This Solution Guide is for Dell EMC® Elastic Cloud Storage (ECS™) and CloudBoost. This Dell EMC Solution provides a cloud enabling technology for Dell EMC Data Protection and third party software for long term retention.

August, 2017
Table of Contents
SOLUTION OVERVIEW ........................................................................................................................................................ 3
NETWORKER WITH CLOUDBOOST AND ECS .................................................................................................................. 3
WHY ECS & CLOUDBOOST ................................................................................................................................................. 5
SALES GUIDANCE................................................................................................................................................................ 5
TECHNICAL REFERENCE ARCHITECTURE ...................................................................................................................... 6
  INTRODUCTION................................................................................................................................................................ 6
  CONFIGURATIONS ........................................................................................................................................................... 7
  SOFTWARE ARCHITECTURE ........................................................................................................................................ 12
  BEST PRACTICES .......................................................................................................................................................... 14
  PERFORMANCE DATA ................................................................................................................................................... 15
  PERFORMANCE AND DEDUPE..................................................................................................................................... 15
  SIZING GUIDANCE ......................................................................................................................................................... 16
  SECURITY GUIDANCE ................................................................................................................................................... 17

Dell EMC believes the information in this publication is accurate as of its publication date. The information is subject to change
without notice.

The information in this publication is provided “as is.” Dell EMC makes no representations or warranties of any kind with respect to
the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular
purpose.

Use, copying, and distribution of any Dell EMC software described in this publication requires an applicable software license.
SOLUTION OVERVIEW

Leading companies use Dell EMC’s Data Protection Solutions to simplify, accelerate and scale their backup and recovery environments.

Meeting long-term retention requirements (LTR) of data protection workloads has now shifted to include object storage in addition to tape.

Dell EMC’s CloudBoost enables data protection workloads from the Data Protection Suite or Symantec NetBackup to move de-duplicated LTR data to the cloud with three key value points.

- The data is de-duplicated and compressed, minimizing resource consumption (network bandwidth and cloud capacity) and maximizing cost savings (3 cents a month plus de-dupe means your cost to serve is very low).
- The metadata stays local to the CloudBoost appliance. This means micro pruning and retrieval is simple and easy, saving bandwidth and storage utilization. It also greatly enhances security since encryption keys (part of metadata) are stored uniquely in the CloudBoost appliance behind the customer’s firewall.
- Since the metadata is local, DP Search indexes this data, so it can be searched as well.

Trends Driving Demand:

- Cloud economics have changed the way customers think about storage.
- Customer desire to get out of the highly risky tape museum business, the costs of refreshing media and libraries when old versions become unreadable, and the desire to remove manual tape handling costs and risks
- The alternative of low cost and stable object storage

NETWORKER WITH CLOUDBOOST AND ECS

ECS is a turnkey, software-defined cloud-scale object storage platform that combines the cost advantages of commodity infrastructure with the reliability, availability and serviceability of traditional arrays. ECS features native multi-tenancy to deliver Storage-as-a-Service that can support multiple tenants and applications on a single, easy to manage cloud storage platform. ECS brings the cost profile, simplicity and scale of public cloud services to everyone with the trust, reliability and support you expect from Dell EMC. Benefits include:
Low Cost - High Scalability

ECS helps you cost effectively meet current application or data requirements and effortlessly scale to meet today's ever increasing storage demands. ECS can be up to 48% lower TCO when compared to public cloud storage providers.

Unlimited Application and User Support

Supports unlimited applications and users on a single cloud-scale storage system that can scale from Petabyte to Exabyte size.

Geo-Distribution

Features a multi-site, active/active architecture with a single global namespace that allows you to manage a geographically distributed environment as a single logical resource.

Policy Driven Geo-Protection

Through patent pending technology, ECS efficiently stores data with a unique approach that optimizes data production, efficiency and performance without compromise.

Future-proofed Archives

ECS is designed with an eye to the future and incorporates the industry’s most modern and widely used object interfaces including OpenStack Swift, Amazon S3, and Dell EMC's own Atmos REST and Centera CAS. ECS provides incredible deployment flexibility and helps to future proof your investment.

CloudBoost

EMC CloudBoost is cloud-enabling technology that facilitates secure, automatic, efficient long-term retention of backups in private or public cloud. CloudBoost eliminates the risks of tape, reduces cost, and increases IT and business agility by extending the reach of EMC Data Protection Suite to resilient, hyper-scalable cloud storage.

High Performance

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 to 32TB further speeds long-term retention backup and restore operations.

Cost-effective and Flexible

Select the virtual or physical CloudBoost model and cache size to fit your specific needs and budget. CloudBoost supports leading private and public clouds such as Dell EMC Elastic Cloud Storage (ECS) and integrates with Dell EMC Data Protection Suite. Moreover, the 2TB CloudBoost virtual edition is included at no additional charge with every Data Protection Suite license.

Secure

CloudBoost delivers enterprise-grade security even when data is stored or transferred outside your firewall. Data is segmented and encrypted at all times, in-flight and at-rest, and all data transfers occur over TLS. Add an EMC ECS private cloud for even greater control.

Scalable

Every CloudBoost virtual or physical instance can manage up to 6 petabytes of logical data in the cloud. CloudBoost is deployed in your datacenter, remote or branch office (ROBO), of wherever data is being backed up. Administrators manage one or many CloudBoost instances and cloud profiles from the Web-based Dell EMC Cloud Portal, a secure, single-pane-of-glass console.
WHY ECS & CLOUDBOOST

ECS is a cloud-scale object storage platform that when combined with CloudBoost allow customers of the Data Protection Suite (Avamar & NetWorker) or Symantec/Veritas NetBackup to perform LTR of data to the Cloud in a cost effective, secure fashion. This allows customers to not only stay away from risky tape based solutions but control the data in their own data center for ingress and egress of information.

1. **Archive** – Leading companies use Dell EMCs Data Protection Solutions to simplify, accelerate and scale their backup and recovery environments. Meeting long-term retention requirements (LTR) of data protection workloads has now shifted to include object storage in addition to tape. Dell EMCs CloudBoost enables data protection workloads from Data Protection Suite to move deduplicated LTR data to the cloud with three key value points. 1. The data is deduplicated and compressed, minimizing resource consumption (network bandwidth and cloud capacity) and maximizing cost savings (3 cents a month plus de-dupe means your cost to serve is very low). 2. The metadata stays local to the CloudBoost appliance. This means micro pruning and retrieval is simple and easy, saving bandwidth and storage utilization. It also greatly enhances security since encryption keys (part of metadata) are stored uniquely in the CloudBoost appliance behind the customer’s firewall. 3. Since the metadata is local, DP Search indexes this data, so it can be searched as well.

2. **Market Trends** – The main trends driving customers to use CloudBoost for LTR with ECS are:
   - Customer desire to get out of the highly risky tape museum business, the costs of refreshing media and libraries when old versions become unreadable, and the desire to remove manual tape handling costs and risks
   - The alternative of low cost and stable object storage combined with deduplication for LTR data.

3. **Cloud-scale** - ECS delivers object-based cloud storage at scale that can support an unlimited number and types of applications at a cost lower than public cloud storage providers. With high storage efficiency, multi-tenancy and worldwide access to data archives, ECS is ideal for archiving structured and unstructured data of any kind so it can be preserved, discovered or monetized as business needs demand.

4. **Flexibility** – ECS is a turnkey solution and is available in capacity points that meet the needs of enterprises of all sizes. Whether your environment is less than 150TBs or scales into Exabyte size, ECS provides multiple unstructured configuration options to suit your storage needs for today and for the future. Regardless of where you start you can be assured that capacity can be seamlessly scaled.

5. **Optimized Storage** – ECS is the latest object storage platform from Dell EMC incorporating all of the great features of Centera and Atmos while having a design center squarely focused on 3rd platform applications. ECS embraces the Open Source world with comprehensive implementations of Open Stack Swift and Amazon S3. ECS is the storage platform to take you on the journey from the 2nd to the 3rd platform.

SALES GUIDANCE

1. Target existing or new customers of Data Protection Suite (NetWorker and Avamar) or Symantec/Veritas NetBackup
   a. Top priority is NetWorker, since CloudBoost integrates tightly and seamlessly with NetWorker.
   b. Next priority is Avamar (the supported workflow is Avamar to NetWorker to CloudBoost).
   c. Remaining priority is NetBackup, where CloudBoost should be used when there’s no viable selling motion with Data Protection Suite in order to retain account control/presence and drag along sales of ECS.
   d. Prospect only in existing install base of these products or new Data Protection Suite or NetWorker sales opportunities.
2. Most all of the current install base for NetWorker, Avamar (Data Protection Suite) and Symantec NetBackup does LTR to tape. Prepare to engage in cost justification discussions.

3. CloudBoost at the most fundamental level eliminates the risks of tape while maintaining or lowering TCO by moving LTR backup copies from tape to cloud.

4. Supported Private Cloud Destination: ECS, ATMOS, OpenStack Swift – focus is all around ECS.

5. Supported Public cloud destinations of CloudBoost LTR: AWS S3 (not Glacier), Microsoft Azure, Google Cloud Storage, AT&T Synaptic.

6. You biggest sales hurdles will revolve around doing cost/reliability justification around Tape and Public Cloud providers vs. ECS.

7. Seek out customers that are forward thinking and looking to leverage Platform 3/Cloud scale economics in a solution that is:
   a. Durable
   b. High Performance
   c. Scalable
   d. Secure
   e. Flexible – Physical or Virtual appliance
   f. Easy Management
   g. Cost Effective

TECHNICAL REFERENCE ARCHITECTURE

INTRODUCTION

Dell EMC CloudBoost is based on technology from Maginatics, which EMC acquired in October, 2014. It is virtual or physical technology that cloud-enables Dell EMC Data Protection Suite and third-party Veritas (formerly Symantec) NetBackup.

CloudBoost

CloudBoost enables long-term retention (LTR) backup copies in public or private cloud (object storage), which can be far more reliable and less costly than tape. CloudBoost ingests backup clones from protection storage; encrypts, deduplicates, and compresses the data; and copies it to public or private cloud for secure, cost-effective, efficient long-term retention. The protection storage from which CloudBoost ingests backups is typically a Data Domain appliance; but with NetWorker it can be any AFTD, and with NetBackup it can be an NBU backup target such as MSDP. With CloudBoost, in-cloud LTR backup copies are managed via the native protection software interface (typically NetWorker Management Console) and can be restored, in full or incrementally, just like a local backup.

CloudBoost performs a “copy to cloud” operation from backups that are currently protected on a NetWorker local target, a Data Domain appliance, or – in the case of NetBackup – an NBU backup target such as MSDP. CloudBoost is NOT intended as a replacement for operational backup and recovery in general or Data Domain in particular. Operational backup and recovery is best suited to current Avamar/NetWorker and Data Domain solutions.

CloudBoost enables extended storage provisioning for data that must be protected on a lower cost tier of storage (i.e., long term retention in public or private cloud) while maintaining online performance and enterprise security and avoiding the disadvantages of tape, including the risk of data loss or corruption.

CloudBoost 2.0
CloudBoost 2.0 is the second major release of CloudBoost. It introduces the same features as version 1.0 but adds a hardware appliance and local caching capabilities. It also improves capacity scale from 400TB to 6PB, a 15X increase from version 1.0. Additionally, compression is enabled to provide better data reduction rates. The following is a list of the new features in Version 2:

<table>
<thead>
<tr>
<th>Feature</th>
<th>CloudBoost v1.0</th>
<th>CloudBoost v2.0</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalability</td>
<td>400TB logical</td>
<td>6PB logical</td>
<td>15X greater capacity per appliance</td>
</tr>
<tr>
<td>Compression</td>
<td>No</td>
<td>Yes</td>
<td>2x greater efficiency</td>
</tr>
<tr>
<td>Throughput</td>
<td>45-50MB/sec (random non-dedupe dataset)</td>
<td>120-150MB/sec (same dataset)</td>
<td>2x-3x higher throughput</td>
</tr>
<tr>
<td>Site Cache (Local Data Cache)</td>
<td>No</td>
<td>2TB or 6TB (virtual)</td>
<td>Faster backup and recovery. Cached objects accessible when network is down</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10TB or 32TB (physical)</td>
<td></td>
</tr>
<tr>
<td>Form Factor</td>
<td>Virtual</td>
<td>Physical and Virtual</td>
<td>Added flexibility</td>
</tr>
<tr>
<td>Centralized Management</td>
<td>No</td>
<td>Yes</td>
<td>EMC Cloud Portal: 10 pane management of multiple CloudBoost appliances</td>
</tr>
<tr>
<td>ESRS Integration</td>
<td>No</td>
<td>Yes</td>
<td>Register/unregister, Keepalive, Alerts, PhoneHome, etc.</td>
</tr>
<tr>
<td>Support for 3rd Party Protection Software</td>
<td>No</td>
<td>Yes (Veritas NetBackup)</td>
<td>Better In-Cloud LTR for existing NBU environments</td>
</tr>
</tbody>
</table>

**CONFIGURATIONS**

The CloudBoost appliance is comprised of two principal components. CloudBoost incorporates a NetWorker storage node and Advanced File Type Device to deliver a physical or virtual appliance that enables NetWorker to effectively clone data to cloud.

The primary interface to NetWorker is the Virtual File System where it overlays the File System presented by the AFTD. This enables CloudBoost to clone directly to the AFTD (the VFS) and send data to cloud via Maginatics IP. The CloudBoost Client is the primary agent that exposes a file system interface to NetWorker. The Metadata server (Virtual Filer) manages CloudBoost metadata including encryption keys, dedupe hashes, file to chunk mapping. The Cache helps accelerate backups, which is useful in cases where customer has a thin pipe to the object store.
CloudBoost 2 is available in both physical and virtual form factors. Virtual and physical appliances differ only in cache size. Both can support up to 6PB of logical capacity.

Note that logical capacity is different from the Front-end TB (FETB). For example, if a customer has 100TB of FETB, after 10 clones the CloudBoost appliance is managing 1PB of logical capacity.

- The CloudBoost physical edition:
  - CloudBoost 100-10 (10TB local cache)
  - CloudBoost 100-30 (32TB local cache)
- The CloudBoost virtual edition:
  - CloudBoost 100-2 (2TB local cache)
  - CloudBoost 100-6 (6TB local cache)

Differences among the CloudBoost physical editions are summarized in the table below:

<table>
<thead>
<tr>
<th>Category</th>
<th>EMC CloudBoost 100-10</th>
<th>EMC CloudBoost 100-30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>2U</td>
<td>2U + 3U</td>
</tr>
<tr>
<td>Usable data cache capacity</td>
<td>10 TB</td>
<td>32 TB</td>
</tr>
<tr>
<td>Disk bays</td>
<td>12</td>
<td>12 (+15 External DAE)</td>
</tr>
<tr>
<td>Processor</td>
<td>1X Intel Xeon 2600 Series</td>
<td></td>
</tr>
<tr>
<td>RAM</td>
<td>Up to 64 GB (8GB RDIMM)</td>
<td></td>
</tr>
<tr>
<td>Disk drives</td>
<td>12</td>
<td>27</td>
</tr>
<tr>
<td>Redundant power &amp; cooling</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>10Gb NIC ports</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6Gb/s SAS ports (DAE connect)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Mini-SAS with 4 ports of 6Gb/s (DAE connect)</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**CloudPortal**

CloudBoost V2.0 is managed using the Dell EMC Cloud Portal (the platform is internally called Cloud Portal and Services, or CPS). The Cloud Portal is a secure, Dell EMC-hosted, single-pane-of-glass console for centrally managing and administering every CloudBoost appliance within the enterprise. It replaces the separate local UI that was used for CloudBoost V1.0 and enables registration, monitoring and management of multiple appliances, cloud profile creation, reporting, logging, alerts and more. Through CPS, each CloudBoost appliance reports on the health of the appliance and cloud storage usage. However, as before, there is no integration between the CloudBoost UI (now Cloud Portal) and the NetWorker or NetBackup Management console.
The CPS is a platform that provides Dell EMC product teams the ability to easily enable administration functionality for their products through the EMC Cloud Portal. As of Q4 2015, key CPS functionality for CloudBoost includes:

- Appliance registration
- Cloud Profile creation
- Identity and Access Management (IAM)
- Initial appliance configuration
- Reporting
- Events and notifications
- Upgrade
- Disaster Recovery (of a CloudBoost appliance)

Through the CPS, the administrator specifies policies whereby CloudBoost metadata is copied, encrypted (using a separate encryption process from that used by CloudBoost to store data chunks in the cloud), and backed up periodically to the same cloud target as the backup data. In case of failure of a CloudBoost appliance, the metadata can be restored from the cloud to another CloudBoost appliance using the simple CloudBoost Disaster Recovery workflow accessible through the Cloud Portal.

**NetWorker**

The workflow for NetWorker integration remains unchanged from version 1. CloudBoost is tightly integrated with 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.

The backup admin creates a backup to either a Data Domain system or an AFTD. They then create a local mount of the object store using the AFTD device type and perform a NetWorker cloning operation to it. Because the NetWorker infrastructure is being leveraged, the backup information is stored in the NetWorker central catalog and the cloning process can be scheduled with retention policies assigned to it. Data is rehydrated when it is cloned from its local target to CloudBoost, resulting in a single rehydration of data before it hits the object store.
Recovery from CloudBoost, although not a direct path from Cloud to Client, it is still a one-step operation that goes from Object Storage through the CloudBoost virtual appliance (via the internal storage node). Backup target is not involved in the recovery.

CloudBoost will be supported with NetWorker 8.1.x, 8.2.x and NetWorker 9.0. However, certain NW 9 clients are not supported.

Supported NetWorker 9 clients:

- File system client
- NMDA client
- NMSAP client
- NMMeditech client
- NSM client

NMM 9 client is not supported with CloudBoost v2. This means CloudBoost 2 on NetWorker 9 cannot be used with Microsoft applications. Support for NMM (and therefore Microsoft apps) with CloudBoost 2 and NetWorker 9 is planned for after NetWorker 9 GA. The VBA 9 client (solution for VM protection) is expected to be supported with CloudBoost v2 by the time NetWorker 9 GAs.

All NetWorker 9 clients require a NW 9 server. Additionally, NMM 9 clients require a NW 9 storage node as well. CloudBoost 2.0 uses an embedded NetWorker 8.2.1 storage node. After NW 9 GA, the current roadmap upgrades the embedded NW storage node within CloudBoost to support NMM 9 clients. Timeline planned is H1 2016.

**Avamar to NetWorker**

Currently there is no “pure” Avamar/CloudBoost workflow. An Avamar (or Avamar + DD) deployment requires a 2-step process to enable long-term retention to the Cloud because Avamar workflow relies on NetWorker to send its data to cloud. Avamar exports its data to a staging server using Avamar Data Migration Enabler (ADMe) so it can be ingested by NetWorker. The NetWorker workflow then exports data to cloud via cloning. There is a no “pure” Avamar/CloudBoost workflow at this time.

For Avamar, the process leverages Avamar Data Migration Enabler, which performs a redirected restore to a staging server. NetWorker then ingests the data from the staging server, from which it is then cloned to cloud. The result is two rehydrations of client backup data before it hits the object store.
Leveraging the NetWorker flow is required for many compliance use cases because the transfer of data is maintained and recorded by a separate backup product (in this case, NetWorker).

**NetBackup**

CloudBoost represents an “ace in the hole” in accounts where there is no viable selling motion for NetWorker or NetWorker/Avamar because the customer is “married” to NBU. In this scenario, CloudBoost offers a better path-to-cloud for NBU users than NBU’s native technology, as explained below; so CloudBoost can help our account team maintain account control/presence while also possibly dragging along the sale of EMC object storage such as ECS.

To use CloudBoost, a NetBackup user must have an NBU Media Server running Windows 2012. Unlike Dell EMC NetWorker (which requires only the CloudBoost appliance), deploying CloudBoost with NBU requires installing a CloudBoost Client (Maginatics Client) on the NBU Media Server, which presents a NAS share to the media server using NBU’s “Advanced Disk” option. A duplicate flow in NBU is used to rehydrate data out of Data Domain (or equivalent) to the CloudBoost NAS share. As with other CloudBoost deployments, the appliance chunks, encrypts and deduplicates the data before pushing it to the object store.

The following diagram illustrates the deployment:
The NetBackup workflow requires a CloudBoost Client to be installed on the NetBackup media server. The solution presents a NAS share to the media server using the “Advanced Disk” option. Duplication workflow 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.

**SOFTWARE ARCHITECTURE**

On the backend (where data is transferred to the object store), CloudBoost supports the leading object storage protocols. On the front end (where CloudBoost connects to a protection storage node), CloudBoost employs an agent (CloudBoost Client) -- embedded with NetWorker and deployed as a software agent with NetBackup -- which enables CloudBoost to communicate via the protection software’s native protocol.

A single CloudBoost appliance can manage 6PB of logical data. Once this limit is reached, the customer must add another CloudBoost appliance. However, each CloudBoost has its own namespace, so any deduplication benefits would not carry over from one appliance to another.

**Object Protocols**

Private object stores supported by CloudBoost include Dell EMC ECS, Dell EMC Atmos, and OpenStack Swift. Public object stores supported by CloudBoost include AT&T Synaptic Storage (powered by Atmos), Amazon Web Services S3, Google Cloud Storage (including Nearline), and Microsoft Azure.

CloudBoost does not support Amazon Glacier. RTO of data from Glacier is undesirable and increases TCO of the solution. Further reading is encouraged on Tom Poole’s [Competitive Wiki](#) and [Pricing Guide](#) on AWS Glacier and Mark Gould’s [Inside Dell EMC Post](#), “Is it really $.01?” and his [Cloud TCO Calculator](#) (new TCO calculator coming soon). For technical reasons, Glacier support would require the backup software (NetWorker, Avamar, and DDBoost) to implement Quality of Service, which is not implemented today. (Without QoS, a restore operation from Glacier might for example have to run for 10+ hours without timeout.)

In addition to Standard Google Cloud Storage, CloudBoost does support Google Nearline storage, which is similar to AWS Glacier, except that Nearline offers ~3 second rather than ~3 hour retrieval SLAs.

**Workflow**

A single CloudBoost appliance can only connect to one object store. Multi-cloud namespace support is in-plan for future versions of CloudBoost. For now, the mapping of CloudBoost appliance to target object store is one-to-one. Data is not sent to cloud in the native backup format of an Avamar, NetWorker, NBU, or Data Domain target. This effectively means CloudBoost is **NOT** enabling Avamar replication of Avamar data/metadata chunks to cloud, or Data Domain-managed file replication of CDSF containers to cloud, etc.
Backup data is rehydrated from a local NetWorker, Avamar, Data Domain or NBU target to CloudBoost, where it is re-chunked and re-deduped using proprietary CloudBoost (Maginatics) algorithms before being sent to the cloud. Therefore, one will observe some variation in the amount of data sent to or stored in to cloud relative to data sent to an Avamar or Data Domain system.

The following diagram illustrates the data and metadata paths for CloudBoost:

Backup data is deduped source-side before being sent over the network to cloud, similar to the workflow Avamar employs. The CloudBoost Client will segment all data entering the virtual file system and store the encrypted data to an object store as data chunks. The metadata needed to address and decrypt these chunks is then stored on a virtual filer (the index engine of sorts). To reconstitute a file, CloudBoost will decrypt and reassemble all data chunks stored in cloud (or the “Site Cache” local data cache), with the operation again arbitrated by the virtual filer. Dedupe is source-side-verified by the virtual filer before being sent to cloud.

Data is restored to clients through CloudBoost. This affords the ability to appear much like a direct-from-cloud restore; however, the data path actually involves the NetWorker storage node, so the transfer from Cloud to CloudBoost occurs over a WAN (unless the object store and CloudBoost appliance are co-located, in which case it’s a LAN), while the transfer from CloudBoost to clients occurs over a LAN.

CloudBoost itself does not manage retention policies – it relies instead on the protection software to make deletions to the virtual file system as appropriate. For NetWorker, the standard backup retention policies will apply and act as retention policies for backups cloned to cloud. There is no secondary “Extended Retention” policy as there is in Avamar. NetWorker will manage the retention, and CloudBoost will perform the Garbage Collection of data stored in cloud. Once a backup has expired in NetWorker, CloudBoost will trigger this backup to be cleaned up in its next garbage collection cycle.

**WAN Optimizations**

CloudBoost employs variable-length chunking, but the average chunk size is 256KB. This represents a design optimization across total addressable storage, target object store capabilities, and WAN and storage efficiency. As a result, the deduplication factor with data sets comprised of smaller files will typically see lower ratios while those with larger files such as VMDKs can achieve higher ratios.

CloudBoost v2.0 includes a local data cache, also referred to as a “Site Cache.” As mentioned earlier, the choice of Site Cache size varies between the virtual and physical form factors. The virtual form factor comes with either a 2TB or 6TB Site Cache, while the physical form factor comes with either a 10TB or a 32TB Site Cache. Upgrading the 2TB virtual edition to 6TB is a simple software upgrade. Upgrading the 10TB physical edition to 32TB requires field installation of another component (DAE).

The Site Cache, an integral component of CloudBoost (physical or virtual), functions as a large, persistent cache for data transiting the CloudBoost appliance to or from the cloud target. Site Cache is included with every CloudBoost 2
The Site Cache, an integral component of CloudBoost (physical or virtual), functions as a large, persistent cache for data transiting the CloudBoost appliance to or from the cloud target. Site Cache is included with every CloudBoost 2 appliance but can be turned off when it would add overhead without adding additional value; i.e., when CloudBoost and the cloud target are connected by a high bandwidth, low latency, reliable network (typically a LAN).

Whenever CloudBoost connects to the cloud target via a WAN or other low bandwidth, high latency or unreliable network, Site Cache can speed both backup and restore LTR operations. Typical WAN scenarios are remote and branch offices (ROBO) pushing data to a public cloud or a central private cloud over WAN, and data centers pushing backups to a public cloud over WAN. Site Cache allows administrators to move backups to the CloudBoost very quickly over the LAN. From there, data is pushed to the cloud in the background at whatever throughput is possible given the network characteristics.

Another benefit Site Cache conveys is the ability to quickly access data resident in the cache even when the network connection is down. However, note that these benefits are limited to situations where the sought-after data is cache-resident. The Site Cache uses FIFO (first in, first out) eviction, so a given restore request will miss the cache if the cache has been thrashed. But clean data resident in the Site Cache is restored at LAN speed, even if the connection to the cloud is unavailable.

Site Cache always preserves strong data consistency. It is also designed to transition from a read-write to a read-only state when running out of space, when the network link between Site Cache (CloudBoost) and cloud storage is lost completely or becomes very limited (making the Site Cache egress rate zero or very low compared to its ingress rate), or upon disk failure.

Site Cache is not appropriate when cloud storage service is LAN_accessible (i.e., over a reliable high bandwidth, low latency connection).

**BEST PRACTICES**

**Load Balancer**

CloudBoost connects to ECS using a single hostname or IP Address. Therefore, a CloudBoost appliance can connect to either a single ECS node, or a load balancer that will distribute the communication between all ECS nodes. Best practices recommend using a scalable load balancer so CloudBoost appliances do not need to be tied 1:1 with ECS nodes.

**Multiple Clone Sessions**

For parallelism, we recommend creating multiple AFTD devices under the /mnt/magfs/ base mount point within CloudBoost, and using one clone session per device. One session per device is recommended for optimal de-duplication. Multiple clone sessions to the same device can result in lower de-duplication ratios and longer clone times.

**CloudBoost Data Migration**

Data cannot be migrated from one object store to another today using CloudBoost. To change cloud targets, a re-deployment of the CloudBoost appliance is needed. Follow instructions provided in the CloudBoost with NetWorker Integration Guide. To retrieve the data, the best practice is to download to a local NetWorker server.

**Virtual Appliance Minimums**

CloudBoost virtual appliance is packaged as an OVA for ESX 5 or later. Minimum requirements are:

- RAM: 64GB (32GB with Site Cache disabled)
- CPU: 16 vCPU (8 vCPU with Site Cache disabled)
- Disk: Metadata Default is 10GB, which must be resized prior to boot. 1GB can support 2TB of logical data (e.g., 100GB of disk will support approx. 200TB of namespace capacity)
Bandwidth Minimums

For proper operation, it is recommended to have a minimum bandwidth of at least 10 MB/s to the cloud with a maximum latency of less than 100 ms RTT for the CloudBoost solution. Extremely low bandwidth links may result in backup and restore timeouts.

PERFORMANCE DATA

Performance Tests

Test Setup

The following results were obtained using virtual CloudBoost appliances with appliance cache disabled. The ECS tests connected to a single U300 ECS system with either 1 or 4 active nodes. When connected to a single ECS node, the CloudBoost appliances used the physical IP of the ECS nodes without a load balancer. CloudBoost concurrency to ECS was tuned to 128 threads per ECS node using the CloudBoost tuner.

ECS Results

Using a dataset with 2x data reduction – 50% compression, no dedupe
- A single CloudBoost appliance to 1 ECS node: 194 MB/s ingress by a CloudBoost appliance
- 4 CloudBoost appliances to 4 ECS nodes: 400 MB/s ingress by CloudBoost, 200 MB/s ingress into ECS
- Results in ~1.4 TB/hour (~33.6 TB/day) of throughput to cloud

Using a dataset with no compression or dedupe
- A single CloudBoost appliance to 1 ECS node: 120 MB/s
- 4 CloudBoost appliances to 4 ECS nodes: 290 MB/s
- Results in ~1 TB/hour (~24 TB/day) of throughput to cloud

AWS Results

Using a dataset with uncompressible data, and no dedupe, and Cache disabled
- Single CloudBoost appliance to AWS S3 – 190 MB/s (~1.5 Gbps)
- WAN will most likely become the bottleneck before this

Using a dataset with 50% compressible, non-dedupe data, and Cache enabled
- 25 MB/s ingress into cache with a 10TB physical appliance
- 50 MB/s ingress into cache with a 32TB physical appliance

PERFORMANCE AND DEDUPE

Test Setup

The following tests used virtual CloudBoost appliances with appliance cache disabled using a storage pool across 2 x U700 ECS systems with 1 or 4 active nodes. All CloudBoost appliances used the physical IP of the ECS nodes without a load balancer. CloudBoost concurrency to ECS was tuned to 128 threads per ECS node using the CloudBoost tuner. Performance and Dedupe Results
Using a dataset with 2x data reduction – 50% compression, no dedupe

- A single CloudBoost appliance to 1 ECS node: 200 MB/s ingest by a CloudBoost appliance
- 4 CloudBoost appliances to 4 ECS nodes: 420 MB/s ingest (105 MB/s per CloudBoost)
- Results in ~1.4 TB/hour (~33.6 TB/day) of throughput to cloud

Using a dataset with no compression or dedupe

- A single CloudBoost appliance to 1 ECS node: 150 MB/s
- 4 CloudBoost appliances to 4 ECS nodes: 380 MB/s (95 MB/s per CloudBoost)
- Results in ~1.3 TB/hour (~31.2 TB/day) of throughput to cloud
- Use only for datasets where you expect poor to no data reduction

Monthly clones to cloud with 15% monthly change rate and retention period of 6 months (6 clones)

- Filesystems – 4x (75% dedupe)
- MSSQL Database – 4x-6x (75% - 84% dedupe)
- VMware Image – 20x (95% dedupe)

SIZING GUIDANCE

Bandwidth

WAN bandwidth is the most common bottleneck. A properly-resourced CloudBoost appliance can saturate a 1 GB/s link with 30 ms RTT latency without hitting any limits within the VM itself. Object store ingest limits are another potential bottleneck. In some cases it is possible to reach the objects/sec limit that can be sustained by a single logical container in the object store. Additionally, an Avamar solution is constrained by the restore rate ADM(e) can sustain from the Avamar Data Store or the Avamar and Data Domain integrated system. For the Avamar Data Store, the restore rate ranges from 37.5 GB/hr/node on the M1200 to 75 GB/hr/node on the M2400. For Avamar and the Data Domain system, please review the Data Domain system specification sheets.

Bandwidth to the object store will also dictate the choice of appliance type. Customers with strong connectivity to the object store, with bandwidth over 400Mbps (50MB/) to the object store, it is recommended to connect directly to the object store with appliance cache disabled. Select the smallest virtual or physical appliance. Generally customers using an on-premises ECS should have strong connectivity. Example: a customer connecting to an on-prem ECS. Recommend choosing the smallest virtual appliance in this case and disabling cache. If customer prefers a physical form factor, choose the smallest physical appliance and turn cache off.

Customers with weak connectivity to the object store on the other hand, with bandwidth under 200 Mbps (25 MB/s), will benefit from any form of caching. Choose an appliance cache size based on backup dataset size. Generally, customers using a public object store should benefit from cache depending on their WAN link.

Customers with medium connectivity (200 Mbps< Bandwidth< 400 Mbps) can still benefit from caching if they select the largest physical appliance. If they choose any other appliance with a smaller cache size, it is recommended they connect direct to the object store with cache disabled.

Physical or Virtual Appliance

Choose a Physical appliance if the customer does not have access to VMware which is required for the virtual appliance, or the customer does not have the disk resources to size the VM, or the customer prefers not to deal with the sizing of the virtual appliance.

Choose a Virtual appliance if the customer does not want to manage hardware and prefers a software only footprint, or the customer wants the flexibility to start with a small logical capacity and cache size and expand as needs grow

Virtual Appliance Resource Sizing
With Appliance Cache enabled, best performance can be obtained through the following settings:

- CPUs: 16 vCPUs
- Memory: 64GB
- Metadata Disk: 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 Disk:
  - Minimum of 200GB required for cache. Supports up to 2TB or 6TB of cache depending on the virtual appliance purchased.
  - Cache should be sized based on dataset size so backups can benefit from caching
- With Appliance Cache Disabled:
  - CPUs: 8
  - Memory: 32GB
  - Metadata Disk: requirements are the same. No space needed for cache.

In general, with a load balancer the number of CloudBoost appliances can be calculated as (Ingest required into CloudBoost) / (194 MB/s ingest per CB). The number of U300 systems can be calculated as [(Ingest required into CloudBoost) / (Data Reduction rate)] / (290 MB/s per U300).

**Example: Performance Sizing**

Consider a dataset with 4x data reduction that requires ingest of 1000MB/s into CloudBoost: Using the formula above, with a load balancer a single U300 can sustain 290MB/s of data ingress from CloudBoost. For this dataset with 4x data reduction, the U300 sees 250 MB/s of data ingress from CloudBoost. The Number of U300 systems = (1000 / 4) / 290, or about 1 U300 system. Each CloudBoost appliance can support 194 MB/s of data ingest. Therefore, the number of CloudBoost appliances needed = 1000 / 194, or about 6 CloudBoost appliances. This configuration would require a load balancer to connect the 6 CloudBoost appliances to the 4-node U300

Without a load balancer, 4 CloudBoost communicating with a single U300 node should be able to achieve 400 MB/s ingest. The number of “4 CB + U300” systems required = (1000 MB/s) / 400 MB/s, or about 3 U300 nodes + 12 CloudBoost VMs. The CloudBoost appliances tied 1:1 to the physical IP of one of the U300 nodes.

**SECURITY GUIDANCE**

Encryption keys are managed by the system and are not user configurable. All data is “chunked,” each chunk is individually encrypted with its own AES-256 bit key, and all keys are stored in the CloudBoost appliance itself (save for the fact that the keys are included in the metadata that is itself encrypted and stored in the cloud for CloudBoost disaster recovery purposes). All data handled by CloudBoost is encrypted in-flight and at-rest.

To allow recovery to a secondary appliance in case of failure of the primary appliance, all metadata including encryption keys are backed up to the cloud periodically. Unlike competitors like Nasuni who place this burden of backing up encryption keys entirely on the customer, CB backs up encryption keys periodically to the object store. All backups of CB metadata are sent to the object store encrypted and the master key for these metadata backups is stored within the Cloud Portal. This master key today is automatically applied during recovery of an appliance.

Some customers have expressed concerns about the master key being at risk if the cloud portal credentials are compromised or EMC having access to all the customer data since it hosts the Cloud Portal and has the master encryption key.
To address this concern, CB plans to use asymmetric keys for encryption of the metadata backups. The master key for CB metadata backups will use a private/public key pair where unless the customer provides his portion of the key, an appliance cannot be recovered on failure. This ensures that a compromise of the Cloud Portal credentials does not put customer data at risk. It also ensures nobody including EMC has access to the data except for the customer.

Other things being planned to improve security of the Cloud Portal include Role-based access, automatic account lock-out, notifying a customer when recovery of an appliance is initiated.