White Paper

MICROSOFT® WINDOWS™ SERVER FAILOVER CLUSTERING WITH EMC® VPLEX™

BEST PRACTICES PLANNING

Abstract

This white paper describes Microsoft Windows Server Failover Clustering, with functionalities and features leveraging EMC VPLEX storage virtualization and high availability features, along with EMC PowerPath® software running on servers, to provide a fully-automated disaster recovery solution.

February 2014

Part Number **H11992**

Table of Contents

Executive summary

High Availability (HA) comes in many flavors. For many important business applications, highly reliable and redundant hardware provides sufficient uptime. For other business needs, the ability for a critical application to fail over to another server in the same data center is sufficient. However, neither of these server availability strategies will help in the event of truly catastrophic server loss.

For some business applications, even an event as unlikely as a fire, flood, or earthquake can pose an intolerable amount of risk to business operations. For truly essential workloads, distance can provide the only hedge against catastrophe. By failing server workloads over to servers separated by hundreds of miles, truly disastrous data loss and application downtime can be prevented.

EMC VPLEX with Microsoft Windows Server Failover Clustering (WSFC) and technologies provides the answer.

Introduction

This white paper discusses the VPLEX models and features to provide high availability data across geographically dispersed locations. It describes how to use Windows Server Failover Clustering along with applications such as Hyper-V and SQL Server 2012 in a VPLEX Metro environment. Together these technologies provide fully-automated failover mechanisms designed to protect customer data and provide data mobility for planned or unplanned downtimes.

Audience

This white paper is intended for technology architects, storage administrators, and system administrators who are responsible for architecting, creating, managing IT environments that utilize EMC VPLEX technologies. The white paper assumes the reader is familiar with EMC VPLEX, EMC PowerPath, Microsoft Windows Server Technologies, and Multi-Site Clusters.

EMC VPLEX

EMC VPLEX delivers data mobility and availability across arrays and sites. VPLEX is a unique virtual storage technology that enables mission-critical applications to remain up and running during any of a variety of planned and unplanned downtime scenarios. VPLEX permits painless, non-disruptive data movement, taking technologies such as Windows Server Failover Clustering along with Hyper-V and allowing them to function transparently across multiple heterogeneous storage arrays and across distance. The following figure provides an example of an EMC GeoSynchrony™ operating environment.

Figure 1 - GeoSynchrony Operating Environment

VPLEX family

The following are the EMC VPLEX family offerings:

- **VPLEX Local:** VPLEX local allows centralized management of all arrays in the data center. Storage management is simplified allowing for improved storage utilization across all storage arrays. Data mobility and availability is enhanced using VPLEX local.
- **VPLEX Metro:** VPLEX Metro uses EMC Federated AccessAnywhere technology, allowing block level access to data between two sites within synchronous distances. VPLEX Metro with its active/ active configuration provides high levels of availability, mobility and resource utilization within and across datacenters at synchronous distances.
- **VPLEX GEO**: VPLEX Geo uses EMC Federated AccessAnywhere technology, allowing block level access to data between two sites within asynchronous distances. VPLEX GEO, in conjunction with Windows Server Failover Clustering, provides Highly Available Applications and Data Mobility within and across datacenters at asynchronous distances.

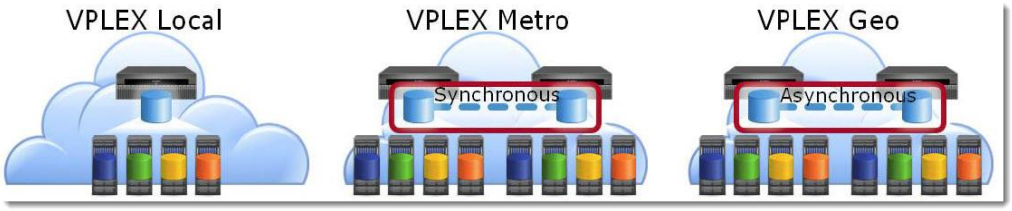

Figure 2 - VPLEX Offerings

VPLEX Features

Data Mobility

Move applications, virtual machines, and data in and between data centers without impacting users.

Figure 3 - Data Mobility Example

Continuous Availability

Deliver application and data availability within the data center and over distance with full infrastructure utilization and zero downtime.

Figure 4 - Continuous Availability Example

Multi-Site Clusters

Server clusters support a single cluster spanning multiple sites.

Figure 5 - Multi-Site Cluster Example

Migrations

Non-disruptive data migrations enable risk-free, faster tech refreshes and load balancing.

Figure 6 - Migrations Example

HA Infrastructure

HA Infrastructure reduces recover time objective (RTO).

VPLEX is inherently a high availability solution that provides application and data mobility. In Figure 7, we added Windows Server Failover Clustering with CSVs and Hyper-V Server virtualization to the VPLEX storage solution to create a highly available environment with storage, servers, and application protection.

Application and Data Mobility

In the following example, Windows Hyper-V Virtual Machines are shown moving (failover, failback) between datacenters with no downtime or impact to users. VPLEX facilitates the movement of virtual machines and the applications running on them between data centers.

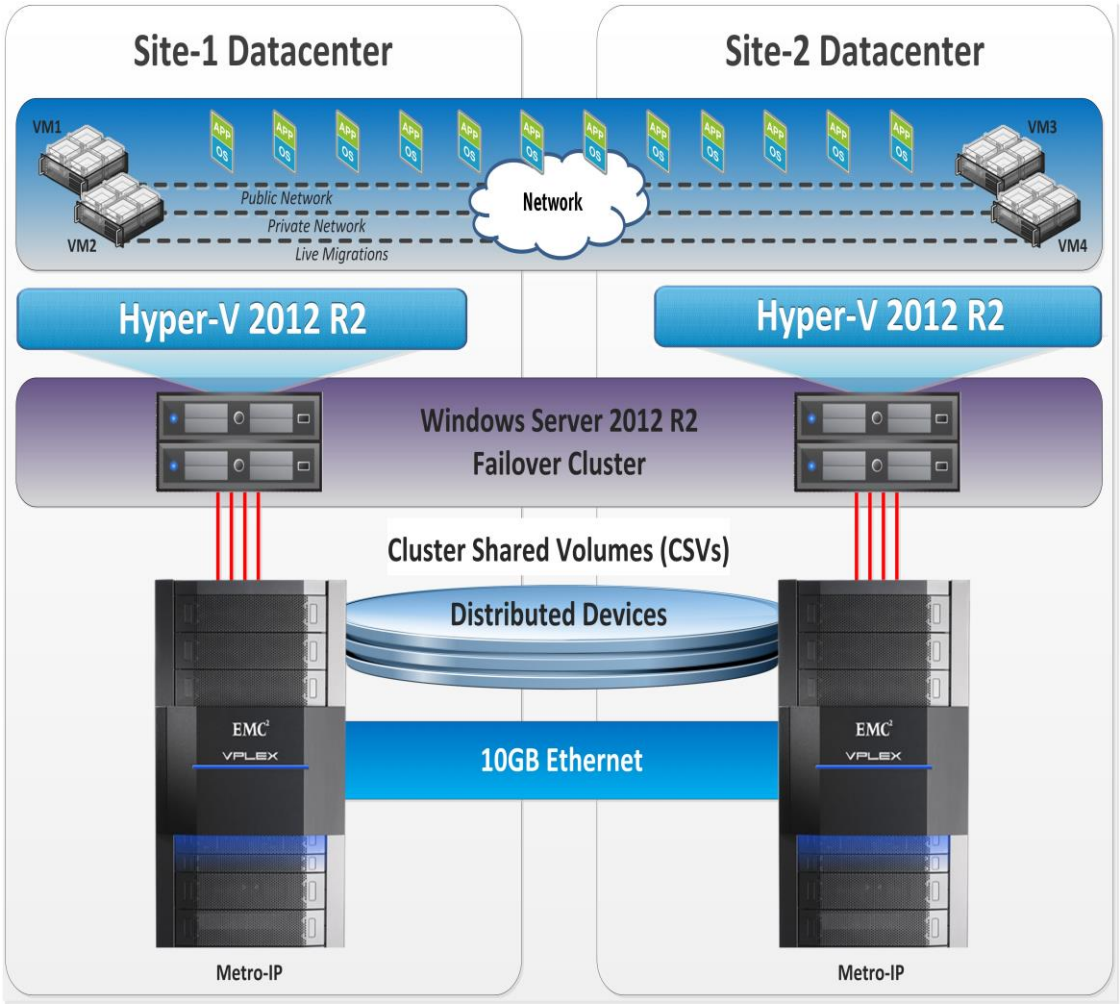

Figure 7 – VPLEX Application and Data Mobility Example

Federated AccessAnywhere

VPLEX utilizes a break-through technology called Federated AccessAnywhere. Federated AccessAnywhere provides cache coherency to consistently present a single copy of data across VPLEX clusters. This single consistent view allows for data to be shared, accessed and relocated over distance and between sites.

Using VPLEX Detach Rules

For every VPLEX distributed virtual volume placed into a consistency group, the user can configure a detach rule that determines which cluster will continue to service I/O in case of inter-site link failure. The preference can be selected as follows:

- Prefer A: Cluster A services I/O
- Prefer B: Cluster B services I/O

- No Automatic Winner: No one services I/O
- Active Cluster Wins: If active, detaches and services I/O

VPLEX Witness

The VPLEX Witness is responsible for helping VPLEX clusters distinguish between VPLEX cluster failures and inter-cluster partition failures. The witness observes health-check heartbeats to both clusters over the IP network and notifies clusters about its observations. All distributed virtual volumes are still configured with the preferred detach rules but, in the instance of a cluster or connectivity failure the VPLEX Witness will force the majority rule to take precedence over the preference rule. This means that in order for a given cluster to continue processing I/O it must either be connected to a peer cluster or the VPLEX Witness. The static preference plays a role only in the case of an inter-cluster network partition when both clusters remain connected to the VPLEX Witness.

VPLEX Witness is deployed as a virtual machine and may be hosted by either Microsoft Hyper-V or a VMware ESXi hypervisor.

NOTE: If both clusters were active, or both clusters were passive at the time of the link failure, I/O is suspended on both clusters to ensure data integrity. This is only happens with the preference set to "Active Cluster Wins".

For more information, refer to the *EMC VPLEX Metro Witness Technology and High Availability TechBook*, available on https://support.emc.com.

Figure 8 - VPLEX Witness Deployed with a Hyper-V Virtual Machine

VPLEX Clustering Architecture

EMC VPLEX represents the next-generation architecture for data mobility and information access. The new architecture is based on EMC's more than 20 years of expertise in designing, implementing, and perfecting enterprise-class intelligent cache and distributed data protection solutions.

As shown in the following figure, VPLEX is a solution for federating both EMC and non- EMC storage. VPLEX resides between the servers and heterogeneous storage assets and introduces a new architecture with unique characteristics:

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

VPLEX uses a unique clustering architecture to help customers break the boundaries of the data center and allow servers at multiple data centers to

have concurrent read and write access to shared block storage devices. A VPLEX cluster can scale up through the addition of more engines, and scale out by connecting multiple clusters to form a VPLEX Metro configuration.

A VPLEX Metro system consists of two VPLEX clusters joined together. These VPLEX clusters can be located in the same datacenter or across two different datacenter sites within synchronous distances (approximately up to 60 miles or 100 kilometers apart).

VPLEX Metro configurations help users to transparently move and share workloads, consolidate data centers, and optimize resource utilization across data centers. In addition, VPLEX clusters provide nondisruptive data mobility, heterogeneous storage management, and improved application availability.

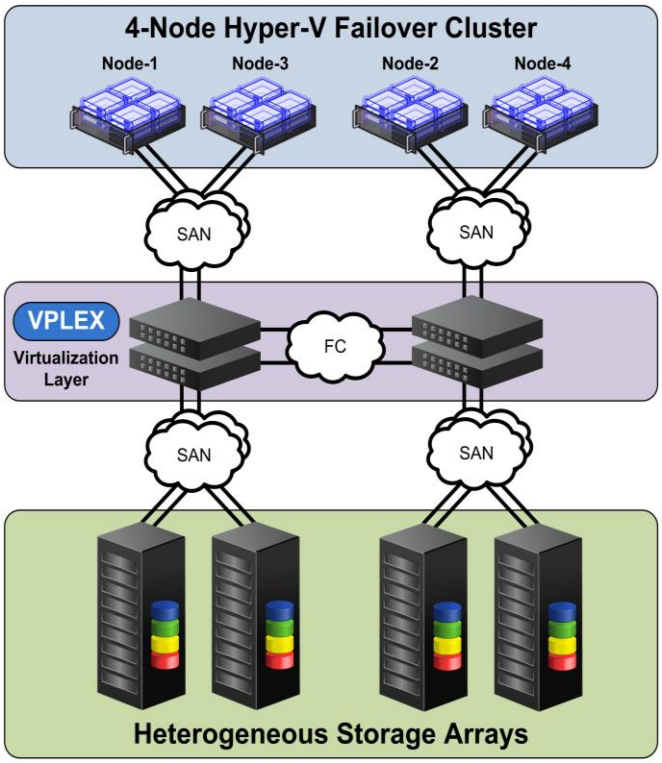

Figure 9 - Overview of EMC VPLEX Clusters

A VPLEX cluster is composed of one, two, or four engines. The engine is responsible for virtualizing the I/O stream, and connects to hosts and storage using Fibre Channel connections as the data transport. A single-engine VPLEX cluster consists of the following major components:

- Two directors, which run the GeoSynchrony software
- Dedicated 8Gb FC front-end, back-end, and Local/WAN COM

- One Standby Power Supply, which provides backup power to sustain the engine through transient power loss

Each cluster also consists of:

- A management server that provides a GUI and CLI interface to manage a VPLEX cluster
- An EMC standard 40U cabinet to hold all of the equipment of the cluster

Additionally, clusters containing more than one engine also have:

- A pair of Fibre Channel switches used for inter-director communication between various engines
- A pair of Universal Power Supplies that provide backup power for the Fibre Channel switches and allow the system to ride through transient power loss

EMC PowerPath

PowerPath is a server-resident software solution designed to enhance performance and application availability. PowerPath combines automatic load balancing, path failover, and multiple path I/O capabilities into one integrated package. PowerPath enhances application availability by providing load balancing and automatic path failover and recovery functionality. PowerPath supports servers, including WSFC servers connected to EMC and qualified third-party arrays. PowerPath for Windows Multipathing has the following new features:

- Automatic Host-Array Registration
- Support for Microsoft Windows Server
- Support for VNX2 arrays.
- New VPLEX Class for Virtualized Volumes

PowerPath Configuration Checker

PowerPath Configuration Checker (PPCC) is a software program that verifies that a host is configured to E-Lab Interoperability Navigator standards with the hardware and software required for PowerPath multipathing features (failover and load-balancing functions, licensing, and policies). Prior to installing or upgrading PowerPath, download the latest version of EMC Reports available on EMC Online Support and then run PPCC. This ensures that the system version used by PPCC includes the latest configuration information.

PowerPath Autostandby Feature for VPLEX

PowerPath for Windows supports setting PowerPath to automatically put paths into autostandby that have intermittent I/O failures (also called

flaky paths) and automatically select autostandby for high-latency paths in VPLEX cross-connected Metro configurations.

Note: This is a feature that no other multipath can deliver.

These two new autostandby functions are called IOsPerFailure-based (asb:iopf) autostandby and proximity-based (asb:prox) autostandby:

- The IOsPerFailure-based autostandby puts a path into autostandby if that path has fewer than the specified number of I/Os between failures. This puts undependable paths into reserve.
- The proximity-based autostandby puts a path into autostandby if it is associated with the designated remote/non-preferred VPLEX cluster within a VPLEX Metro system (for hosts with cross-connected, distributed volumes). PowerPath groups paths internally by VPLEX cluster. The VPLEX cluster with the lowest minimum path latency is designated as the local/preferred VPLEX cluster, while the other VPLEX cluster within the VPLEX Metro system is designated as the remote/non-preferred. A path associated with the local/preferred VPLEX cluster is put in active mode, while a path associated with the remote/non-preferred VPLEX cluster is put in autostandby mode.

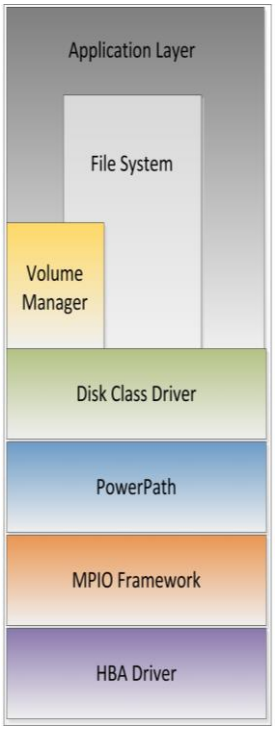

Figure 10 - MPIO Stack with PowerPath

Windows Server Failover Clustering (WSFC)

A failover cluster consists of a group of independent computers that work together to increase the availability and scalability of clustered roles. These clustered roles contain applications and services that are made to be highly available. The clustered servers (called nodes) are physically connected by cables and logically grouped by failover cluster software. If one or more of the cluster nodes should fail, the application or service will move (or failover) to alternate nodes as to continue providing those services without disruption. This is done by proactively monitoring each cluster resource to verify that they are working properly. If there is an issue, then those targeted resources are restarted and/or moved to another node.

Multi-Site Clustering

Using the VPLEX and WSFC solution to build your Multi-Site Clusters is crucial for ensuring Continuous operations and Availability. When combining WSFC and VPLEX technologies you are enabling proactive operations such as disaster avoidance (DA) which can help prevent data corruption or inconsistent results to clients during certain, disaster recovery (DR) and application/service tolerance scenarios. Using WSFC stretched across two separate geo's with VPLEX would add the following dynamics:

- VPLEX supports multiple storage arrays, with at least two storage arrays deployed at each site. This ensures that in the event of failure of any one site, the other site or sites will have the required copies of the data that they may maintain continuous availability of critical applications and/or services.
- VPLEX enables automated restart of application resources on cluster nodes that are connected to VPLEX virtualized storage in surviving sites. This is done through a combination of WSFC policies and VPLEX Witness settings that we will cover in a later section.

Because of their extreme disaster tolerance, multi-site clusters should be thought of as both a high-availability solution and a disaster recovery solution. The automatic failover of WSFC multi-site clustering means that your application services and data being managed by WSFC is available in moments upon failure of your primary site. What is more, automatic failover means that you can quickly failback to your primary site once your servers there have been restored.

Failover Cluster Validation Tests

Before you create a failover cluster, or when you add a node to an existing cluster, we recommend that you validate your configuration by running all tests in the Validate a Configuration Wizard. By running these tests, you can confirm that your hardware and settings are compatible with failover clustering. For example, the tests validate that the servers in the cluster solution are connected correctly to the networks and storage, and that they contain identical software updates. Microsoft supports a failover cluster solution only if the complete configuration (servers, network, and storage) can pass all tests in the Validate a Configuration Wizard. Cluster validation is intended to do the following:

- Find hardware or configuration issues.
- Help ensure that the clustering solution will be dependable.
- Provides a way to validate changes to an existing cluster.
- Perform diagnostic tests on an existing cluster.
- Improved performance for testing storage.
- Targeted validation of new LUNs. Allows specifying a new LUN (disk), rather than testing all LUNs when validating storage.
- Integration with PowerShell and WMI.
- Test support for CSVs, and for Hyper-V and virtual machines.
- Validation test for replicated hardware. (for support multi-sites)

Important Note: With Windows Server Failover Cluster, the validate storage tests may not discover VPLEX distributed devices, when the failover cluster nodes are located across multiple sites.

This is because storage validation test will only select shared LUNs. Microsoft has determined that a LUN is determined to be shared if its disk signatures, device identification number, and storage array serial number are the same on all cluster nodes. However, when you have deployed a VPLEX Distributed Device across your datacenters, these LUNs have the same disk signatures and device identification number, but the storage array serial number are different. Therefore, they are not recognized as shared LUNs.

The following is an example of a cluster storage validation error:

```
Cluster validation message:

List Potential Cluster Disks

Description: List disks that will be validated for cluster
compatibility. Clustered disks which are online at any node will be
excluded.

Start: 11/17/2013 5:59:01 PM.

Physical disk 84d2b21a is visible from only one node and will not be
tested. Validation requires that the disk be visible from at least two
```

```
nodes. The disk is reported as visible at node: WNH6-
H5.elabqual.emc.com

Physical disk 6f473a9f is visible from only one node and will not be
tested. Validation requires that the disk be visible from at least two
nodes. The disk is reported as visible at node: WNH6-
H13.elabqual.emc.com
```

To resolve the issue, run all the cluster validation tests before you configure distributed devices to the Multi-Site Servers and take note that Microsoft does NOT require validation tests to be ran against disks in a Multi-Site Cluster. However, if the validation test is needed afterward for support situations, LUNs that are not selected for storage validation tests are supported by Microsoft and EMC VPLEX Distributed Devices.

Please refer to the following Microsoft KB articles for more information:

For Windows 2008 and 2008 R2, see:
http://support.microsoft.com/kb/943984

For Windows 2012 and 2012 R2, see:
http://support.microsoft.com/kb/2914974

Validation Tests for Storage

There are some caveats to setting up multi-site clusters with VPLEX and WSFC so let's take a look at some of those aspects. First, with a multisite cluster there is an assumption that there are an even amount of nodes located at Site-1 that are connected to one VPLEX and that there are an equal amount of nodes located at Site-2 that are connected to another VPLEX. The two VPLEXs will form either a Metro or Geo cluster and are attached to at least two back-end storage arrays at each site. The VPLEX will ensure that all data is replicated between sites so that both sets of storage have the exact same data.

If your cluster has already been deployed it may not be advisable to take a production disk offline due to the availability impact it might have on the clustered roles that use it. In this situation, you can run validation tests (including storage tests) by creating or choosing a new Distributed Device (DD) from the VPLEX and presenting it to all nodes in the cluster. By testing this DD, you can avoid disruption to clustered roles that are already online within the cluster and still test the underlying storage subsystem.

If a failover cluster passed the full set of validation tests and has no future hardware or software changes, it will continue to be a supported. However, if you perform routine updates to software components such as HBA firmware and/or drivers, or PowerPath it will be necessary to rerun the Configuration Wizard to ensure that the current configuration of the

failover cluster is still supported. The following guidelines can help you decide when this is necessary:

- All components of the storage stack should be identical across all nodes in the cluster. (see Figure-11) It is required that multipath I/O (MPIO) software and Device Specific Module (DSM) software components be identical across all Failover Cluster Nodes. It's also recommend that the host bus adapters (HBAs), drivers, and firmware are identical for each Node too. If you must use dissimilar HBAs, you should verify with the storage vendor that you are following their supported or recommended configurations.
- A best practice is to keep a small LUN available to allow the Validate a Configuration Wizard to run tests on available storage without negatively impacting clustered roles. If Microsoft customer support requests that you run a full set of cluster validation tests, the wizard allows you to select that disk for the storage tests to verify that the storage is working properly.

Planning the Quorum for Failover Clustering

The concept of using a shared quorum device in Windows Server Failover Clustering hasn't changed over the years, but maintaining cluster quorum has evolved from a shared quorum device to include newer more robust options. With Multi-Site Failover Clusters powered by VPLEX, the concept of quorum now refers to the number of "votes" that the failover cluster must equate to form a majority. This includes all of the WSFC Nodes and either a "file share witness" located at a 3rd site (possibly with the VPLEX Witness) or a "disk witness" which would use a VPLEX Distributed Device (that is synchronously replicated between sites) to store a copy of the cluster database across all failover nodes. Both solutions will provide the necessary votes needed to obtain majority for WSFC membership.

VPLEX allows Multi-Site Failover Clusters to be deployed in a way that automates the failover of applications in situations where the following occurs:

- Communication has failed between sites.
- Complete site failure that prevents applications from running.

Of the four quorum models available in WSFC, two are best suited to multi-site clustering with VPLEX: *Node and File Share Witness* and *Node and Disk Witness*. The node and file share witness configuration enables the creation of up to 64-Node clusters with no shared disk. This is the preferred choice for multi-site clustering with VPLEX and Microsoft. In this quorum configuration, a file share acts as a *witness*, a tie-breaking vote to the votes provided by the Multi-Site Cluster nodes. With the addition of this 3rd site file share witness, WSFC will total all of the votes, meaning

that connectivity can be lost by any of the nodes (or the witness itself) and the cluster can continue functioning.

Note: The "file share witness" keeps track of which "node" has the most current version of the cluster database, but does not keep a copy of that database. VPLEX Distributed Devices will allow for the use of a shared disk across datacenters which can enable the ability to use the "disk witness" model so that you will always have a backup of the cluster database.

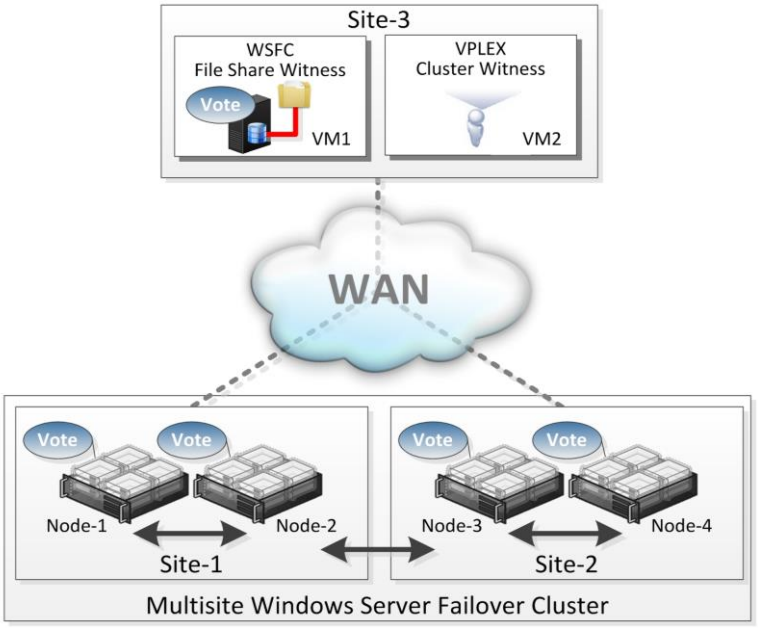

Figure 11 - Node & File Share Majority Quorum Model

The node-majority quorum configuration can work when there is an odd number of nodes at each site. Take the case of a multi-site cluster consisting of five nodes, three of which reside at Site-1 and the remaining two at Site-2. With a break in communication between the two sites, Site-1 can still communicate with three nodes (which is greater than 50 percent of the total), so all of the nodes at Site-1 stay up.

The nodes in Site-2 are able to communicate with each other, but no one else. Since the two nodes at Site-2 cannot communicate with the majority, they drop out of cluster membership. (Were Site-1 is to go down in this case, in order to bring up the cluster at Site-2, it would require manual intervention to override the non-majority.

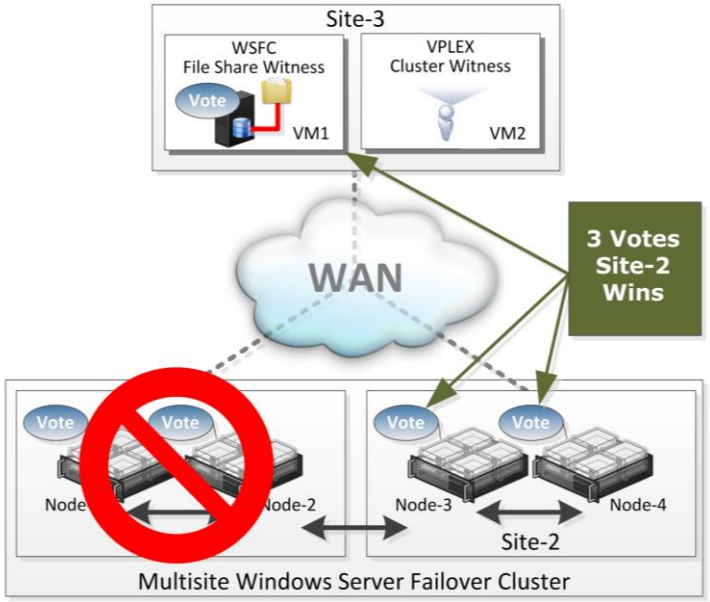

Figure 12 - Node and File Share Witness Failure Scenario

The VPLEX uses SCSI timeouts, Not Ready states, and a few other methods to identify the issues that cause split-brain, but is not actively preventing split-brain from happening. This is where the importance of the VPLEX Witness becomes relevant. VPLEX Witness is designed to handle the continuous availability of the storage between datacenters while the Node & File Share Witness is designed to prevent split-brain and maintain your Highly Available Applications deployed with WSFC. When deployed together at a 3rd Site, these two features will offer the best solution for both worlds.

Quorum Configuration Options

The quorum model for WSFC is flexible. If you need to modify the quorum configuration for your cluster, you can use the Configure Cluster Quorum Wizard or the Failover Clusters Windows PowerShell cmdlets. The following list of quorum configuration options are available when running the configure cluster quorum wizard with Windows 2012 R2:

- **Use default quorum configuration:** The WSFC Cluster automatically assigns a vote to each node and dynamically manages the node votes, a technology called Dynamic quorum, available starting with Windows 2012. If it is suitable for your cluster, and there is cluster shared storage available, the cluster selects a WSFC disk witness. Starting with Windows 2012 R2, a feature called Dynamic witness will adjust the use of the witness resources depending on the number of active nodes voting in the cluster. This option is recommended in most cases, because the cluster software

automatically chooses a quorum and witness configuration that provides the highest availability for your cluster.

- **Select the quorum witness:** You can add, change, or remove a WSFC witness resource. You can configure a file share or disk witness. The cluster automatically assigns a vote to each node and dynamically manages the node votes.
- **Advanced quorum configuration:** You should select this option only when you have application-specific or site-specific requirements for configuring the quorum. You can modify the quorum witness, add or remove node votes, and choose whether the cluster dynamically manages node votes. By default, votes are assigned to all nodes, and the node votes are dynamically managed.

Depending on the quorum configuration option that you choose for your VPLEX Multi-Site Cluster and your specific WSFC settings, the WSFC Cluster will be configured in one of the following two quorum modes:

- **Node majority with witness (disk or file share):** Nodes have votes. In addition, a quorum witness has a vote. The cluster quorum is the majority of voting nodes in the active cluster membership plus a witness vote. A quorum witness can be a designated disk witness or a designated file share witness.

Witness Configuration

As a general rule when you configure a quorum, the voting elements in the cluster should be an odd number. Therefore, if the cluster contains an even number of voting nodes, you should configure a disk witness or a file share witness. The cluster will be able to sustain one additional node down. In addition, adding a witness vote enables the cluster to continue running if half the cluster nodes simultaneously go down or are disconnected.

A disk witness is usually recommended if all nodes can see the disk. A file share witness is recommended when you need to consider multisite disaster recovery with replicated storage. The following table provides additional information and considerations about the quorum witness types.

Disk Witness: The disk witness is a dedicated LUN that stores a copy of the cluster database and is most useful in clusters that have shared storage that is not replicated. The requirements for a disk witness are as follows:

- LUN size must be a minimum of 512 MB
- Must be dedicated to cluster use (no role assignments)
- Must pass storage validation tests and added to cluster storage
- Cannot be a Cluster Shared Volume (CSV)
- Basic NTFS or ReFS formatted disk with a single volume

- Does not require a drive letter assignment
- May use hardware RAID for fault tolerance and resiliency
- Must be excluded from backups and antivirus activities

File Share Witness: The file share witness is an SMB file share that is configured on a file server running Windows Server. The files share witness does not store a copy of the cluster database, but only maintains cluster information in a witness log file. The file share witness is the preferred method used for multisite clusters with replicated storage. The requirements for a file share witness are as follows:

- Requires a minimum of 5 MB of free space.
- File share must be dedicated to cluster use only and may not be used to store user or application data.
- A single file server can be configured with file share witnesses for multiple clusters.
- The cluster name must have write permissions for computer object.
- As with the VPLEX Cluster Witness, the file server hosting the file share witness must be located at a third site. This allows for any cluster site to survive if SAN or WAN communication is lost. Otherwise, if the file server was located at the same site, that site becomes the primary site, and it is the only site that can reach the file share.
- The file share witness may me run on a virtual machine.
- For HA requirements, it's highly recommended by EMC and Microsoft that the file share witness should be configured at a 3rd Site to prevent unnecessary disruption of services.

Dynamic Quorum Management:

In WSFC, as an advanced quorum configuration option, you can choose to enable dynamic quorum management by cluster. When this option is enabled, the cluster dynamically manages the vote assignment to nodes, based on the state of each node. Votes are automatically removed from nodes that leave active cluster membership, and a vote is automatically assigned when a node rejoins the cluster. Dynamic quorum management is enabled by default.

With dynamic quorum management, it is also possible for a cluster to run on the last surviving cluster node. By dynamically adjusting the quorum majority requirement, the cluster can sustain sequential node shutdowns to a single node.

Additional considerations:

- Dynamic quorum management does not prevent loss of quorum with multiple simultaneous failures of the voting members.

- To continue running, the cluster must always have a quorum majority at the time of a node shutdown or failure.
- If you have explicitly removed the vote of a node, the cluster will not dynamically add or remove that vote.
- Microsoft recommends to always configure a witness, even with an odd number of nodes.

Note: WSFC will automatically determine when to use the dynamic witness or not. This is a new feature with Windows 2012 R2.

Tie Breaker for 50% Node Splits

As an enhancement to dynamic quorum functionality, a cluster can now dynamically adjust a running node's vote to keep the total number of votes at an odd number. This functionality works seamlessly with dynamic witness. To maintain an odd number of votes, a cluster will first adjust the quorum witness vote through dynamic witness. However, if a quorum witness is not available, the cluster can adjust a node's vote. For example:

- You have a six node cluster with a file share witness. The cluster stretches across two sites with three nodes in each site. The cluster has a total of seven votes.
- The file share witness fails. Because the cluster uses dynamic witness, the cluster automatically removes the witness vote. The cluster now has a total of six votes.
- To maintain an odd number of votes, the cluster randomly picks a node to remove its quorum vote. One site now has two votes, and the other site has three.
- A network issue disrupts communication between the two sites. Therefore, the cluster is evenly split into two sets of three nodes each. The partition in the site with two votes goes down. The partition in the site with three votes continues to function.

If for some reason you don't like the fact that Windows will randomly choose which of the two sites will stay online in this scenario, you might want to use the "LowerQuorumPriorityNodeID" setting against a node in the site you'd prefer to go offline. This property will determine which site survives if there is a 50% node split where neither site has quorum. Instead of the cluster randomly picking a node to remove its quorum vote, you can use this setting to predetermine which node will have its vote removed allowing one side of the cluster to continue to run in the case of a 50% node split where neither side would normally have quorum.

Note: To minimize downtime and manual intervention, it is a best practice to configure WSFC and VPLEX in parallel; that is, for a given resource the

VPLEX preferred site and the WSFC preferred node(s) should be set to the same physical location.

Automatic failover

This is where the cluster configuration consists of two or more sites that can host clustered roles. If a failure occurs at Site-1, Site-2, or the Site-3 Witness… the clustered roles are expected to automatically fail over to the remaining site. Therefore, the cluster quorum must be configured so that any site can sustain a complete site failure. The following summarizes the recommended settings:

- **Number of node votes per site:** Should be equal.(and even)
- **Dynamic quorum management:** Should be enabled.
- **Witness configuration:** WSFC File Share Witness is highly recommended to be configured in a 3rd site that is separated from the other WSFC sites.
- **Workloads:** Workloads can be configured on any of the sites

Note: An assumption is being made that the file share witness has been configured at a 3rd site along with the VPLEX Witness. This will give each site an equal opportunity to survive outages and/or disruptions in service.

Conclusion

By using VPLEX distributed virtual volumes, data can be transparently made available to all notes in a Windows Failover Cluster divided across two physical locations. The Windows Failover Cluster can be defined either as a local cluster or a geographically extended cluster by the host cluster software. VPLEX supports either scenario.

Like a host clustering product, VPLEX is architected to react to failures in a way that will minimize the risk of data corruption. With VPLEX, the concept of a "preferred site" ensures that one site will have exclusive access to the data in the event of an inter-site failure.

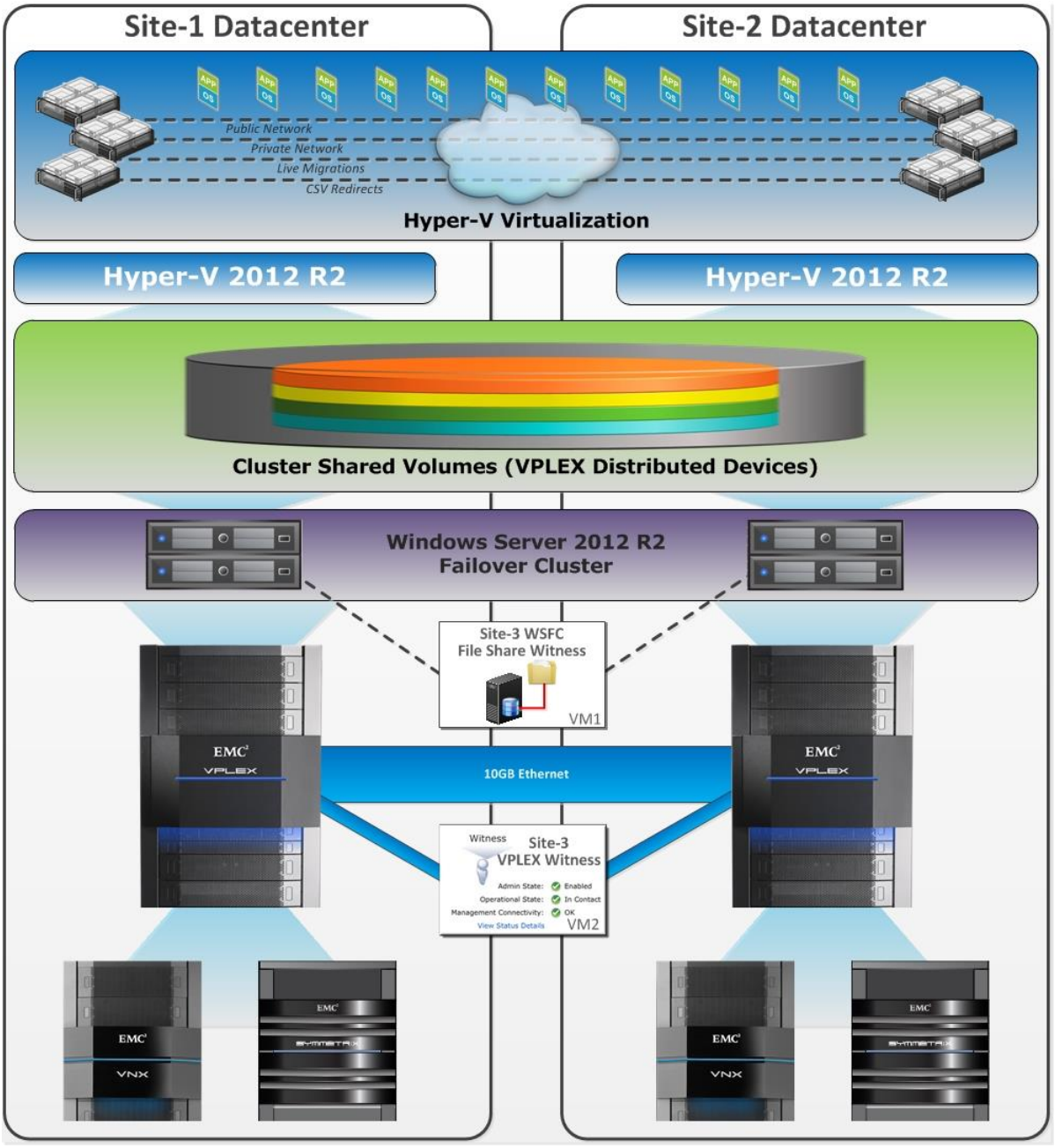

Figure 13 - Complete VPLEX and WSFC Solution

Appendix - A

Windows PowerShell® is a task-based command-line shell and scripting language designed especially for system administration. Windows PowerShell commands, called cmdlets, let you manage the computers from the command line.

Windows PowerShell includes the following features:

• Cmdlets for performing common system administration tasks, such as managing the registry, services, processes, and event logs, and using Windows Management Instrumentation (WMI).

• A task-based scripting language and support for existing scripts and command-line tools.

• Consistent design. Because cmdlets and system data stores use common syntax and naming conventions, data can be shared easily and the output from one cmdlet can be used as the input to another.

PowerShell Cmdlets for Failover Clusters

Get-Cluster: Gets information about failover clusters.

Get-ClusterGroup: Gets information about clustered roles, or resource groups, in a failover cluster.

Get-ClusterLog: Creates a log file for all nodes in a failover cluster.

Get-ClusterNode: Gets information about nodes in a failover cluster.

Get-ClusterResourceDependencyReport: Generates a report that lists the dependencies between resources in a failover cluster.

Get-ClusterSharedVolume: Gets information about Cluster Shared Volumes (CSVs) in a failover cluster.

Get-ClusterSharedVolumeState: Gets the state of Cluster Shared Volumes in a cluster.

Move-ClusterGroup: Moves a clustered role, a resource group, from one node to another in a failover cluster.

Move-ClusterResource: Moves a clustered resource from one clustered role to another within a failover cluster.

Move-ClusterSharedVolume: Moves a Cluster Shared Volume (CSV) to ownership by a different node in a failover cluster.

Move-ClusterVirtualMachineRole: Moves the ownership of a clustered virtual machine to a different node.

Start-ClusterGroup: Brings one or more clustered services and applications, also known as resource groups, online on a cluster.

Start-ClusterResource: Brings a resource online in a failover cluster.

Stop-ClusterGroup: Takes one or more clustered services and applications, also known as resource groups, offline on a cluster.

Stop-ClusterNode: Stops the Cluster service on a node in a cluster.

Stop-ClusterResource: Takes a resource offline in a failover cluster.

Test-Cluster: Runs validation tests for failover cluster hardware and settings.

Note: This is just a short list of the cmdlets available for failover clustering, please use the following command for a complete list:

```
Get-Command -Module FailoverClusters
```

References

The following in available on EMC Online support at https://support.emc.com:

- *EMC VPLEX Metro Witness Technology and High Availability TechBook*
- *Implementation and Planning Best Practices for EMC VPLEX Technical Notes*
- *Long-Distance Application Mobility Enabled by EMC VPLEX GEO White Paper*
- *EMC VPLEX with GeoSynchrony 5.3 Product Guides*
- *EMC Host Connectivity Guide for Windows*, https://elabnavigator.emc.com
- *Failover Clustering in Windows 2012 R2*, http://technet.microsoft.com/