EMC VPLEX WITH CLOUDARRAY

Integration, Best Practices, and Use Cases

ABSTRACT

With organizations around the world facing stringent compliance regulations, increasing quantities of data, and decreased IT spending, there is growing need for cost effective, online storage that can be used to scale out current infrastructures in order to meet business requirements. EMC CloudArray is the gateway that enables users to take advantage of cloud storage with the benefits of near limitless capacity, reduced management, and lower cost. VPLEX provides support for CloudArray allowing applications to automatically and seamlessly tier cold data to cost appropriate private and public cloud environments. This paper explains how VPLEX integrates with CloudArray.

February, 2016

To learn more about how EMC products, services, and solutions can help solve your business and IT challenges, contact your local representative or authorized reseller, visit www.emc.com, or explore and compare products in the EMC Store

Part Number H14857

TABLE OF CONTENTS

EXECUTIVE SUMMARY 4
AUDIENCE 4

CLOUDARRAY OVERVIEW 4
CloudArray Appliance 4
CloudArray Features 5

VPLEX OVERVIEW 5
EMC VPLEX Architecture 6

EMC VPLEX FAMILY 6

VPLEX HARDWARE PLATFORM 6
VPLEX GeoSynchrony 5.5 System Configuration Limit Summary 7
VPLEX Terminology 8

CONFIGURING YOUR SAN FOR VPLEX WITH CLOUDARRAY 8
VPLEX Provisioning 10
Ensuring Data Integrity 10

VPLEX WITH CLOUD ARRAY USE CASE EXAMPLES 11

VPLEX WITH CLOUDARRAY BEST PRACTICES 13
Network 13
CloudArray Cache 13
Configuring CloudArray Cache with VPLEX 14
General Considerations and Limitations 15
Software and GeoSynchrony Requirements 15
Conclusion 16

EXECUTIVE SUMMARY

With the exponential growth of data being seen in enterprise environments, cloud storage is rapidly becoming a viable method for online, cost effective storage. Organizations are looking to reduce the high costs of IT spending and decrease footprint without sacrificing capability, scalability, availability and replication feature sets. VPLEX solves these problems by allowing EMC CloudArray to be accessed as a traditional block storage tier. This allows users to take advantage of applications that use business logic to identify and move infrequently accessed or cold data off of primary storage to cost appropriate storage tiers built from private or public cloud providers. VPLEX integration with CloudArray helps accelerate time to value from investments in cloud storage.

VPLEX integrates with CloudArray allowing tiering to the cloud or object storage in the form of a public or private cloud. Cloud enabled VPLEX storage lowers the total cost of ownership by moving infrequently and seldom used data to lower cost cloud storage. CloudArray addresses all the requirements for connecting to cloud based storage providers. Designed to combine the resource efficiency of the cloud with traditional, on premise storage, CloudArray enables organizations to scale their enterprise storage capacity without increasing physical footprint.

Whether used for archive/backup, file, or secondary data, CloudArray ensures instant accessibility allowing VPLEX users to take advantage of the cloud's greater scalability, lower costs, and reduced maintenance.

AUDIENCE

This white paper is intended for:

- Customers, including IT planners, storage architects, and administrators involved in evaluating, acquiring, managing, operating, or designing and implementing a **VPLEX** and EMC CloudArray solution.
- EMC staff and partners, for guidance and development of proposals.

CloudArray Overview

EMC CloudArray is a storage software technology that easily integrates cloud-based storage into traditional enterprise IT environments. CloudArray is designed to combine the resource efficiency of the cloud with traditional, on-site storage, allowing organizations to scale their infrastructure and plan for future data growth. CloudArray makes cloud object storage look, act, and feel like local disk, seamlessly integrating with existing applications, giving a virtually unlimited tier of storage in one easy package.

CloudArray Appliance

Integration of VPLEX and CloudArray requires a hardware CloudArray hardware appliance. The appliance supplies the physical connection capability from the VPLEX to cloud storage by presenting the appliance as an external device with fibre channel controller cards that allows the VPLEX to connect with the appliance.

The hardware CloudArray appliance is a 2U server that consists of:

• 20TB or 40TB usable local cache. 12x2TB or 12x4TB drives in a RAID6 configuration

• 192GB RAM

• 2x2 port 8Gb fibre channel cards which will be configured in add-in slots on the physical appliance

Figure 1 shows a rear view of the CloudArray physical appliance.

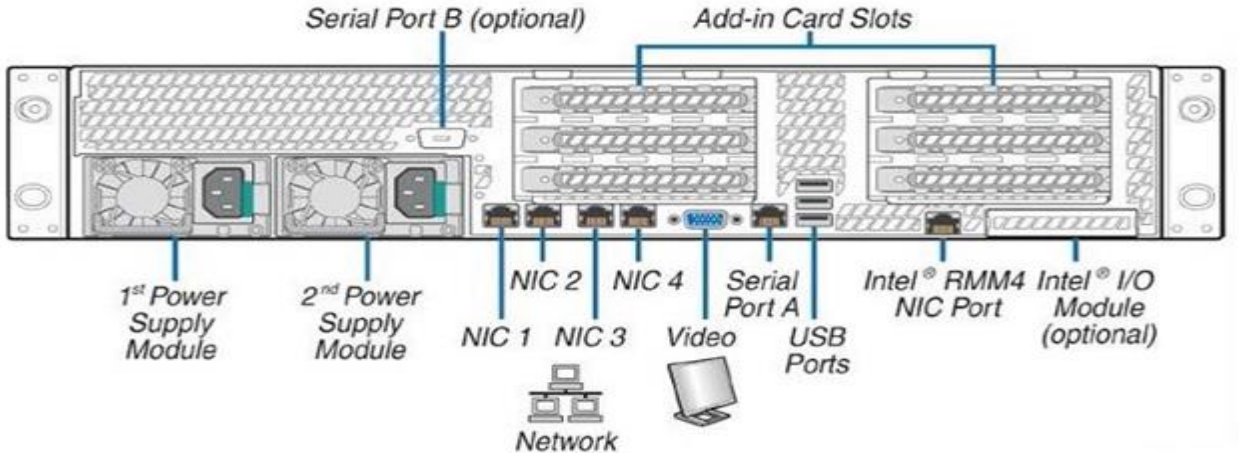

Figure 1. CloudArray Physical Appliance Rear View

CloudArray Features

EMC CloudArray is more than just a gateway to the cloud. CloudArray provides a complete storage solution that encompasses a range of features and allows VPLEX users to quickly begin connecting existing applications to cloud storage.

CloudArray uses local caching to reduce bandwidth and latency. VPLEX connects to the 20TB or 40TB CloudArray disk-based cache through fibre channel connections. This cache can be used for multiple data volumes according to use cases and stores a portion of users' most recently-accessed data locally to maintain performance while simultaneously replicating to the cloud.

In addition to the local caching to reduce bandwidth and latency, CloudArray also offers 256 AES inflight and at-rest encryption, data compression and bandwidth scheduling to optimize network utilization.

VPLEX Overview

EMC VPLEX virtualizes block storage from arrays such as EMC CLARiiON, CloudArray, VMAX, XtremIO and over 65 3rd party arrays by abstracting physical storage to create virtual volumes. The logical components of each VPLEX virtual volume are shown below in Figure 2.

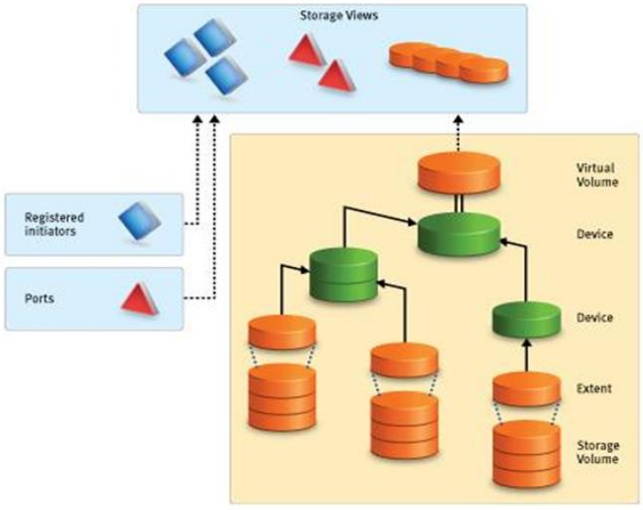

Figure 2. VPLEX Virtualization Components

EMC VPLEX Architecture

VPLEX is a virtual storage solution designed for use with heterogeneous host operating system environments and with both EMC and non-EMC block storage arrays. VPLEX physically resides between the servers and the block storage arrays. It introduces a virtual storage layer which enables VPLEX to the following desirable storage characteristics:

- Scale-out clustering hardware, which lets customers to start small and grow big with predictable service levels
- Advanced data caching utilizing large-scale SDRAM cache to improve performance and reduce I/O latency and array contention
- Distributed cache coherence for automatic sharing, balancing, and failover of I/O across the cluster
- Consistent view of one or more LUNs across VPLEX clusters separated either by a few feet within a data center or across synchronous distances (up to 10ms), enabling new models of high availability and workload relocation

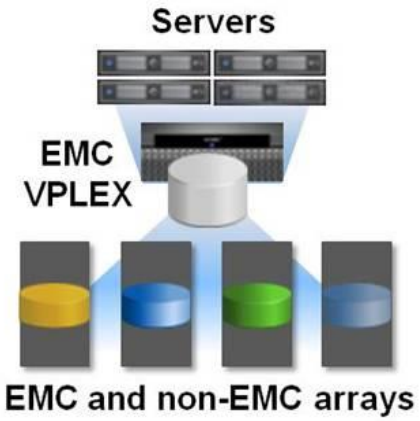

Figure 3. High level VPLEX architecture

EMC VPLEX Family

VPLEX Local: Provides non-disruptive and fluid mobility, high availability, and a single point of management for heterogeneous block storage systems within a single data center.

VPLEX Metro: Provides non-disruptive and fluid mobility, high availability, and a single point of management for heterogeneous block storage systems within and across data centers at synchronous distance. The VPLEX Metro offering also includes the unique capability to remotely export virtual storage across datacenters without the need for physical storage at the remote site.

VPLEX Hardware Platform

A VPLEX system with GeoSynchrony is composed of one or two VPLEX clusters: one cluster for VPLEX Local systems and two clusters for VPLEX Metro and VPLEX Geo systems. These clusters provide the VPLEX AccessAnywhere capabilities.

Each VPLEX cluster consists of:

- A VPLEX Management Console
- One, two, or four engines. SAN Connectivity is show in Figure 4 below.
- One standby power supply for each engine.

In configurations with more than one engine, the cluster also contains:

- A pair of Fibre Channel switches
- An uninterruptible power supply for each Fibre Channel switch
- As engines are added, cached is also added along with additional front-end, back-end, and wan-com connectivity capacity as indicated in Table 2 below.

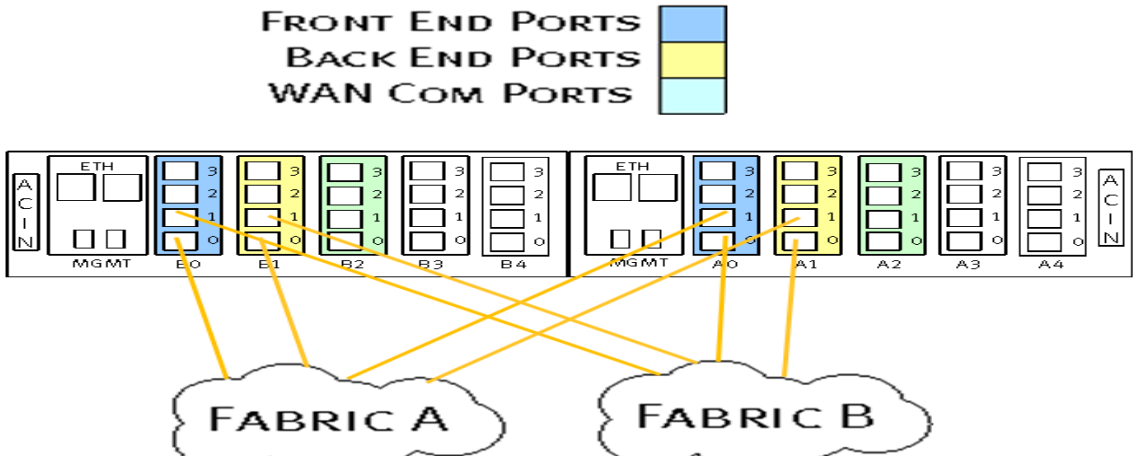

Figure 4. VPLEX VS2 Single Engine to SAN Connectivity

VPLEX GeoSynchrony 5.5 System Configuration Limit Summary

Capacity	Local	Metro
Maximum virtualized capacity	No Known Limit	No Known Limit
Maximum virtual volumes	10,000	16,000
Maximum storage elements	10,000	16,000
Minimum/maximum virtual volume size	100MB/32TB	100MB/32TB
Minimum/maximum storage volume size	No VPLEX Limit / 32TB	No VPLEX Limit / 32TB
Number of host initiators	1600	3200

Table 1

Engine Type	Model	Cache [GB]	FC speed [Gb/s]	Engines	FC Ports
VPLEX VS2	Single	72	8	1	16
	Dual	144	8	2	32
	Quad	288	8	4	64

Table 2

Table 1 and Table 2 show the current limits and hardware specifications for the VS2 hardware version. Each VS2 port can supply full line rate (8 Gbps) of throughput. Several of the VPLEX maximums are determined by the limits of the externally connected physical storage frames and therefore unlimited in terms of VPLEX itself. The comprehensive list of configuration limits are published in the GeoSynchrony 5.5 Release Notes which are available on Powerlink.EMC.com.

VPLEX Terminology

Storage volume	LUN or unit of storage used by VPLEX presented by the back-end arrays.
Extent	All or part of a storage volume used by VPLEX.
VPLEX device	Logical layer where protection scheme applied to an extent or group of extents used by VPLEX.
Virtual volume	Unit of storage presented from VPLEX front-end ports to hosts.
Front-end port	VPLEX director port connected to host initiators (acts as a target).
Back-end port	VPLEX director port connected to storage arrays (acts as an initiator).
VPLEX Director	The central processing and intelligence of the VPLEX solution. There are redundant (A and B) directors in each VPLEX Engine.
VPLEX Engine	Consists of two directors and is the unit of scale for the VPLEX solution.
VPLEX cluster	A collection of VPLEX engines in one rack, using redundant, private Fibre Channel connections as the cluster interconnect.
Storage Array	An intelligent, highly available, scalable, hyper consolidated, multi-tiered block storage frame. For example, EMC CX4, CloudArray, VNX, Symmetrix DMX4, VMAX, VMAX3, and XtremIO.

Configuring your SAN for VPLEX with CloudArray

Configuring VPLEX Directors

VPLEX Directors are fully redundant, and, when necessary, remaining healthy VPLEX directors continue I/O for any director(s) that have failed. This enables continuous I/O servicing by the VPLEX cluster. VPLEX requires a minimum of four paths to each external device per director. This means that at least two ports on each director (1 to SAN A and 1 to SAN B) must be zoned to a minimum of two CloudArray FC ports. This rule applies to the 1, 2, and 4 engine configurations of VPLEX.

Zoning

When configuring zoning for VPLEX with CloudArray using a single-target per single-initiator (1:1) zoning scheme is preferred. If the FC switch zone count limitation has been reached, it is possible to use single-target per multiple-initiator (1:Many) zoning.

The zoning examples provided below allow the servicing of the components within a VPLEX environment without incurring data unavailability and are required in order for the configuration to be supported. The potential service activities include:

- Cable changes and individual FC port servicing
- VPLEX director and I/O module replacement
- CloudArray firmware upgrade
- SAN fabric servicing

Proper zoning also ensures that a failing switch or storage controller will not cause a director to lose all connectivity to cloud array which would lead to the director isolating itself from the cluster.

Figure 5 below show an example of how to meet the dual-fabric configuration requirement which represents the most common VPLEX SAN configuration. These zoning requirements are in addition to existing connectivity requirements consisting of two, physically independent SCSI I-T nexus per VPLEX director port for each storage volume (4 total I-T nexus per director). These base

connectivity requirements are checked by VPLEX ndu and configuration pre-checks during the configuration and upgrade processes. If the zoning does not pass VPLEX ndu pre- checks, attempts to upgrade VPLEX GeoSynchrony code will fail.

Dual Fabric with two external Fibre Channel Controller ports

Though single-fabric connectivity is technically feasible, the VPLEX requirement for redundancy is met by using dual fabrics. At least one VPLEX director BE port must be connected to each SAN fabric.

As an example, take a VPLEX environment configured with a single-engine VPLEX with a single CloudArray appliance, C1, which contains two dual-port fibre channel cards, FC1 and FC2.

Figure 5 and 6 show an examples of dual-fabric zoning with this configuration.

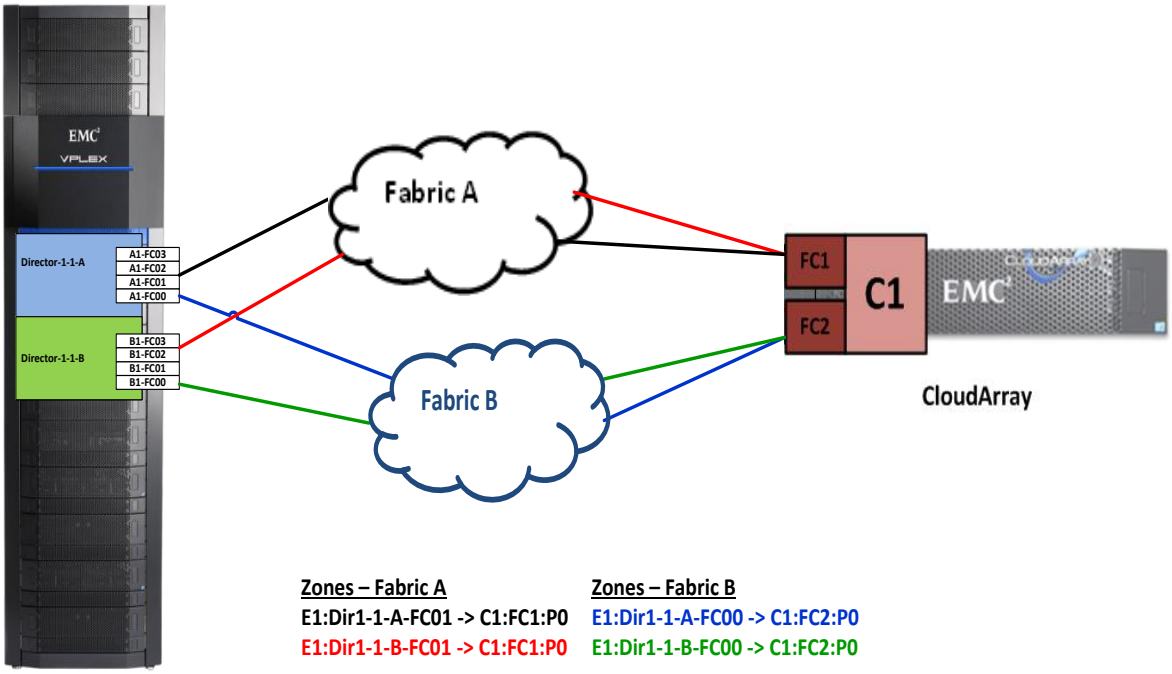

Figure 5. Single Engine VPLEX with Dual Fabric Zoning and 2 FC ports on Cloud Array

Note: *Figure 5 shows the logical, not physical connections from the switch to the CloudArray. In the diagram, from the VPLEX, there are four physical connections to the fabric, one from each director BE port to the switch(es). However, there are only two physical connections, one from each of the two CloudArray fibre channel adapters, from the fabric(s) to each CloudArray.*

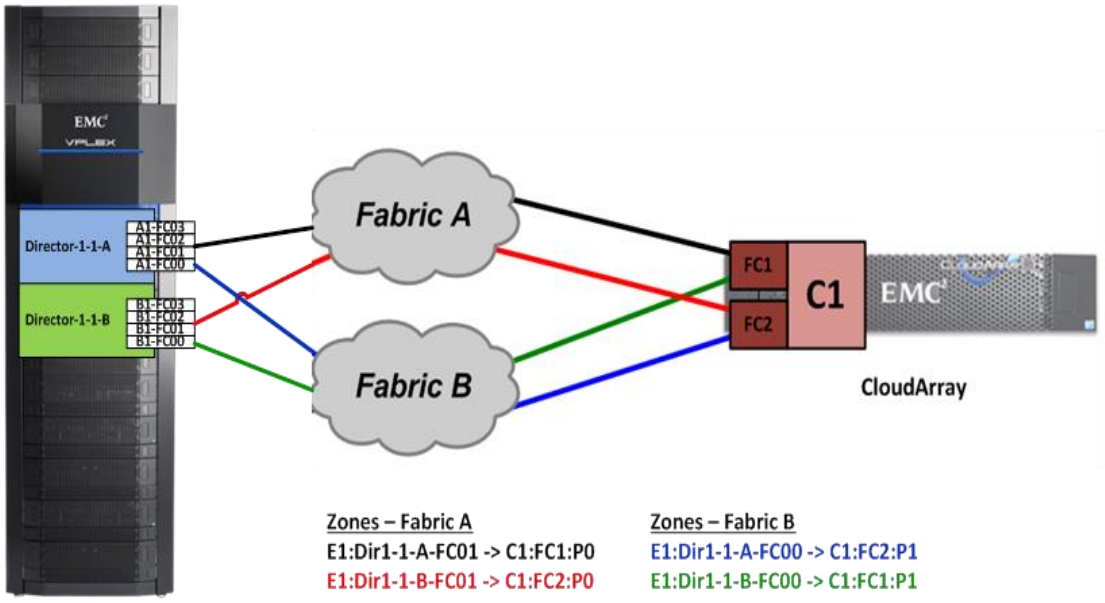

Figure 6. Single Engine VPLEX with Dual Fabric Zoning and 4 FC ports on CloudArray

VPLEX Provisioning

VPLEX provisioning with CloudArray volumes follows the standard process used for all storage volumes. As shown in Figure 7 VPLEX administrators can assign access to the entire capacity that exists on the CloudArray in a 1:1 relationship between the virtualized volumes on the CloudArray and the virtual volumes VPLEX presents back to hosts. Alternatively, as CloudArray storage volumes are often configured as large, multi-terabyte storage devices (due to the limit of 128 CloudArray devices), extents may be used to sub-divide large storage volumes into smaller virtual volumes when required by the application or use case.

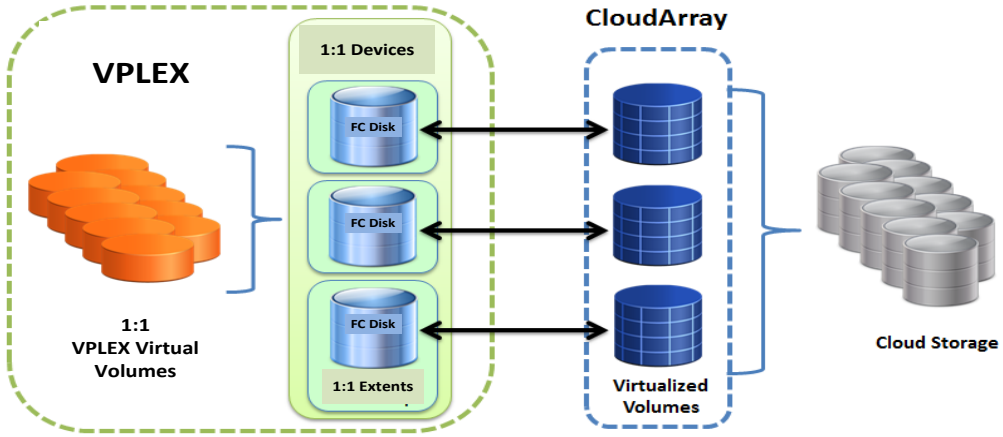

Figure 7. VPLEX 1:1 Provisioning with Cloud Array

Ensuring Data Integrity

CloudArray's encryption is a critical component in ensuring data integrity. CloudArray segments its cache into cache pages and, as part of the encryption process, generates and assigns a unique hash to each cache page. The hash remains with the cache page until that page is retrieved for access by a requesting initiator. When the page is decrypted, the hash must match the value generated by the decryption algorithm. If the hash does not match, then the page is declared corrupt. This process helps to prevent any data corruption from propagating to an end user.

VPLEX with Cloud Array Use Case Examples

Without a doubt there are many powerful examples of use cases for VPLEX with CloudArray. The following are just a few of the many ways VPLEX with CloudArray can provide tangible benefits in terms of improved cost efficiency of data storage, improved array performance, reduced capital expenditure, and decreased operational expenses within your IT environment today.

Mobility within and across Cloud providers

Local and remote data migration and mobility functionality is supported on VPLEX volumes built using CloudArray storage. VPLEX administrators can non-disruptively move cold and/or infrequently used data from high cost, high performance, tier-1 storage arrays using VPLEX data mobility. In addition, mobility within and across various public and private cloud providers is simple and non-disruptive using VPLEX's non-disruptive mobility.

Managing and Migrating Existing Data

VPLEX users can quickly and easily move existing data applications to the cloud once CloudArray is configured. Archive data can also be written directly to the cloud through VPLEX virtual volumes. CloudArray users can move an application to the cloud or back from the cloud to array-based storage. This enables legacy applications to be non-disruptively migrated to the cloud and for cold/infrequently used data to be purged from tier 1 storage arrays freeing their high performance and high value space.

VPLEX Cloud Tiering: Database Snaps

Database administrators use snapshots to protect data and store these snaps on high-cost, high-performance primary storage. As snapshots age they become less likely to be used and therefore can moved to cloud storage to help cut costs while maintaining accessibility. Using EMC VPLEX and CloudArray a VPLEX administrator can advantage of cost efficient cloud storage and save on capital expenditures.

Advantages:

- Tiering old snapshots to cheaper cloud storage
- Maintain full accessibility at all time in case of recovery
- Meet all recovery objectives while cutting total costs

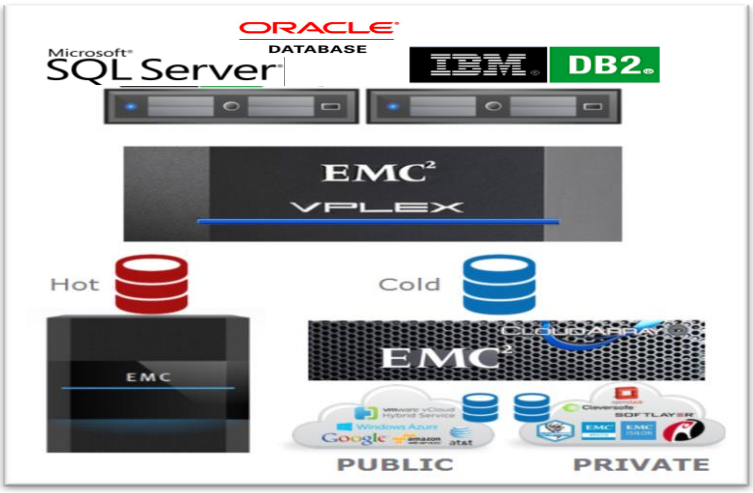

Figure 8. Cloud Tiering with Databases

VPLEX Cloud Tiering: Database Archives

Databases consume high-cost, high performance storage for table space data and temporary copies of data. Some of that data is aged and/or infrequently accessed and would be more cost efficiently stored on cloud storage. Using EMC VPLEX and CloudArray enables database administrators to leverage the cost efficiencies of cloud storage in a simple, automated, and straightforward way to move seldom used table spaces or aging temporary copies of table spaces to the cloud. Many of these capabilities are built into contemporary database offerings from major vendors such as IBM, Microsoft, Oracle, and SAP. These databases provide intelligent, business value driven movement of cold/archive data into the cloud. Figure 8 (above) illustrates some of the database vendors that are able to leverage the capabilities provided by VPLEX with CloudArray.

Advantages:

- Simplified tiering to cheaper cloud
- Pay only for what storage you use and expand as you grow
- Reduce capital expenditures by delaying primary storage upgrades

VPLEX Cloud Tiering: Splunk

Splunk is an application that collects datacenter logs and performs queries on them, collecting large amounts of data over time. However, because the most recent information is more relevant within its queries/analytics, Splunk has the ability to tier data into hot/warm/cold/frozen tiers, using only the relevant data. As shown in Figure 9 (below) using EMC VPLEX and CloudArray allows Splunk administrators to automatically tier cold/frozen data to cost efficient cloud storage and to realize significant capital expenditure savings.

Advantages:

- Simplified Splunk tiering to cost efficient cloud for less frequently accessed data
- Pay only for what storage you use, but keep all data available
- Convert capital expenses to more business favorable operational expenses

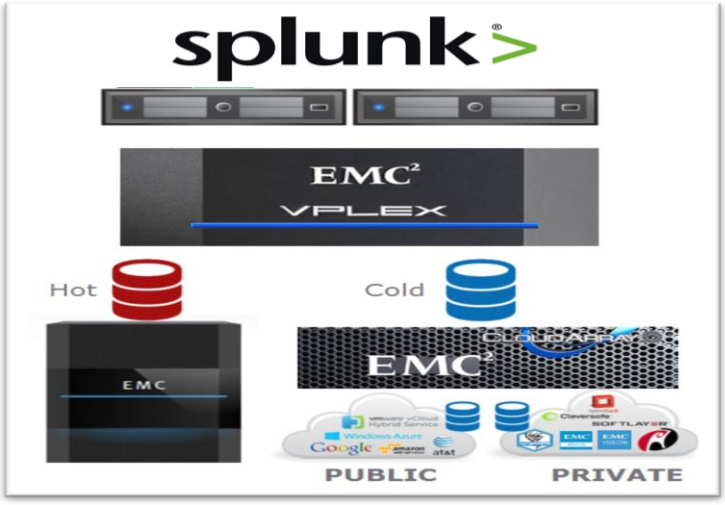

Figure 9. Splunk Cloud Tiering with VPLEX

VPLEX Cloud Tiering: Microsoft Exchange

Microsoft Exchange is typically configured to store all mailboxes on a single tier of high-cost, high-performance storage. Very often, however, email messages residing within Exchange mailboxes are old and/or infrequently accessed and can be archived. For Exchange administrators that want to reduce the number of active e-mail items in primary mailboxes, Exchange Server 2010 introduces integrated e-mail archiving and retention policies. This includes a new Personal Archive—a specialized mailbox associated with a user's primary mailbox. When using this archive, all e-mail data is still stored on the Exchange Server storage devices. The combination of the user's primary and archive mailboxes can be considered together as the user's large mailbox. Using EMC VPLEX and CloudArray allows Microsoft Exchange administrators to take advantage of cost efficient cloud storage for email archives. This reduces the consumption of Tier 1 storage and moves less frequently accessed emails to cost appropriate cloud storage.

Advantages:

- Move mailboxes to cheaper cloud for less expensive archiving
- Keep all mailboxes easily accessible in case archives need to be quickly accessed
- Reduce capital expenditures by delaying primary storage upgrades

VPLEX with CloudArray Best Practices

Network

Optimizing your IP network for cloud storage use is critical to the success of your CloudArray implementation. The following recommendations explain how to optimize your network connection for cloud storage without impacting applications sharing the same internet connection.

- Low bandwidth connections (less than 5 Mbps) can result in communication problems between CloudArray and the cloud.
- In its unaltered state, CloudArray attempts to use all available bandwidth. If you're sharing the connection with other applications, you'll want to mediate this by using the CloudArray bandwidth throttler under the Networking tab. Avoid throttling to near or below 5 Mbps per the previous suggestion.
- Consider the amount of data that will be transferred to and from the cloud daily by CloudArray as well as the amount of data used by other applications on a shared internet connection. CloudArray is a bandwidth-intensive product. You may need to work with your ISP to increase the available bandwidth.
- In addition to the cloud storage provider(s) you wish to access, CloudArray must have outbound access to cloudarray.com on ports 80 and 443 and inbound access on 80, 443, 8080(GUI) and 3260(iSCSI).

CloudArray Cache

One of CloudArray's most unique features is its dynamic caching technology. This sophistication provides customers a wealth of options for customers to customize an implementation that will meet their needs – regardless of what they might be. This section explains the importance of the cache and covers the critical considerations and options available to you.

The cache is required for proper operation of CloudArray. All data that passes through CloudArray is stored in the cache for some period of time. Data written to your volumes and shares is stored in the cache. Over time the data is replicated to the cloud. Once replicated, it may potentially be removed from the cache by CloudArray.

CloudArray appliance disk cache currently consists of 12 x 2TB or 4TB 7200 RPM SATA drives. The disk drives are configured as a single pool of storage with Raid-6 (10+2) protection. The disk cache is preconfigured during installation. The performance characteristics of the disk cache are best suited for sequential read and write operations. Fully cached cloud array disks can be used for higher IOPS applications, but it is important to note that the disk cache spindles are shared and high IOPS from one cache or application can impact another.

There are various strategies for determining this. In nearly all instances, your fibre channel SAN will have much greater bandwidth than your access to the internet. The internet is typically the bottleneck in moving data to the cloud. The cache masks this bottleneck by providing high-speed local storage. With the exception of the "Fully Cached" strategy outlined below, you should avoid the cache becoming more than 80% unreplicated. The strategies outlined below explain different use cases and highlight our general recommendations for caching as a result.

Scenario 1: Fully Cached (1:1)

In this configuration, the size of the cache is greater than or equal to the aggregate size of all volumes using the cache. This solution provides two copies of the data, one local and one in the cloud. It also provides optimal performance because data will not need to be read from the cloud under most circumstances.

Scenario 2: Archive Storage (1:X)

Use cases include any application where data is uploaded to the cloud but rarely read back from the cloud.

Data is streamed to the cloud through the cache and is rarely read. In this case you want the cache to be large enough to buffer any writes to the cloud regardless of size of the volume(s). For example, assume you have 100 Mbps of usable upload bandwidth to the cloud and 800 Mbps of usable bandwidth between CloudArray and your iSCSI initiator. Now imagine that you upload a 100GB file every day. It will take about 12 minutes to write the file from the iSCSI initiator into the CloudArray cache. It will take about 2.5 hours to copy that 100GB of data to the cloud. If the cache is at least 100GB then the iSCSI initiator will be able to operate at full speed without having to wait for the cloud. As your workload grows, your cache size will have to grow also. An application that performs differential backups is often not an Archival use case because differential backups often require reading what has been previously written to the backup.

Configuring CloudArray Cache with VPLEX

When configuring the CloudArray cache for fibre channel storage presented to VPLEX there are two recommended approaches:

1. Start off with fewer large caches (2 – 4 TB each) and round robin use of each cache for alternating CloudArray volumes exposed to VPLEX. This way each CloudArray cache will be relatively evenly balanced and one cache doesn't get overused when there is abundant cache space available elsewhere. The key is to avoid creating high VPLEX virtual volume to CloudArray storage cache ratios. In practice, ratios of 10:1 are feasible once storage has been fully migrated to the Cloud and is not accessed that frequently. During cloud ingestion, however, 1:1 or even over cached configurations work best to help buffer the incoming writes produced by VPLEX.

   Or

2. Create several large caches and assign storage to each of them based on the application, business unit, SLA, or cost center consuming the cloud storage. This way one application cannot monopolize the disk cache from other applications that are also using CloudArray. The key, as with option 1, is to avoid high virtual volume to cache ratios for active applications and for mobility within and migration to the cloud. Ideally, a ratio of 1:1 can be maintained for I/O bursts or use cases that have unpredictable read/write behavior and/or more stringent performance requirements.

Note: *Cloud array uses an 8 GB transfer size to move data to and from the cloud. This means that at any given time 8 GB of data may be locked and accessed by normal CloudArray operations. For this reason, when configuring CloudArray cache for 1:1 cache to disk ratios it is advisable to add 8 GB to the total desired cache size.*

CloudArray I/O Behavior

As the percentage of cache pages that are unreplicated approaches 100%, the cache will throttle access to CloudArray to postpone reaching the 100% mark. When all the cache pages are unreplicated, then all volumes using that cache are taken offline until a sufficient number of pages have been replicated to the cloud.

Creating an appropriate cache size for your requirements is the first step to avoiding the scenario outlined above. When CloudArray cache starts to become full (at ~80%) an automatic IO throttling mechanism will kick in to artificially slow I/O operations to allow CloudArray additional time to free up cache space by flushing cache to the cloud. When CloudArray cache becomes completely full the CloudArray storage volumes will become non-responsive and marked dead by VPLEX. VPLEX will attempt to auto-resurrect them every 5 minutes.

When CloudArray loses connectivity to the Cloud storage provider it can no longer retrieve read miss data. When there is no network connection to the Cloud VPLEX write I/O can continue until the CloudArray cache is full. On first read miss or cache full condition with no Cloud connection, the storage volume will go offline and be marked dead by VPLEX.

Avoiding Full Cache

To minimize the chance of VPLEX marking CloudArray storage volumes dead, it is important to create CloudArray caches and cache to disk ratios that correspond with the intended storage use. The I/O and utilization level of various applications (OLTP, Migration into Cloud, Archive, or Occasional access) running on CloudArray storage is the key. For migration into the cloud a 1:1 cache ratio is strongly recommended. Once data is migrated into the cloud and in a steady state (archive mode) or used for occasional access then higher cache to disk ratios are perfectly acceptable.

Note: *A read miss from CloudArray cache can potentially mean several hundred milliseconds to retrieve data from the cloud storage provider. Confirm your application runs acceptably for any > 1:1 cache ratios you select to ensure read miss performance is acceptable.*

Using 1:1 (or better) cache to disk ratios with VPLEX volumes does not guarantee high performance. The overall performance of CloudArray storage volumes starts to be impacted as the quantity of CloudArray backed VPLEX volumes exceed 15-20. For this reason, it's important to plan how many concurrent workloads and how much of a workload will be placed onto the CloudArray appliance at any given time.

Note: *The performance for each CloudArray environment will vary based on many factors. These factors include the performance capability of the cloud storage provider, the quality of the IP network, the average I/O size, the I/O pattern, the read/write mix, and total number of I/O per second*.

Additional guidelines that affect cache size

Plan for outages - You should plan for occasional network or cloud provider outages that will require the data to remain in cache for an extended period of time. We recommend that you overprovision the cache storage to avoid a high ratio of unreplicated pages. A good rule of thumb would be to configure the cache to be approximately twice the size of your anticipated cache.

Read miss rate - The cache hit rate should not be lower than 99% where read performance is a concern. If you have read performance issues and the read hit rate is below 99% then consider adding capacity to your cache.

CloudArray cache page size and performance

Cache page size is a significant factor in CloudArray performance. Larger page sizes work well for applications that perform large sequential writes to the volume. Most backup applications are in this category. Smaller page sizes are preferred when small random reads are needed. This is typically seen when the CloudArray volume is used for a file system with generic user data. Larger pages also have smaller CloudArray RAM requirements, see below. Using a page size smaller than 1MB is rarely appropriate.

Cache pages greater than 1MB require an internal sector size of 4KB rather than the typical 512 bytes. *The larger sector size is not recommended for transactional workloads such as database or typical file system applications.*

CloudArray Compression and Encryption

There is a CloudArray appliance CPU cost associated with compression. Consider disabling compression for policies with volumes where a majority of the data is pre-compressed or incompressible.

General Considerations and Limitations

- VPLEX supports dual fabric connectivity to CloudArray only. Direct attach is not supported.
- VPLEX supports maximum of one (1) CloudArray appliance per VPLEX Local cluster. Maximum two (2) total for VPLEX Metro.
- VPLEX local or remote mirroring (raid-1) with CloudArray storage volumes and non-CloudArray storage volumes is not supported. The temporary raid-1 devices used for VPLEX mobility and migration is, however, supported.
- A maximum of 128 CloudArray storage volumes can be configured
- Four (4) logical paths per director per CloudArray storage volume are required.
- Claiming CloudArray storage volumes can only be done from the VPLEX CLI. VPLEX 5.5 GUI currently does not support claiming CloudArray storage.
- The minimum recommended cache size is 2 TB.
- CloudArray currently supports a maximum of 31 individual caches.
- The maximum CloudArray storage volume size that can be presented to VPLEX is 32TB.
- A supported cloud provider must be used in a CloudArray configuration. For more information on compatible providers, go to http://www.emc.com/collateral/technical-documentation/h14387-emc-cloudarray-support-matrix.pdf
- The EMC SolVe Desktop application (available @ https://support.emc.com) contains the most up to date configuration and connectivity instructions for VPLEX and VPLEX with CloudArray.

Software and GeoSynchrony Requirements

VPLEX and EMC CloudArray require the following software versions:

- VPLEX GeoSynchrony 5.5 or higher
- CloudArray 6.0 or higher

Conclusion

VPLEX with CloudArray offers easily integrated cloud-based storage into traditional enterprise IT environments. CloudArray utilizes a hardware appliance that supplies physical connection capabilities from VPLEX to cloud storage using a fibre channel SAN. The CloudArray appliance leverages 20TB or 40TB local disk-based cache that can be used as multiple storage volumes, storing the user's most recently accessed data within the CloudArray appliance while replicating to the cloud. VPLEX with CloudArray combines the efficiency of cloud storage along with onsite primary storage. VPLEX provides non-disruptive data mobility into, out of, and across private and public cloud providers -- enabling organizations to easily and efficiently move cold and/or seldom used data from high value tier-1 storage arrays to cost efficient/cost appropriate cloud storage. Cloud storage provide near limitless scaling and storage capacity, reduced management, and lower cost.

If you would like further information about VPLEX with CloudArray please contact your EMC Account representative or visit http://www.emc.com/storage/cloudarray/index.htm