EMC CLOUDARRAY
PRODUCT DESCRIPTION GUIDE

INTRODUCTION

IT organizations today grapple with two critical data storage challenges: the exponential growth of data and an increasing need to keep more data for longer. Together, these two factors strain existing data centers, available staff, and stagnant IT budgets, forcing companies to seek innovative solutions that can support on-premise infrastructure investments and requirements while providing the flexibility for uncertain future data growth.

The growing availability of both public and private cloud storage has given organizations of all sizes access to a wide range of elastic storage options that make it easier and simpler to manage storage growth and complexity. For most companies, on-premise data centers remain critical for performance, security or complex architecture requirements, but second-tier, archive and backup data can be shifted to lower cost, yet still accessible, options such as public or private cloud storage.

This paper explains how organizations can use CloudArray cloud-integrated storage to enable hybrid cloud storage, marrying the lower cost and scalability of cloud storage with the high-performance, high-availability requirements of production systems.
# TABLE OF CONTENTS

**INTRODUCTION** ................................................. 1

**CLOUDARRAY OVERVIEW** ........................................ 3

**FEATURES AND FUNCTIONALITY** ................................. 3

**CLOUD STORAGE CONNECTIVITY** ............................... 3
  Protocol Management ............................................. 3
  "Native soil" and specialized cloud storage connectivity ........ 4
  Supporting a hybrid environment ................................ 4

**INTELLIGENT DYNAMIC CACHE** ............................... 4
  Cache Configuration ............................................. 5
  Cache management, sizing and migration ....................... 6
  Cache in the context of cloud replication .................... 6
  Cache in practice ............................................... 6
  Fast cache reload ............................................... 7

**SECURITY** ................................................................ 7

**DATA PROTECTION AND DISASTER RECOVERY** .............. 7
  Snapshot scheduler .............................................. 7
  Backup connectivity ............................................ 8
  On-demand disaster recovery .................................. 8
  Live fire drill mode ............................................. 8

**DATA MOVEMENT** .................................................... 8
  Minimizing Network Impact for day-to-day operations ........ 8
  Cloud-to-cloud migration ...................................... 9
  Secure bulk ingestion .......................................... 9

**CLOUDARRAY MANAGEMENT** ..................................... 9

**FORM FACTOR** ...................................................... 9

**CLOUDARRAY ADMINISTRATION** ............................... 10

**SUMMARY** ........ ERROR! BOOKMARK NOT DEFINED.
CLOUDARRAY OVERVIEW

CloudArray enables organizations to maximize on-premise infrastructure for production systems while offloading second-tier data to public and private cloud storage.

Applications interact with the CloudArray appliance via SAN (iSCSI) and/or NAS (NFS/CIFS/SMB/SMB2) volumes. CloudArray caches data locally for rapid access and automatically replicates all data to a public or private cloud.

This approach eliminates the custom API integration and lengthy learning curve typically required by cloud storage integration, making it easy to connect existing infrastructure to the cloud with minimal disruption or delay.

In addition to making object storage accessible via SAN and NAS interfaces, CloudArray also provides a wide range of features designed to address the common concerns and challenges that typify a cloud storage deployment including security, performance, network bandwidth, ingestion, data reduction, data protection and cloud-to-cloud migration.

FEATURES AND FUNCTIONALITY

Cloud Storage Connectivity

The object-based storage that typifies cloud deployments provides a number of advantages (such as greater scalability and improved manageability). Yet, integration of object-based storage into an organization’s existing environment requires custom integration with providers’ web services-based APIs. Not surprisingly, this customization blunts the inherent advantages of cloud storage and delays implementation by requiring highly skilled developers, extensive testing, and considerable risk management.

PROTOCOL MANAGEMENT

CloudArray has been programmed to work directly with the APIs of more than 20 public and private cloud storage providers. On the front end, CloudArray converts the
cloud's object-based storage to one or more local volumes that can be mounted on existing servers using familiar protocols such as iSCSI, NFS, CIFS, SMB or SMB2. In the case of iSCSI access, CloudArray presents raw, block-level volumes to the servers where they can be mounted and formatted. For file-based systems, CloudArray presents standard file shares. Administrators can create many different volumes on each CloudArray instance, each volume using either SAN or NAS protocols based on requirements, and each of which can be individually configured to meet the performance and availability requirements of the server.

"NATIVE SOIL" AND SPECIALIZED CLOUD STORAGE CONNECTIVITY
For performance, security and regulatory reasons, organizations may elect to move some data to a private cloud they control, or alternatively connect to public cloud storage vendors that have been vetted based on their geographic location or industry-specific capabilities.

Administrators can easily associate specific volumes with a more restrictive cloud provider (whether public or private).

SUPPORTING A HYBRID ENVIRONMENT
CloudArray has the unique ability to support multiple clouds and multiple protocols simultaneously from the same instance. That capability means that the administrator can, for instance, simultaneously manage:

- A CIFS-based archive volume connected to a low-cost, low-availability public cloud
- An iSCSI-based backup volume connected to a different, standard service public cloud
- And an NFS-based file share connected to a highly secure private cloud

Intelligent dynamic cache
Cloud skeptics often cite latency as a key barrier to cloud storage adoption. This criticism is justified as the use of raw cloud storage by itself – especially for large files or production systems – can certainly introduce unacceptable delays.

To address this, CloudArray takes advantage of local storage (termed “cache”). The cache serves two purposes: 1) to provide local performance for active data, minimizing – and in many cases eliminating – the effects of cloud latency, and 2) to serve as a buffer for read-write operations.
In a CloudArray physical appliance, the total amount of cache available is pre-determined based on hardware configuration, whereas in a virtual appliance the cache can be any local storage accessible by the CloudArray software.

In fact, a CloudArray virtual appliance can be mapped to multiple different storage types based on requirements. For example, solid state drives (SSDs) for a cache associated with high-performance or production systems, and SATA drives for archive or other data with lower performance and availability requirements.

Ultimately, the performance of a particular volume depends on two key factors: the amount of data kept locally in the cache (and whether the requested data is on cache) and the type of disk used for the cache.

**CACHE CONFIGURATION**

Each data volume can be associated with its own cache (a “dedicated cache”) for better performance or alternatively operate off of a communal pool. In addition, the amount of cache assigned to each volume can be individually configured as well.

For infrequently accessed volumes used for archive or lightly accessed file sharing, cache sizes of 1-10 percent of total capacity are often sufficient. This cache can come from a cache pool that is shared across volumes. All of the data is available to the user, but only a small portion is also kept locally. This method makes it easy to scale out the total amount of storage available without expanding local footprint.

For volumes with higher performance requirements, a larger, dedicated cache is often optimal. In this case, you may want to size the cache up to 100 percent of the size of the volume, ensuring that you have one full copy on site and one full copy in the cloud. This provides significant accessibility and performance advantages. Since all data as all operations will be performed against the local cache rather than the cloud volume. As a result, performance will be based on the speed of the storage used for cache.

In many cases, cache size in between these two extremes is ideal. For example, a volume dedicated to backups might have a cache just large enough for the last full backup. That way, day-to-day operations, such as file recovery or test procedures will take place locally. Yet, in the background, a copy of the backup is being sent offsite, as are all older backups. The offsite copy of the data can be easily accessed in the event of a disaster.

Regardless of its size, the cache will always contain the most recently accessed data across its associated volumes. Less frequently accessed data will be tiered solely to the cloud. All data – whether locally cached or not – will be presented to and accessed by users in the same way, though obviously cached data will perform faster than data that needs to be retrieved from the cloud.
CACHE MANAGEMENT, SIZING AND MIGRATION

As needs change, administrators will often find that the originally configured cache size is no longer appropriate. In these instances, cache size can easily be grown via the CloudArray user interface as long as there is available disk. If, for example, you have a volume that is pushing larger amounts of data to the cache, you may want to grow your cache proportionally to the new volume size.

On the other hand, you may have configured a very large cache to support a rapid data transfer of archive data to the cloud, but once all the data has been ingested, you may want to scale that back to an “everyday level” of one percent. You can do that either by reducing the disks associated with your cache or by migrating the volume to a smaller cache.

As data evolves, organizations can use volume migration functionality to move “hot” volumes to larger and/or faster cache (i.e. SSD) and cold/archive volumes to smaller and/or slower cache (i.e. SATA).

CACHE IN THE CONTEXT OF CLOUD REPLICATION

Since data replication to the cloud happens asynchronously, CloudArray also uses the cache to hold “dirty data” (data that has been written to the cache but not yet replicated to the cloud). The local cache must be large enough to offset the difference between the amount of incoming data and the amount of bandwidth available for replication to the cloud.

CACHE IN PRACTICE

Let’s take a look at how this looks in a practical environment.

On the left are the users and applications that need to access storage. In this example, Acme Company that has four different types of use cases – files, backups, archive data and primary storage. Each data type has different requirements.

For data the company uses all the time, like files and other primary data, they want their data stored on a fast disk. But in the case of primary data, Acme wants ALL of its application data to have fast, local performance, so it has configured that volume with a dedicated 100 percent cache, which means that a copy of all the data is kept locally.

For files, on the other hand, Acme actively uses only about 30% of its data, so it keeps an on-site cache of 30 percent. CloudArray can figure out which data Acme is most likely to need (based on what’s been accessed most recently), keep that data in the cache and push everything else up to the cloud.

For data such as backups and archives, on the other hand, Acme doesn’t need super fast disks. Instead, it uses cheaper, slower disks for the on-site cache, with a cache size that is just a small proportion to the overall volume size.

When Acme needs more storage, it simply resizes an existing data volume or adds a new volume to CloudArray. In addition, when a volume ages out of constant usage (for example, a project has been completed), administrators can migrate that volume.
from an active policy associated with a fast disk and a large cache to an archive-type policy with a small cache based on slower disks.

**FAST CACHE RELOAD**

Finally, CloudArray also provides the administrator the option of populating empty cache space with data from the cloud at key points in the management of the appliance. Following events such as disaster recovery, exposing snapshots and volume migration, administrators can pre-populate the cache with cloud data, providing faster response times for applications reading shares and volumes.

**Security**

In addition to the strong physical site security that cloud storage providers (CSPs) deploy at their data centers, many CSPs also provide data encryption capabilities as a standard set of their service. However, for many organizations, such protection does not satisfy even moderate security requirements since the CSP owns and manages the encryption keys rather than data owner.

As a result, CloudArray provides a multi-layered AES-256 bit encryption strategy to protect data. Raw data and metadata are encrypted separately, with two different sets of keys. Those keys are always kept locally, rather than with the data in the cloud provider, so the CSP never has access to the keys. Keys can either be solely managed by the end user locally or a copy of the keys can be sent to the CloudArray portal in case of disaster. However, the keys themselves are encrypted through password protection, providing further assurance that access to data will remain limited to authorized entities.

**Data protection and disaster recovery**

A common use case for cloud storage is for offsite data protection and improved disaster recovery. CloudArray includes a number of specific features to make backup and disaster recovery procedures easier and more reliable.

**SNAPSHOT SCHEDULER**

CloudArray provides space-efficient snapshots and advanced scheduling and retention functionality, to provide secondary – and in some cases primary – protection for data in the cloud. When a volume snapshot is taken, all subsequent changes to the volume are only applied to the primary copy, while the original data remains intact for the snapshot copy. The snapshots are both space and bandwidth-efficient and do not experience the overhead of traditional “copy-on-write” snapshot technology.

CloudArray snapshots reside in the cloud and can be accessed for read and write purposes as soon as the snapshot completes. Snapshots are invoked on the appliance, using the GUI built in to the administration console. Using the snapshot scheduling GUI, age-based retention policies can be established as well. To access a snapshot, the snapshot is “exposed” by mounting and presenting it as a new volume. In this way, data from snapshots can also be used for more detailed analysis without impacting production data.

CloudArray snapshots provide a 20x capacity savings over traditional full backups, reducing space requirements while providing continuous, offsite data protection in the cloud and point-in-time recovery capabilities.

Snapshots can be used either in place of or in concert with traditional backup strategies. Many customers choose to use a schedule of snapshots to backup large file servers or application data that would otherwise take days to backup.
BACKUP CONNECTIVITY
Since CloudArray can present both SAN and NAS interfaces, organizations can use CloudArray as a target for nearly any backup software. The backup server sends data to CloudArray; CloudArray loads it into its local cache, and in the background uploads it to cloud storage. There’s no need to replace current backup software or modify existing backup policies.

CLOUDARRAY RECOVERY
For many companies, cloud storage represents an opportunity to significantly improve and simplify their disaster recovery strategies. To support this use case, CloudArray has a large number of recovery-focused features to ensure organizations can gain access to their data quickly in the event of a disaster.

CloudArray’s ability to operate as a virtual machine means that if an organization experiences a site disaster or some kind of hardware failure, administrators can spin up a new instance of CloudArray within minutes on whatever they have available – a secondary box, a secondary site or even in a cloud compute environment.

Administrators login to the CloudArray portal to download the virtual machine. During the installation process, administrators will be presented with the option to restore an existing implementation using a backup configuration file stored in the CloudArray portal. Within minutes, a fully functional replica of the affected CloudArray license will be available with all the volumes, capabilities and configurations as the original.

When cloud compute is used as a recovery site, a cloud-specific image of CloudArray is available for deployment and can be used to interact with either on-premise or cloud-based applications. By using CloudArray as part of a cloud compute-based recovery strategy, organizations can save significantly on the ongoing costs of maintaining a separate recovery site.

Live fire drill mode
In addition, CloudArray has a live fire drill mode, which enables organizations to test their disaster recovery strategy without impacting production systems. When using this mode, administrators can recover a live CloudArray on another site and confirm read/write access to volumes without shutting down production systems. None of these changes are replicated back to the cloud, guarding against data corruption and ensuring that the production system remains the master.

Data Movement
MINIMIZING NETWORK IMPACT FOR DAY-TO-DAY OPERATIONS
For many organizations, cloud storage usage comes with concerns about bandwidth monopolization. CloudArray includes sophisticated bandwidth throttling to optimize network performance.
With a simple graphical interface, administrators can schedule bandwidth allocation to CloudArray based on their requirements. For example, during production hours, when bandwidth usage is highest, the CloudArray appliance can be throttled back to avoid resource contention. Administrators can use the interface to create, display or alter recurring schedules with 15-minute granularity.

In addition, CloudArray "chunks" data together before sending, thereby reducing overall transaction and network charges, as well as minimizing network impact. The size of those chunks can be adjusted for optimal performance.

CloudArray also applies compression to each cache block before it is sent to the cloud. CloudArray customers typically report a 50% or more reduction in data volume size. This means that 100GB of data would actually consume somewhere around 50GB in the cloud and proportionally reducing cloud storage monthly recurring fees.

CLOUD-TO-CLOUD MIGRATION

Many organizations seek to diversify or change cloud providers due to changing requirements, competitive pricing or a decision to store data in long-term archive. CloudArray enables data migration between cloud providers with minimal disruption and absolutely no changes required to the storage presented to applications from CloudArray. The cloud migration feature within CloudArray migrates data via on-premises cache. CloudArray also supports direct data migration between cloud providers and through cloud compute.

SECURE BULK INGESTION

Organizations that have a large amount of data that they want to move to the cloud typically need to do a one-time bulk ingestion. In addition to the 20+ cloud providers that CloudArray supports, CloudArray can also use an NFS box as a "cloud" target. As a result, customers can use CloudArray to load data at high-speed in the security of their own data center onto an on-premises NFS device. The fully encrypted object-based data can then be shipped to a cloud provider without the need to share encryption keys. Once loaded, administrators can use CloudArray’s cloud migration capabilities to access the data directly in the cloud.

CloudArray management

Form Factor

CloudArray is offered as both a physical appliance and a virtual appliance that can be deployed either on site or in a cloud compute environment. CloudArray uses local storage ("cache") for performance acceleration. User-defined policies dictate which data and how much will be kept in the local cache. Each CloudArray can manage multiple policies and caches simultaneously, each of which can be individually configured to support the needs of different use cases.
The physical appliance is a dedicated machine that comes pre-integrated with the CloudArray software. The size of the hardware appliance determines the total amount of local cache available. The appliance’s storage can be carved into multiple, smaller caches to allow for greater granularity and flexibility.

CloudArray can also be easily installed as a virtual appliance using a hypervisor's console. Supported hypervisors today include VMware vSphere ESXi and Hyper-V.

**CloudArray Administration**

Each instance of CloudArray can be individually configured using a browser-based user interface.

Administrators use this interface to:
- Create, modify or expand volumes, file shares, and caches
- Monitor and display CloudArray health, performance, and cache status
- Apply software updates
- Schedule and configure snapshots and bandwidth throttling

In addition to the local configuration controls, administrators can access an online portal (www.cloudarray.com) that enables them to:
- Download CloudArray virtual appliances
- Configure alerts and access CloudArray product documentation
- Store a copy of the CloudArray configuration file for disaster recovery retrieval

Ultimately, CloudArray provides an easy way to move inactive and secondary data off of primary storage and into low-cost cloud storage. This makes it easier for organizations to provision storage and manage their requirements against growing data demands.

CloudArray’s robust feature set combine customers’ performance and security requirements with the cloud's low-cost, low-maintenance model.

To learn more, visit [http://www.EMC.com/CloudArray](http://www.EMC.com/CloudArray)