VPLEX GEOSYNCHRONY V5.0 TECHNICAL DIFFERENCES

Welcome to the VPLEX GeoSynchrony v5.0 Technical Differences course.

Click on the Supporting Materials tab to access the Student Resource Guide and course navigation information.

Course Overview

Description	This course will present the technical differences in the VPLEX GeoSynchrony v5.0 product. These differences will include the new supported use cases. It will also cover the new VS2 hardware and GeoSynchrony software. It will delve into the new features associated with VPLEX GeoSynchrony v5.0 such as consistency groups, consistency group mobility, the VPLEX Witness, the VPLEX Element Manager API, thin awareness, cache vaulting, GUI performance monitoring, and ALUA support.
Audience	This course is designed for those already knowledgeable with VPLEX GeoSynchrony v4.X and needs to understand the differences in the new VPLEX GeoSynchrony v5.0 product.
Objectives	Upon completion of this course, you should be able to: • List the changes between the VS1 and VS2 hardware • Differentiate between VPLEX Local, Metro and Geo clusters configurations • Present the use cases for using a VPLEX Local, Metro and Geo • Describe and implement the new features associated with VPLEX GeoSynchrony v5.0 such as Consistency Groups, the VPLEX Witness, the VPLEX Element Manager API, Thin Awareness, ALUA support, WAN-COM Performance Monitoring, and cache vaulting

EMC believes the information in this course is accurate as of its publication date. It is based on pre-GA product information, which is subject to change without notice. For the most current information, see the EMC Support Matrix and product release notes on Powerlink.

EMC²
where information lives®

This course will present the technical differences in the VPLEX GeoSynchrony v5.0 product. These differences will include the new supported use cases. It will also cover the new VS2 hardware and GeoSynchrony software. It will delve into the new features associated with VPLEX GeoSynchrony v5.0 such as consistency groups, consistency group mobility, the VPLEX Witness, the VPLEX Element Manager API, thin awareness, cache vaulting, GUI performance monitoring, and ALUA support.

This course is designed for those already knowledgeable with VPLEX GeoSynchrony v4.X and needs to understand the differences in the new VPLEX GeoSynchrony v5.0 product.

VPLEX GeoSynchrony Features Matrix

Shown is a matrix that lists all of the major features that were added to the code from GeoSynchrony v4.X to GeoSynchrony v5.0. Within the matrix you can on a GeoSynchrony version tab to display the features added for the selected version of code. Hovering you mouse over a feature will display more information on the feature.

Launch

Shown is a matrix that lists all of the major features that were added to the code from GeoSynchrony v4.X to GeoSynchrony v5.0. Within the matrix you can on a GeoSynchrony version tab to display the features added for the selected version of code. Hovering you mouse over a feature will display more information on the feature.

Module 1: VPLEX GeoSynchrony v5.0 Overview and Use Cases

This module provides an overview of the VPLEX GeoSynchrony 5.0 product and its use cases.

Upon completion of this module, you should be able to:

- Describe how VPLEX fits into an environment
- Describe the use cases for a VPLEX GeoSynchrony v5.0 Local, Metro and Geo deployment
- List the benefits of the Metro HA deployment with and without the cross-cluster connect

This module provides an overview of the VPLEX GeoSynchrony v5.0 product and its use cases.

Architected for Global Applications
VPLEX Family: Today and Future

Next Generation Data Mobility and Access
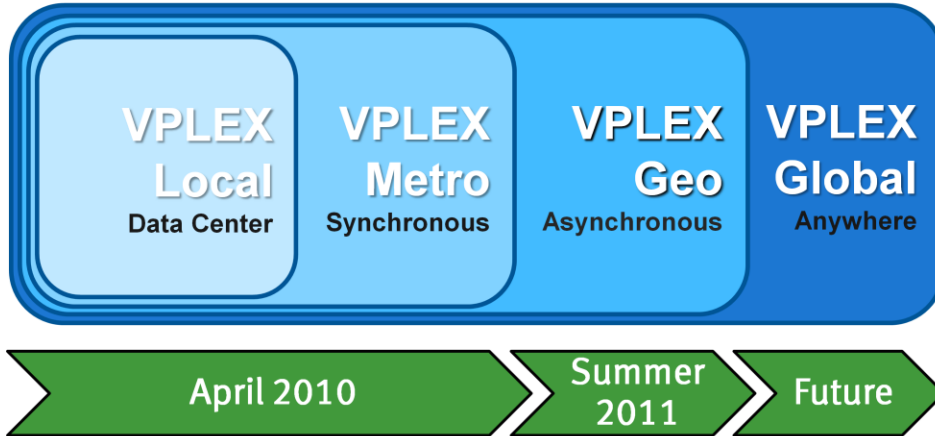
| VPLEX Local Data Center | VPLEX Metro Synchronous | VPLEX Geo Asynchronous | VPLEX Global Anywhere |

April 2010 → Summer 2011 → Future

VPLEX GEOSYNCHRONY v5.0 Technical Differences 5

Here's a look at the VPLEX roadmap. In 2010, we introduced the first two products in the VPLEX GeoSynchrony family: VPLEX Local, for use within a single data center, and VPLEX Metro, which connects two VPLEX clusters together for deployment within or between data centers, up to synchronous distances apart. This quarter we'll be introducing the family's newest member, VPLEX Geo, which extends VPLEX's capabilities for connecting two clusters together across asynchronous distances. VPLEX Global is planned for a future release, and will support more than two clusters connected together simultaneously, over synchronous or asynchronous distances.

VPLEX Family – Introduced May 2010
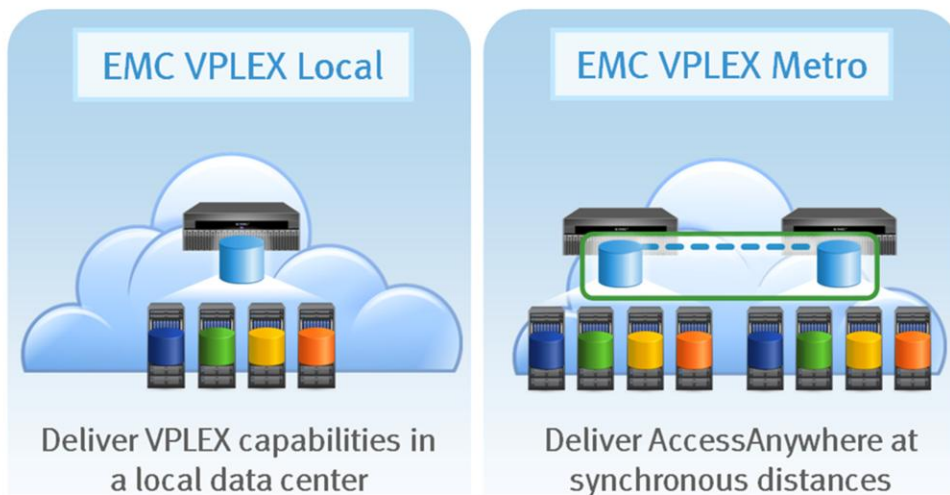
EMC VPLEX Local

EMC VPLEX Metro

Deliver VPLEX capabilities in a local data center

Deliver AccessAnywhere at synchronous distances

VPLEX GEOSYNCHRONY v5.0 Technical Differences 6

VPLEX Local is designed for use within a single data center, while VPLEX Metro allows you to connect two VPLEX clusters together within synchronous distances to move or share data between them.

A common use case for VPLEX Local includes non-disruptive data mobility between EMC and non-EMC platforms. It allows users to implement a VPLEX Local to federate storage arrays, simplify storage management, and speed up data movement between arrays. A VPLEX Local also provides a way to increase High Availability and local storage resiliency for arrays in a single site. VPLEX GeoSynchrony allows storage to be mirrored across mixed platforms without requiring host resources. Leveraging this capability can increase protection and high availability for critical applications while leveraging a user's existing storage resources. VPLEX Local can also be used to simplify management of multi-array storage environments, with tools to provision and allocate virtualized storage devices to standardized LUN presentation and management. It will be possible to upgrade a VPLEX local to a VPLEX Metro or Geo at a later time.

VPLEX Metro allows users to move applications over distance and, in combination with server virtualization, allows users to transparently move and relocate Virtual Machines and their corresponding applications and data over distance. This provides a unique capability allowing users to relocate, share and balance infrastructure resources between data centers. A VPLEX Metro also delivers increased high availability and resiliency by allowing users to mirror volumes within and across locations. This provides non-stop application availability in the event of a component failure. A VPLEX Metro also allows data to be distributed, shared and collaborated on across sites. A single copy of data can be accessed from multiple users across two locations. This allows instant access to information in real time, and eliminates the operational overhead and time required to copy and distribute data across locations.

6

VPLEX Family – Coming Summer 2011

There are several new capabilities being announced for VPLEX this quarter, including:

The new VPLEX VS2 hardware, which is more powerful and more efficient.

The latest release of VPLEX's operating system, GeoSynchrony v5.0, which will run on the existing VPLEX VS1 hardware currently deployed, as well as the new VS2 hardware.

Other new GeoSynchrony v5.0 capabilities include the following:

- The ability to move or copy virtually provisioned or thinly provisioned devices with VPLEX; when a virtually provisioned or thin volume is on the array, VPLEX previously converted the volume into a thick device and then moved it, allowing the user to manually reclaim the space after the device was copied over to the target array. VPLEX can move a VP or thin volume across EMC and non-EMC arrays to another array non-disruptively and still preserve its virtually provisioned nature, without any changes to the device. This is a key benefit for customers looking to optimize their storage capacity with virtual provisioning, as it allows users to only consume storage as needed. VPLEX supports VP-Copy for all EMC arrays with virtual provisioning and also many supported non-EMC arrays with thin provisioning.

- VPLEX also delivers several new array qualifications, including support for ALUA-based arrays. This enables VPLEX to support the full range of mid-tier arrays, including ALUA-based arrays from NetApp and HP.

- Also a major new innovation for VPLEX Metro and VPLEX Geo deployments is the VPLEX Witness. The witness will help to provide High Availability through a witness that will run on a virtual machine and identify cluster failure situations. This helps to keep applications online for true automated High Availability. The VPLEX Witness will be discussed in more detail in the slides to come.

- Lastly, VPLEX is able to offer customer support over asynchronous distances with VPLEX Geo. With VPLEX Geo, customers can take advantage of AccessAnywhere capabilities between two sites, even if those sites are more than 100k apart.

Mobility
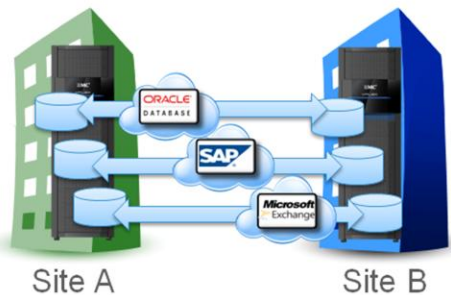
VPLEX
Enables Apps and Data to be moved over distance in seconds

Site A   Site B

VPLEX Advantage
- Move anytime, to anything, anywhere
- Single interface for multivendor arrays
- Data remains available during relocation

VPLEX GEOSYNCHRONY v5.0 Technical Differences   8

When customers use VPLEX for application and data mobility, they have several advantages over traditional solutions. First, VPLEX is designed as a cluster, and two clusters can be connected together – forming either a Metro-Plex with VPLEX Metro, or a Geo-Plex using VPLEX Geo. Once this is established, customers can have immediate access to their data – because VPLEX can present the same data at each cluster's location simultaneously.

Second, customers can automatically balance loads through VPLEX, using storage and compute resources from either cluster's location. And when combined with server virtualization, VPLEX allows users to transparently move and relocate Virtual Machines and their corresponding applications and data over distance. This provides a unique capability allowing users to relocate, share and balance infrastructure resources between sites; this can occur within a campus, between data centers up to 100km or 5ms in round trip latency apart, or further apart across asynchronous distances with VPLEX Geo. Note that for VPLEX Geo, the distance supported far exceeds 100km, up to a 50 ms in round-trip latency.

Third, with VPLEX customers no longer need to spend significant amounts of time and resources preparing to move their storage. They also don't have to take a forced outage and restart the application after the move is completed. Instead, a move can be made instantly between sites, over distance, and the data remains online and available during the move – no outage or downtime is required. VPLEX also provides a single, all-inclusive interface for both EMC and non-EMC arrays; so even if your customer has a mixed storage environment, VPLEX still provides an easy, all-encompassing solution for customers.

Availability

VPLEX
Delivers continuous storage availability

Multi-site Cluster

VPLEX WITNESS
VM

AccessAnywhere

Active — Active

VPLEX Advantage
• Enables stretched clusters with zero RTO
• Automated recovery with near-zero RPO
• Support for sync and async distances

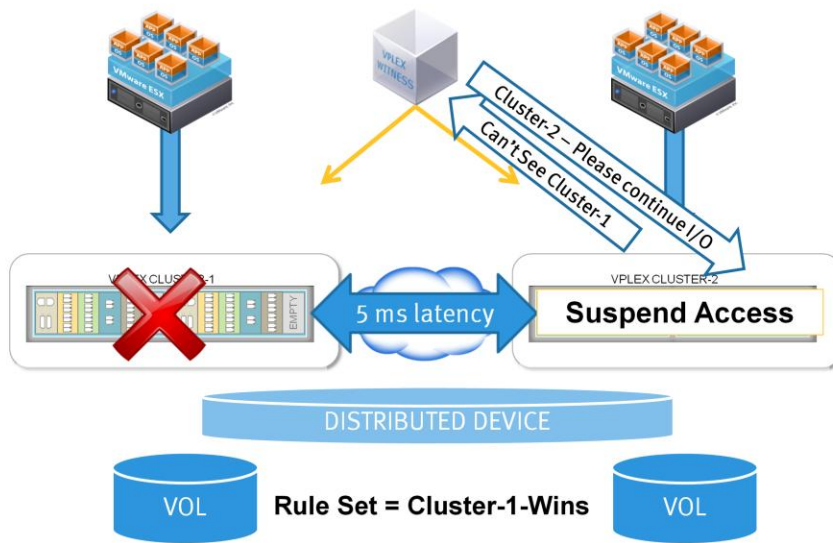VPLEX GEOSYNCHRONY v5.0 Technical Differences   9

With VPLEX, you can offer customers true High Availability – this means operations continue and data remains online even if a failure occurs. No manual intervention is required, no recovery or restart is necessary within synchronous distances with VPLEX Metro. In fact, you can think of VPLEX Metro as providing Disaster Avoidance instead of just Disaster Recovery – because there is literally no disaster event from which to recover.

When two VPLEX clusters are connected together with VPLEX Metro or VPLEX Geo, VPLEX gives customers shared data access between sites, meaning the same data, not a copy but the same data, exists at more than one location simultaneously. Instead of using costly and complex processes, and having to choose RPOs and RTOs, VPLEX can withstand a component failure, a site failure or loss of communication between sites and still keep the application and data online and available. This is true High Availability, with VPLEX deployed in a real active-active configuration, that also includes a new VPLEX Witness. The VPLEX Witness exists in a different failure domain from either VPLEX cluster to arbitrate between sites, if necessary, when a failure occurs.

Note that when deploying VPLEX Metro, because it is within synchronous distances, you can offer customers a High Availability solution with zero RPO AND zero RTO. With VPLEX Geo, because a greater distance can be supported between the two clusters and up to a 50 millisecond response time, zero RPO is not guaranteed, but customers can still achieve near-zero RTO with VPLEX Geo and the failover could still be automated. Note that while VPLEX High Availability gives customers several advantages around continuous uptime and provides Disaster Recovery capabilities, VPLEX should not necessarily be positioned in customer scenarios where their SOLE requirement is Disaster Recovery. For those customers whose sole requirement is Disaster Recovery, then traditional solutions such as SRDF or RecoverPoint may be the most appropriate.

With VPLEX, you're also giving customers an architecture that is 100% non-disruptive for all hardware and software upgrades – forever. So with VPLEX, you can provide a solution that allows instant access to information in real-time, eliminates operational overhead and reduces complexity, and reduces the time required to copy and distribute data across locations.

9

VMware Metro HA: Without Cross-Connect

Cluster-2 – Please continue I/O

Can't See Cluster-1

VPLEX CLUSTER-1   5 ms latency   VPLEX CLUSTER-2
EMPTY   **Suspend Access**

DISTRIBUTED DEVICE

VOL   **Rule Set = Cluster-1-Wins**   VOL

VPLEX GEOSYNCHRONY v5.0 Technical Differences   10
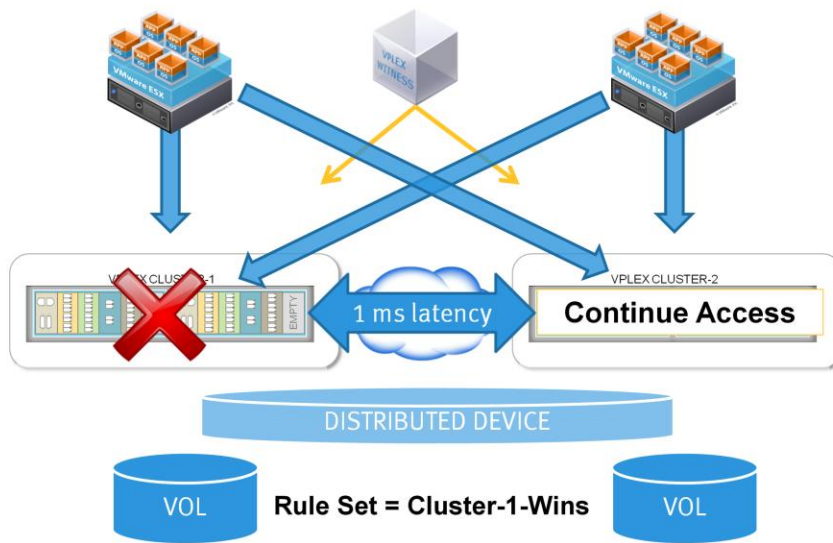
The Metro HA deployment is one of major use cases of VPLEX GeoSynchrony v5.0 when the availability use case is spoken about. There are two variations of the Metro HA deployment, one includes the cross-cluster connection and the other does not include the cross-cluster connection. Shown here is the variation without the cross-cluster connection.

VPLEX GeoSynchrony v4.X did not include the VPLEX Witness and without the Witness, there was potential for a system-wide DU event. With VPLEX GeoSynchrony v5.0, VPLEX can help to prevent a system-wide DU event. In this example the user is using VPLEX GeoSynchrony v4.X and has configured the rule-set to be cluster-1-wins which means that if there is a failure, cluster-1 will allow I/O to continue and cluster-2 must suspend I/O access. This idea is ok as long as the user's applications are running on cluster-1 and cluster-1 did not fail. An issue arises when cluster-1 fails as it results in cluster-2 suspending I/O access and also causing a system wide DU event. With the introduction of the VPLEX Witness in GeoSynchrony v5.0 the system-wide DU event can be prevented. The VPLEX Witness resides in a separate failure domain and listens for messages from the two clusters. The clusters send the VPLEX Witness heath report messages twice every second. The witness does nothing as long as the clusters are sending healthily messages to it. If Cluster-1 fails, Cluster-2 will send a message to the VPLEX Witness saying that it can't communicate with Cluster-1. The VPLEX Witness will start a count down of 10 seconds. If 10 seconds passes and a healthy message is not sent to the VPLEX Witness about cluster-1, the VPLEX Witness sends a message out to tell cluster-2 to continue I/O and to tell cluster-1 to suspend. In this case Cluster-2 will continue I/O and there will not be a system-wide DU event. However, in this case the applications at cluster-1 cannot continue I/O, but at least the applications at cluster-2 are running perfectly fine and there isn't a system-wide DU event. In this deployment there is a 5ms round-trip latency requirement between the clusters. More information about the VPLEX Witness will be described in the upcoming slides.

VMware Metro HA: Cross-Cluster Connect Deployment

VPLEX CLUSTER-1 — 1 ms latency — VPLEX CLUSTER-2 **Continue Access**

DISTRIBUTED DEVICE

VOL **Rule Set = Cluster-1-Wins** VOL

The other variation of the Metro HA deployment includes the cross-cluster connection.   The cross-cluster connection is where one or more HBAs from the hosts connect to cluster-1 and also cluster-2. In the previous deployment, the applications at cluster-1 could not read or write to the storage because the cluster was down.   The cross-connection with the VPLEX Witness resolves this issue as the applications at Cluster-1 can still access the storage through access at Cluster-2. One of the requirements for the cross-cluster connection is that the round-trip latency between the two clusters must be 1ms or less. The round-trip latency between the hosts and clusters must also be 1ms or less. Therefore, this type of deployment should only be used when the latency is low. The VPLEX Witness only provides advice to distributed volumes that are part of a consistency group. It does not provide advice to distributed volumes that are not part of a consistency group.
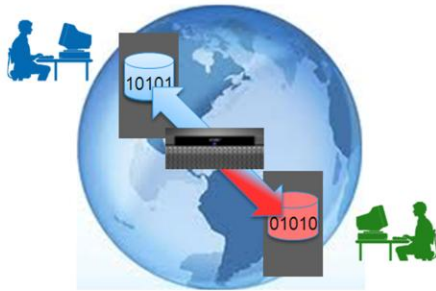
With VPLEX, the same data can be accessible to all users at all times – even if they are at different sites. The data is literally shared, not copied, so that a change made in one site shows up right away at the other site. This is a huge benefit for customers who have always had to rely on shipping large log files and data sets back and forth across sites, and then had to wait for updates to be made in another location before they could resume working on it again.

VPLEX literally makes the data available in both locations, and because VPLEX is so intelligent, it doesn't need to ship the entire file back and forth like other solutions – it only sends the changed updates as they are made, thus greatly reducing bandwidth costs and offering significant savings over other solutions. Deploying VPLEX in conjunction with third party WAN optimization solutions can deliver even greater benefits. And with VPLEX's AccessAnywhere, the data remains consistent, online and available – always.

Module 2: VPLEX GeoSynchrony 5.0 Hardware Overview

This module describes the new VPLEX GeoSynchrony VS2 hardware platform and how it differs from the previous VS1 platform. The VS2 platform will be described in detail.

Upon completion of this module, you should be able to:

- Analyze the delta between the VPLEX GeoSynchrony VS1 and VS2 hardware
- Identify and describe the components of the VS2 platform and how they function in the overall VPLEX GeoSynchrony design
- Identify VPLEX GeoSynchrony v5.0 front-end, back-end, LOCAL-COM, and WAN-COM ports

EMC²
where information lives'

This module describes the new VPLEX GeoSynchrony VS2 hardware platform and how it differs from the previous VS1 platform. The VS2 platform will be described in detail.
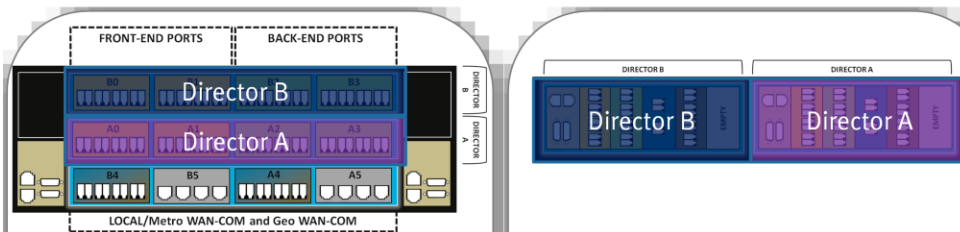
Engine Comparison: Overview

FRONT-END PORTS   BACK-END PORTS

Director B

Director A

LOCAL/Metro WAN-COM and Geo WAN-COM

DIRECTOR B
DIRECTOR A

DIRECTOR B   DIRECTOR A

Director B   EMPTY   Director A   EMPTY

VS1 Engines
- Larger footprint - 4-RU
- Directors are horizontally-oriented
- 4 front-end and 4 back-end I/O Modules (IOM), each with four 8-Gbps Fibre Channel Ports
- If all ports are occupied, average throughput will be approximately 1.6-Gbps because of oversubscription
- IOM B4 and A4 house two LOCAL-COM ports and two Metro WAN-COM ports
- IOM B5 and A5 house four 1-GigE ports; only the first two are active

VS2 Engines
- Smaller footprint - 2-RU
- Directors are situated side-by side
- 2 front-end and 2 back-end I/O Modules (IOM), each with four 8-Gbps Fibre Channel Ports
- All ports are non-oversubscribed
- IOM B2 and A2 are **either** unpopulated (for VPLEX Local), populated with a Metro WAN-COM FC card (for VPLEX Metro), or a Geo WAN-COM 10-GE card (for VPLEX Geo)
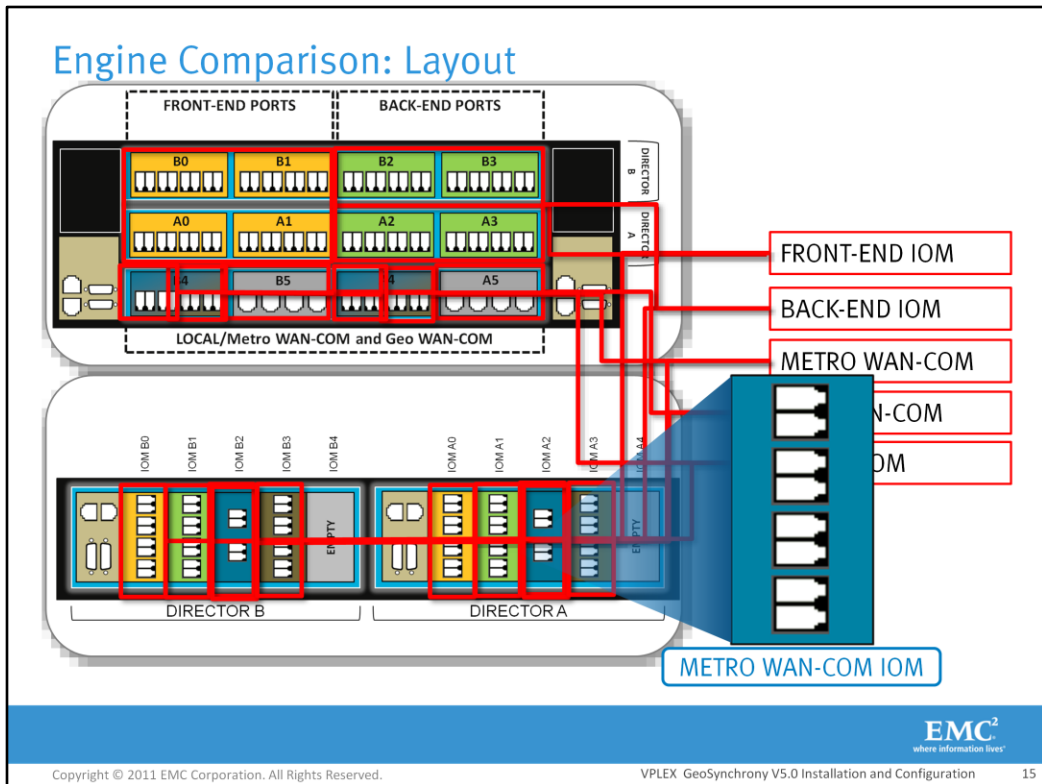- IOM B3 and A3 is used for LOCAL-COM; only the bottom two are active

EMC²
where information lives®

 Shown is a side-by-side comparison of the VS1 engine and the new VS2 engine. The VS1 engines required 4U (rack-units) of space, whereas the VS2 engines only requires 2U. In addition, the directors are situated horizontally on the old engines , while they are side-by-side in the new engines. The old engines had more front-end and back-end ports than the new engines. However, they did not operate at full line-rate if all ports were used. Rather, they averaged approximately 1.6-Gbps per port, if fully-populated. The new engines offer greater performance and throughput, offering eight front-end and eight back-end ports, each running at 8-Gbps non-oversubscribed.

The VS1 engines used I/O modules B4 and A4 for LOCAL-COM and Metro WAN-COM. The first two ports were always used for LOCAL-COM and the next two were only used for Metro WAN-COM. If deploying VPLEX Geo on a VS1 engine, I/O module B5 and A5 will be used. The I/O modules in each engine house four 1-GigE ports; two of which will be used. The new engine offers separate I/O modules for each function. I/O modules B2 and A2 house either a VPLEX Metro WAN-COM card, with four 8-Gbps ports or a VPLEX Geo WAN-COM card with two 10-GigE ports. If the engine is being used in a VPLEX Local, slots B2 and A2 are empty. I/O module A3 and B3 are used for LOCAL-COM. Although the new LOCAL-COM I/O modules each have four 8-Gbps Fibre Channel ports, only the first two (bottom two) ports are used.

Engine Comparison: Layout

FRONT-END PORTS   BACK-END PORTS

FRONT-END IOM
BACK-END IOM
METRO WAN-COM
N-COM
OM
METRO WAN-COM IOM

VPLEX GeoSynchrony V5.0 Installation and Configuration   15

The layout of the VS1 and VS2 engines are very different.

There are four total front-end I/O modules on the VS1 engine, that contain 8-Gbps Fibre Channel ports.

On the VS2 engine, there are two front-end I/O modules that contain four 8-Gbps Fibre Channel ports.

The reduction in the number of total ports is not detrimental to the performance of the hardware because the VS2 engine is non-oversubscribed. The VS1 engine, if fully-loaded averages about 1.6-Gbps per port due to oversubscription. On the VS2 hardware, all ports run at the full line rate of 8-Gbps. Overall port throughput is much better on a fully-loaded VS2 engine than a fully-loaded VS1 engine.
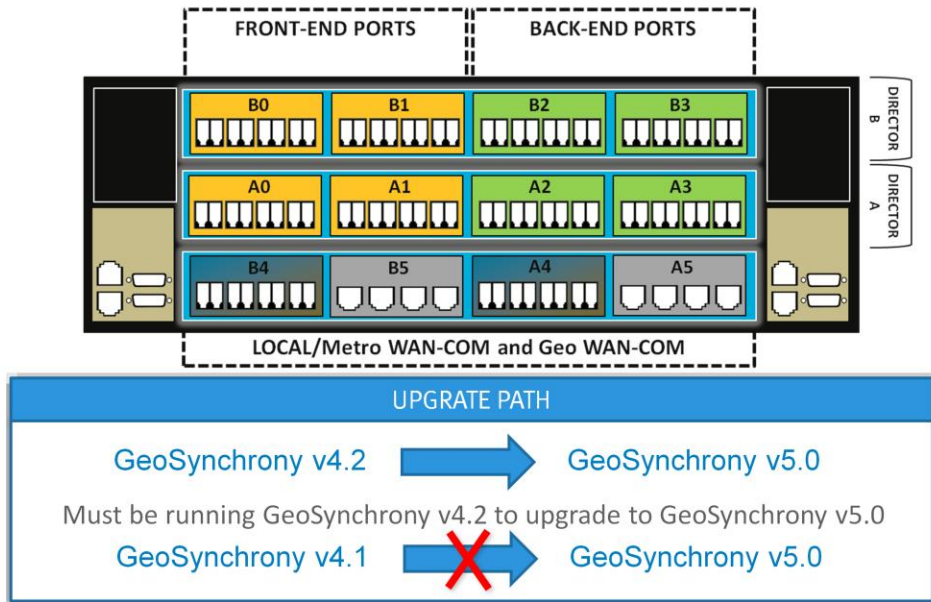
There are four back-end I/O modules on the VS1 engine, compared to the VS2 engine. The ports in the VS2 engines are also non-oversubscribed at 8-Gbps.

The VS1 engine uses two I/O module for LOCAL-COM and WAN-COM. The two ports on the right are for the WAN-COM. On the VS2 engine, there is a dedicated Metro WAN-COM card that is installed in I/O module B2 and A2. Although this card contains four 8-Gbps Fibre Channel ports, only two of these ports are available for use.

The VS1 engine contains two I/O modules, each with four 1 GigE ports for Geo WAN-COM. Only the first two ports in each card are usable. The VS2 hardware has two I/O modules, each with two 10 GigE ports. VS1 Geo WAN-COM ports start with 'GE' whereas VS2 Geo WAN-COM ports start with 'XG.'

LOCAL-COM communication on the VS1 engine runs over the first two ports in I/O module B4 and I/O module A4. The VS2 hardware has an I/O module dedicated to LOCAL-COM, populated in B3 and A3. Although the LOCAL-COM card on the new engine has four 8-Gbps ports, only two are used.

  15

Compatibility: Legacy VS1 Engines

FRONT-END PORTS   BACK-END PORTS

DIRECTOR B

B0   B1   B2   B3

DIRECTOR A

A0   A1   A2   A3

B4   B5   A4   A5

LOCAL/Metro WAN-COM and Geo WAN-COM

UPGRADE PATH

GeoSynchrony v4.2 → GeoSynchrony v5.0

Must be running GeoSynchrony v4.2 to upgrade to GeoSynchrony v5.0

GeoSynchrony v4.1 ✗→ GeoSynchrony v5.0
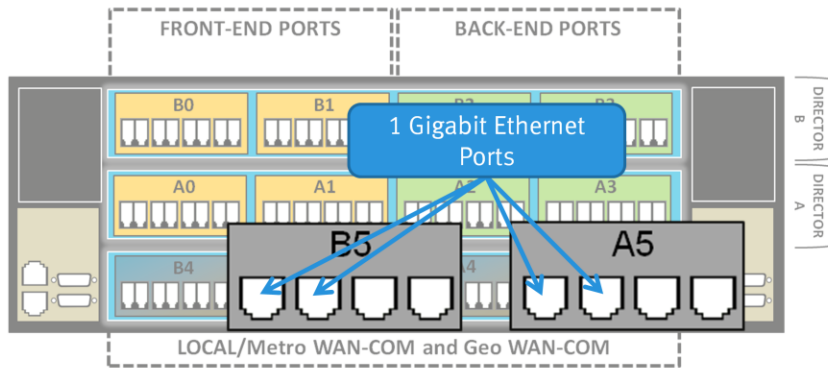
VPLEX GEOSYNCHRONY v5.0 Technical Differences   16

GeoSynchrony v5.0 is compatible with the VS1 hardware. However, there is no direct upgrade path from GeoSynchrony v4.1 (or prior version) to GeoSynchrony v5.0. Rather, the VS1 engines must be first upgraded to GeoSynchrony v4.2 prior to upgrading to GeoSynchrony v5.0.

    16

Compatibility: Legacy VS1 Engines

FRONT-END PORTS | BACK-END PORTS

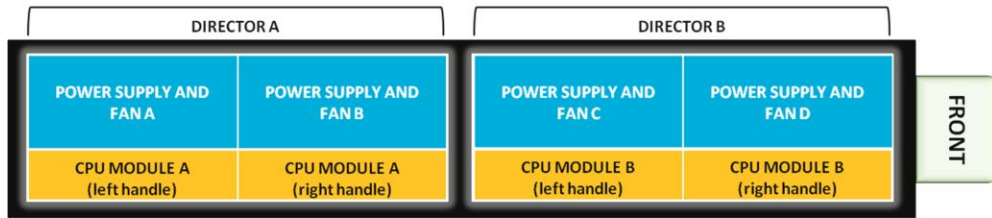1 Gigabit Ethernet Ports

LOCAL/Metro WAN-COM and Geo WAN-COM

- I/O Modules (IOM) B5 and A5 become active for Geo WAN-COM once upgrade to GeoSynchrony v5.0
- Only the first two ports are used in B5 and A5

VPLEX GEOSYNCHRONY v5.0 Technical Differences   17

Following an upgrade to GeoSynchrony v5.0, I/O Modules B5 and A5 become available for Geo WAN-COM connectivity. I/O modules B5 and A5 each contain four 1-GigE ports. Only the first two ports in the I/O modules are used. Both the VPLEX Metro WAN-COM ports and the VPLEX Geo WAN-COM ports cannot be used at the same time.

VPLEX VS2 Architecture: Front View

DIRECTOR A		DIRECTOR B	
POWER SUPPLY AND FAN A	POWER SUPPLY AND FAN B	POWER SUPPLY AND FAN C	POWER SUPPLY AND FAN D
CPU MODULE A (left handle)	CPU MODULE A (right handle)	CPU MODULE B (left handle)	CPU MODULE B (right handle)

FRONT

- Two co...rector
  - A an...
  - C an...
- One CP...
  - CPU module A belongs to Director A
  - CPU module B belongs to Director B
 - CPU modules house the memory and one SSD

U02

U01

VPLEX GEOSYNCHRONY v5.0 Technical Differences
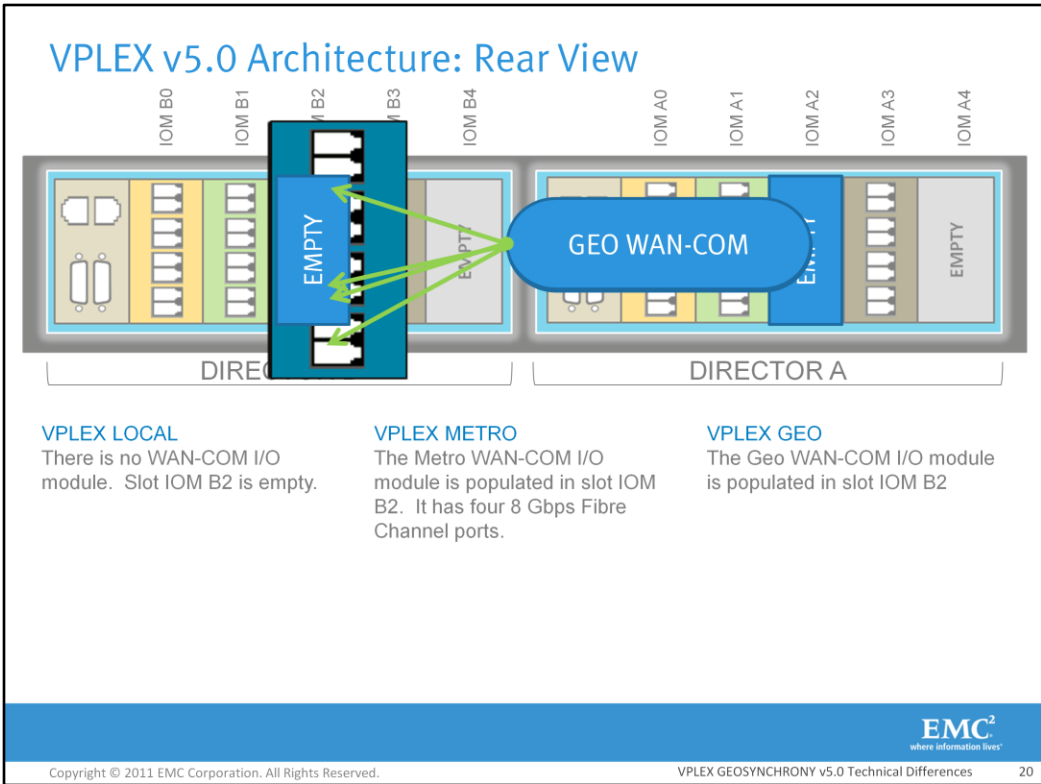
18

EMC²
where information lives*

The VS2 hardware has a smaller footprint than the VS1 hardware , taking up only 2U of space.   On the front of the VS2 engines are four power supply and fan units as well as two CPU modules. Power Supply and Fan A and B belong to director A on the left and Power Supply and Fan C and D belong to director B on the right. CPU module A belongs to director A and CPU module B belongs to director B. Each CPU module contains one Solid State Drive (SSD) card, which contains the GeoSynchrony code. In addition, each CPU module also contains the memory and CPU.

VPLEX v5.0 Architecture: Rear View

(Slide 19, VPLEX GEOSYNCHRONY v5.0 Technical Differences)
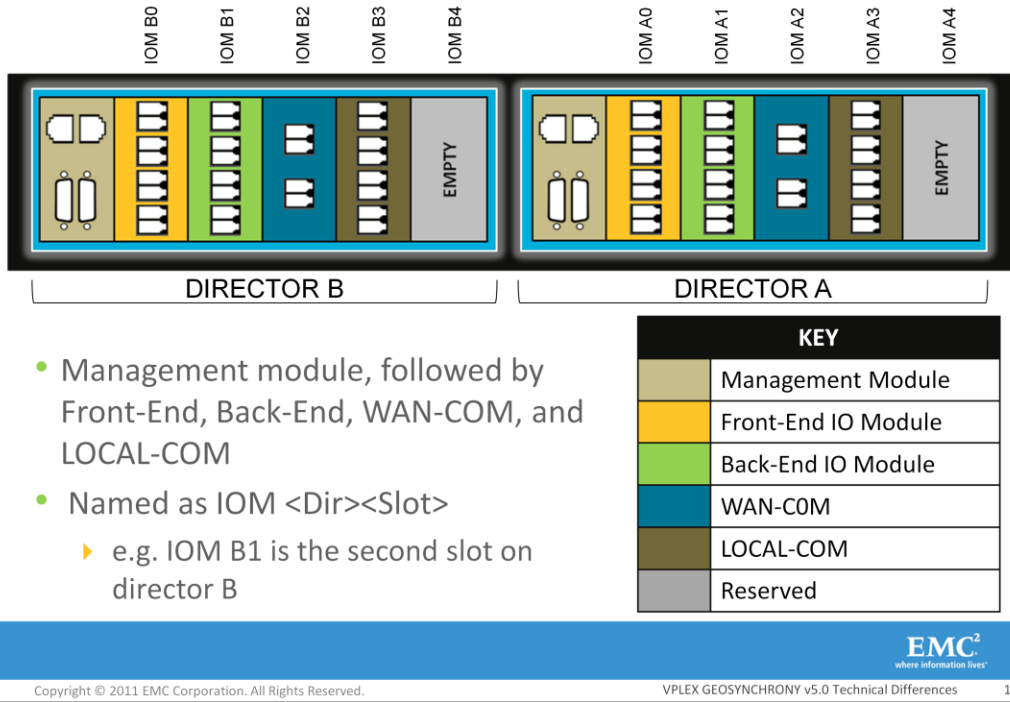
All I/O modules can be accessed from the rear of the engine. From the rear, director B is on the left and director A is on the right. Each director contains a Management module, followed by a Front-End, Back-End, WAN-COM (if applicable), LOCAL-COM I/O module, and an empty slot reserved for future use.

Each slot is labeled as **IOM <Director [A | B]><Slot [0-4]>**, when viewed in the CLI.

VPLEX v5.0 Architecture: Rear View

VPLEX LOCAL
There is no WAN-COM I/O
module. Slot IOM B2 is empty.

VPLEX METRO
The Metro WAN-COM I/O
module is populated in slot IOM
B2. It has four 8 Gbps Fibre
Channel ports.

VPLEX GEO
The Geo WAN-COM I/O module
is populated in slot IOM B2

VPLEX GEOSYNCHRONY v5.0 Technical Differences   20

The I/O modules in slots B2 and A2 depend on the VPLEX system ordered.

If the system is a VPLEX Local, slots B2 and A2 will not be populated with an I/O module. There is no card populated in these slots because there is no inter-cluster connectivity required for a single-cluster VPLEX.

If the system is a VPLEX Metro, slots B2 and A2 are populated with a Metro WAN-COM card. The Metro WAN-COM card has four 8-Gbps Fibre Channel ports; two of which are usable.

For VPLEX Geo systems, slots B2 and A2 will contain Geo WAN-COM cards.

VPLEX I/O Module Tour

This is an overview of a VS2 director. Please click on an I/O module to learn more about the I/O module.

Launch

This is an overview of a VS2 director. Please click on an I/O module to learn more about the I/O module.

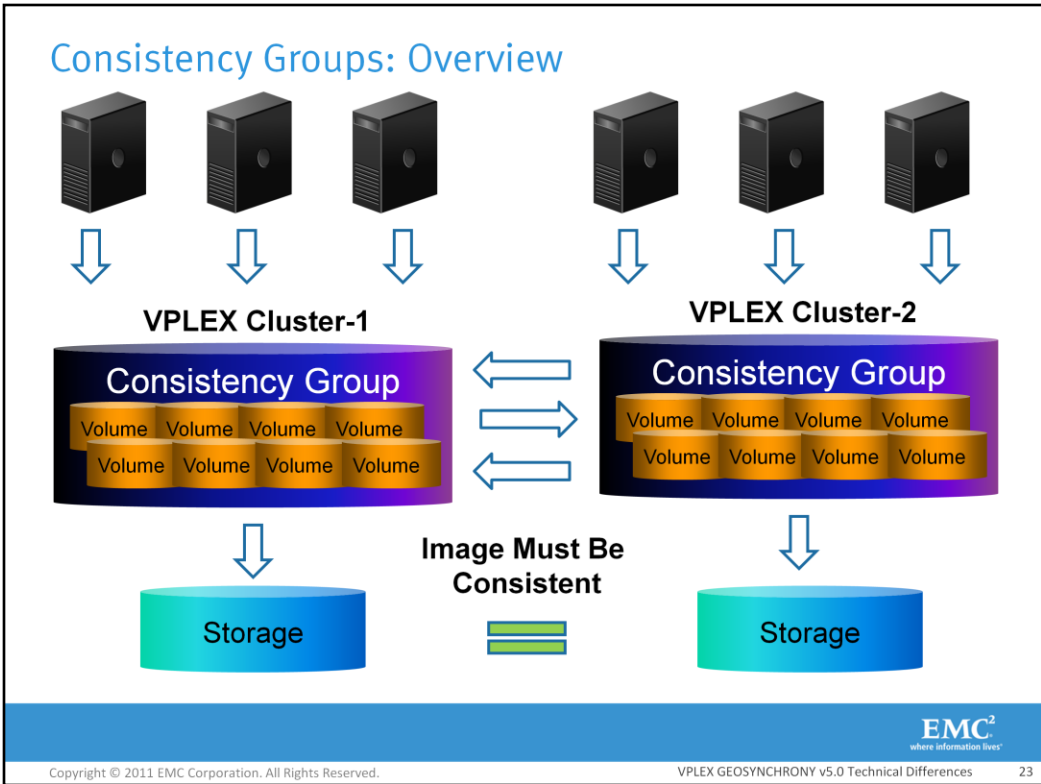## Module 3: VPLEX GeoSynchrony 5.0 Features

This module covers the new features included in VPLEX GeoSynchrony v5.0.

Upon completion of this module, you should be able to:

- Describe the path that I/O takes when synchronous or asynchronous consistency groups are being utilized
- Create and move consistency groups from cluster to cluster
- Configure VPLEX to respect the thinness of a thinly provisioned volume
- Describe how the VPLEX Witness helps to prevent system wide Data Unavailable (DU) events
- Deploy the VPLEX Witness into a customers environment
- Utilize the VPLEX Element Manager API to aid in executing scripts remotely on a VPLEX GeoSynchrony v5.0 system
- Explain how VPLEX cache vaulting protects cache in during power loss situations
- Utilize the VPLEX Management Console Performance Monitor

EMC²
where information lives

This module covers the new features included in VPLEX GeoSynchrony v5.0.

Consistency Groups: Overview

VPLEX Cluster-1

Consistency Group

| Volume | Volume | Volume | Volume |
| Volume | Volume | Volume | Volume |

VPLEX Cluster-2

Consistency Group

| Volume | Volume | Volume | Volume |
| Volume | Volume | Volume | Volume |

Image Must Be Consistent

Storage

Storage

VPLEX GEOSYNCHRONY v5.0 Technical Differences   23

A consistency group is a set of virtual volumes that are grouped together that require write order consistency. Volumes in the group require the same I/O behavior in the event of an inter-cluster link outage or cluster outage. Consistency groups commit their deltas to disk in a coordinated fashion to maintain consistency. In a consistency group, the data on disk is guaranteed to represent a consistent point in time. A set of LUNs for a database application requires a consistent image on disk at all times. Shown here is an asynchronous consistency group that spans two clusters in a VPLEX Geo. The hosts at both clusters write to the VPLEX distributed volumes in the consistency group. Writes get written to the back-end storage in a consistent fashion. This guarantees that the image on the back end storage is an exact copy at both clusters.

Local Synchronous Consistency Groups
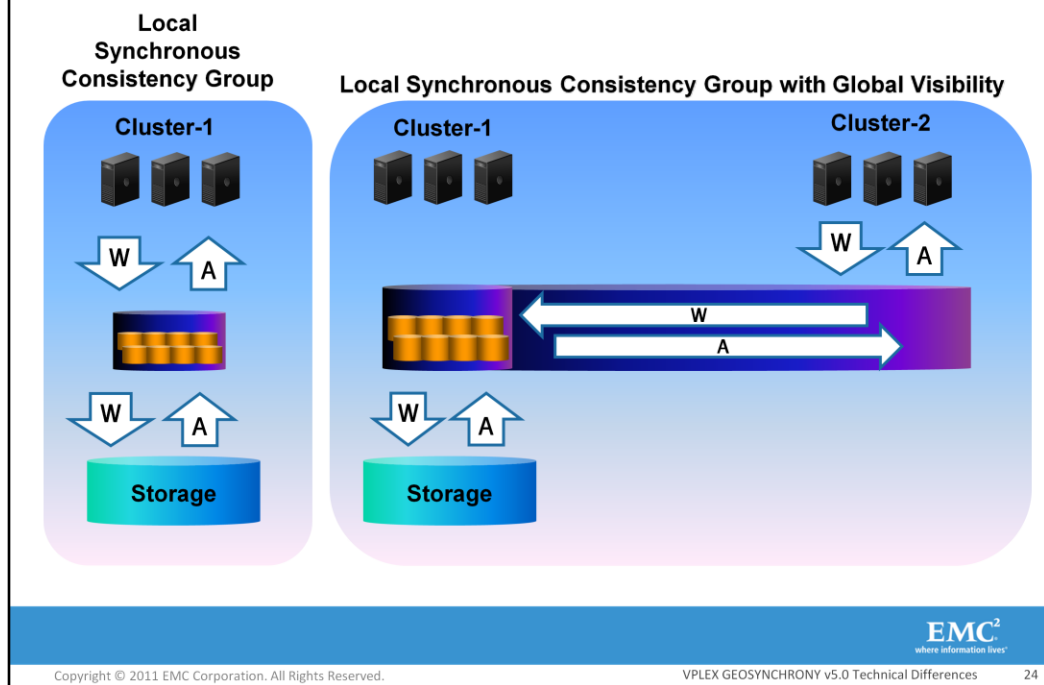
Local synchronous consistency groups are new in VPLEX GeoSynchrony v5.0. A local synchronous consistency group is a logical grouping of volumes that are grouped together in order to organize the volumes.   I/O that is performed on local synchronous consistency groups is always synchronous. This means that I/O must be written through VPLEX to the back-end cache of the storage array prior to sending the acknowledgement to the host). They can be beneficial for applications that are local to the cluster and need some type of organization. Local synchronous consistency groups can only be read and written to by their local cluster unless they are local synchronous consistency groups with global visibility.

 Local synchronous consistency groups with global visibility allow the remote cluster to also read and write to the consistency group. However, all reads and writes to the consistency group from the remote cluster must pass over the inter-cluster link in order to access the volumes in the consistency group. This allows the remote cluster to have instant on demand access to the consistency group, but also adds additional latency for the remote cluster. Local synchronous consistency groups with global visibility are only supported in a VPLEX Metro environment. Only local volumes can be placed into the local synchronous consistency group with global visibility. Local synchronous consistency groups with global visibility cannot be set to asynchronous cache mode. I/O that goes to local synchronous consistency groups with global visibility will always be synchronous.

Local Synchronous Consistency Group Mobility

Entire consistency groups can be moved from one cluster to another cluster. This feature can be used to relocate an entire data center. Consistency Group Mobility moves the volumes within a local consistency group to the target cluster. Depending upon the cluster type (Metro or Geo) consistency groups can be moved synchronously or asynchronously. Synchronous mobility is only supported in a VPLEX Metro environment. During a synchronous mobility operation writes from the application will be synchronously replicated to the remote site. Writes are copied to the remote site before they are acknowledged back to the host. Asynchronous consistency group mobility is only supported in VPLEX Geo environments. During the mobility operation, writes from the application will be asynchronously replicated to the remote site. Writes are acknowledged to the host before they are copied to the remote site. Only 25 sub-jobs or devices can be moved at a time in both types of mobility. However, like device mobility an unlimited number of devices can be queued to run.
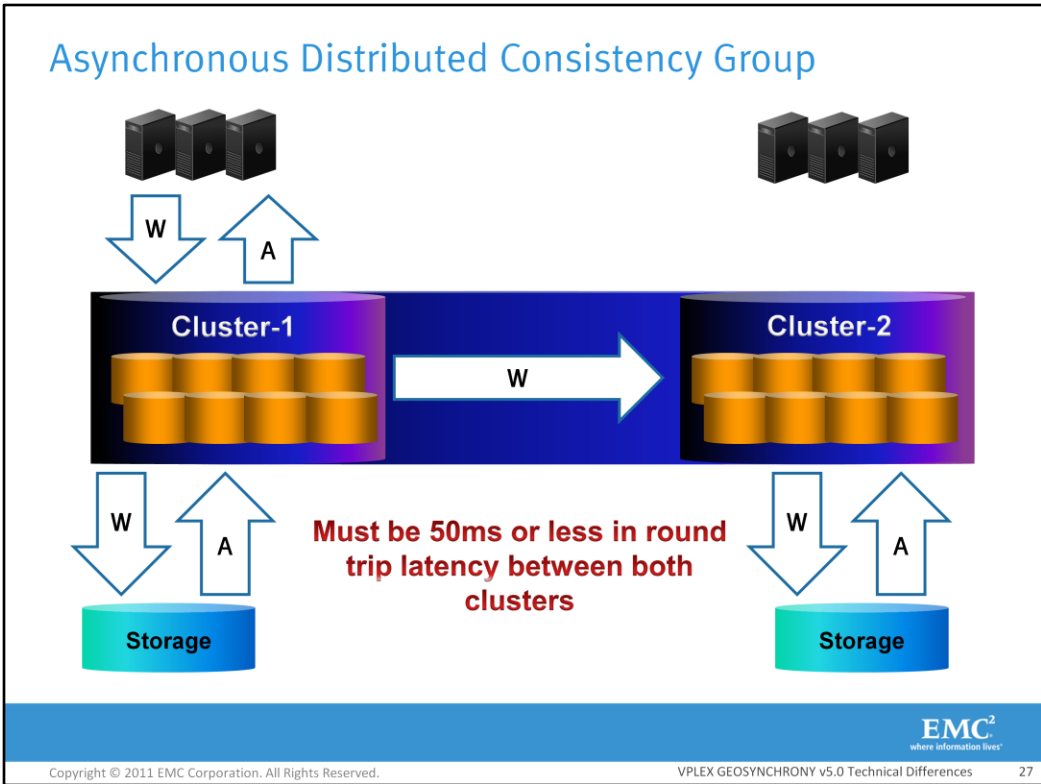
Synchronous Distributed Consistency Group
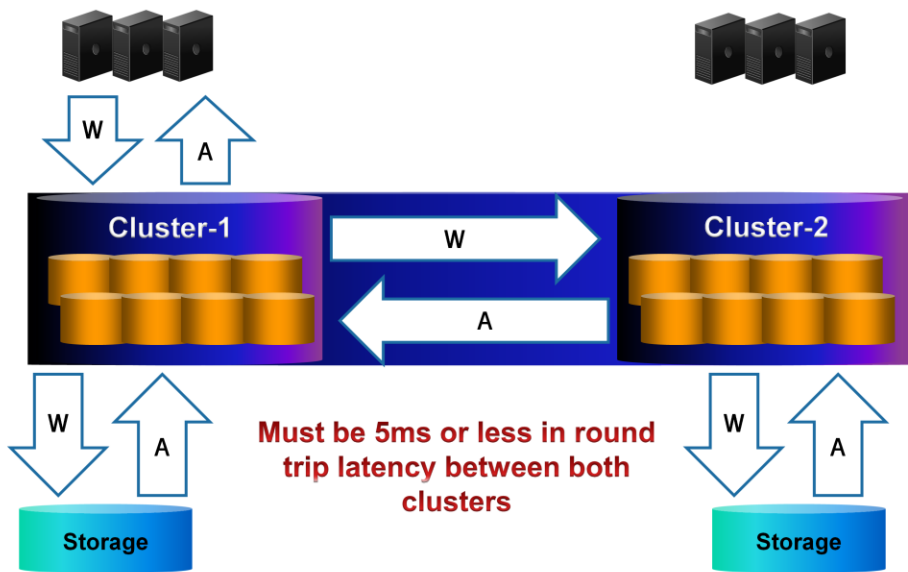
A synchronous distributed consistency group differs from a local synchronous consistency group in the respect that the physical storage resides at both clusters. When writing to a synchronous distributed consistency group, I/O must be written to the cache of the array at both clusters prior to sending the acknowledgement to the host. In this example, the host writes to the VPLEX at Cluster-1 and the write passes through VPLEX at cluster-1 to the back-end array. It also gets sent across the inter-cluster link to the remote cluster where it eventually gets sent to the back-end storage at the remote cluster. The acknowledgement gets sent back to both VPLEX clusters from the arrays. It also gets sent from the remote VPLEX to the local cluster. Finally, the host is acknowledged. A synchronous distributed consistency group is similar to a distributed volume. Synchronous distributed consistency groups can only be created in a VPLEX Metro. Only distributed volumes can be added to synchronous distributed consistency groups. The VPLEX Witness is able to provide guidance to synchronous distributed consistency groups.

Asynchronous Distributed Consistency Group

Cluster-1

Cluster-2

W

Must be 50ms or less in round trip latency between both clusters

Storage

Storage

An asynchronous distributed consistency group differs from a synchronous distributed consistency group in that the host gets acknowledged after the write reaches VPLEX's cache and has been mirrored to another director.   In this example the host writes I/O to the VPLEX cluster. The I/O is mirrored to another director within the cluster and then an acknowledgement is sent to the host. Writes collect at the cluster in the form of a delta which is a grouping of writes.   At a later point in time deltas get exchanged between the two clusters. The clusters send communication messages back and forth between each other to facilitate the exchange.   Lastly, a global merged delta is written from the clusters to the back-end array cache and an acknowledgment is sent to VPLEX. Asynchronous distributed consistency groups are only supported in a VPLEX Geo environment and the maximum round trip latency between cluster must be 50 ms or less.

This is an example of an active/passive clusters where writes are only being written from hosts at one cluster. The I/O flow for consistency groups will be described in more detail. The VPLEX Witness is able to provide guidance to asynchronous distributed consistency groups.

Distributed Consistency Groups: I/O Behavior

Synchronous Distributed Consistency Group

Rule Sets	I/O Behavior after an Inter-cluster Link Failure or Cluster Failure
No-Automatic-Winner	I/O suspends at both clusters
Cluster-1-Wins	Cluster-1 continues I/O and Cluster-2 suspends I/O
Cluster-2-Wins	Cluster-2 continues I/O and Cluster-1 suspends I/O.

Asynchronous Distributed Consistency Group

Rule Sets	I/O Behavior after an Inter-cluster Link Failure or Cluster Failure
No-Automatic-Winner	I/O suspends at both clusters
Cluster-1-Wins	Cluster-1 continues I/O and Cluster-2 suspends I/O. Both clusters suspend I/O if cluster-2 is active during the failure.
Cluster-2-Wins	Cluster-2 continues I/O and Cluster-1 suspends I/O. Both clusters suspend I/O if Cluster-1 is active during the failure.
Active-Cluster-Wins	I/O continues at the cluster that was active during the failure. If both clusters were active, I/O suspends at both clusters. If both clusters were passive, I/O continues at the clusters that wrote I/O to the consistency group last.

 An administrator can control the behavior of I/O during a failure. A failure here is considered to be an inter-cluster link failure or a complete cluster failure. The behavior during a failure is controlled by predefined rules that are configured on the consistency groups by the administrator. For synchronous distributed consistency groups there are three different rules that can be configured. These are No Automatic winner, Cluster-1-Wins, and Cluster-2-Wins. No-Automatic-Winner means that if there is a failure, both clusters suspend I/O. Cluster-1-Wins, means that during a failure I/O will continue at Cluster-1 and I/O will suspend at Cluster-2. Cluster-2-Wins means that during a failure, Cluster-2 will continue I/O and Cluster-1 will suspend.

    All of the rules that can be configured on synchronous distributed consistency groups can also be configured on asynchronous distributed consistency groups. Asynchronous distributed consistency groups also have one additional rule that may be used called the Active-Cluster-Wins rule. When using the Active-Cluster-Wins rule, I/O will continue at the cluster that was active. Active means that the cluster has data in cache that has yet to be written to the back-end storage (dirty data). I/O must suspend at both clusters if both clusters were active during the failure. This is because the cache image is inconsistent on both clusters and must be rolled back to point where both clusters had a consistent image in order to continue I/O. VPLEX performs the rollback automatically and suspends the volumes if this is the case. This is true when there is a site failure or when there is a long inter-cluster link failure. Short inter-cluster link failures will not cause the rule-sets to fire and as a result a data rollback will not be required. This behavior also applies to the Cluster-1-Wins and Cluster-2-Wins rules when dealing with asynchronous consistency groups. If both clusters are passive (There is no data in cache that needs to be written to BE storage) when using the active cluster wins rule, the cluster that was written to last will be the cluster that is allows to continue I/O.

Local Synchronous Consistency Groups and Individual Distributed Volumes: I/O Behaviors

Local Synchronous Consistency Group With Global Visibility

Rule Sets	I/O Behavior after a WAN-COM Failure or Cluster Failure
No-Automatic-Winner	I/O suspends at both clusters
Cluster-<N>-Wins	I/O will continue at the local cluster where the storage resides. The other cluster will suspend I/O.

Individual Distributed Volume

Rule Sets	I/O Behavior after a WAN-COM Failure or Cluster Failure
Cluster-1-detaches	Cluster-1 continues I/O and Cluster-2 suspends I/O
Cluster-2-detaches	Cluster-2 continues I/O and Cluster-1 suspends I/O.

EMC²
where information lives·

Local synchronous consistency groups with global visibility have two rules that can be applied, "No-Automatic-Winner and Cluster-<N>-Wins" where <N> is the local cluster number. The Cluster-<N>-Wins means that the local cluster where the storage resides will be able to continue I/O during a failure and the remote cluster will suspend I/O.

Two rules can be assigned to individual distributed volumes. These are Cluster-1-Detaches and Cluster-2-Detaches. Cluster-1-Detaches means the same thing as Cluster-1-Wins and Cluster-2-Detaches means the same thing as Cluster-2-Wins. These rules have been part of VPLEX GeoSynchrony since the beginning and are listed here for completeness.

Asynchronous Consistency Group Attributes

```
VPlexcli:/clusters/cluster-1/consistency-groups/async_consist> ll

Attributes:
Name                    Value
--------------------    --------------------------------------------------------
active-clusters         [cluster-1]
cache-mode              asynchronous
detach-rule             winner cluster-1 after 5s
operational-status      [(cluster-1,{ summary:: ok, details:: [] }), (cluster-2,{
                        summary:: ok, details:: [] })]
passive-clusters        [cluster-2]
storage-at-clusters     [cluster-1, cluster-2]
virtual-volumes         [dd_vol]
visibility              [cluster-1, cluster-2]


Contexts:
Name
--------
advanced
```

These are the attributes of an asynchronous consistency group.   Active clusters shows which clusters are active or which clusters have received writes from hosts where the writes have not been written to the back-end array. In this consistency group cluster-1 is the only cluster that is active.   Cache mode is the mode of cache for the consistency group. The cache-mode can be either synchronous or asynchronous depending upon the cluster type (Metro or Geo). The detach-rule is the predefined rule set by the administrator that will be acted upon if there is a cluster failure or inter-cluster partition failure. In this example Cluster-1 will continue I/O if there is a failure.   The operational-status attribute will provide information on the state of the cluster and if the administrator needs to run a recovery command to enable access to the consistency group.   The passive clusters attribute displays the clusters that have been passive. Passive means that all of the data in cache has been written out to the back-end array. In this example cluster-2 is the only passive cluster.   The storage at clusters attribute defines where the storage is physically located for the consistency group. In this example the storage is located at both cluster-1 and cluster-2. This attribute is an attribute that the administrator can set and prevents devices not matching the storage-at-cluster value from being added to the consistency group. It is required to be set for distributed consistency groups. If it is unset, any device can be added to the consistency group, but the consistency group cannot be changed to an async consistency group. Therefore, only distributed volumes could be placed into an asynchronous consistency group.   The virtual volumes attribute displays the virtual volumes that are part of the consistency group. Virtual volumes can only be part of one consistency group at a time. The visibility attribute defines the clusters that are capable of viewing and accessing the volumes in the consistency group.

Synchronous Consistency Group Attributes

```
VPlexcli:/clusters/cluster-1/consistency-groups/sync_consist> ll

Attributes:
Name                  Value
-------------------   -------------------------------------------------------
active-clusters       []
cache-mode            synchronous
detach-rule           no-automatic-winner
operational-status    [(cluster-1,{ summary:: ok, details:: [] }), (cluster-2,{
                      summary:: ok, details:: [] })]
passive-clusters      []
storage-at-clusters   [cluster-1, cluster-2]
virtual-volumes       [next-dd_vol]
visibility            [cluster-1, cluster-2]


Contexts:
Name
--------
advanced
```

Synchronous consistency groups do not have as many attributes as asynchronous consistency groups. Synchronous consistency groups do not have the active-clusters or passive-clusters attributes because unlike asynchronous consistency groups, they will never have dirty data in cache. In synchronous consistency groups all I/O is written out to the back-end array prior to sending an acknowledgement to the host.

Asynchronous Consistency Group: Advanced Attributes

```
VPlexcli:/clusters/cluster-1/consistency-groups/async_consist/advanced> ll
Name                          Value
----------------------------  --------
auto-resume-at-loser          false
current-queue-depth           1
current-rollback-data         4K
default-closeout-time         0.5min
delta-size                    16M
local-read-override           true
max-possible-rollback-data    608M
maximum-queue-depth           20
potential-winner              -
write-pacing                  inactive
```

Consistency groups also have advanced attributes which are located in their advanced directory. The auto-resume-at-loser attribute defines if the losing cluster will automatically resume I/O at the loser after an inter-cluster partition heals. By default this attribute is set to false in order to give the administrator a chance to halt the application at the losing cluster prior to resuming I/O. This prevents the host at the losing cluster from having its data image sporadically change from under it. Hosts might cache data and the cached data could be from a point in the future. This data can be problematic and can lead to inconsistencies. This is the reason that a rollback is required along with an application restart which guarantees that the applications data reflects the rolled-back point in time recovery. For cross-cluster-connect configurations the auto-resume-at-loser attribute should be set to true in a VPLEX Metro and it should be set to false in a VPLEX Geo.

   The current-queue-depth attribute defines how many delta are in the closed queue at the cluster. In this example there is only 1 delta. The current-rollback-data attribute defines the amount of data in the closed queue at the cluster that would be lost if the cluster had to perform a data rollback. The default-close-out time attributed represents the maximum amount of time to wait before closing a delta. If the time is set to 0, VPLEX will close the delta when the delta size has been reached. The delta size attribute is a hard set value and represents the maximum amount of data that can collect for an asynchronous consistency on a director. Once a director reaches 16 MB in size for a consistency group, it will close the delta and place the delta on the close queue to be exchanged. The local-read-override attribute is set to true by default and means that a director will prefer to read from back-end storage over pulling the data across the inter-cluster link from the cache of a director in the other cluster. By default VPLEX will always prefer to get data from a peer director at the local site. Local-read-override is only applicable to non-dirty pages in cache (read cache). If reads across the inter-cluster link are faster than reading locally, then the attributed should be set to false. In almost all cases, this should be left as true. The maximum-possible-rollback-data attributed represents the maximum amount of data that could be lost if there is inter-cluster link outage and a cluster must rollback its view of data. The maximum-queue-depth attribute represents the maximum number of deltas that can exist in the closed queue.   The potential-winner attribute only applies to asynchronous consistency groups when the active-cluster-wins rule is applied to a consistency group. The attribute displays the cluster that is most likely to continue I/O if there is an inter-cluster partition.   The write-pacing attribute cannot be disabled and is an attribute used to slow down the pace of I/O to the consistency group in order to prevent too much I/O from going across the inter-cluster link. If write pacing is not currently being applied to I/O associated to the consistency group, the value of this property will be inactive. If write pacing is being applied to I/O associated with the consistency group, the value of this property will be active. Write pacing will be activated when the underlying consistency group utilization reaches 70%.
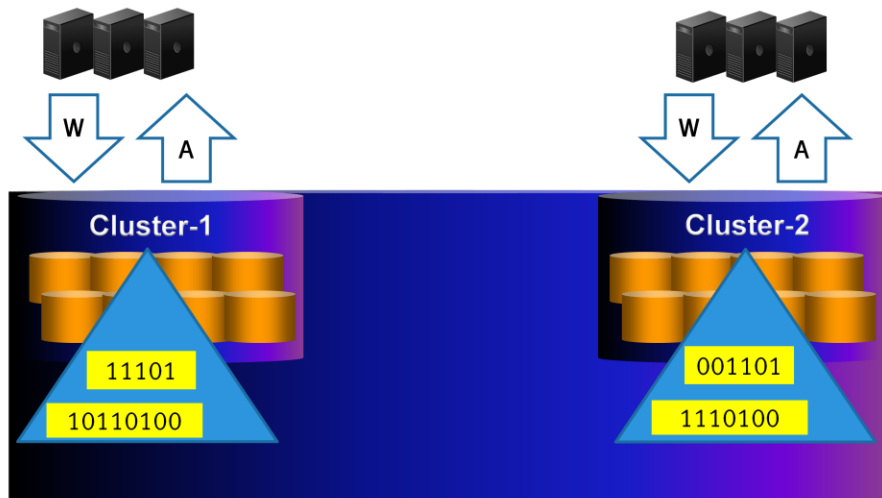
Synchronous Consistency Group: Advanced Attributes

```
VPlexcli:/clusters/cluster-1/consistency-groups/sync_consist/advanced> ll
Name                          Value
--------------------------    --------
auto-resume-at-loser          false   <------------------------------------
current-queue-depth           -
current-rollback-data         -
default-closeout-time         -
delta-size                    -
local-read-override           true    <------------------------------------
max-possible-rollback-data    -
maximum-queue-depth           -
potential-winner              -
write-pacing                  disabled <-----------------------------------
```

Synchronous consistency groups also have advanced attributes.   However, they do not have as many advanced attributes as asynchronous consistency groups do as they will never have dirty cache. The only advanced attributes that are used are the ones shown with the arrows. Write pacing is always disabled for synchronous consistency groups as all I/O is synchronous. The attributes with the "-" are never used.

Async Consistency Group: Open Delta Phase

Cluster-1
11101
10110100

Cluster-2
001101
1110100

VPLEX GEOSYNCHRONY v5.0 Technical Differences   34

The Open Delta Phase is the first state of the write process and it is where writes are collected in what is known as a delta which is a set of writes. Each cluster has one delta in the Open Delta phase at all times.   In this example the hosts at each cluster write to VPLEX and the writes are collected into a delta. Each asynchronous consistency group has its own set of deltas. Writes are mirrored to the cache of another director in the same cluster. Once the writes have been mirrored, an acknowledgement is sent to the host. If any one director fails, a copy is still maintained. There can only be one delta per asynchronous consistency group per cluster in the Open Delta Phase. Each director involved in the Open Delta Phase contains parts of the delta.

Async Consistency Group: Closed Delta Phase
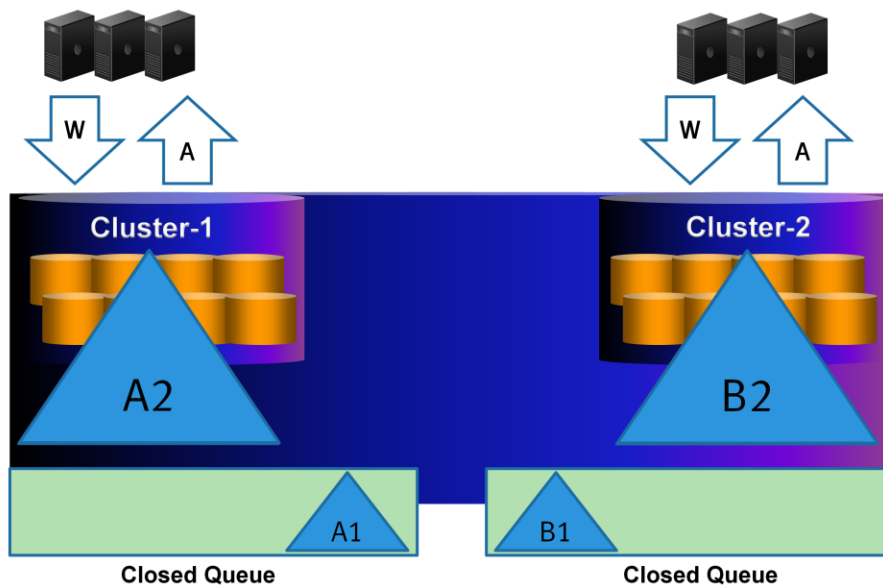
The Closed Delta Phase occurs once a director contains 16MB of data for the asynchronous group or the default closeout time has been reached. The 16MB maximum delta size cannot be modified. The default closeout time is 30 seconds by default and can be modified by the user. Setting the default closeout time will influence the size of the delta. This value is meant to give the user a way to limit the maximum possible data loss in the event of an inter-cluster link failure. The maximum RPO = (maximum-queue-depth * default closeout-time). Shorter delta queues decrease the maximum RPO in the event of an inter-cluster link failure or cluster failure. Once the default closeout time or maximum delta size is met, the delta is transferred to the closed queue and another delta appears in the Open Delta Phase. Multiple deltas can be waiting in the closed delta queue. The closed delta queue is a FIFO queue. Deltas are held in the closed queue until they can be exchanged. The default closeout time can be modified on each asynchronous consistency group. The closing of the delta is coordinated across all of the directors within a cluster. Write order consistency is maintained on delta boundaries, not individual writes. One consistency group with write order consistency is not dependent on another consistency group with write order consistency.

Async Consistency Group: Exchanging Delta Phase
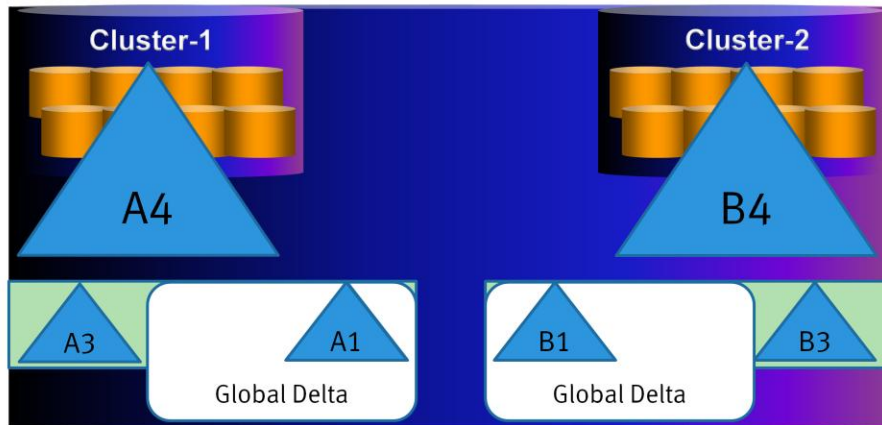
Deltas are gathered up at both clusters and then they are exchanged between the clusters. This is called the Exchanging Delta Phase. Only one delta per consistency group per cluster can be exchanged at one time. The clusters send each other communication messages to coordinate the exchanging of the deltas. Once the deltas have been exchanged, the combination of both deltas is called a global delta. Only one copy of the data is pulled across during the exchange. Each director involved in writing pages to the consistency group is also involved in the exchange. If there is a director failure during the exchange, the directors at the receiving cluster will automatically retrieve the pages that failed to arrive from the sending cluster.

Async Consistency Group: Commit Delta Phase
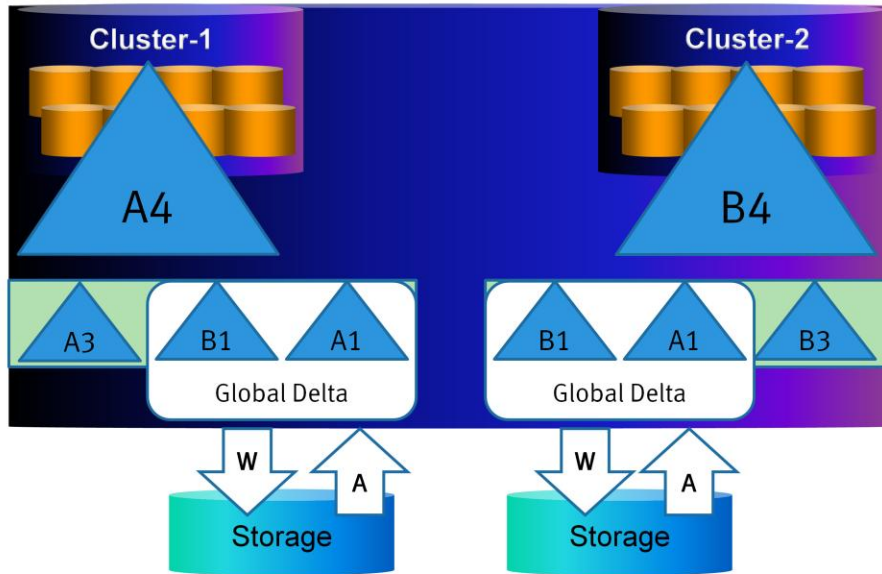
The commit phase is where the global deltas are written to the storage. 16 asynchronous consistency groups can be active at once in various states of (open, closing, exchanging, and committing).   In this example a global delta is committed in both clusters at the same time.

Asynchronous I/O: Active/Active Clusters
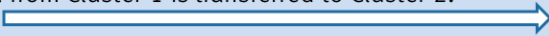
	Cluster-1		Cluster-2
Open:	11	Writes accumulate in an open delta	10
Closing	Messages exchanged to coordinate the closure on both clusters		
Closed	Delta is closed and moved to the closed queue		
Exchanging	11	Deltas are exchanged between the clusters	10
Exchange Complete	11 / 10	Each cluster now has a complete copy of the global delta	10 / 11
Committing	Storage	Writes in the global delta are committed to disk	Storage
Committed	Storage	All the writes in the global delta have been committed to disk	Storage

This table lists the steps on how deltas are exchanged between two clusters when there is active/active I/O on both clusters. Active/Active I/O to both clusters will always involve a data rollback unless the administrator chooses to let I/O suspend and not allow I/O to continue until the inter-cluster link comes back up.

Asynchronous I/O: Active/Passive Clusters

	Cluster-1		Cluster-2
Open:		Writes accumulate in an open delta	
Closing / Closed	The delta from the active cluster (Cluster-1) is moved to the closed queue. Cluster-2 is passive and the delta is empty, but still closed.		
Exchanging	The delta from Cluster-1 is transferred to Cluster-2.		
Exchange Complete		Cluster-2 now has a copy of the delta from Cluster-1. Both clusters now have a global delta.	
Committing	Storage	Writes in the global delta are committed to disk	Storage
Committed	Storage	All the writes in the global delta have been committed to disk	Storage

VPLEX GEOSYNCHRONY v5.0 Technical Differences 39

This table lists the steps on how deltas are exchanged between two clusters in an active/passive cluster.   If there is an inter-cluster link failure between an active/passive pair where Cluster-2 is passive and Cluster-1 is active and then the inter-cluster link heals, Cluster-1 does not have to rollback because it has the complete set of data. This behavior applies to the three rules "Cluster-1-Wins, Cluster-2-Wins, and Active-Cluster-Wins".

Active/Active Clusters: Deltas Lost

Phase	LOST
OPEN	YES
CLOSING	YES
CLOSED	YES
EXCHANGING	YES
COMMITTING	NO
COMMITTED	NO

If there is an inter-cluster link failure in an active/active cluster, the deltas that will be lost are the deltas in the open, closing, closed and exchanging phases. The only deltas that are not lost are the deltas in the committing and committed phases. A short duration link outage can also be handled without losing deltas by suspending I/O (causing the VM to suspend). This would result in a DU situation rather than a DL situation. The user has a choice between a data rollback or DU during the link outage. If there is no detach rule, the consistency group will suspend but there is no data loss.

Active/Passive Clusters: Deltas Lost

Phase	LOST
OPEN	NO
CLOSING	NO
CLOSED	NO
EXCHANGING	NO
COMMITTING	NO
COMMITTED	NO

VPLEX GEOSYNCHRONY v5.0 Technical Differences 41

There is no data rollback during an inter-cluster link failure of an active/passive cluster.

## Inter-cluster Link: Failure Effects on Async Consistency Groups

Detach Rule	C1 is Active (I/O) C2 is Passive (No I/O)	C1 is Passive (No I/O) C2 is Active (I/O)	C1 is Active (I/O) C2 is Active (I/O)
Cluster-1 Wins	I/O is allowed on C1	Requires resume-after-rollback at C1	Requires resume-after-rollback at C1
	I/O suspends at C2	Requires resume-after-rollback at C2	Requires resume-after-rollback at C2
	No data loss / No data rollback	Requires a Data Rollback	Requires a Data Rollback
Cluster-2 Wins	Requires resume-after-rollback at C1	I/O suspends at C1	Requires resume-after-rollback at C1
	Requires resume-after-rollback at C2	I/O is allowed at C2	Requires resume-after-rollback at C2
	Requires a Data Rollback	No data loss / no data rollback	Requires a Data Rollback
Active Cluster Wins	I/O is allowed at C1	I/O suspends at C1	Requires resume-after-rollback at C1
	I/O suspends at C2	I/O is allowed at C2	Requires resume-after-rollback at C2
	No data loss / no data rollback	No data loss / no data rollback	Requires a Data Rollback
No Automatic Winner	I/O suspends at C1	I/O suspends at C1	I/O suspends at C1
	I/O suspends at C2	I/O suspends at C2	I/O suspends at C2
	No data loss / no data rollback	No data loss / no data rollback	No data loss / No data rollback

VPLEX GEOSYNCHRONY v5.0 Technical Differences   42

EMC² where information lives

This tables displays the outcomes of a inter-cluster link failure for the various detach rules and I/O profiles. It is important to note that if the losing cluster is active during a failure, there will be a data rollback situation. Writes that were in the open, closed, and exchanging delta phases from the losing cluster will be lost. The best outcome is to have the winning cluster be the only cluster that receives I/O prior to an inter-cluster link failure. The next few slides will describe various failure scenarios and the steps that the administrator can perform to recover from the failure.

Async Consistency Groups: Multiple Director Failure

- Handles the failure of a single director within an engine
- Data loss is possible if
  - More that one director fails
  - A director fails and an inter-cluster link outage occurs while committing a global delta
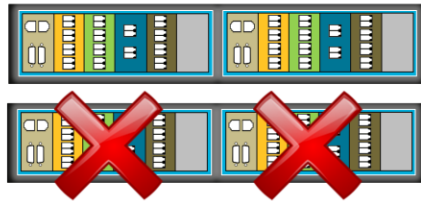
VPLEX GEOSYNCHRONY v5.0 Technical Differences   43

VPLEX can handle the failure of a director in a cluster within an engine without losing any consistency. This is because asynchronous data is mirrored on both directors within an engine of a cluster. This means that more than one director could fail without losing consistency as long as the directors that failed were on separate engines within the cluster. The delta # for a page is tracked and allows recovered pages to be reinserted into the proper delta. Protection is not guaranteed if more than one director fails in the same engine at the same time.

A data loss situation can occur in a select few situations. In all data loss scenarios VPLEX suspends I/O for only the asynchronous consistency groups that were affected by the failure. It will occur if both directors within an engine fail at the same time. This is because cached writes and replicated copies may both have been lost. In this situation VPLEX assumes this to be true for all cached writes. As a result, VPLEX suspends I/O access to all affected asynchronous consistency groups and sends a Call Home event. If an engine fails and there were no writes to that engine for a particular async consistency group, the consistency group will not be suspended. Therefore, in a quad configuration where async consistency groups were only exported through engines 1 and 2, the loss of engine 3 would not cause a data loss failure for the async consistency group.

It is also possible for data loss to occur if a director fails and an inter-cluster link outage occurs while the director is committing a global delta. This is because the winning cluster may lose data that was exchanged from the losing cluster. As a result, the consistency group on the cluster has an inconsistent image on disk. The winning cluster will be marked as having a data loss failure and manual intervention will be required. VPLEX will suspend access to the consistency group while the inter-cluster-link is down. The volumes in the consistency group will be resynchronized once the inter-cluster-link is re-established. Resolving/confirming a data loss failure mode can only be done when the inter-cluster link comes back up. Once the data loss failure has been confirmed and resolved, I/O will be re-enabled and resynchronization will occur.

After a data loss failure, the administrator must specify which cluster is most likely to have a consistent image on-disk. The image on-disk that the administrator selects will be used going forward. I/O will be resumed at both clusters after the administrator chooses the image to use. Failure/recovery commands are described in more detail in the VPLEX GeoSynchrony v5.0 Troubleshooting course.

Consistency Group: Limits

Item	#
Maximum # of volumes in an Async Consistency Group	1000
Maximum # of volumes in an Sync Consistency Group	1000
Maximum # of Synchronous Consistency Groups	Unlimited
Maximum # of Asynchronous Consistency Groups	16
Volumes that can be added to a Distributed Asynchronous Consistency Group	Distributed Volumes
Volumes that can be added to a Distributed Synchronous Consistency Group	Distributed Volumes
Volumes that can be added to a Local Synchronous Consistency Group	Local Volumes
Volumes that can be added to a Local Synchronous Consistency Group with Global Visibility	Local Volumes and Local Volumes with Global Visibility

This table lists the current consistency group restrictions and limitations. Currently, a maximum of 16 asynchronous consistency groups can be created. Please take a moment to review the rest of the information in this table.

Consistency Group: Best Practices

Grouping volumes into a consistency group	Group all volumes in use by a specific application, host, cluster, or cluster of hosts
Applying detach rules	Apply to consistency groups and not to individual volumes. This avoids human error causing one volume to be set inconsistently and unavailable at the required location during a link outage
Applications per consistency group	It is best practice to only use one application per consistency group because issues affecting one volume in the group will affect the whole group

Listed here are common consistency group best practices. Please take a moment to read the best practices.

Async Consistency Groups: Cache Vaulting
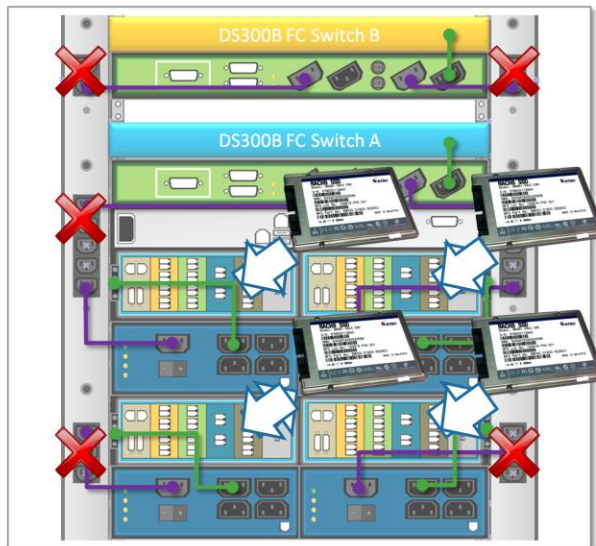
VPLEX GEOSYNCHRONY v5.0 Technical Differences   46

Cache vaulting will protect async cache in case of power failure, by copying dirty data in cache to director internal SSDs. This process protects async dirty data that has not had a chance to be written to the back-end array. In order for VPLEX to vault cache both power zones must fail. The directors monitor the standby power supplies and can identify when a power zone has failed. The standby power supplies are monitored every 10 seconds.   Once VPLEX detects that both power zones have failed, it will wait 30 second to ensure that the power loss was not a short power disruption. VPLEX will continue to service I/O for 30 seconds. If power is restored prior to 30 seconds, cache vaulting will not occur and I/O will continue as normal. If 30 seconds passes and it has been determined by VPLEX that both power zones are down, I/O will suspend at the VPLEX front-end ports and each director will begin to vault its dirty cache to its internal SSD. During the vaulting all migrations/rebuilds are suspended and configuration changes are prevented during power disruptions. It is possible for a director to vault up to 3.2 GB of cache. It will take a maximum of 300 seconds for a director to vault its dirty cache to the SSD. After vaulting is complete for all directors, the directors shut themselves down gracefully.   When power returns the VPLEX directors reboot and detect that the directors contain dirty data in the vault. The dirty data is returned to cache and it is also flushed to the array back-end storage. Once all of the dirty cache has been unvaulted to the directors, I/O resumes at the VPLEX front-end ports. Note, dirty data is defined as data in VPLEX cache that has not been written to the back-end storage array. Cache vaulting is only used for asynchronous distributed consistency groups. It is not used for individual distributed volumes or synchronous distributed consistency groups.

Cache Vaulting: Recovery Scenarios

- Detach Rule is set to (No Winner)
  - Host I/O can continue without data loss after the unvault
- Detach Rule is set to (Winning Cluster Continues)
  - Losing cluster loses power
    - Unvaulted data will be discarded when the losing cluster is restored

If the pre-defined detach rule is set to no winner, host I/O will continue without data loss after the dirty data has been unvaulted from all of the directors. If the pre-defined rule set to cluster-1-wins or cluster-2-wins and the losing cluster loses power then data that is unvaulted from the losing cluster will be discarded when power to the cluster returns. Even though, "data that is unvaulted from the losing cluster will be discarded when power to the cluster returns", there is no data loss because the losing cluster will get any missing data from the winning cluster once power is restored.

Cache Vaulting States

Stat	Description
Vault Inactive	Vault/Unvault is not in progress
Vault Requested	Initial vault request has been received
Stop front-end I/O	Stop new front-end I/O
Vault Writing	Writing vault data
Vault Written	Vault Data has been written
Unvaulting	Reading vaulted data to the local director's memory
Unvault Complete	Unvaulting of data to local director's memory has been completed
Vault Recovering	Recovering all the vaulted data to VPLEX global cache from the local

EMC²
where information lives

These are the different cache vaulting states. Please take a moment to read the descriptions for the various states.

Cache Vaulting: vault status

Check the Status of the Vault Process

```
VPlexcli:/> vault status -c cluster-2 --verbose

==========================================================================
Vault/Unvault status summary
==========================================================================
Cluster:/clusters/cluster-2
        Total number of bytes remaining to vault in the cluster: 707977216 bytes
        Estimated time remaining for cluster's vault completion: 9 seconds

==========================================================================
Director level vault/unvault status summary
==========================================================================
/engines/engine-2-1/directors/director-2-1-B:
        state: Vault Writing - Writing vault data to vault disk
        Total number of bytes to vault: 295649280 bytes
        Total number of bytes vaulted: 111587328 bytes
        Total number of bytes remaining to vault: 184061952 bytes
        Percent vaulted: 37%
        Average vault rate: 55439360 bytes/second
        Estimated time remaining to vault complete: 3 seconds

/engines/engine-2-1/directors/director-2-1-A:
        state: Vault Writing - Writing vault data to vault disk
        Total number of bytes to vault: 631287808 bytes
        Total number of bytes vaulted: 107372544 bytes
        Total number of bytes remaining to vault: 523915264 bytes
        Percent vaulted: 17%
        Average vault rate: 53317632 bytes/second
        Estimated time remaining to vault complete: 9 seconds
```

VPLEX GEOSYNCHRONY v5.0 Technical Differences   49

The vault status command displays the current vault/unvault status of the cluster. This will display the total number of bytes to be vaulted in the cluster and it will also display the time to complete the vault.

Cache Vaulting: vault go

```
VPlexcli:/> vault go -c cluster-1
vault started
```
Manually Issue a Vault
```
VPlexcli:/> 128.221.252.36/cpu0/log:5988:W/"00601636c9cb171553-2":54404:<0>2011/
04/18 15:05:03.57: utl/O ABORT: /import/griphooked_local8/backedup/yy_swreleases
/firmware/D10_20.1.rc35/yy/snac/vault/vaultSiteVault.c:vaultDirectorShutdown/206
: vault complete
```

The vault go command allows an administrator to manually schedule a vault. This will vault the dirty data in all of the directors cache at the chosen cluster. This command is helpful if the administrator needs to shutdown the cluster in a hurry. The vault go command simulates a real vault scenarios, but it does not wait 30 seconds to start vaulting as it vaults immediately.

VPLEX Management Console: Adding a Mirror

In VPLEX GeoSynchrony 4.X the only way to add a mirror to an existing volume was to use the VPLEX CLI device attach mirror command. In VPLEX GeoSynchrony 5.X it is now possible to add a mirror to a local volume or device through the VPLEX GUI. The volume to add to the mirror to must be selected and then the Add Local Mirror or Add Remote Mirror must be selected. This feature will allow the user to create both local mirrors and remote mirrors. Adding a remote mirror creates a distributed volume.

When adding a remote mirror, the user has three options. The user can add the resulting distributed volume to an existing distributed synchronous consistency group, distributed asynchronous consistency group, or the user can simple choose to keep the volume as an individual distributed volume. Adding the volume to an asynchronous distributed consistency group will not have an impact on performance. All writes will be in the asynchronous cache mode while the distributed mirror is synchronized across the clusters. If the user chooses to keep the resulting volume as an individual distributed volume, the user must choose a rule to apply to the volume.

VPLEX Management Console: Distributed Volumes
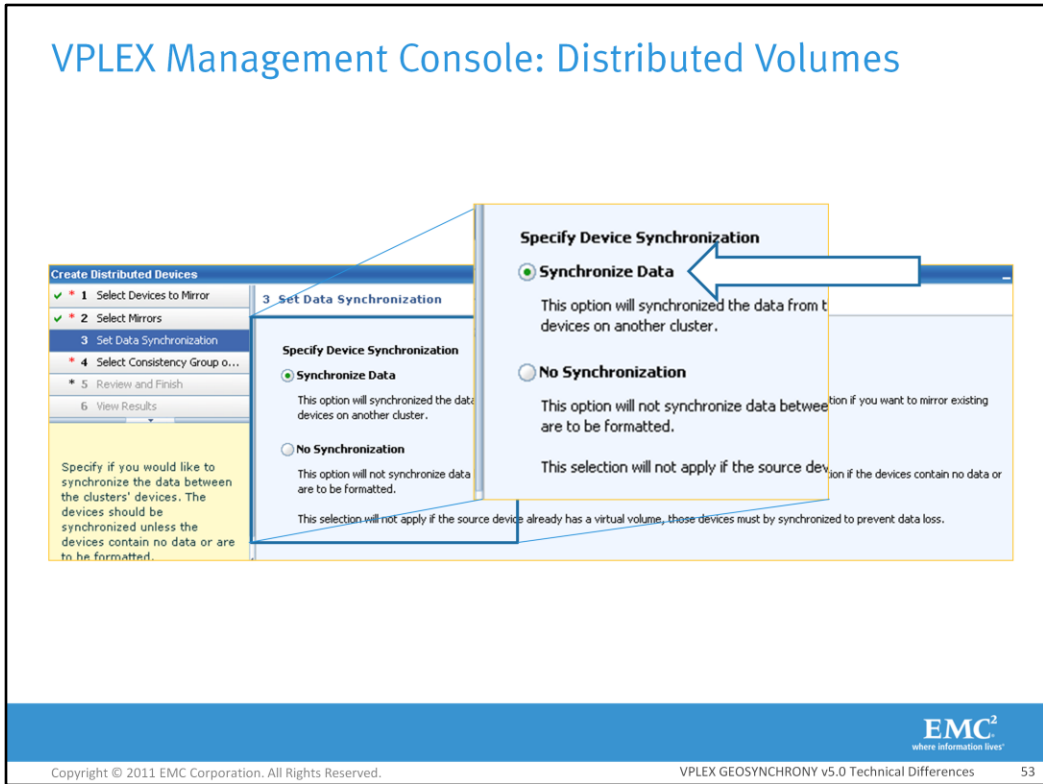
In VPLEX GeoSynchrony v5.0, the user now has the option of synchronizing the source and target devices of a distributed volume when creating a distributed volume. This will cause a rebuild and ensure that the data at the source is the same as the data at the target prior to using the distributed volume. The user should synchronize distributed volumes unless the user knows that the devices do not contain previous data. If the user knows that there is no data on the devices, the user can save time by choosing not to synchronize the devices. Adding a remote mirror to an existing virtual volume will always cause synchronization from the source device to the new mirror device.

VPLEX Management Console: Performance Monitoring

VPLEX GEOSYNCHRONY v5.0 Technical Differences   54
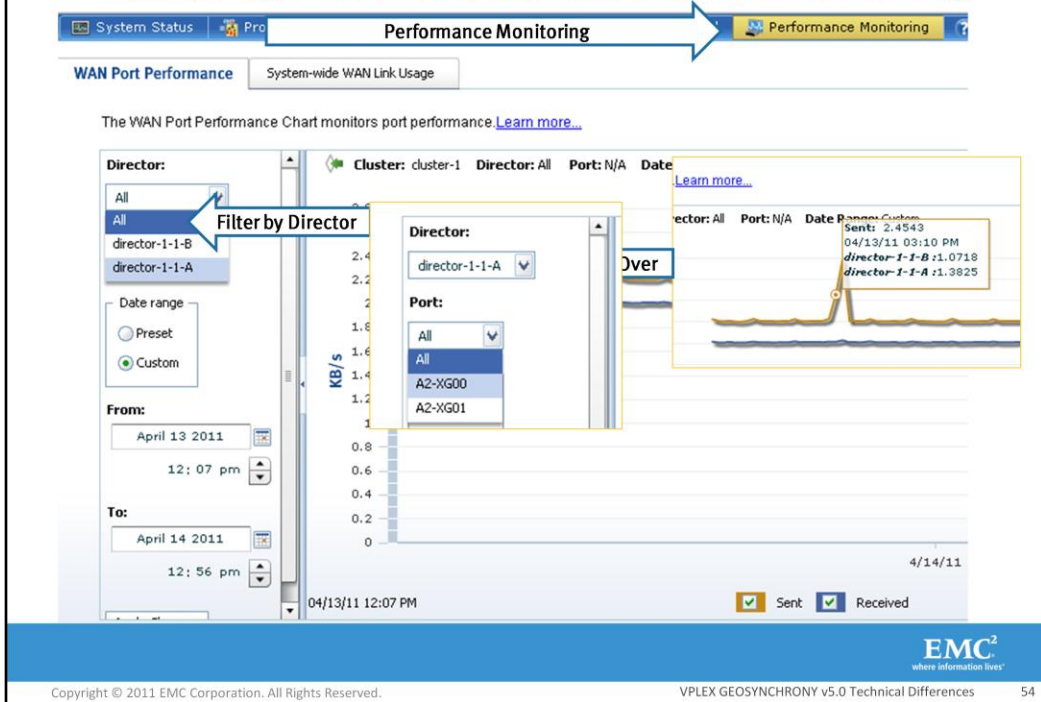
VPLEX GeoSynchrony v5.0 now includes inter-cluster performance monitoring in a new Performance Monitoring tab. This tab displays performance information for all inter-cluster Ethernet ports and Fibre Channel ports on the directors in a cluster. It provides a graphical view of data transmission over a single port or all ports on a director at specific time intervals.   Shown here the user can filter the graph to only show performance data from a specific director. Once a director has been selected, the user can filter the data by port. The chart allows the user to view current or historical performance data. This allows the user to monitor peaks in workload, detect performance issues, or view what was happening in the system when a specific problem occurred. The chart only displays data for the ports in the cluster that the user is logged onto. To simultaneously view port performance for another cluster, open a second browser session and connect to the second cluster. The performance chart data can be filtered by hour, day or week, Kilobytes per second (KB/s) sent or received, and packets per second (Packets/s) sent or received. Hovering over one of the lines in the graph will cause the graph to display the exact amount of data sent or received.

Performance Monitoring: Filtering by Date
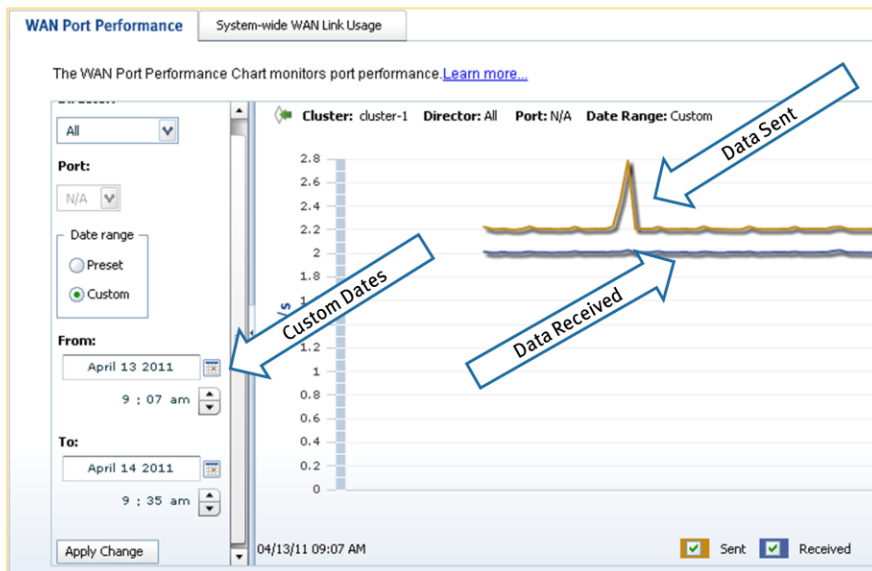
The date ranges can be refreshed automatically or manually. By default, the performance chart shows data for specific time intervals during the current day.   The performance monitoring data can be filtered to display a date range.   Shown here is the amount of data sent over the inter-cluster link between the dates specified.   Shown here is the amount of data the cluster received over the inter-cluster link.

Performance Monitoring: System-Wide Performance
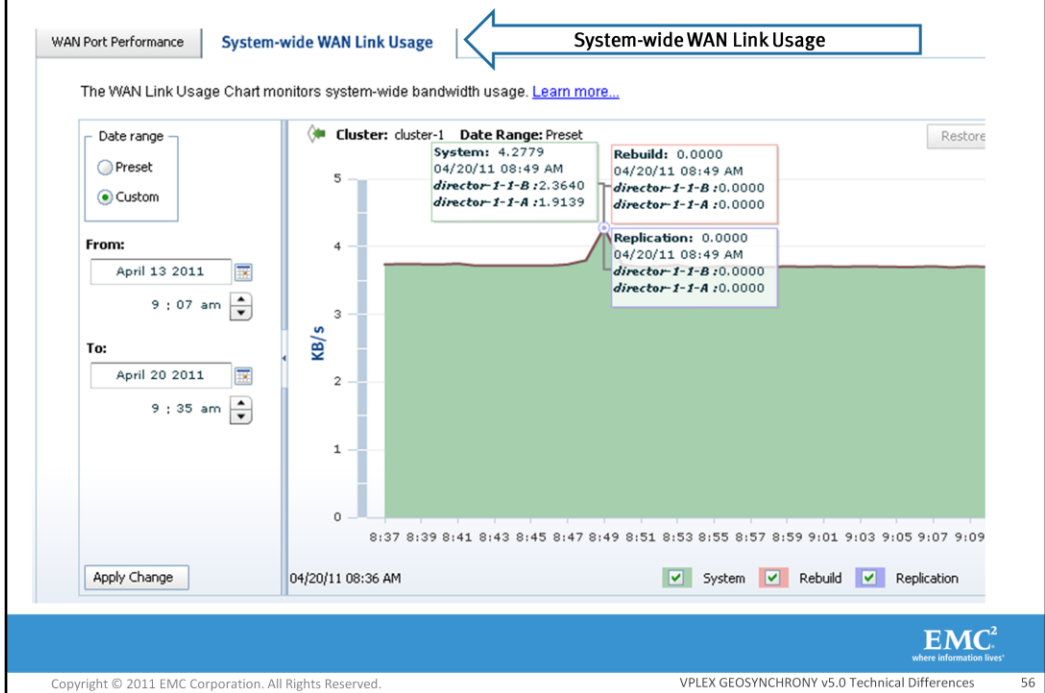
The System-wide WAN Link usage tab allows the user to monitor the amount of bandwidth being used for normal system operations, rebuilds, and replications. System Operation data is data used for normal day-to-day system operations. Rebuild data displays distributed volume rebuild traffic and replication data shows the data that is being replicated across the inter-cluster link from the host applications.

Thin-Awareness for VPLEX

Thinly Provisioned
Storage Volume

GeoSynchrony 5.X

Thickly Provisioned

Thinly Provisioned
Storage Volume

VPLEX GEOSYNCHRONY v5.0 Technical Differences   57

EMC²
where information lives'
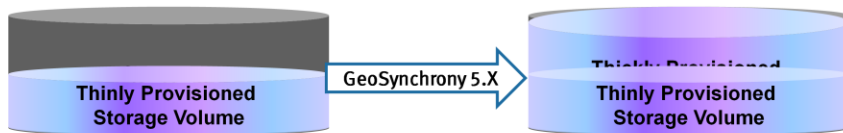
VPLEX is now aware of thinly provisioned back-end storage volumes.   In VPLEX GeoSynchrony 4.X, VPLEX was not thin aware and as a result if the user moved a device using VPLEX mobility, the device would lose its thinness and become a thick device and as a result it lost all of the benefits of being thinly provisioned.   In VPLEX GeoSynchrony v5.0, VPLEX must be told that the storage volume is a thinly provisioned device. In 5.0, VPLEX respects the thinness and only copies the blocks of storage that it needs to.

Thin Awareness: Claiming the Volume as Thin

Claiming Storage Volume to be Thinly Provisioned

Storage volumes can be claimed as thin. In order to claim a storage volume to be thin, the thin attribute must be set to true during the claiming process. This is possible by selecting the option, "use thin rebuilds". Once the "use thin rebuilds" option has been selected, the volumes will be respected as thin volumes when moved. At this point in time VPLEX cannot, "automatically" detect if a volume is thinly provisioned and as a result, the user must tell VPLEX that the volume is thin.

VPLEX Thin Awareness: Setting a Claimed Device to Thin

A storage volume that has already been claimed can also be configured so that VPLEX treats it as a thin volume. To accomplish this, the user must navigate to the desired storage volume and click, "Set Rebuild Type". This will give the user the option to change the rebuild type to thin so that VPLEX treats it as a thin volume.

VPLEX Thin Awareness: Displaying Thin Volumes

Once a storage volume has been configured so that VPLEX treats it as a thinly provisioned volume, the volume will appear in the VPLEX Management Console GUI and CLI with the thin attribute set to true.

VPLEX Witness: Overview

- Monitors heartbeat on both Clusters
- Observes health check heart beats to both clusters
- Provides guidance to both clusters based on its observations
- Eliminates the risk of data corruption and minimizes the risk of DU
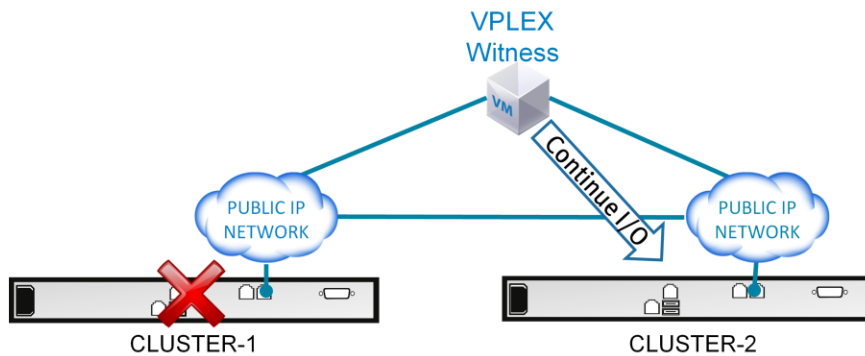
VPLEX
Witness

VM

Continue I/O

PUBLIC IP
NETWORK

PUBLIC IP
NETWORK

CLUSTER-1

CLUSTER-2

VPLEX GEOSYNCHRONY v5.0 Technical Differences   61

The VPLEX Witness Server is an optional server used to minimize the risk of a system-wide data unavailability situation. It is not required when using VPLEX GeoSynchrony v5.0. The VPLEX Witness Server is relevant only for VPLEX Metro and VPLEX Geo clusters; a VPLEX Witness Server is not necessary for a VPLEX Local cluster. The VPLEX Witness Server is a virtual machine, launched on a customer's ESX or ESXi server, which monitors the status of both clusters. If there is a failure, the VPLEX Witness Server will make a recommendation on which cluster should continue I/O. The VPLEX Witness Server is intended to improve failure handling in VPLEX. A VPN tunnel is established between each cluster and the VPLEX Witness Server.

In this example, if Cluster-1 were to incur a failure , the VPLEX Witness Server would be told by Cluster-2 that Cluster-1 is not reachable. It would then send a recommendation to Cluster-2 to continue servicing I/O and to suspend I/O at Cluster-1. If the VPLEX Witness Server was not deployed and the same failure occurred and the predefined detach rule was 'Cluster-1-Wins,' a complete system wide DU would occur.

All consistency groups are still configured with predefined rule sets even if the VPLEX Witness is utilized with VPLEX .

However, in the presence of cluster or connectivity failures, the VPLEX Witness forces the majority rule to take precedence over the pre-defined rule-sets. This means that in order for a cluster to continue I/O it must be either connected to a peer cluster or the VPLEX Witness. The term guidance here is crucial. In case of partition, the VPLEX Witness guides both clusters to proceed with I/O to according to their predefined rules. However, the cluster may still decide ignore the guidance and suspend I/O even if it is told to continue I/O by the VPLEX Witness. An example of this is if the non-winning cluster was active during the failure. This would mean that the winning cluster would have an inconsistent image and would need to rollback prior to resuming I/O. In this situation, even if the VPLEX Witness told the cluster to continue I/O, the cluster would not.

Benefits of the VPLEX Witness

Predefined Detach rule = Cluster-1-Wins
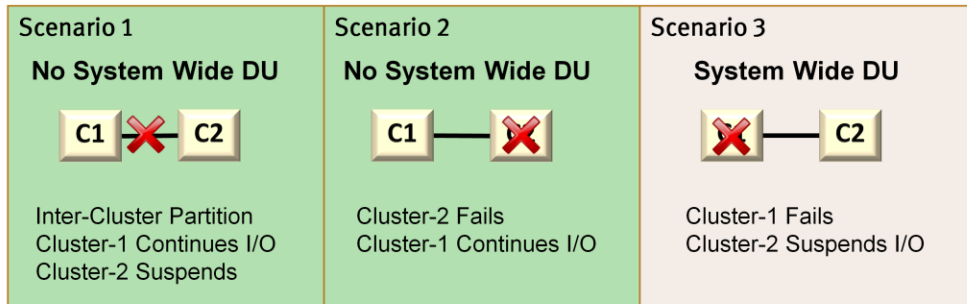
Scenario 1	Scenario 2	Scenario 3
No System Wide DU	**No System Wide DU**	**System Wide DU**
C1 — C2	C1 —	— C2
Inter-Cluster Partition Cluster-1 Continues I/O Cluster-2 Suspends	Cluster-2 Fails Cluster-1 Continues I/O	Cluster-1 Fails Cluster-2 Suspends I/O

VPLEX GEOSYNCHRONY v5.0 Technical Differences   62

Without the VPLEX Witness customers that are using distributed consistency groups are at an increased risk of experiencing a system wide DU event. Without a VPLEX Witness the only way for the clusters to react to a failure is by enabling the pre-defined detach rules that are configured per consistency group.   In this example, the predefined detach rule configured is Cluster-1-Wins. In the first scenario there is an inter-cluster partition failure and in the second scenario cluster-2 fails. In both cases I/O is allowed to continue at one of the clusters.   However, in scenario 3 the predefined winning cluster fails. This causes I/O to suspend at the non-winning cluster. In this scenario there is a system wide DU event. Recovery from this failure will require manual intervention. This is unacceptable for most customers. The VPLEX Witness could have prevented a system wide DU in the last scenario by enabling I/O access to cluster-2. The VPLEX Witness can reduce the Recovery Time Objective (RTO) and Recovery Point Objective (RPO).

VPLEX Witness: Failure Handling Semantics

- Apply to
  - Distributed volumes in ALL consistency groups
- Does not apply to
  - Local volumes
  - Distributed volumes that are not in a consistency group
  - Distributed volumes within a consistency group where the VPLEX Witness is disabled
  - Consistency groups that are using the "No-Automatic-Winner" rule set

EMC²
where information lives®

VPLEX GEOSYNCHRONY v5.0 Technical Differences   63

The VPLEX Witness only observes and sends guidance to distributed volumes in consistency groups. The VPLEX Witness functionality must be enabled for this to occur. All other volumes in the cluster are not impacted by the VPLEX Witness. During failures the VPLEX Witness will provide guidance to distributed volumes in ALL consistency groups.

VPLEX Witness: Components

VPLEX Witness

VPLEX Witness Server

VPLEX Witness Client Component
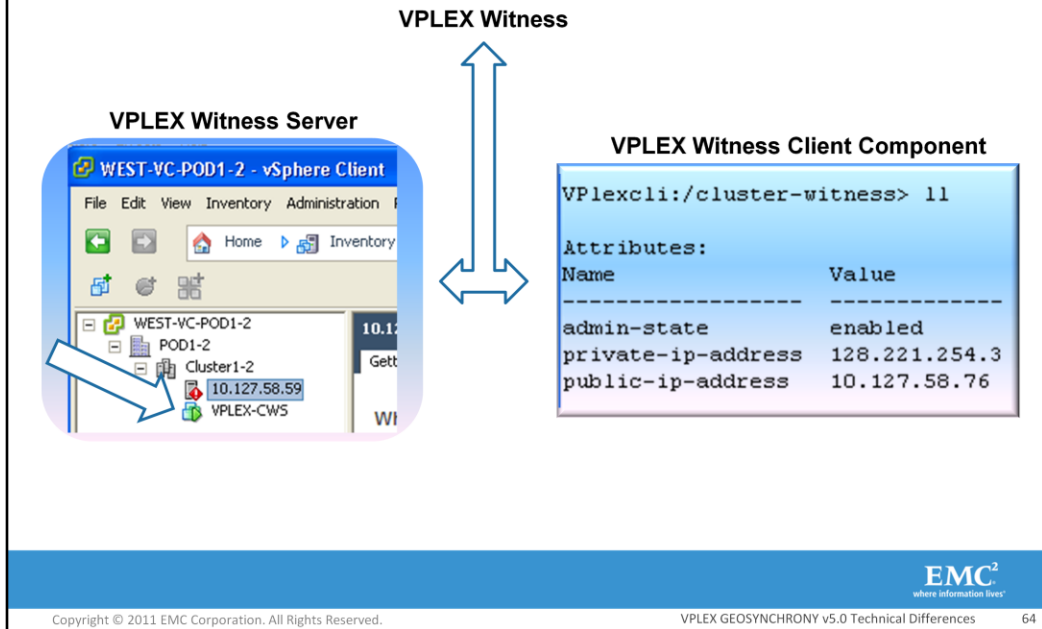
The VPLEX Witness functionality includes both the VPLEX Witness Server and the VPLEX Witness Client Component.

The VPLEX Witness server is a VM that is deployed on a customer provided v4.1 or greater ESX host. The server is configured to function with both a VPLEX Metro and a VPLEX Geo. However, it can only manage a single VPLEX Metro or VPLEX Geo. Multiple VPLEX Witness servers can be deployed on the same ESX server. The VPLEX Witness server runs on the customer's provided IP management network. This network should be separate from the network used for both clusters.

The VPLEX Witness Client component runs on the management server at each cluster. The VPLEX Witness Client Component directory is hidden from the user until it has been configured.

VPLEX Witness: Rules
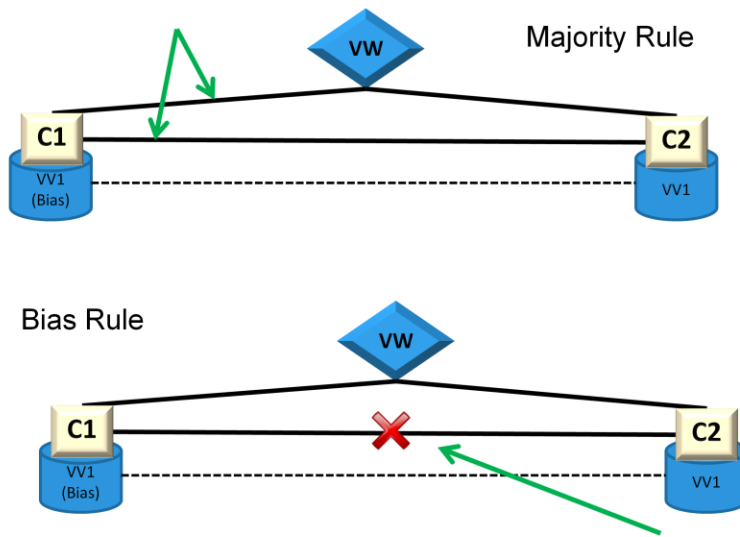
Majority Rule

Bias Rule

VPLEX GEOSYNCHRONY v5.0 Technical Differences   65

VPLEX with the help of the VPLEX Witness utilizes two rules to determine which cluster(s) are able to perform I/O. These rules are the **Majority Rule** and the **Bias Rule**. The **Majority Rule** means that in order for a cluster to continue processing I/O, it must be either connected to a peer cluster or the VPLEX Witness. However, the Majority Rule is only acted on when the VPLEX Witness first sees a failure and provides guidance. After that point the VPLEX Witness sticks to the same guidance that it provides previously (even if it no longer matches its current observation) until it observes a complete recovery on the cluster side. The Majority Rule is always used unless there is an inter-cluster partition where both clusters cannot see each other, but can still see the VPLEX Witness. In this case the **Bias Rule** is used.   The Bias Rule means that one cluster will be able to perform I/O to all the distributed volume(s) in the consistency groups for which the cluster has bias while the other cluster will suspend I/O to the distributed volume(s). This bias is configured by the administrator upon the creation of a consistency group.

The VPLEX Witness server states are sticky. This means that once the VPLEX Witness observes a failure and provided its guidance it will stick to this failure until both clusters report complete recovery to the VPLEX Witness. This is crucial in order to avoid data corruption. As a result you may have a scenario where Cluster 1 becomes isolated and the VPLEX Witness tells cluster 2 to continue I/O and then Cluster-2 then becomes isolated. However, because it has previously received guidance to proceed from the server, it will proceed even while it is isolated. In the meantime, if Cluster-1 were to reconnect with the server, the server will tell it to stop! In this case, because of event timing we ended up in a situation where cluster 1 is connected to the VPLEX Witness but it is suspending while cluster 2 is isolated but it is proceeding.

VPLEX Witness: Primary Failure Types

Local Cluster Isolation
I/O Suspends | CW | I/O Continues
C1 — C2

Remote Cluster Isolation
I/O Continues | CW | I/O Suspends
C1 — C2

Inter-Cluster Partition
Predefined rules govern which cluster will continue I/O
CW
C1 — C2

Loss of Contact with VPLEX Witness
I/O Continues | CW | I/O Continues
C1 — C2
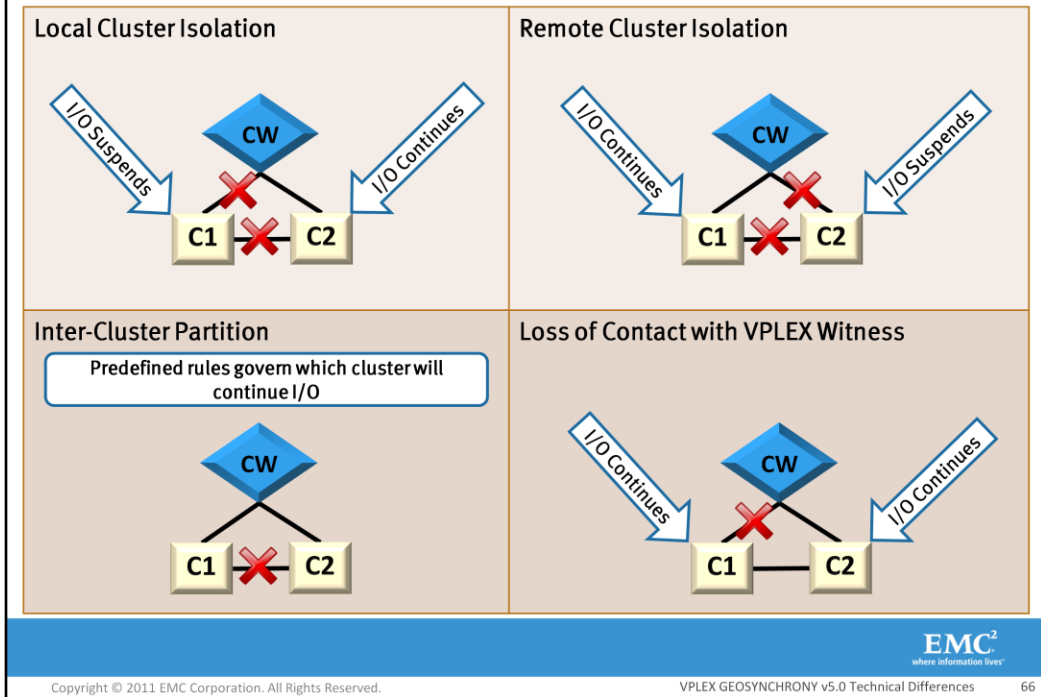
There are four main failure types that could occur. These are Local Cluster Isolation, Remote Cluster Isolation, Inter-Cluster Partition, and Loss of Contact with the VPLEX Witness. Local cluster isolation occurs when the local cluster loses contact with both the remote cluster and the VPLEX Witness. In the example shown Cluster-1 suspends I/O and the VPLEX Witness guides Cluster-2 to continue I/O.

In the Remote Cluster Isolation scenario the local cluster has lost contact with the remote cluster. However, the local cluster still has access to the VPLEX Witness and thus continues I/O. In the example shown Cluster-1 continues I/O as it is still in contact with the VPLEX Witness. The remote cluster must suspend I/O.

In the case of a inter-cluster partition where both clusters lose contact with each other, but still have access to the VPLEX Witness, the predefined rules takes effect. This is the only situation where the static bias rules or active bias rules are used.

In the scenario where one of the clusters loses contact with the VPLEX Witness, but is still in contact with the other cluster, there is no change in I/O. The cluster that lost connectivity with the VPLEX Witness simply issues a Call Home.

VPLEX Witness: One Component Failure Scenarios
No System Wide DU
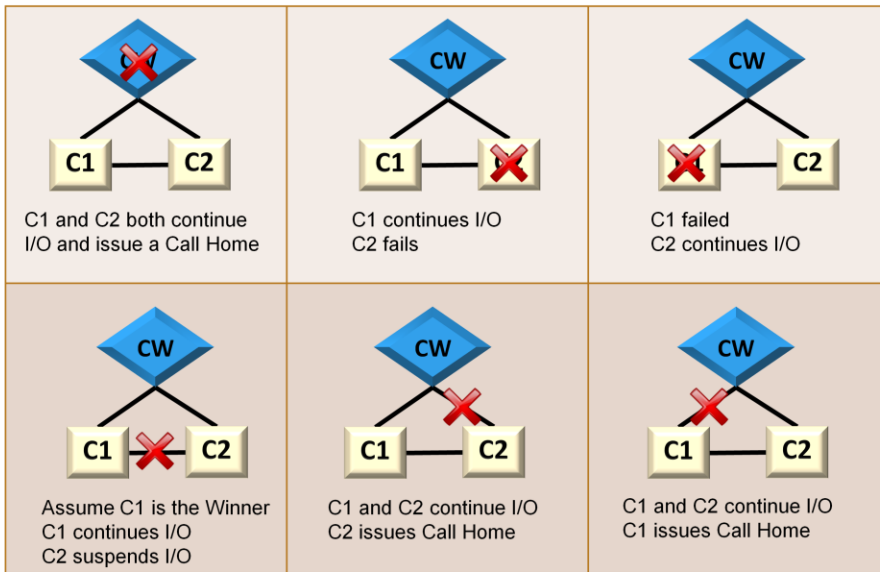
C1 and C2 both continue I/O and issue a Call Home	C1 continues I/O C2 fails	C1 failed C2 continues I/O
Assume C1 is the Winner C1 continues I/O C2 suspends I/O	C1 and C2 continue I/O C2 issues Call Home	C1 and C2 continue I/O C1 issues Call Home

VPLEX GEOSYNCHRONY v5.0 Technical Differences   67

Shown are VPLEX failure handling scenarios that do not cause a system wide DU. The only scenario listed here where there might be a system wide DU is the left scenario at the bottom. This is because the application may not be on the winning cluster when the pre-defined bias rules are enabled.

VPLEX Witness: Two Component Failure Scenarios No System Wide DU

C1 continues I/O;
C2 suspends I/O and
issues Call Home

C1 suspends I/O and
issues Call Home
C2 continues I/O

C1 and C2 both continue
I/O and issue Call Home
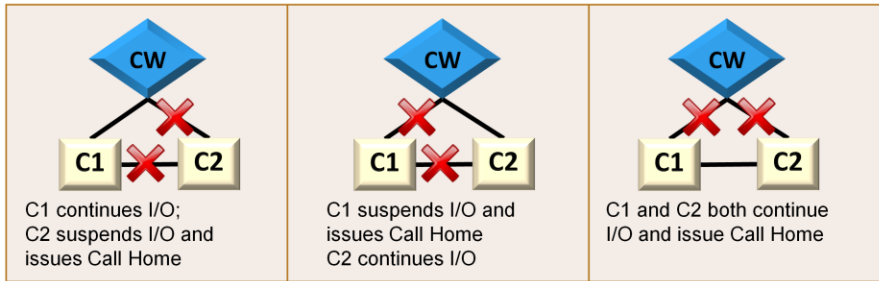
Shown here are two component failure scenarios that do not cause a system wide DU event. Please take a moment to review the scenarios shown.

VPLEX Witness Failure Handling – System Wide DU
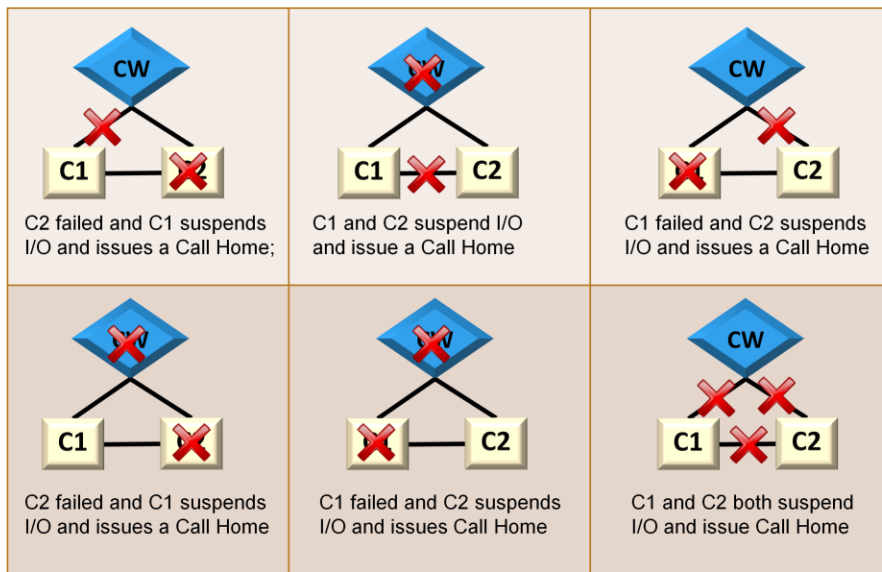
CW — C1 — C2 (C2 failed)	CW — C1 — C2	CW — C1 — C2 (C1 failed)
C2 failed and C1 suspends I/O and issues a Call Home;	C1 and C2 suspend I/O and issue a Call Home	C1 failed and C2 suspends I/O and issues a Call Home
CW — C1 — C2	CW — C1 — C2	CW — C1 — C2
C2 failed and C1 suspends I/O and issues a Call Home	C1 failed and C2 suspends I/O and issues Call Home	C1 and C2 both suspend I/O and issue Call Home

VPLEX GEOSYNCHRONY v5.0 Technical Differences   69

It is very important to properly deploy the VPLEX Witness. The most important task in deploying the VPLEX Witness is to deploy it in a separate failure domain. These are the VPLEX failure handling examples where a system wide DU will occur. The top three scenarios are not likely to occur even if the VPLEX Witness is properly deployed. If the VPLEX Witness is not deployed correctly, it may cause a VPLEX Metro or Geo system to perform worse than a deployment without the VPLEX Witness. For example, in the last scenarios in the bottom row both clusters will suspend and there will be a system-wide DU on all consistency groups. In this scenario it would have been a better choice to not use a VPLEX Witness as it causes more harm when not properly deployed. The bottom three scenarios are likely to occur if the VPLEX Witness is not properly deployed.

69

VPLEX Witness: Pre-deployment Physical Requirement

- Required to be deployed in a third failure domain
  - Cluster-1 = Failure domain #1
  - VPLEX Witness = Failure domain #2
  - Cluster-2 = Failure domain #3
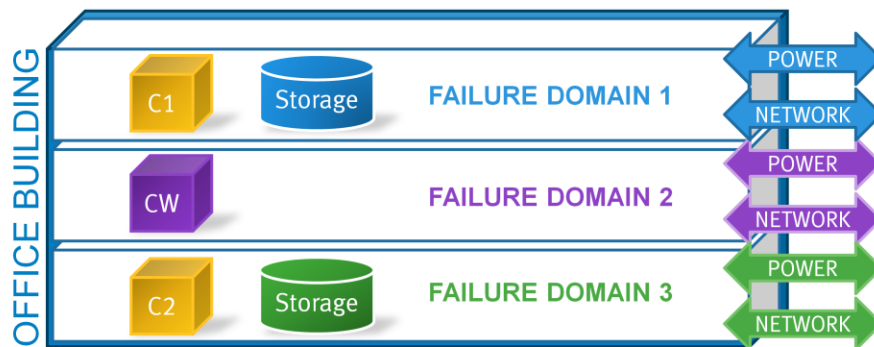
VPLEX GEOSYNCHRONY v5.0 Technical Differences   70

In order for the VPLEX Witness Server to be effective, it must be deployed in a third failure domain, with independent power and networking from Cluster-1 and Cluster 2. A failure domain is a collection of entities that are affected by the same fault. The VPLEX Witness should be deployed in such a way that it is not affected by the individual clusters. This includes faults related to physical location, power, networking connectivity, etc.   The VPLEX Witness must use a power source that is independent from either cluster. The IP Management network (connecting both Management Servers and the VPLEX Witness Server) must be physically separate from either of the inter-cluster networks.

In the scenario displayed, Cluster-1 is deployed on the third floor of an office building and Cluster-2 is deployed on the first floor. To decrease the potential for a single domain failure taking out both a cluster and the VPLEX Witness Server, the VPLEX Witness should be deployed in a third failure domain on the second floor with independent power and networking.

VPLEX Witness Pre-deployment: Network Requirements

VPLEX Management Server   VPLEX Witness Server
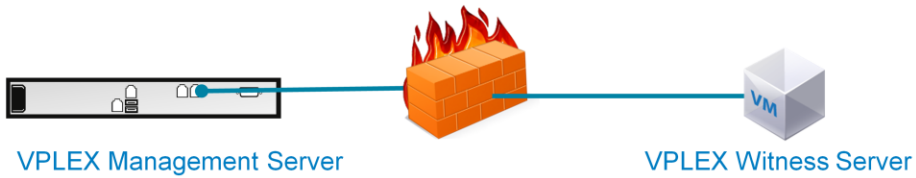
Protocols That Must be allowed on The Firewall	Subnets that Must NOT be Routatable on the IP Management Network
Internet Key Exchange (IKE): UDP port 500	128.221.252.0/24
Encapsulating Security Payload (ESP): IP protocol number 50	128.221.253.0/24
Authentication Header (AH): IP protocol number 51	128.221.254.0/24
NAT Traversal in the IKE (IPsec NAT-T): UDP port 4500	

The VPLEX Witness traffic as well as the VPLEX management traffic leverages VPN tunnels established on top of IPsec. This is why IPsec tunnels must not be prevented. The subnets listed must not be routable on the management network. If any of the networks are accessible from the IP management network, EMC Customer Support should be contacted. If it is possible to route VPLEX Witness packets on the IP Management network, they may not go through the expected management network which will violate all of our assumptions about failure domains.

VPLEX Witness Pre-deployment: Quality of Service

Requirement	Description
Bandwidth	2-3 Kbps per cluster
MTU	1500 bytes or larger
Round Trip Latency	Should not exceed 1 second

These are the VPLEX Witness quality of service recommendations. Please take a moment to read them. The EMC Ping tool can used to help ensure that these requirements are met. A typical VPLEX Witness deployment will generate around 2-3 Kbps of duplex VPLEX Witness IP traffic (transmitted over IP management network) per cluster

VPLEX Witness Host: Hardware Requirements

Requirement	Description
VPLEX Witness Server Host (ESX Server)	•Enabled BIOS Virtualization Technology (VT) extension •Access to the host must be secured by a password assigned by the customer •Recommended to connect the host to an Uninterruptable Power Source (UPS)
CPU	•64 bit Intel or AMD-based •Capable of running 64 bit VM images
Supported 64 Bit Processors	•AMD Athlon 64, Revision D or later •AMD Opteron revision E or later •AMD Turion 64, revision E or later •AMD Sempron 64-bit-capable, revision D or later (experimental support) •Intel EM64 VT-capable processors (experimental support)
CPU Utilization	•Allocate one vCPU for the VPLEX Witness Server VM

EMC²
where information lives·

These are the VPLEX Witness host (ESX Server) hardware requirements. Please take a moment to read over the requirements.

VPLEX Witness Host: Hardware Requirements (Cont)

Requirement	Description
Network Interface Card	•Must have a 1 GigE NIC •Ethernet port must be connected to the IP Management network and must be configured with a static public IP address •Must be possible to allocate two Virtual Network Adapters for use by the VPLEX Witness Server VM
RAM Utilization	•Allocate 512MB for VPLEX Witness Server VM
Hard Disk Storage Utilization	•Allocate 2.5GB of storage space for the VPLEX Witness Server VM

These are also VPLEX Witness host (ESX Server) hardware requirements. Please take a moment to read over the requirements.

VPLEX Witness Deployment: Demonstration

This video will demonstrate how to deploy and enable the VPLEX Witness.

Launch

This video will demonstrate how to deploy and enable the VPLEX Witness.

VPLEX Witness: Deployment Assumptions
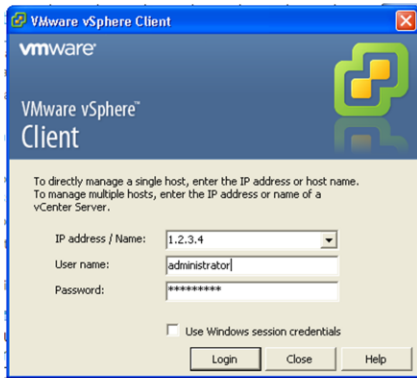
vSphere Client 4.1 or >

ESX 4.1 or > Host

VPLEX Witness OVF File

VPLEX GEOSYNCHRONY v5.0 Technical Differences   76
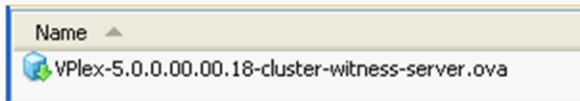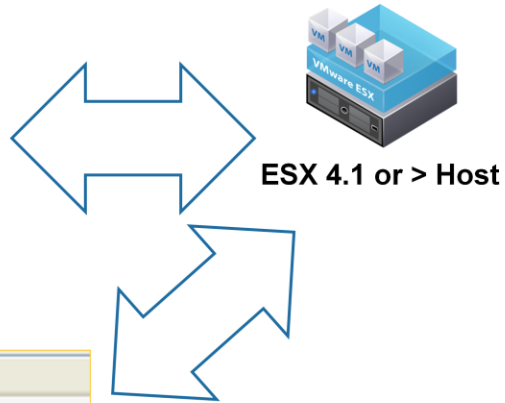
The VPLEX Witness Server is a Linux process that runs under a SLES-11 guest-OS based VM. The VM is packaged as OVA (Open Virtualization Format Archive files) that can potentially be deployed on all virtualization platforms that support Open Virtualization Format.

Initially, the deployment will only be targeted on ESX servers. The installer should have an intermediate level of knowledge about VMware vSphere products and offerings. They should also have access to an ESX v4.1 host with enough storage and computing resources and should have installed a vSphere client v4.1 on a computer with the proper privileges and with access to the ESX host in the deployment. The installer should also have access to the VPLEX Witness OVA file. The VPLEX Witness OVA file will be available on Powerlink. The VPLEX Witness has not been qualified to run in a VMware HA cluster (yet).

VPLEX Witness CLI Directory

```
VPlexcli:/cluster-witness> ll

Attributes:
Name                 Value
-----------------    -------------
admin-state          enabled          <
private-ip-address   128.221.254.3    <
public-ip-address    10.127.58.76     <



Contexts:
Name          Description
----------    -------------------------
components    Cluster Witness Components
```
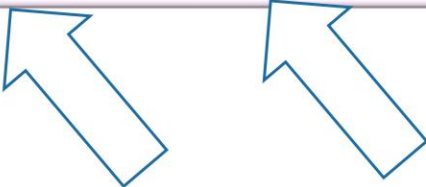
EMC²
where information lives®

The VPLEX Witness context is hidden until it is enabled using the cluster-witness configure command. Within this directory, there are four attributes, "admin-state, private-ip-address, and public-ip-address. The private-ip-address is the VPLEX Witness server VM's private IP address and will always be 128.221.254.3. The clusters will not be able to communicate with the VPLEX Witness until the VPN has been created between the two entities. The public-ip-address is the public IP address of the VPLEX Witness server and is assignable upon the deployment of the VPLEX Witness. The admin-state identifies if the VPLEX Witness is enabled or disabled. If the VPLEX Witness is enabled, it participates in observations and handling of inter-cluster partition and cluster failures. If the VPLEX Witness is disabled, the clusters use their pre-defined rules to govern which cluster is allowed to continue I/O and which cluster must suspend I/O.

VPLEX Witness Components CLI Directory

```
VPlexcli:/cluster-witness/components> ll
Name       ID  Admin State  Operational State    Mgmt Connectivity
---------  --  -----------  -------------------  ------------------
cluster-1  1   enabled      in-contact           ok
cluster-2  2   enabled      in-contact           ok
server     -   enabled      clusters-in-contact  ok
```

The components directory resides within the VPLEX Witness directory.   The user can identify the operation status of the remote cluster and the VPLEX Witness server VM. They can also tell if they are in contact with the remote cluster and the VPLEX Witness server VM.

VPLEX Witness Diagnostic Information

```
VPlexcli:/cluster-witness/components/cluster-1> ll     Cluster Diagnostic Info
Name                       Value
----------------------     ----------------------------------------------------
admin-state                enabled
diagnostic                 INFO: Current state of cluster-1 is in-contact (last
                           state change: 0 days, 76510 secs ago; last message
                           from server: 0 days, 0 secs ago.)
id                         1
management-connectivity    ok
operational-state          in-contact
```

```
VPlexcli:/cluster-witness/components/server> ll     Server Diagnostic Info
Name                       Value
----------------------     ----------------------------------------------------
---
admin-state                enabled
diagnostic                 INFO: Current state is clusters-in-contact (last st
ate
                           change: 0 days, 76627 secs ago.) {last time of
                           communication with cluster-2: 0 days, 0 secs ago.)
                           (last time of communication with cluster-1: 0 days,
 0
                           secs ago.)
id                         -
management-connectivity    ok
operational-state          clusters-in-contact
```

VPLEX GEOSYNCHRONY v5.0 Technical Differences   79

Within the components directory of the VPLEX Witness there are three directories. One directory is for the VPLEX Witness server, one is for Cluster-1, and one is for Cluster-2. Within the subdirectory diagnostic information will be displayed. In the cluster directory such as cluster-1 in the example, the diagnostic information shows that cluster-1 is in contact with the VPLEX Witness. It also shows the amount of time that it has been in-contact with the Witness Server. In the VPLEX Witness server directory the diagnostic information will display the clusters that the VPLEX Witness is in communication with. It will also display the amount of time that the server has been in contact with the clusters. If there are any problems trying to connect to these components error messages will be displayed here.

VPLEX Witness: Operational States

State	Meaning
ok	VPLEX Witness server is in-contact with both clusters
unreachable	Several VPLEX Witness components are not reachable to reconcile the status
unknown	Several VPLEX Witness components show their state as unknown to reconcile the status
partition detected	A cluster partition has been previously detected, clusters are proceeding according to their configured rule-set.
cluster-unreachability-detected	A cluster has been detected to be isolated, the other cluster is going to proceed with I/O regardless of the configured rule-set

These are the VPLEX Witness operational states. Please take a moment to review them.

VPLEX Witness: Diagnostic Strings

String	Meaning
Cannot establish connectivity to query diagnostic information	Occurs if the VPLEX Witness Server or one of the Clusters is unreachable
Local cluster-xx hasn't yet established connectivity with the server or remote cluster-yy hasn't yet established connectivity with the server	Occurs if the VPLEX Witness Server does not report the state of one of the clusters at all i.e. (The Cluster has never connected to the VPLEX Witness)
Local cluster-xx has been out of touch from the server for X days, Y secs ago or remote cluster-yy has been out of touch from the server for X days, Y secs	Occurs if the VPLEX Witness Server reports that it has not received messages from a given cluster for longer than 60 seconds
VPLEX Witness server has been out of touch for X days, Y secs ago	Occurs either cluster reports that they have not received messages from the VPLEX Witness Server for longer than 60 seconds
Cluster Witness is not enabled, so no diagnostic information is available	Occurs if the VPLEX Witness Server or VPLEX Witness client is not enabled

These are some of the diagnostic strings that could be displayed within the sub-components section. Please take a moment to look them over.

VPLEX: Asymmetric Logical Unit (ALUA) Support

- Characteristics of a path are called access states
- Active/Optimized: Provide higher bandwidth
- Active/Non-Optimized: Do not provide higher bandwidth
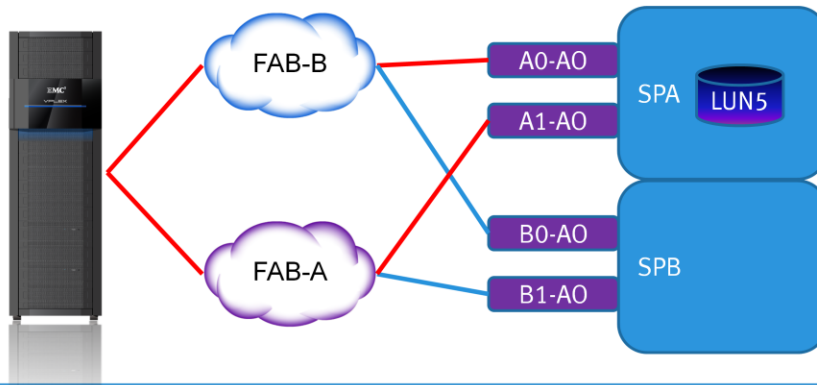
ALUA is a feature that many new active/passive arrays are capable of providing. VPLEX GeoSynchrony v5.0 can now take advantage of arrays that have ALUA support. In active/passive arrays, logical units or LUNs are normally exposed through several array ports on different paths and the characteristics of the paths might be different. ALUA calls these paths characteristics access states. ALUA provides a framework for managing these access states. T10 defines a number of access states. The most important access states are active/optimized and active/non-optimized. Active optimized paths usually provide higher bandwidth than active non-optimized paths. Active/optimized paths are paths that go to the service processor of the array that owns the LUN. In this picture there are two active optimized paths which are shown in red. These ports provide greater performance than the active/non-optimized paths shown in blue. I/O that goes to the active non-optimized ports must be transferred to the service processor that owns the LUN internally. This transfer increases latency and has an impact on the array. VPLEX is able to detect the active optimized paths and the active/non-optimized paths and performs round robin load balancing across all of the active optimized paths. Since VPLEX is aware of the active/optimized paths it is able to provide the user with better performance to the LUN. There are two variations of ALUA, implicit and explicit and these variations will be covered next.

ALUA Variations: Implicit ALUA
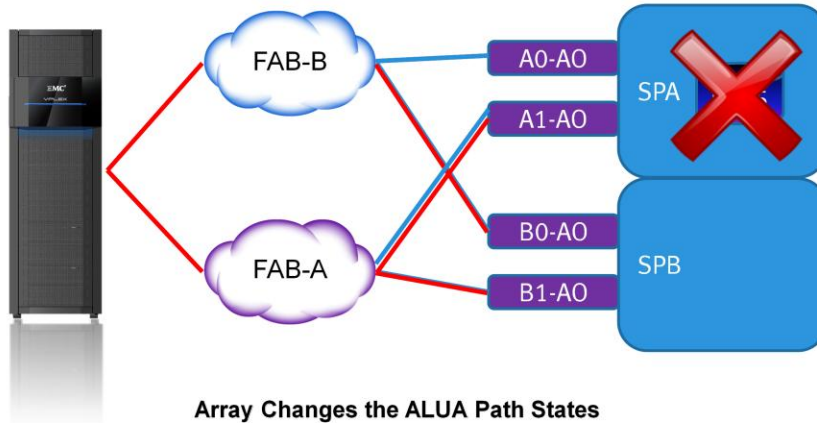
FAB-B — A0-AO / A1-AO — SPA ✗

FAB-A — B0-AO / B1-AO — SPB

Array Changes the ALUA Path States

VPLEX GEOSYNCHRONY v5.0 Technical Differences 83

With implicit ALUA, the array is in control of changing the access states of the paths. In this example the controller that owns the LUN can fail.   If this occurs the array will trespass the LUN from the failed controller to the other controller.   Through Implicit ALUA the array will change the active/non-optimized ports into active/optimized ports.

ALUA Variations: Explicit ALUA
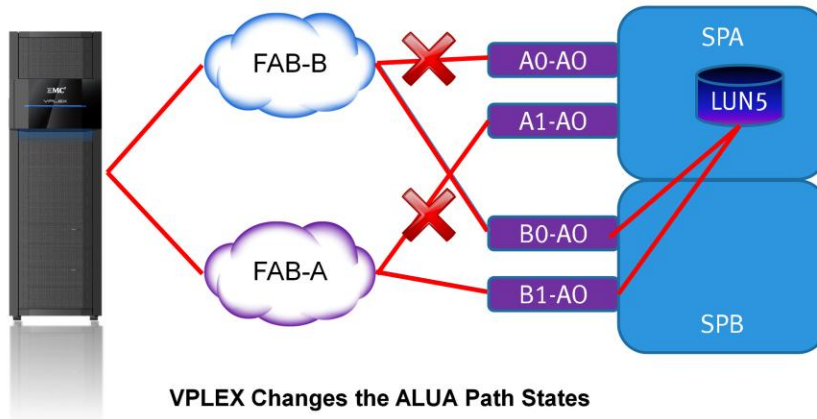
VPLEX Changes the ALUA Path States

In explicit ALUA the host or VPLEX is able to change the ALUA path states.   If the active/optimized paths fail VPLEX can cause the active/non-optimized paths to become active optimized paths and as result increase the performance. I/O is able to go between the controllers to access the LUN through a very fast bus. There is no trespassing of the LUN in this case and SPA still owns the LUN.   When VPLEX notices that the paths are once again back up, it will revert the modified active/optimized paths back to active non-optimized paths.

VPLEX Element Manager API: Overview

- Allows a user written script to:
  - List VPLEX directories and attributes
  - Change VPLEX Attributes
  - Execute VPLEX CLI Commands
- service and admin roles are supported

VPLEX Response:
Here is the listing of storage views and virtual volumes within the views

Perl Script:
Get a listing of the storage views and the virtual volumes in the views

VPLEX GEOSYNCHRONY v5.0 Technical Differences 85
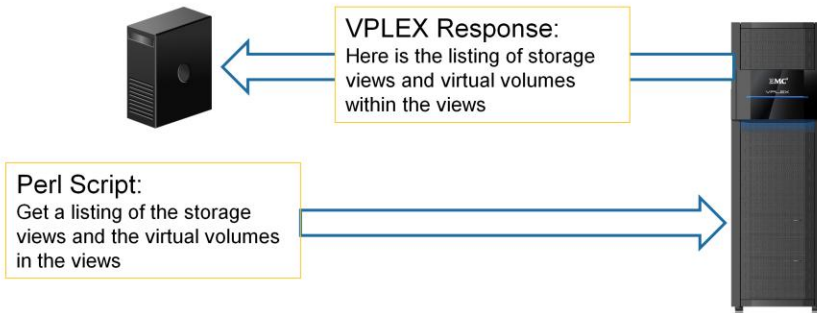
The VPLEX Element Manager API is a RESTful API that provides access to the resources within VPLEX. The API is accessible at the base URL of https://<mgmt-server-public-ip>/vplex. The API uses the default https port 443, but it is not necessary to include the port number in the request. Communications are secured through HTTPS and data is exchanged in the JSON format for interoperability. In VPLEX GeoSynchrony 4.x there wasn't an easy way to script for VPLEX. In VPLEX GeoSynchrony 4.X users had to SSH into the Management Server and then log into the VPLEX CLI just to execute a command. This made scripting very difficult. In VPLEX GeoSynchrony 5.0 users now have much easier way to execute scripts on VPLEX. For example, a user could create a Perl script to retrieve a listing of all of the VPLEX storage views and the volumes within each view. This script could be embedded in a web-based application. A set of authentication headers must be included with each request. The authentication headers must include the username and password in each request. Each request is a separate session and requests do not span multiple sessions.

VPLEX Element Manager API: Supported Requests

HTTP Request	Contexts	Attributes	Commands
GET	Get a listing of a directory	Read the value of an attribute	
PUT		Update the value of an attribute	
POST			Execute a command

There are three types of resources that supported using the VPLEX Element Manager API. The resources supported are contexts or VPLEX CLI directories, attributes of objects within VPLEX, and VPLEX CLI commands. These resources can be acted upon through the VPLEX Element Manager API by three different types of requests, GETs, PUTs, and POSTs. A GET request can get a listing of the components within a VPLEX CLI directory or it can update the value of an attribute. A GET request is similar to entering the ls command within a VPLEX CLI directory. A put request is a write operation that will update the value of an attribute within the VPLEX CLI. A POST request is able execute VPLEX CLI commands. A listing of supported VPLEX CLI commands are listed in the VPLEX Element Manager API documentation. This guide should be consulted for more information on the VPLEX Element Manager API. VPLEX CLI navigation commands such as cd, pushd, popd, and session based commands (capture and monitor) are not supported. VPLEX Element Manager API request should return successfully with an HTTP 200 return code. HTTP 4XX return codes mean that the request was invalid and HTTP 5XX return codes mean that there were errors while handling the requests.

VPLEX Element Manager API: Demo

This video will demonstrate the three different VPLEX Element Manager API requests that are possible.

Launch

EMC²
where information lives®

This video will demonstrate the three different VPLEX Element Manager API requests that are possible.

VPLEX GeoSynchrony v5.0 Interoperability Support

	VPLEX Local	VPLEX Metro	VPLEX Geo
Native	AIX 5300/6100 – PowerHA HPUX 11iv2 – A11.19 HPUX 11iv3 – A11.12 VMware – ESX 3.5, 4.0 VMware – ESX 4.1 Win 2K3 – MSCS Win 2K8 – Failover Cluster	AIX 5300/6100 – PowerHA HPUX 11iv2 – A11.19 HPUX 11iv3 – A11.12 VMware – ESX 4.1 Win 2K3 – MSCS Win 2K8 – Failover Cluster	 VMware – ESX 4.1 Win 2K8 – Failover Cluster
Veritas VCS & VxCFS	RHEL 5.3, 5.4 – 5.0 MP3 RP2 SLES 10, 11 – 5.1 Sol 9 – 5.0 MP3 RP2 Sol 10 – 5.0 MP3 RP2	 Sol 10 – 5.0 MP3 RP2	
Oracle RAC	AIX 6100 – 10g, 11g RHEL 5.4 – 10g, 11g SLES 11 – 10g, 11g Win 2K3 – 10g, 11g Win 2K8 – 10g, 11g	AIX 6100 – 10g, 11g RHEL 5.4 – 10g, 11g Win 2K8 – 10g, 11g	
Non-Native CFS	RHEL 5.4 – OCFS v1.4.1 SLES 11 – OCFS v1.4.1 Win 2K8 – Sambolic Melio Win Hyper V – Sanbolic Melio	RHEL 5.4 – OCFS v1.4.1 SLES 11 – OCFS v1.4.1 Win 2K8 – Sambolic Melio Win Hyper V – Sanbolic Melio	StorNext Win 2K8 – Sanbolic Melio Win Hyper V - Sanbolic Melio

Displayed here are the current clusters and cluster file systems (CFS) that will be supported at GA. Please take a moment to review them.

Course Summary

- VPLEX has ALUA support which increases performance when accessing arrays that are ALUA enabled.

- Users are able to run remote scripts on VPLEX through the VPLEX Element Manager API.

- Synchronous and Asynchronous consistency groups can now be created in VPLEX which provides organization of the volumes and a crash consistent image.

- The VPLEX Witness can be deployed with a VPLEX Metro or Geo to help prevent against system-wide DU events.

- The new VS2 engine has a smaller form factor and more bandwidth than the previous VS1 engine.

This concludes the instruction; proceed to the course assessment.
After launching the assessment, you must complete it before returning to the course.
The course will automatically move to your Transcript within 48 hours after passing the assessment.

EMC²
where information lives'

Copyright © 2011 EMC Corporation. All Rights Reserved.
VPLEX GEOSYNCHRONY v5.0 Technical Differences   89

Listed are the key points covered in this course. This includes the instruction, please proceed to the course assessment.

Copyright © 2011 EMC Corporation. Do not Copy - All Rights Reserved.   89

VPLEX GeoSynchrony v5.0 Technical Differences Course

An administrator has configured the rule Active-Cluster-Wins on their asynchronous consistency group. There is an inter-cluster link failure, but only Cluster-2 is active during the failure.

What is the outcome?

○ Both clusters continue I/O after the failure

○ Cluster-2 continues I/O and Cluster-1 suspends I/O

○ Both clusters suspend I/O and there is a data loss and data rollback situation

○ Cluster-1 continues I/O and Cluster-2 suspends I/O

PROPERTIES

On passing, 'Finish' button:	**Goes to Next Slide**
On failing, 'Finish' button:	**Goes to Next Slide**
Allow user to leave quiz:	**After user has completed quiz**
User may view slides after quiz:	**At any time**
User may attempt quiz:	**Unlimited times**

Properties...

Edit in Quizmaker

Course Feedback (eLearning Only)

FEEDBACK

1. As a result of this training, I will be more

- ○ Agree Completely
- ○ Agree Somewhat
- ○ Neutral. Neither Agree nor Disagree
- ○ Disagree Somewhat
- ○ Disagree Completely
- ○ Not Applicable

Provide Feedback

EMC^2
where information lives

VPLEX GEOSYNCHRONY v5.0 Technical Differences