial CloudBoost registration and cloud profile creation, reporting and alerts, CloudBoost disaster recovery, and upgrade management. At the same time, administrators continue to use their native Data Protection Suite management console to schedule backups, set retention policies, perform recoveries, and perform all other aspects of the backup and restore process.



In addition to improving performance and scalability, CloudBoost's split-plane architecture separates data from metadata, enabling native support for a wide range of object stores. Private clouds supported by CloudBoost include Dell EMC Elastic Cloud Storage (ECS), Dell EMC Atmos, and OpenStack Swift. Public clouds supported by CloudBoost include AT&T Synaptic, Amazon Web Services S3, Google Cloud Storage (including Nearline), and Microsoft Azure.



Dell EMC Data Protection Suite capacity license holders receive the virtual edition of CloudBoost with 2TB local data cache at no additional cost, so Data Protection Suite customers can begin to enjoy the benefits of cloud-enablement out of the box. This includes the ability to use DP Search to query simultaneously across backups located in the cloud as well as in local protection storage. Backup administrators continue to use their native protection software console to manage backup and restore of LTR copies in the cloud.

CHOOSING THE APPLIANCE

There are a number of important factors to consider when planning the deployment of Dell EMC CloudBoost. Understanding these options will help you select the best CloudBoost appliance form factor for your particular needs.

APPLIANCE FORM FACTORS

Virtual and physical CloudBoost appliances differ only in cache size. All can support up to 6PB of logical capacity in the cloud. Note that logical capacity is different from Front-end TB (FETB) capacity. If a customer has 100TB of FETB, after 10 clones the CloudBoost appliance would be managing 1PB of logical capacity. Valid form factors of the virtual and physical CloudBoost appliances are:

- **Virtual appliance form factors**
    - 2TB cache
    - 6TB cache
- **Physical appliance form factors**
    - 10TB cache
    - 32TB cache

BANDWIDTH TO THE OBJECT STORE

Bandwidth to the object store is the primary factor that dictates the choice of a particular CloudBoost model. Customers with strong connectivity to the object store will not benefit from cache and are better off going directly to the object store. For customers connecting to an on-premises ECS system, we recommend choosing the smallest virtual appliance and disabling cache. If a customer prefers a physical form factor, choose the smallest physical appliance and turn cache off.

Customers with medium connectivity (200 Mbps - 400 Mbps) can still benefit from caching if they select the largest cache size. If they choose any other appliance with a smaller cache, we would recommend going direct to the object store with caching disabled. These three scenarios are further illustrated in the chart below.

Bandwidth to Object Store	Appliance Recommendations
Strong Connectivity >400Mbps	Connect directly to the object store with appliance cache disabled. Generally Customers using an on-premise ECS will have strong connectivity.
Medium Connectivity 200-400Mbps	Enable appliance cache. Select the 32TB physical appliance
Weak Connectivity <200Mbps	Enable appliance cache. Choose appliance cache size based on backup dataset size. Generally public object store should benefit from cache.

VIRTUAL OR PHYSICAL APPLIANCE

As stated earlier, both the virtual and physical CloudBoost appliances support up to 6PB of logical capacity in the cloud. Dell EMC provides customers the choice of using virtual or physical appliances depending on which fits their infrastructure preferences. Please note that virtual appliances require VMware® ESX 5.0 or higher. The choices and our recommendations are summarized as follows:

- **Customers should choose a virtual appliance when**
    - Customer does not want to manage hardware and prefers a software only footprint
    - Customer wants the flexibility of starting with a small logical capacity or cache size & expand as needs grow
- **Customers should choose a physical appliance when**
    - Customer does not have access to a VMware ESX host
    - Customer does not have the disk resources to size the VM
    - Customer prefers not to deal with sizing the virtual appliance

VIRTUAL APPLIANCE RESOURCE SIZING

The CloudBoost virtual appliance sizing recommendations depend on whether you are enabling appliance cache or not. The sizing recommendations are illustrated in detail in the following two examples:

- **With appliance cache enabled**
    - 16 vCPUs, 64GB memory
    - Disk space
        - Metadata
            - The amount of space provisioned for metadata directly impacts the logical capacity addressable by the appliance. Ratio of metadata space to logical capacity is 1:4000 – example: 100GB of metadata allows the appliance to address 400TB of logical capacity.
            - 1.5TB metadata space needed to address the maximum logical capacity of 6PB.

- Minimum of 100GB metadata is required.

- Cache

  - Minimum of 192GB/64GB required for the 6TB/2TB virtual appliance, respectively. Cache should be sized based on dataset size so backups can benefit from caching.

- **With appliance cache disabled**

  - 8 vCPUs, 32GB memory

  - Disk space

    - Metadata

      - The amount of space provisioned for metadata directly impacts the logical capacity addressable by the appliance. Ratio of metadata space to logical capacity is 1:4000 – example: 100GB of metadata allows the appliance to address 400TB of logical capacity.

      - 1.5TB metadata space needed to address the maximum logical capacity of 6PB.

      - Minimum of 100GB metadata is required.

GROWING VIRTUAL MACHINE RESOURCES

The CloudBoost virtual appliance can start small and grow as needed. Both logical capacity and appliance cache sizes can be increased. Logical capacity managed by the appliance is determined by the amount of metadata space. The ratio of metadata space to logical capacity is 1:4000. For example, 100GB of metadata allows the appliance to address 400TB of logical capacity. Maximum metadata size is 1.5TB, which supports 6PB of logical capacity, the most that can be managed by a single CloudBoost appliance instance. Metadata disk can be resized from the Dell EMC Cloud Portal. Only expansion of the existing disk is allowed, not addition of new disks. The appliance will need to be rebooted after resizing.

The amount of appliance cache can also be increased by adding more vdisks from vCenter. Existing disks cannot be expanded. Valid numbers of vDisks for site cache are: 1, 2, 4, 8, 16, and 32. All disks used for cache must also be the same size. For the 6TB VM, ensure you add a minimum size of 192GB per disk, so you can scale to the maximum of 6TB (192GB x 32 disks). For the 2TB VM, ensure you add a minimum size of 64GB per disk, so you can scale to the maximum of 2TB (64GB x 32 disks). The CloudBoost appliance requires a reboot after adding vdisks to increase cache.

SITE CACHE CONSIDERATIONS

CloudBoost supports multiple cache sizes as explained earlier. A local cache within the appliance can improve backup performance in cases where the appliance has low bandwidth to the target object store. The cache can support ingest performance of 25-50 MB/s. For sites that have a strong connection of greater than 50 MB/s (400 Mbps) to the object store, we recommend disabling cache and connecting directly to the object store. This should generally be the case when using a private cloud solution with an on-premises object store such as ECS, Atmos or OpenStack Swift. However, cache would be beneficial when connecting to public object stores over a thin WAN pipe such as T1-T4, OC-1 or OC-3 connections. It would allow faster ingest of backups and could also improve restore times when data is in cache. See section on “Choosing the Appliance” for guidelines on how to select the right appliance based on bandwidth and other needs.

CloudBoost virtual appliance allows provisioning 1, 2, 4, 8, 16, or 32 disks for site cache. We recommend adding at least 4-8 disks to get sufficient performance from cache.

CREATING A CLOUD PROFILE

Setting up a Cloud Profile is required before CloudBoost can send objects to the object store. This is done through the Dell EMC Cloud Portal. Currently CloudBoost supports seven private and public cloud providers. To set up a new Cloud Profile, you will need to provide the Access Key, the Full Token ID/Secret Access, the Storage Region (if supported), and the Endpoint (if supported). These are provided by the cloud provider.

REGISTRATION AND CONFIGURATION

The CloudBoost appliance is managed using a Cloud-based portal. The CloudBoost appliance must first be registered with the Dell EMC Cloud Portal using a claim code obtained through the CloudBoost CLI. After it has been registered, the appliance can then be configured and managed. Configuration information includes the appliance name, a cloud profile with an optional bucket name, enabling or disabling site cache, SSL Certificate, an NTP server (recommended), and backup frequency. "Backup frequency" here refers to backups of metadata that resides on the CloudBoost appliance. All metadata including dedupe hashes, encryption keys and pointers to objects are backed up periodically to the same object store as the data. This enables recovery to a second appliance in the event of a failure of the primary appliance. All appliance backups are encrypted before sending to the target object store.

SECURITY CONSIDERATIONS

CLOUDBOOST SECURITY FEATURES

CloudBoost delivers enterprise-grade security even when data is stored or transferred outside your firewall. CloudBoost segments each file into many small "chunks" and encrypts each chunk with its own independent AES-256 key. Data remains chunked and encrypted in-flight and at-rest, and all data transfers occur over TLS. CloudBoost allows enterprises to seamlessly leverage their existing systems, polices and tools for data protection while augmenting the traditional enterprise environment with features and capabilities that ensure end-to-end security in an era of cloud computing and unfettered mobility. The pillars that support this architecture are explained in more detail in the following sections.

PHYSICAL AND NETWORK SECURITY

CloudBoost stores all metadata including encryption keys primarily on the appliance within the confines of the customer's data center. However to avoid the appliance becoming a single point of failure, the metadata is backed up to the same target object store as the data periodically to allow recovery to a second appliance in the event of a failure of the primary appliance. Backups of the metadata are encrypted before sending to the object store, and are stored in a completely different container from the data itself. The key for metadata backups is stored securely within the Cloud Portal and automatically applied during recovery/DR of the appliance, so customers do not have to deal with the additional step of locating and providing the key in the midst of trying to recover from the failure of an appliance. The Cloud Portal does not allow initiating recovery of the appliance unless it detects that the primary appliance has failed, and as an additional security step will email account holders to inform them of a DR event.

FINE-GRAINED ENCRYPTION

Upon entry into the system, every file is broken into many small, variable sized "chunks," each of which is individually encrypted with its own independent AES-256 key. In addition, all data and metadata transfers take place over HTTPS. With CloudBoost, there is no master key that can compromise all data in the system in the event of a theft or loss. Instead, each plaintext chunk is individually encrypted with its own independent AES-256 key. This key is applied to the raw chunk using the AES/CBC-256 cipher, generating the encrypted chunks. The one-way cryptographic hash function is then applied again, this time to the encrypted chunk, to generate the "chunk reference." The encryption process is best illustrated using the following example; i.e., that of a file being backed up by NetWorker and sent to the cloud for long-term retention. The steps are as follows:

1. **File Segmentation** – The file is segmented into many small pieces or "chunks" of variable size. Besides enhancing performance by enabling granular inline deduplication and allowing threaded transfers of files between the appliance and object store, file "chunking" improves security by forcing an attacker to compromise not just a single key, but many keys, in order to decrypt a single file. Chunk sizes are determined algorithmically to optimize deduplication efficiency.

2. **Encryption** – With CloudBoost, there is no master key that can compromise all data in the system in the event of a theft or loss. Instead, each plaintext chunk is individually encrypted with its own independent AES-256 key. This key is applied to the raw chunk using the AES/CBC-256 cipher, generating the encrypted chunks. The one-way cryptographic hash function is then applied again, this time to the encrypted chunk, to generate the "chunk reference."

3. **Data Storage** – Each <chunk-reference, encrypted chunk> tuple represents a <key, value> pair which is stored in the local encrypted cache of the CloudBoost Agent to enable deduplication and to enhance performance (if the same key-value pair is already resident in the cache, this new copy is ignored). The chunk-reference is also transmitted along with other metadata (e.g., the associated file name) over HTTPS to CloudBoost, which persists the information and uses it to create a "chunk map" that maps file names to their constituent encrypted chunks. Both Agent and Server are located in the CloudBoost appliance. Upon receipt of the chunk-reference, CloudBoost computes a location in the cloud to which the associated encrypted chunk will be stored and generates a "pointer" to that location in the form of a cryptographically signed, time-limited Universal

Resource Identifier (URI) which is sent over HTTPS. Communicating over a secure channel with the object store using the latter's native REST interface, CloudBoost transfers the encrypted chunk to the cloud with a simple PUT command. The object store validates the URI by checking the signature and expiration time before storing the encrypted chunk. If the time limit is exceeded, CloudBoost must generate a new URI.

DATA INTEGRITY

A robust out-of-band data verification process protects against malicious clients or network failures (e.g., HTTP proxy errors) causing data corruption. Chunks written to the object store are quarantined prior to being verified. CloudBoost mechanisms ensure that all data entering the system is stored accurately, reliably and consistently.

CLOUDBOOST SECURITY RECOMMENDATIONS

As detailed above, the concept of data security and integrity is inherent in the CloudBoost design. Data is chunked, encrypted in-flight and at-rest, transferred across secure connections and additionally protected in-cloud behind signed URIs. Security should never be assumed with any networked solution, however, and we recommend adhering to best practices for security to protect your deployment.

If the following ports are not configured before you first configure the CloudBoost appliance, it will be necessary to reboot the CloudBoost appliance.

It is not recommended to route outbound http traffic from the CloudBoost appliance through a proxy. This can create a performance bottleneck. In environments where outbound http traffic is restricted, it is recommended to create an exception for the appliance in the firewall after consultation with the IT security team.

Firewall port requirements			
From	To	TCP Port	Description
Administrator workstation	Storage node on the CloudBoost appliance	22	SSH for maintenance and troubleshooting
Storage node on the CloudBoost appliance	Cloud storage (public or private)	443	HTTPS to access object store (if supported)
Storage node on the CloudBoost appliance	EMC Cloud Portal	443	HTTPS to EMC Cloud Portal and Cloud Portal Services/APIs
CloudBoost client	Storage node on the CloudBoost appliance	443	When Veritas NetBackup is deployed with CloudBoost, this is necessary when the CloudBoost client is on the Windows media server.
Administrator workstation	Storage node on the CloudBoost appliance	4444	HTTPS to EMC Cloud Portal/API
Storage node on the CloudBoost appliance	site cache servers	8443	Incoming HTTPS port for all data read and write traffic
Storage node on the CloudBoost appliance	ESRS gateway	9443	Communication from CloudBoost appliance to the EMC Secure Remote Services gateway

CloudBoost utilizes SSL. To support rapid testing CloudBoost supports the use of self-signed certificates. Self-signed SSL certificates are less secure than those signed by a trusted Certificate Authority and should only be used for test deployments. In production environments, you should use a wildcard SSL certificate signed by a trusted Certificate Authority. Wildcard certificates are public key certificates that can be used with multiple sub-domains. Only a single level of sub-domain matching is supported.

DISASTER RECOVERY

If a CloudBoost appliance fails, you can recover to another CloudBoost appliance using automatically-taken backups (every 12 hours) that are stored in the same object store as the data. As stated earlier, this DR protection removes the single-point-of-failure risk from your deployments.

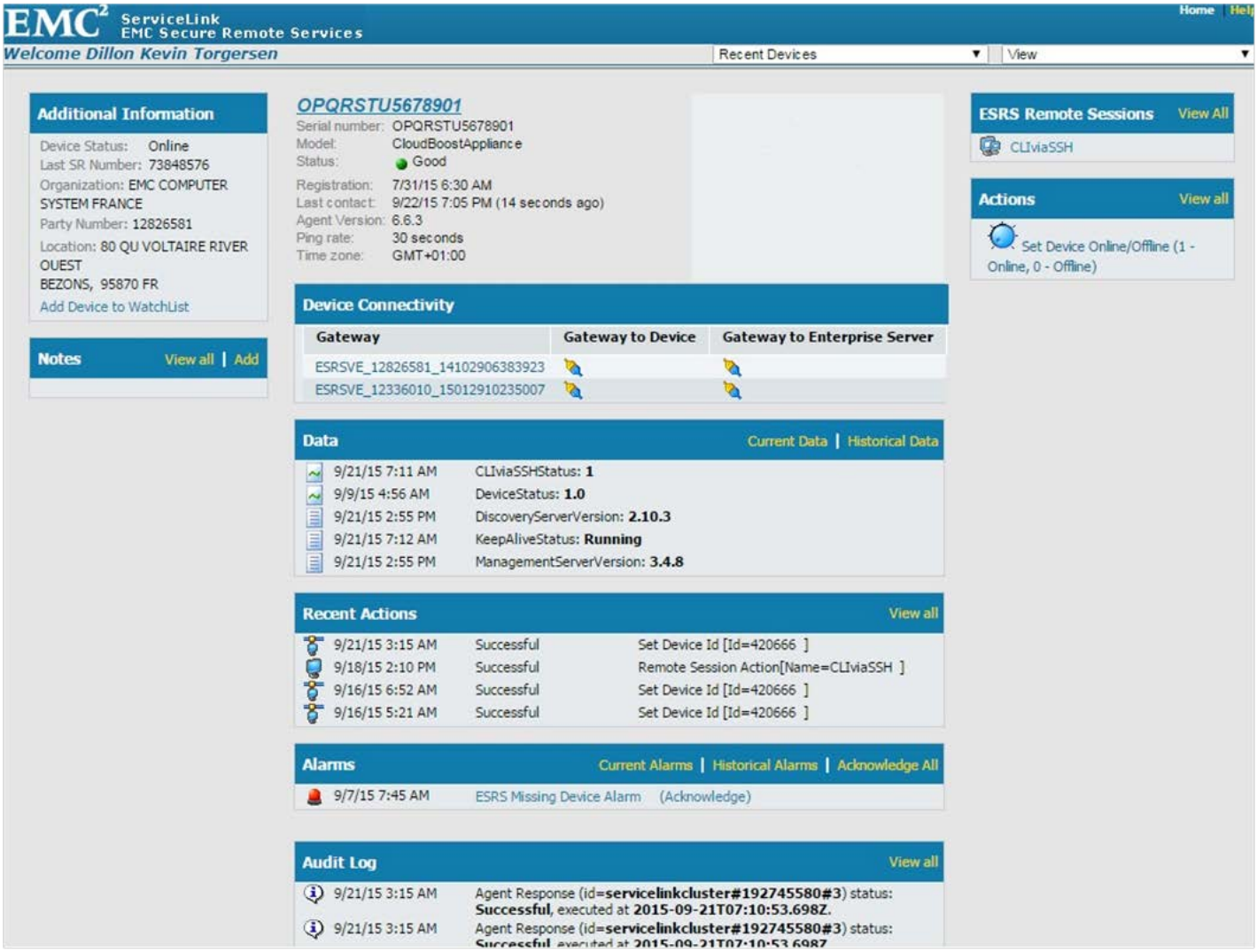
The recovery process is as simple as selecting a running appliance to recover to. This new appliance must be running, registered to the Cloud Portal but not configured. The new appliance must additionally be running the same CloudBoost version as the failed appliance.

The Cloud Portal will prevent a DR event being enacted when the primary appliance is in a healthy state to ensure reliability and availability of your environment. When a DR event is triggered the Cloud Portal will trigger an email to the registered account holders informing them of the DR event.

The Cloud Portal will display a RECOVER button next to any failed appliance. Clicking this button will begin the DR process. You select the recovery appliance from a list of available, registered but not configured, appliances. The recovery target appliance will adopt the FQDN and display name of the failed appliance.

DELL EMC SECURE REMOTE SERVICES (ESRS)

Dell EMC Secure Remote Services (ESRS) is available for advanced Dell EMC Support with CloudBoost. Customers can deploy ESRS in their own environment or through a Dell EMC public ESRS Gateway. Customers benefit from many advantages of ESRS. A screen shot showing ESRS information and the many customer advantages it provides is illustrated below.

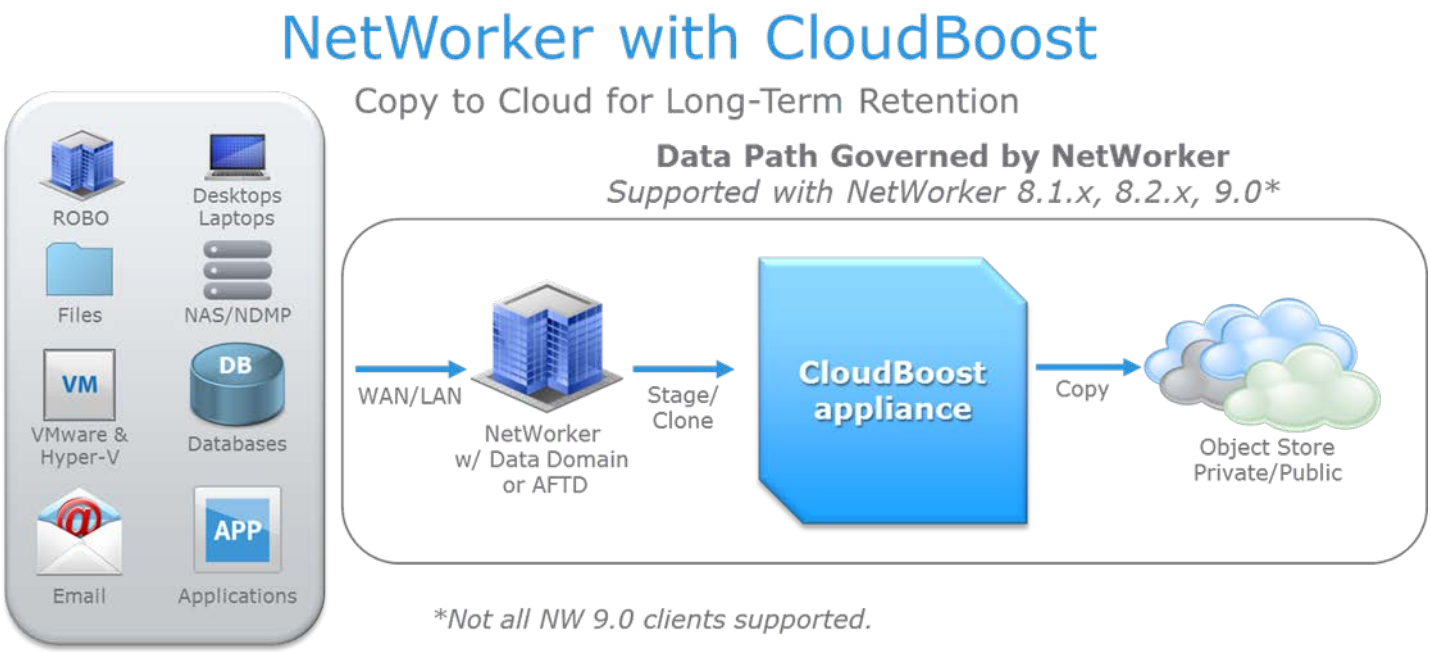


ESRS provides:

- Health Status
- Recent Actions
- Alarms
- Audit Logging
- Location
- Owner Organization
- Time Zone
- Additional Notes
- Remote Access

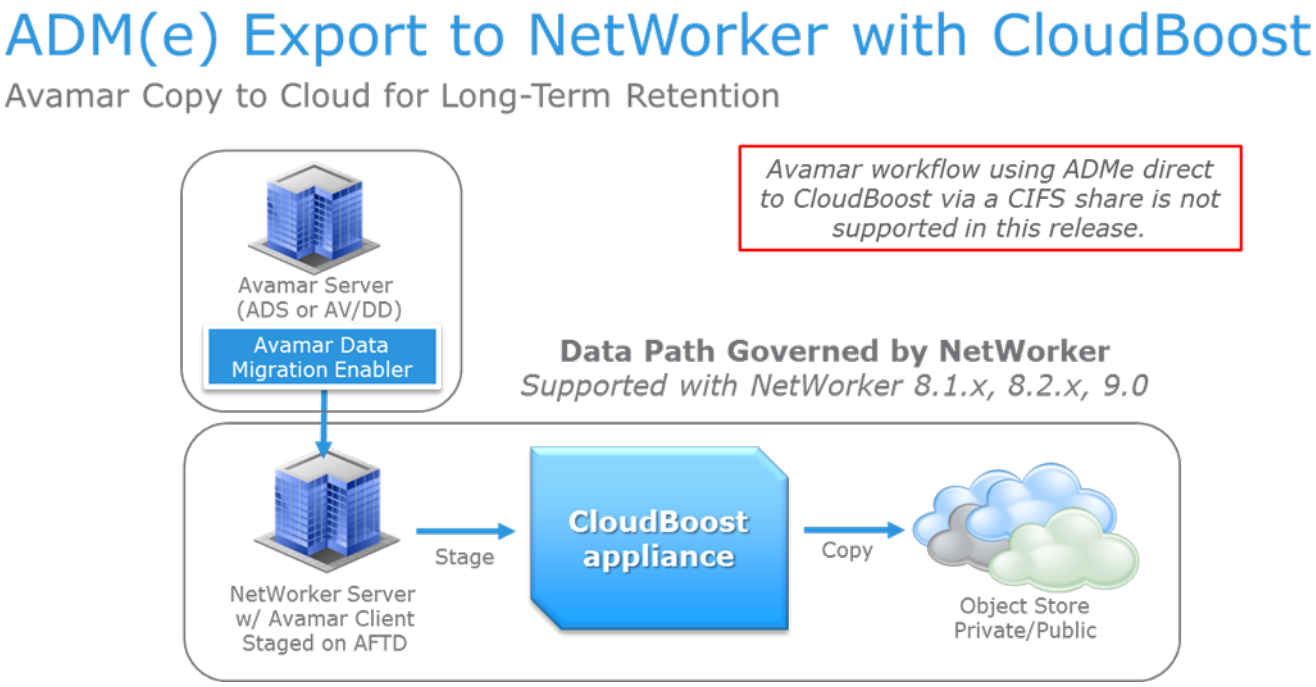
DELL EMC NETWORKER CONSIDERATIONS

CloudBoost is tightly integrated with Dell EMC NetWorker and, thus, uses the same workflows the backup administrator is already familiar with. With NetWorker, you continue to protect all the workloads you do today, including virtualized data, databases, applications, and snapshots. You also maintain application consistent backup data when using the NetWorker Modules. Nothing on the NetWorker administration side changes.



DELL EMC AVAMAR CONSIDERATIONS

As you can see in the diagram below, for CloudBoost with an Avamar® (or Avamar + Data Domain) deployment, the workflow is ADMe > Windows Staging server with CloudBoost client > CloudBoost appliance >Object Store. Leveraging the NetWorker flow is required for many compliance use cases because the transfer of data is maintained and recorded by the backup product (in this case, NetWorker). Both Avamar and NetWorker are included in Data Protection Suite.



CUSTOMER BENEFITS

Customers will realize a number of important benefits from deploying Dell EMC CloudBoost including:

- Using object storage as the LTR target eliminates risk of data loss
- Cloud economics plus CloudBoost deduplication, compression and WAN optimizations reduce TCO of long-term retention
- CloudBoost enables efficient LTR with ROBO and/or public cloud
- CloudBoost delivers enterprise-grade security with any object store

CONCLUSION

Leading organizations worldwide use Dell EMC's Data Protection Solutions to simplify, accelerate, and scale their backup and recovery environments. Systems for long-term retention (LTR) of backups are a critical element of that environment.

Because tape is inherently risky and surprisingly costly, object storage is rapidly replacing tape as the LTR target of choice. IT organizations wish to get out of the risky "tape museum business," eliminate the costs of refreshing media and libraries when old versions become unreadable, and end tape handling costs and risks.

Dell EMC CloudBoost enables data protection workloads from Data Protection Suite or Symantec/Veritas NetBackup to leverage the economics and agility of cloud for long-term retention. CloudBoost compliments the native cost-effectiveness of object storage by further consumption of network and cloud storage resources through a combination of source-side deduplication, compression, and WAN optimization. CloudBoost also protects the safety, privacy, and integrity of data stored in public or private cloud with a multi-pronged approach to data security.

Nonetheless, there is much that administrators can do to ensure that CloudBoost fully delivers on its capabilities. This white paper has set out a series of considerations and best practices to help IT organizations reap the full benefits of cloud-based LTR enabled by CloudBoost.

(APPENDIX) NETBACKUP CONSIDERATIONS

CloudBoost can also be used for long term retention of backup data with Veritas NetBackup v7.6. Here are some important notes when using CloudBoost with NetBackup:

- The CloudBoost client is installed on the NetBackup media server.
- It is supported on NetBackup media servers running on Windows 2012.
- Presents a NAS share to the media server using the “Advanced Disk” option.
- The deduplication workflow in NetBackup is used to rehydrate the data out of Data Domain to the CloudBoost NAS share.
- The CloudBoost appliance sends the data to the object store deduplicated and encrypted.
- NBU Storage Lifecycle Policies (SLP) can be used instead of manual duplications to clone data to cloud.
- The NetBackup with CloudBoost solution does not use OST. It uses an “Advanced Disk” integration with NetBackup. CloudBoost client deployed on a NBU media server presents a NAS share which is added as an Advanced Disk target to NBU.
- The NAS share on the media server is a pass-through to the object store. It is not a landing or staging area from where CloudBoost picks up the data to then clone to cloud. The NAS share is essentially mounting the underlying object store as a file system on the NBU media server.
- You can use either manual NBU duplications or NBU Storage Lifecycle Policies to clone data to cloud for long-term retention.
- The NetBackup media server can be a VM. There is no need for a physical server. We recommend 8 GB memory, 8 vCPUs.

The NetBackup with CloudBoost solution using their “Advanced Disk” integration is illustrated below.