ABSTRACT

As the enterprise evaluates a move to public cloud, one of the services that often holds them back is enterprise Data Protection for public cloud environments. In this paper we will preview how Dell EMC Data Protection solutions can be deployed in public cloud environments to address enterprise grade Data Protection services for enterprise application workloads.

November, 2016
# TABLE OF CONTENTS

## INTRODUCTION

## CLOUD USE CASES
- Backup to Cloud .......................................................... 6
- Long Term Retention Backups to Cloud .......................... 7
- Backup in Cloud .......................................................... 8
- Backup Across Clouds .................................................. 9
- Backup in Cloud and Across Clouds .............................. 10

## SOLUTION COMPONENTS
- Networker Management Console ................................ 11
- Networker Server ...................................................... 11
- Cloudboost Server .................................................... 11
- Cloud Object Storage ............................................... 14
- Networker Client ...................................................... 16
- Cloudboost Plugin .................................................... 17
- Networker Application Modules ................................. 18

## DESIGN CONSIDERATIONS
- Solution Requirements .............................................. 18
- Networking ............................................................. 19
- Scalability and Performance ...................................... 20
- Management and Reporting ...................................... 20
- Audit and Compliance .............................................. 20
- Data Reduction ....................................................... 21
- Data Security and Integrity ........................................ 21
- Data Caching .......................................................... 22
- Authentication and Authorization ............................... 22
- VM Import and Export .............................................. 23
- Service Isolation ...................................................... 23
- Disaster Recovery .................................................... 23
- Cloud Costs ............................................................ 24

## PUBLIC CLOUD CONSIDERATIONS
- Amazon Web Services ................................................ 24
  Performance ........................................................... 24
INTRODUCTION

With the rise of public cloud computing, companies are evaluating how they can leverage public cloud resources to become more agile and competitive in the marketplace. As companies proceed down this path it is natural to become overwhelmed by the risk and uncertainty associated with treading these uncharted waters. After all, a path to the public cloud involves giving up certain freedoms, such as ownership of IT assets and controls to respond to any situation.

Another way to describe this phenomenon is to understand who has the balance of power. When consuming public cloud resources the balance of power shifts to the public cloud provider. This may be an uncomfortable thought for some, and may even force one to withdraw from the journey. However, there is a way to navigate the uncertainty that public clouds introduce.

We call it Data Protection Everywhere

Data Protection Everywhere is about ensuring you maintain custody and control over your data assets irrespective of where or how they are stored. Data Protection is a key tenet of sound IT strategy that has withstood the test of time. In this new era of public cloud computing, the need for Data Protection is magnified and warrants special attention given the freedoms you are required to forego.

Dell EMC is the world leader in Data Protection solutions. With the release of this technology preview, we are excited to showcase how Dell EMC’s Software-Defined Data Protection solutions will be able to help you gain control of your journey to public cloud, giving you the same level of comfort, control and protection to confidently embark on your journey to any cloud.

The remainder of this paper outlines a number of approaches for protecting workloads running in public cloud environments. These approaches have specifically been verified against Amazon Web Services (AWS) and Microsoft Azure. In the future, we intend to expand the scope to include consideration for other public cloud providers.
CLOUD USE CASES

This paper introduces five use cases available to support an organization’s cloud-enabled Data Protection requirements with Dell EMC’s Data Protection solutions. This section describes each use case along with the advantages, disadvantages and requirements to help the reader align their cloud-based Data Protection requirements with the most suitable use case.

BACKUP TO CLOUD

The Backup to Cloud use case is designed for situations where the user has on-premises infrastructure and would like to use public cloud object storage for all backup workloads, including short term backups for operational recovery, and long term backups for compliance.

This architecture consists of NetWorker and CloudBoost servers (virtual or physical) located on-premises.

The CloudBoost server deduplicates, compresses and encrypts the data before sending it to the object storage.

The CloudBoost server can optionally be attached to data storage that caches backup data as it moves through the CloudBoost server to object storage. The cache can be used to complete backups faster while masking the affects of bandwidth constrained WAN connections (below 200 Mbps). Similarly, and subject to the size of the environment requiring backups, the backup cache can be sized to retain the most frequent backup copies on-premises for fast restore.

In the event of an irrevocable failure of on-premises systems, the NetWorker and CloudBoost servers can be instantiated in the public cloud environment to virtual machine instances, and backups recovered directly from the object storage to the disaster recovery virtual machine instances.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast restore from locally cached short-term backup copies</td>
<td>Restoration speed of backup copies from a cloud providers object storage is entirely dependent on the network connectivity and performance of the provider’s service at the time of restore.</td>
<td>Minimum one NetWorker server for on-premises cloud (physical or virtual).</td>
</tr>
<tr>
<td>Ability to restore systems and applications from backups stored in the cloud, to public cloud compute instances</td>
<td>There is no second independent copy of backup data stored outside of the public cloud provider’s control and infrastructure.</td>
<td>Minimum one CloudBoost server for on-premises cloud (physical or virtual).</td>
</tr>
<tr>
<td>Consumption of low-cost deduplicated public cloud object storage for all backup copies</td>
<td>Potential for data loss in the event the public cloud provider’s service malfunctions or is compromised.</td>
<td>Minimum 100 GB data storage for CloudBoost metadata (virtual only).</td>
</tr>
<tr>
<td>High retrieval cost to move backup copies from one public cloud provider to another or to an on-premises solution if the need arises.</td>
<td>HTTPS connectivity between CloudBoost server and object storage.</td>
<td>HTTPS connectivity between CloudBoost server and EMC cloud portal.</td>
</tr>
</tbody>
</table>
LONG TERM RETENTION BACKUPS TO CLOUD

The Long Term Retention Backups to Cloud use case is designed for situations where the user has existing on-premises infrastructure and would like to use public cloud object storage for long term retention and compliance requirements. Backup copies required for short term operational recovery remain on-premises for fast restore; and optionally, a disaster recovery site may be established for contingency purposes. Ideal candidates for this use case include backup environments that use tape for long term retention and compliance. In this case the tape storage is replaced with public cloud object storage.

This architecture consists of NetWorker and CloudBoost servers located on-premises in addition to local storage, Data Domain protection storage for example, to support short term operational recovery requirements.

The CloudBoost server can be attached to data storage in order to cache backup data as it is copied from the local storage through the CloudBoost server to public cloud object storage.

The CloudBoost server deduplicates, compresses and encrypts the data before sending it to the public cloud object storage.

This process is controlled by the NetWorker policy engine and allows backup and cloning operations to occur concurrently. The optional backup cache allows backups to complete faster while masking the effects of bandwidth-constrained WAN connections. In this illustration long term retention copies to public cloud object storage are created from the disaster recovery site (Site B). Alternatively, if the CloudBoost server is located at the production site (Site A), copies to public cloud object storage could also originate from the production site.

In the event of an irrevocable failure to on-premises production systems, the surviving NetWorker and CloudBoost servers in the disaster recovery site assume ownership of backup and restore activities for both short and long term backup copies.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast restore from on-premises short-term backup copies</td>
<td>There is no second independent copy of the long term retention backup copies stored outside of the public cloud provider’s control and infrastructure</td>
<td>Minimum two NetWorker servers for on-premises cloud (virtual or physical)</td>
</tr>
<tr>
<td>Efficient on-premises short-term backup copies stored and replicated using Data Domain deduplication</td>
<td>Potential for long term retention data loss in the event the public cloud provider’s service malfunctions or is compromised</td>
<td>Minimum one CloudBoost server for on-premises cloud (virtual or physical)</td>
</tr>
<tr>
<td>Reliable on-premises short-term backup copies protected by Data Domain’s Data Involucrable Architecture</td>
<td>High retrieval cost to move backup copies from one public cloud provider to another or to an on-premises solution if the need arises</td>
<td>Minimum 100 GB data storage for CloudBoost metadata (virtual only)</td>
</tr>
<tr>
<td>Consumption of low-cost deduplicated public cloud object storage for long term retention backup copies</td>
<td>HTTPS connectivity between CloudBoost server and object storage</td>
<td>HTTPS connectivity between CloudBoost server and EMC cloud portal</td>
</tr>
</tbody>
</table>
BACKUP IN CLOUD

The Backup in Cloud use case is designed for situations where the user has workloads running in public cloud virtual machine instances and would like to use public cloud object storage for all backup workloads, including short term backups for operational recovery and long term retention backups for compliance.

This architecture consists of NetWorker and CloudBoost servers deployed as public cloud virtual machine instances. The CloudBoost server is used to move backup data from NetWorker clients through the CloudBoost server to public cloud object storage. The CloudBoost server deduplicates, compresses and encrypts the data before sending it to the public cloud object storage.

As the distance between the CloudBoost server and public cloud object storage is short, the backup cache is not required. In fact, in this scenario the use of a cache would slow down backup and restore processing.

This use case does not include a mitigation strategy to address an irrevocable failure to the public cloud provider’s services. Consequently, the risk of permanent data loss increases compared to other use cases.
The Backup across Clouds use case is designed for situations where the user is consuming workloads in two independent public clouds, regions or instances and would like to protect each cloud’s workload using offsite backups to the alternate cloud.

This architecture consists of NetWorker and CloudBoost servers deployed on public cloud virtual machine instances. These instances support the short term operational recovery and long term retention requirements of all deployed public cloud workloads.

The CloudBoost servers in each cloud are optionally attached to data storage to cache backup data as it is moved through the CloudBoost server to the remote public cloud object storage. The CloudBoost servers deduplicate, compress and encrypt the data before sending it to the public object storage.

In this illustration the two clouds are labeled public; however, it is possible to use a combination of public and private clouds including private cloud object storage.

In the event of an irrevocable failure in one of the public clouds, the NetWorker and CloudBoost servers can be instantiated in the surviving public cloud environment to assume ownership of backup and restore activities for both short and long term retention copies.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast restore from locally cached short-term backup copies</td>
<td>Higher cost as two independent backup copies are maintained across two public clouds</td>
<td>Minimum two NetWorker servers for public clouds</td>
</tr>
<tr>
<td>Consumption of low-cost deduplicated public cloud object storage for all backup copies</td>
<td>Restore speeds of non-cached backups will depend on inter-cloud network performance</td>
<td>Minimum two CloudBoost servers for public clouds</td>
</tr>
<tr>
<td>Multiple backup copies maintained across independent public clouds reducing potential for data loss</td>
<td></td>
<td>Minimum 100 GB data storage for CloudBoost metadata</td>
</tr>
<tr>
<td>No exit costs as two independent backup copies are maintained across independent public clouds</td>
<td></td>
<td>HTTPS connectivity between CloudBoost server and object storage</td>
</tr>
<tr>
<td>No on-premises infrastructure required</td>
<td></td>
<td>HTTPS connectivity between CloudBoost server and EMC cloud portal</td>
</tr>
</tbody>
</table>
BACKUP IN CLOUD AND ACROSS CLOUDS

The Backup in Cloud and across Clouds use case is designed for situations where the user is consuming workloads in public clouds and would like to protect each workload using local backups for fast restore, and cloned backups across clouds to maintain offsite independent backup copies. The backups stored in both clouds can be used for both short term operational recovery and long term backups for compliance.

This architecture consists of NetWorker and CloudBoost servers deployed on public cloud virtual machine instances to support the short term operational recovery and long term retention requirements of public cloud workloads.

The CloudBoost servers that facilitate cloning of backups to the remote public cloud are optionally attached to data storage in order to cache data as it moves through the CloudBoost server to the remote public cloud object storage. The CloudBoost servers deduplicate, compress and encrypt the data before sending it to the public object storage.

In this illustration the two clouds are labeled public; however, it is possible to use a combination of public and private clouds including private cloud object storage.

In the event of an irrevocable failure in one of the public clouds, the NetWorker and CloudBoost servers can be instantiated in the surviving public cloud environment to assume ownership of backup and restore activities for both short and long term retention copies.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast restore from local backup copies across all retention classes</td>
<td>Higher cost as two independent backup copies are maintained across two independent public clouds</td>
<td>Minimum two NetWorker servers for public clouds</td>
</tr>
<tr>
<td>Consumption of low-cost deduplicated public cloud object storage for all backup copies</td>
<td>Additional CloudBoost servers required</td>
<td>Minimum four CloudBoost servers for public clouds</td>
</tr>
<tr>
<td>Multiple backup copies maintained across independent public clouds reducing potential for data loss</td>
<td></td>
<td>Minimum 100 GB data storage for CloudBoost metadata</td>
</tr>
<tr>
<td>No exit costs as two independent backup copies are maintained across independent public clouds</td>
<td></td>
<td>HTTPS connectivity between CloudBoost server and object storage</td>
</tr>
<tr>
<td>No on-premises infrastructure required</td>
<td></td>
<td>HTTPS connectivity between CloudBoost server and EMC cloud portal</td>
</tr>
</tbody>
</table>
SOLUTION COMPONENTS

This section describes the role of the solution components and relationships to support the various cloud use cases.

Figure 1: Solution components and how they are related

NETWORKER MANAGEMENT CONSOLE

All NetWorker servers and clients are managed from the NetWorker Management Console server. The management console provides reporting and monitoring capabilities for all NetWorker servers and clients and can also provide logging and alerting of events and notifications.

The console server is accessed through a graphical user interface that can be run from any system with a supported web browser and Java Runtime Environment (JRE). Multiple users can access the console and authentication can be integrated with Lightweight Directory Access Protocol (LDAP/S) and Microsoft Active Directory Server (AD).

To reduce the number of virtual machine instances required, the console server can be deployed on a server that is also acting as the NetWorker server

NETWORKER SERVER

The NetWorker Server is the core component of the solution and supports the policy, scheduling and catalog functions of the solution to manage backup and recover processes between the NetWorker clients, CloudBoost servers and cloud object storage. The NetWorker server also coordinates NetWorker Application Modules to create consistent and recoverable backup copies of operating system and application workloads.

The NetWorker server can be deployed on either Windows or Linux systems. For a complete list of supported operating systems refer to the Online NetWorker Compatibility Matrix.

CLOUDBOOST SERVER

The CloudBoost server provides connectivity and access to the cloud object storage for backup data. NetWorker clients communicate with the CloudBoost server to backup and restore from backup copies stored in the cloud object storage.

Each CloudBoost server can be associated with one cloud object storage profile.

Some of the important functions performed by the CloudBoost server include data compression, data deduplication and encryption. As data passes through the CloudBoost server it is deduplicated using a variable block algorithm, followed by compression and encryption before it is written to a cloud object storage bucket or container. This process ensures the minimum amount of storage and network bandwidth is consumed and is designed to yield significant cost savings over solutions that do not provide data compression and data deduplication features.
Once deployed and activated, the CloudBoost server is considered headless and managed via the Dell EMC Cloud Portal. This portal is hosted by Dell EMC so that CloudBoost servers can be deployed and managed across any combination of private and public cloud environments.

The Dell EMC Cloud Portal is accessed via a standard web browser. Some of the management functions available include configuration, monitoring, capacity and performance management, upgrading and recovering CloudBoost servers.

Below are sample screenshots of the Dell EMC Cloud Portal in action.

Figure 2: List of CloudBoost servers
Figure 3: CloudBoost server usage details
CLOUD OBJECT STORAGE

Cloud object storage is high density, low cost storage designed for infrequently accessed data. In this solution, the cloud object storage is used to store backup copies for short term operational recovery requirements and/or long term retention and compliance requirements.

The cloud object storage represents a near infinite amount of storage that can be consumed by the CloudBoost server and optionally CloudBoost plugins, to store and retrieve backup copies.

For a list of supported object stores refer to the CloudBoost compatibility matrix.

Below are sample screenshots of Azure Blob storage and Amazon S3 storage that have been consumed by a CloudBoost server.

Figure 4: List of CloudBoost objects in Azure Storage container
Figure 5: List of CloudBoost objects in AWS S3 bucket
NETWORKER CLIENT

A NetWorker client is a software component deployed in the public cloud virtual machine instances to protect the operating system and application data. Once deployed, the client configuration is performed from the NetWorker Management Console and associated with one or more protection policies. These protection policies determine when clients are protected, what is protected, where backup copies are sent, and how long they are retained. With the NetWorker client deployed, backups and recoveries are sent to a CloudBoost server where data compression, deduplication and encryption activities occur before writing to the cloud object storage.

The diagram below illustrates the data path for NetWorker client deployments that do not include the CloudBoost plugin. The public cloud object storage is accessed via the CloudBoost server.

The CloudBoost server can be deployed with or without a local data cache and will depend on the proximity of the cloud object storage to the server and available bandwidth.
CLOUDBOOST PLUGIN

The CloudBoost plugin is an optional software component that can be enabled in NetWorker clients to maximize the aggregate performance and efficiency of the Data Protection service. With the CloudBoost plugin enabled, the NetWorker client can perform backups and recoveries directly to the cloud object storage, and assumes responsibility for data deduplication, compression and encryption activities before writing to the cloud object storage.

The diagram below illustrates the data path for NetWorker clients deployed with the CloudBoost plugin. The public cloud object storage can be accessed directly or via a web proxy server. For high performance requirements, use of a web proxy is not recommended.

![Dataflow diagram](image)

**Figure 7: Dataflow of NetWorker client with CloudBoost plugin**

To ensure data is deduplicated across clients, the CloudBoost plugin exchanges metadata with the CloudBoost server to check if a block of data has been backed up previously before committing it to cloud object storage. This avoids transmitting duplicate blocks of data and ensures the solution delivers global deduplication efficiency across clients using the same CloudBoost server.

Although optional, there are several performance and security benefits to using the CloudBoost plugin:

- Deduplicates, compresses and encrypts the data before it leaves the client, ensuring data is transmitted efficiently and securely
- Sends and recalls data directly to/from the cloud object storage and avoids passing the data through a backup server
- Increases the aggregate performance of the environment by distributing the data processing to the CloudBoost plugins
- NetWorker’s Block Based Backup feature can be used to further increase performance of backup and restores

Another benefit is in the case where backups are created across different public cloud providers. The CloudBoost plugin can be enabled for a remote NetWorker client to facilitate direct access to the local public cloud object storage. This avoids bulk data transfer from one public cloud to another and expedites the restore process.
NETWORKER APPLICATION MODULES

NetWorker application modules are software components deployed alongside NetWorker clients to integrate the Data Protection service with application workloads. It is recommended these modules be deployed to provide application administrators control and visibility of Data Protection services, granular recovery capabilities and to ensure the application state is transaction consistent at the time of backup.

There are two application module bundles that would be appropriate for public cloud use cases.

- NetWorker Module for Microsoft (NMM) – support for Microsoft applications (e.g. MS SQL, Exchange, SharePoint)
- NetWorker Module for Databases and Applications (NMDA) – support for UNIX applications (e.g. Oracle, DB2, Sybase, SAP, MySQL)

For a complete list of all supported applications refer to the Online NetWorker Compatibility Guide.

DESIGN CONSIDERATIONS

For many organizations exploring the consumption of public cloud services, the transition from on-premises solutions to public cloud will raise many questions. This section details some of the design considerations we encountered when evaluating our solution using the AWS and Azure public cloud platforms as targets. While these considerations are specific to the solution and use cases described in this paper, the reader may find they are generally applicable to other Data Protection solutions.

SOLUTION REQUIREMENTS

The following table outlines the minimum NetWorker server requirements for small, medium and large configurations.

For more elaborate sizing and requirements refer to the NetWorker Performance Optimization Planning Guide.

The CloudBoost server is sized relative to the amount of logical backup data under management and whether the data cache is enabled or disabled. If the CloudBoost server is in close proximity to the cloud object storage and exhibits LAN characteristics, then the data cache option should be disabled. This will improve ingress and egress performance and speed up the backup and recovery process.

<table>
<thead>
<tr>
<th>Size of configuration</th>
<th>Minimum requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small configuration (up to 50 clients, approximately 1000 jobs per day)</td>
<td>2 CPU cores, 8 GB RAM, 1 GbE</td>
</tr>
<tr>
<td>Medium configuration (up to 200 clients, up to 10,000 jobs per day)</td>
<td>4 CPU cores, 16 GB RAM, 1 GbE</td>
</tr>
<tr>
<td>Large configuration (up to 500 clients, more than 10,000 jobs per day)</td>
<td>8 CPU cores, 32 GB RAM, 1 GbE</td>
</tr>
</tbody>
</table>

Table 1: NetWorker server sizing

<table>
<thead>
<tr>
<th>Data Cache Disabled</th>
<th>Data Cache Enabled</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 CPU cores, 32 GB RAM, 1-10 GbE</td>
<td>16 CPU cores, 64 GB RAM, 1-10 GbE</td>
</tr>
<tr>
<td>100 GB of metadata storage per 400 TB of logical backup data under management</td>
<td>100 GB of metadata storage per 400 TB of logical backup data under management</td>
</tr>
<tr>
<td>Maximum 6 PB logical data under management per server</td>
<td>Maximum 6 PB logical data under management per server</td>
</tr>
<tr>
<td>Minimum 200 GB and up to 6 TB of data cache storage</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: CloudBoost server sizing
The amount of cloud object storage consumed will depend on the backup retention, frequency and application data types protected. In a typical environment CloudBoost technology will yield a 1.5:1 to 4:1 deduplication ratio for the first backup (e.g., 100 TB of backup will be reduced to between 25 TB and 66 TB). For subsequent backups, the amount of data that changes is the primary contributor to the deduplication rate. A lower change rate will yield higher deduplication rates while a higher change rate will yield lower deduplication rates. Furthermore, shorter intervals between backups will yield higher rates and longer intervals lower rates.

For detailed sizing advice please contact your Dell EMC Systems Engineer.

**NETWORKING**

A common approach to deploying infrastructure services in the public cloud is to create a virtual network segment specifically for infrastructure services. The VMs that support the services are deployed into this segment to provide isolation from other systems and segments.

In the example below the NetWorker and CloudBoost servers are deployed in a services segment. Network access controls are used to enforce which systems and segments can access the services.

NetWorker and CloudBoost both require DNS services. In this example they are hosted locally on a DNS server in the frontend network. Alternatively, DNS services from the cloud provider or external Internet facing providers can be used.

As the needs of the infrastructure services grow, additional servers are deployed in the services segment and workloads are distributed accordingly.
SCALABILITY AND PERFORMANCE

One of the frequently under-estimated characteristics of a public cloud service is the infrastructure supporting the services are often shared amongst unrelated users. In order to ensure each user receives their fair share of resource time, cloud providers employ automated resource management techniques to minimize the impact one user can have on another when consuming the same resources.

This behavior makes measuring and modeling performance highly unpredictable and circumstantial. Given this, it would be inappropriate to claim particular levels of performance when consuming public cloud resources with this solution. However, it is appropriate to share how the solution components scale with additional cloud resources.

The primary components that need to scale in this solution are the NetWorker server and CloudBoost server. The NetWorker server is responsible for tracking real-time activities across the Data Protection service. There is a limit to the number of real-time activities and elements a NetWorker server can track given the available resources. The NetWorker server sizing table in the solution requirements section of this document describes how the virtual machine resources can be scaled up to satisfy additional workload. Once it is determined that scaling up yields diminishing returns, the environment must scale out by adding a subsequent NetWorker server. When this occurs a subsequent CloudBoost server must also be added as there is a one-to-one relationship between a CloudBoost server and a NetWorker server.

In the event the CloudBoost server needs to scale to support higher data volumes within a desired backup window, the environment must scale out by adding a second CloudBoost server and associating it with the existing NetWorker media pool. Once this is done NetWorker will distribute the workloads across the available CloudBoost servers.

Alternatively, if the capacity that a CloudBoost server can manage is exhausted, then the NetWorker device that manages it should be flagged read-only. This will prevent new backups from being written to the full CloudBoost server while still allowing recoveries. A second CloudBoost server should be added to take over the remaining backup workload.

The CloudBoost server that was flagged read-only can be flagged read-write once the amount of data under management falls to acceptable levels caused by old backups expiring.

MANAGEMENT AND REPORTING

The Data Protection service is managed using a Java client downloaded from the NetWorker server over the Java Network Launch Protocol (JNLP). This allows the service to be managed from anywhere so long as access to the required network ports is made available to the desired management clients.

Conversely, it is also possible to deploy a Windows or Linux desktop in the public cloud that is used as a jump server to manage access to the systems. This avoids opening up management activities to the public facing Internet.

Operational reporting is provided by the NetWorker server using the NetWorker Management Console. The NetWorker server collects information from the environment and stores it in a database. The NetWorker Management Console interrogates this database to produce a number of reports on backup status, backup statistics, events, hosts, users and devices.

For a detailed understanding of NetWorker’s management and reporting capabilities refer to the NetWorker Administration Guide.

For more advanced reporting, including reporting across multiple NetWorker servers, clouds and on premise deployments, Dell EMC’s Data Protection Advisor can be used. Data Protection Advisor (DPA) is a comprehensive monitoring, analytics, alerting and reporting platform that provides businesses with full visibility into the utilization, effectiveness and compliance status of their data protection environment.

For more information on DPA refer to http://www.dell.com/.

AUDIT AND COMPLIANCE

The data collected by the NetWorker server facilitates auditing tasks and activities performed by specific users. By default the retention policy for audit data is one year; however, this can be increased as required.

Refer to the NetWorker Administration Guide to learn how to adjust the audit retention period.
DATA REDUCTION

The CloudBoost technology embedded in the CloudBoost server and plugin implements data reduction techniques to significantly reduce the consumption of cloud object storage. Specifically, CloudBoost implements the Rabin-fingerprinting scheme to identify natural break points in the backup stream and then use these break points to eliminate variable blocks of data that have been stored previously by the CloudBoost server.

CloudBoost purposefully uses a large average block size of 256KB as cloud object stores tend to exhibit very high overheads for any write or commit operation. By using larger and fewer blocks, the solution can support significantly higher throughput than would be possible with small block schemes. Once blocks are identified as unique, they are compressed, encrypted and transferred to the cloud object storage and stored as objects. References to the objects are added to the CloudBoost metadata database.

When backups expire and unique blocks become orphaned, the CloudBoost server reclaims space by removing the corresponding objects. This process is very efficient and is specifically designed to reclaim space without recalling objects.

DATA SECURITY AND INTEGRITY

The CloudBoost technology implements both in-flight and at rest encryption for all transactions. CloudBoost uses Secure Socket Layer (SSL) to communicate with the object storage and also produces signed URLs that are only valid for a limited time and cannot be replayed. The signed URLs are used to access the public object storage securely. CloudBoost uses Advanced Encryption Standard (AES-256) to encrypt the data stored in the cloud object storage.

CloudBoost ensures the data is written to the cloud object storage correctly. To achieve this a cryptographic hash is generated for the object. This hash is passed to the cloud object storage along with the data at the time it is created. This allows the cloud object storage to validate that the correct data was received and, as a final check, the CloudBoost server checks that the object is present. For features such as fault detection and healing, CloudBoost relies on the underlying cloud object storage to support these low-level data integrity activities.

The diagram below illustrates the CloudBoost data flow when the CloudBoost plugin is enabled in the NetWorker client.

![Dataflow diagram](image.png)

Figure 9: Dataflow using NetWorker client with CloudBoost plugin enabled
The diagram below illustrates the CloudBoost data flow when the NetWorker client is deployed without the CloudBoost plugin.

**DATA CACHING**

The CloudBoost technology can be deployed with a local data cache that is used to cache backup data destined for the cloud object storage. In public cloud deployments the data cache must be allocated from a virtual disk image and attached to the CloudBoost server at the time of setup.

The data cache should only be used when the CloudBoost server and object storage are separated by WAN characteristics. This will allow the cache to temporarily mask the effects of WAN connectivity.

It is important to recognize the caching feature can only provide short term relief from bandwidth constrained environments. If bandwidth is consistently constrained relative to the amount of data that must be written to the cloud object storage, then the disk cache will eventually reach its maximum space threshold and will trigger CloudBoost to throttle the ingest rate. Once the cache falls below its maximum space threshold the throttling will subside. This process will continue until the backup workload is completed. When sizing the cache it is best to choose a cache that is larger than the size of the backups being transferred unless the available bandwidth to the object storage is above 200 Mbps.

The data cache will reclaim space based on a first-in-first-out basis. This is where the oldest data that entered the cache will be removed to make room for new data. Once a minimum space threshold is met the CloudBoost server will cease removing old data.

The CloudBoost server supports up to 6 TB of data cache.

**AUTHENTICATION AND AUTHORIZATION**

NetWorker Authentication Service provides a NetWorker environment with token-based authentication and single sign on (SSO) support. Token-based authentication enables users to securely connect to the NetWorker Management Console, the NetWorker server, and to perform secure backup and recover operations.

When a NetWorker or NetWorker Management Console operation requires authentication, the requesting process contacts the NetWorker Authentication Service to verify the credentials of the user account that started the request. When the NetWorker Authentication Service successfully verifies the user, the application issues a time-limited, signed, and encrypted SAML token to the user.
requesting process. All the NetWorker components that require authentication can use the token to verify the user, until the token expires.

The NetWorker Authentication Service supports both an internal authentication authority and external authentication authority via LDAP and AD services. The service also provides a hierarchical security model for users and groups, which enables you to define access levels, authentication, and authorization in a multi-tenant configuration.

For further information refer to the NetWorker Security Configuration Guide.

**VM IMPORT AND EXPORT**

The import and export of data to the cloud can be facilitated by adopting the Backup across Clouds use case. This allows remote backups from one cloud to be recovered to a running VM instance in the alternate cloud.

Importing and exporting entire VM instances is best achieved with the cloud provider’s integrated tools. The general approach involves:

- Powering down the VM
- Creating a snapshot of the VM
- Exporting the snapshot to a flat file
- Copying the file to the alternate cloud
- Creating a VM from the file image
- Powering on and reconfiguring the VM

For further information refer to the cloud providers VM import and export instructions.

**SERVICE ISOLATION**

One of the most important principles of a Data Protection strategy is maintaining separation of concerns between the primary systems being protected and the systems providing the protection.

When we think about Data Protection we usually think in terms of recovery point objective and recovery time objective. What we rarely consider are the events we are protecting against and whether the protection methods and controls we employ are effective against these events.

Public cloud environments introduce new levels of uncertainty and risk that we need to account for in the design of Data Protection services.

It is important to ensure the overall strategy employed provides diversification away from primary systems, by taking copies of data and placing them under the control of different and diverse systems. Separating the concerns of the Data Protection service and systems away from the primary systems being protected minimizes events from propagating to the protection copies and compromising the recovery position.

One way to implement separation of concerns for public cloud environments is to allocate the Data Protection resources from a separate cloud account that is isolated from the cloud account hosting the primary application systems. This provides management isolation between the accounts supporting the primary datasets and protection copies.

**DISASTER RECOVERY**

While very unlikely, public cloud environments are not immune to long outages and disaster scenarios. It is prudent to ensure that Data Protection services and the data under management is recoverable from failure scenarios affecting the public cloud provider.

There are three datasets required to rebuild the Data Protection service. These include the NetWorker server bootstrap data and client file indexes, and the CloudBoost server metadata database.

NetWorker bootstrap data and client file indexes are backed up regularly via standard NetWorker policy.

---

1 NetWorker requires the bootstrap data to be backed up to a local device. This can be achieved by installing the CloudBoost plugin on the NetWorker server and using the CloudBoost server as a local backup device.
The CloudBoost metadata is also backed up regularly to the cloud object storage and is configured via the Dell EMC Cloud Portal.

In the event of a disaster that renders the servers inoperable the following steps would be taken:

- Deploy a new CloudBoost server and register it to the Dell EMC Cloud Portal account (but do not configure it)
- Recover the CloudBoost server using the Dell EMC Cloud Portal (select the un-configured CloudBoost server)
- Deploy a new NetWorker server instance and associate it with the CloudBoost server
- Recover the NetWorker bootstrap and client file indexes

For further information refer to the NetWorker Server Disaster Recovery and Availability Best Practices Guide.

CLOUD COSTS
Understanding the cost of cloud services is necessary before embarking on your cloud journey. With respect to Data Protection services the rate of cloud object storage consumption will be quite rapid initially and will not exhibit a slow ramp up period that you may be familiar with when deploying storage for application servers.

When Data Protection services are activated, the first backup copy consumes the greatest amount of storage as it benefits least from data reduction techniques. However, subsequent copies consume a fraction of the total required cloud object storage. Therefore, when budgeting for cloud object storage, it is necessary to budget upfront approximately 30-50% of the total storage required to support the desired data protection and retention policies.

Furthermore, it is important to recognize the criticality of Data Protection services particularly when deployed alongside application servers in public cloud, as the failure of the cloud could render your recovery position untenable. Therefore, it is recommended the cloud services supporting the Data Protection service are placed under premium levels of support to allow you to better respond to unplanned events.

PUBLIC CLOUD CONSIDERATIONS
This section outlines a number of deployment considerations when implementing the use cases outlined in this paper for a variety of public cloud providers.

AMAZON WEB SERVICES

PERFORMANCE
We understand the performance of AWS S3 is throttled which will prematurely limit the performance of CloudBoost to S3. To maximize performance, open a support case with AWS and request that rate limits be removed.

Refer to AWS S3 Request Rate and Performance Considerations.

DATA REDUNDANCY
The durability of data copies stored in AWS S3 is a function of the AWS S3 service. As of this time there are three classes supported.

- Standard: designed for 99.99% availability and 99.999999999% durability
- Standard – Infrequent Access: designed for 99.9% availability and 99.999999999% durability
- Reduced Redundancy: designed for 99.99% availability and 99.99% durability

It is important to recognize that redundancy provided by S3 does not prevent data loss in the event of accidental, intentional or malicious deletion of objects contained in the S3 storage. It is for this reason you should strongly consider whether other forms of isolation are necessary to minimize the risk of permanent data loss.

Reduced redundancy is not appropriate for deduplicated backups due to the expected data loss of 1 out of 10,000 objects per year.
One common approach to preventing data loss is to maintain multiple backup copies of data across diverse and independent storage systems; for example, by storing one copy in S3 and a second copy in Azure Blob Storage. This approach can help minimize the risk that an incident will affect both copies.

For further information refer to AWS S3 services documentation.

DATA STORAGE

The CloudBoost server requires Amazon Elastic Block Store (AWS EBS) for the operating system disk, metadata database and optional data cache. In an AWS environment it is recommend the metadata database and optional data cache be allocated from AWS general purpose or provisioned IOPS (SSD) volumes which provide high IOPs and throughput.

AWS EC2 instance storage volumes are ephemeral and should not be used for the NetWorker Servers or CloudBoost Servers.

When the CloudBoost data cache feature is used, the data written to EBS volumes is only replicated within its Availability Zone to protect from component failure. If by chance the Availability Zone is lost, this data will be unrecoverable. If this is undesirable, do not use the disk cache feature which is only required when the CloudBoost server and object storage are separated by WAN conditions.

DOMAIN NAME SYSTEM

The Data Protection service requires hostnames and IP addresses to resolve via the Domain Name System (DNS). AWS includes support for hosting your own domains on the AWS Route 53 DNS service or DNS servers of your choice.

For further information refer to AWS Route 53 services documentation.

NETWORKING

In an AWS Virtual Private Cloud (VPC) environment, each virtual machine is assigned an IP address from a defined virtual network and a public IP address that is accessible via a load balancer. By default, Internet facing access to the public IP address is enabled for secure shell where the VM is running Linux and Windows Remote Desktop where the VM is running Windows.

In order to set up the CloudBoost server, secure shell access should be used. However, once set up, it can be removed from the CloudBoost server's VM security group using the AWS portal.

To access the NetWorker Management Console from AWS's Internet facing load balancers, the following firewall rules should be enabled against the NetWorker server VM.

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Protocol</th>
<th>Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management Clients</td>
<td>NetWorker Server</td>
<td>TCP</td>
<td>5432, 9000, 9001</td>
</tr>
</tbody>
</table>

Note: The CloudBoost Server is managed via the Dell EMC Cloud Portal which is accessible from the Internet. Therefore the CloudBoost server must continue to have Internet access.

By default AWS S3 storage is accessed by sending data over an Internet connection from the VPC. To avoid the Internet it is possible to configure an AWS endpoint to bypass the Internet. For more information refer to AWS VPC endpoints documentation.

When selecting the machine type to run CloudBoost server, consider the network available to the machine. AWS provides four network types: low to moderate, moderate, high and 10 Gigabit. The high network type will result in Gigabit network speeds. This was verified with iperf3 between clients and the CloudBoost server. This the minimum recommended. If higher network capacity is required, deploy the CloudBoost server on a machine with 10 Gigabit network type. Alternatively, deploy multiple CloudBoost servers and distribute the workload between them.
MICROSOFT AZURE

PERFORMANCE
During the testing phase of this technology preview, we recognized that the limiting factor preventing higher levels of performance was the network connectivity available between the virtual machine instances supporting the NetWorker clients and CloudBoost servers.

Using iperf3 we conducted network performance bandwidth tests between the NetWorker virtual machine instances and CloudBoost virtual machine instances to characterize the available bandwidth between them.

The Azure A2 instances used for the NetWorker clients peaked at ~500 Mbit/sec ingress and egress bandwidth rates. The Azure A4 instances used for the CloudBoost server peaked at ~1000 Mbit/sec ingress and egress bandwidth rates.

It is important to identify and recognize the limits of the virtual machine instances used to support Data Protection services in order to achieve the desired backup and recovery expectations for the volumes of data requiring protection.

In addition to networking constraints, we also noticed the performance of Azure storage used for the operating system and application workloads also exhibit low performance for standard Azure storage.

In order to achieve a reliable service that can be recovered quickly, one should consider using performance Azure storage for important application workloads and spreading these workloads across multiple Azure storage accounts.

For further information refer to Azure Subscription and Service Limits, Quotas, and Constraints documentation.

DATA REDUNDANCY
The durability of data copies stored in Azure Blob storage is a function of the Azure Storage service. As of this time there are three replication modes supported.

- Locally redundant replicates data within the same region
- Geo redundant replicates data to a secondary storage region
- Read-access geo redundant replicates data to a secondary storage region with read-only access enabled

It is important to recognize that redundancy provided by Azure Storage does not prevent data loss in the event of accidental, intentional or malicious deletion of objects contained in the Azure Blob storage. It is for this reason you should strongly consider whether other forms of isolation are necessary to minimize permanent data loss.

One common approach to preventing data loss is to maintain multiple backup copies of data across diverse and independent storage systems; for example, by storing one copy in Azure Blob Storage and a second copy in Amazon Simple Storage Service. This approach can help minimize the risk that an incident will affect both copies.

For further information refer to Azure Storage services documentation.

DATA STORAGE
The CloudBoost server requires Azure Blob storage for the operating system disk, metadata database and optional data cache. In an Azure environment it is recommend that the metadata database and optional data cache be allocated from Azure premium storage disks which provide high IOPs and throughput.

DOMAIN NAME SYSTEM
The Data Protection service requires hostnames and IP addresses to resolve via the Domain Name System. Azure includes support for hosting your own domains on Azure DNS servers or DNS servers of your choice.

For further information refer to Azure DNS services documentation.
NETWORKING

In an Azure environment each virtual machine is assigned a private IP address from a defined virtual network and a public IP address that is accessible via a load balancer. By default, Internet facing access to the public IP address is enabled for secure shell where the VM is running Linux, and Windows Remote Desktop where the VM is running Windows.

In order to set up the CloudBoost server, secure shell access should be used. However, once set up, it can be removed from the CloudBoost server's VM endpoint configuration using the Azure portal.

To access the NetWorker Management Console from Azure’s Internet facing load balancers, the following firewall rules should be enabled against the NetWorker server VM.

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Protocol</th>
<th>Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management Clients</td>
<td>NetWorker Server</td>
<td>TCP</td>
<td>5432, 9000, 9001</td>
</tr>
</tbody>
</table>

Note: The CloudBoost Server is managed via the Dell EMC Cloud Portal, which is accessible from the Internet. Therefore the CloudBoost server must continue to have Internet access.

SUMMARY

This technology preview has shown how Dell EMC Data Protection solutions can be used to protect valuable customer data in the most common public and private cloud use cases. The advantages, disadvantages and requirements of each use case are also discussed to help the reader match their requirements to the most appropriate use case.

ADDITIONAL RESOURCES

Link to NetWorker documentation

Link to CloudBoost documentation

Link to NetWorker with CloudBoost Technical Demo Video

Link to CloudBoost Overview Lightboard Video

Link to Long Term Retention to the Cloud with Dell EMC Cloudboost, TCO and the Cloud – White Paper

Link to Backup to the Cloud with Dell EMC DPS & CloudBoost – Best Practices – White Paper