Implementing EMC VPLEX and Microsoft Hyper-V and SQL Server with Enhanced Failover Clustering Support

Applied Technology

Abstract

This white paper examines deployment and integration of Microsoft Hyper-V and Microsoft SQL Server solutions on EMC® VPLEX™ storage federation systems. Details of integration with VPLEX systems are documented with practical examples for storage and database administrators.

May 2010
# Table of Contents

- **Executive summary** ........................................................................................................................................... 4  
- **Introduction** ..................................................................................................................................................... 4  
  - Audience ........................................................................................................................................................... 4  
- **VPLEX technology overview** .......................................................................................................................... 4  
  - VPLEX Local ...................................................................................................................................................... 5  
  - VPLEX Metro ...................................................................................................................................................... 6  
  - Storage device federation ................................................................................................................................. 7  
- **Connectivity recommendations** ...................................................................................................................... 8  
  - Back-end connectivity for storage .................................................................................................................. 8  
  - Front-end connectivity for hosts .................................................................................................................... 9  
- **Provisioning storage with VPLEX** .................................................................................................................. 11  
- **VPLEX insertion into a SQL Server environment** ........................................................................................ 12  
  - Mapping SQL Server storage objects ........................................................................................................ 13  
  - Claiming storage volumes with VPLEX ........................................................................................................ 15  
  - Defining encapsulated devices ..................................................................................................................... 17  
  - Configuring devices into a VPLEX view ......................................................................................................... 20  
  - Host access and registration ........................................................................................................................ 20  
- **EMC VPLEX and Microsoft Windows Server Failover Clusters** .................................................................... 21  
  - Windows Failover Clustering and Windows Hyper-V ................................................................................ 22  
  - Windows Hyper-V Live Migration ................................................................................................................ 22  
  - Configuring VPLEX storage as Hyper-V pass-through .............................................................................. 22  
- **VPLEX Metro and Cluster Shared Volumes** ..................................................................................................... 26  
  - Defining a geographically dispersed Windows Failover Cluster ................................................................ 26  
  - Creating remote storage devices .................................................................................................................. 28  
  - Defining distributed storage devices .......................................................................................................... 29  
  - Enabling remote node access ...................................................................................................................... 31  
- **Additional VPLEX Metro cluster features** .................................................................................................... 32  
  - Exported volumes .......................................................................................................................................... 32  
- **Conclusion** ....................................................................................................................................................... 34
Executive summary

The EMC® VPLEX™ family of products based on the EMC GeoSynchrony™ operating system provides an extensive range of new features and functionality for the evolving era of cloud computing. The VPLEX system removes physical barriers within, across, and between data centers and allows users to access common, federated storage volumes at differing geographical locations utilizing EMC AccessAnywhere™ technology. This single consistent view across heterogeneous storage systems, and between multiple physical locations, enhances the extensive solution offerings from EMC for Windows Failover Clustering.

With the introduction of storage federation within a single site, VPLEX Local solutions allow customers the ability to utilize all storage resources in a single consistent view. Combined with Hyper-V Live Migration capabilities, administrators are able to execute dynamic migrations and load-balancing operations with no impact to application availability. In multi-site configurations, VPLEX Metro provides support for a single federated view of the storage resources across the multiple sites, and extends the functionality provided by VPLEX Local to provide multi-site disaster recovery solutions. VPLEX Metro also allows customers to dynamically load balance virtual machine resources across the sites and to perform Hyper-V Live Migration capabilities in a seamless manner via the extension of the single federated storage view with AccessAnywhere.

An EMC VPLEX system is thus a natural extension for a virtualization environment based on Microsoft technologies. The capability of the EMC VPLEX family to provide both local and distributed federation allows for transparent cooperation of physical data elements within a single site or across two geographically separated sites, and allows IT administrators to break physical barriers and expand their Windows- and Hyper-V-based cloud offering. The synergies provided by a Hyper-V virtualization offering connected to the EMC VPLEX system thus help customers to reduce total cost of ownership while optimizing efficiencies by providing a dynamic service that can rapidly respond to the changing needs of their business.

Introduction

This white paper examines deployment and integration of Microsoft Hyper-V and Microsoft SQL Server solutions on EMC VPLEX storage federation systems. Details of integration with VPLEX systems are documented with practical examples for storage and database administrators.

Audience

This white paper is intended for Microsoft SQL Server and Windows Hyper-V administrators, storage administrators and architects, customers, and EMC field personnel who want to understand the implementation of new features and functions that can provide additional benefits in an EMC VPLEX environment.

VPLEX technology overview

The EMC VPLEX is an enterprise-class storage area network-based federation solution that aggregates and manages pools of Fibre Channel-attached storage arrays that can be either collocated in a single data center or multiple data centers that are geographically separated by Metropolitan Area Network (MAN) distances. EMC VPLEX Metro provides nondisruptive, heterogeneous data movement and volume management functionality within synchronous distances. With a unique scale-up and scale-out architecture, VPLEX systems’ advanced data caching and distributed cache coherency provide workload resiliency, automatic sharing, and balancing and failover of storage domains, and enable both local and remote data access with predictable service levels.

The Microsoft Windows Server platform enables customers to provide support for the largest of customer business environments. Microsoft SQL Server running within the Microsoft Windows Server environment introduces the ability to implement extremely scalable database environments, providing an industry-leading range of OLTP, data warehousing, and business intelligence solutions. The Microsoft Hyper-V server virtualization technology allows administrators to combine the scalability of both the Windows Server platform and associated Microsoft server products, including Microsoft SQL Server, to provide...
scalable business environments that provide both application scalability requirements as well as IT cost efficiencies. EMC VPLEX enhances the solution offerings for these dynamic and resilient customer environments, and provides full support for Windows Server Failover Clustering in single and geographically dispersed configurations. The utilization of EMC AccessAnywhere for the federated storage devices allows for fully active/active multi-site cluster configurations, enhancing the Hyper-V Cluster Shared Volume deployment model.

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. Local federation provides the transparent cooperation of physical storage elements within a site, while distributed federation extends the concept between two locations across distance. The distributed federation is enabled by a breakthrough technology available with VPLEX, AccessAnywhere, which enables a single copy of data to be shared, accessed, and relocated over distance.

The EMC VPLEX family consists of two offerings:

- **VPLEX Local**: This solution is appropriate for customers that would like federation of homogeneous or heterogeneous storage systems within a data center and for managing data mobility between physical data storage entities.
- **VPLEX Metro**: The solution is for customers that require concurrent access and data mobility across two locations separated by synchronous distances. The VPLEX Metro offering also includes the unique capability where a remote VPLEX Metro site can present LUNs without the need for physical storage for those LUNs at the remote site.

The EMC VPLEX family with the current architectural limits is shown in Figure 1.

---

**Figure 1. EMC VPLEX family offering**

The EMC VPLEX systems also maintain customer expectations for high-end storage in terms of availability. High-end availability is more than just redundancy; it means nondisruptive operations and upgrades, and being “always online.” EMC VPLEX provides:

- The AccessAnywhere implementation, which provides full connectivity of resources across clusters and Metro-Plex configurations
- Data mobility and migration options across heterogeneous storage arrays
- The power to maintain service levels and functionality as consolidation grows
- Simplified control for provisioning in complex environments
- Dynamic load balancing of data between storage array assets

Many of the new features provided by the new EMC VPLEX platform can reduce operational costs for customers deploying SQL Server or Windows Hyper-V solutions, as well as enhance functionality to enable greater benefits. This white paper details those features that provide significant benefits to customers utilizing Microsoft SQL Server and Windows Hyper-V.

**VPLEX Local**

An EMC VPLEX Local configuration is defined by up to four VPLEX Engines, which are integrated into a single cluster image through their fully redundant inter-engine fabric interconnections. VPLEX is designed to seamlessly grow from entry-level configurations, to very large, highly available cluster configurations.
As shown in Figure 2, 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 that 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 with a data center or across synchronous distances, enabling new models of high availability and workload relocation

**VPLEX Metro**

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, shown in Figure 3, can scale up through the addition of more engines, and scale out by connecting multiple clusters to form a VPLEX Metro configuration. In the initial release, a VPLEX Metro system supports up to two clusters, which can be in the same data center or at two different 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.
A VPLEX Cluster is composed of one, two, or four engines. The engine is responsible for federating the I/O stream, and connects to hosts and storage using Fibre Channel connections as the data transport. A small VPLEX Cluster consists of an engine with the following major components:

- Two directors, which run the GeoSynchrony software and connect to storage, hosts, and other directors in the cluster with Fibre Channel and gigabit Ethernet connections
- One Standby Power Supply, which provides backup power to sustain the engine through transient power loss
- Two management modules, which contain interfaces for remote management of a VPLEX Engine

Each cluster also consists of:

- A management server, which manages the cluster and provides an interface from a remote management station
- 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

The “VPLEX Metro” section on page 26 covers this topic in more detail.

**Storage device federation**

EMC VPLEX includes the ability to provide levels of storage federation as shown in Figure 4. Physical storage objects provided by storage arrays at the back end may be defined as encapsulated devices, whereby the underlying storage volume is effectively presented to the hosts configured at the front end as a pass-through storage device. This style of connectivity may be utilized in configurations where VPLEX is injected into an existing configuration. In these cases, data storage on the LUNs may be retained through the use of encapsulation. Even in such configurations, the abstraction layers will allow for local or remote mirrors to be attached or to execute data migrations to target devices.
When encapsulation of entire storage volumes is not required, VPLEX can be utilized to create multiple extents from associated storage volumes. These extents may then be combined into aggregate volumes to provide additional workload distribution and protection. VPLEX supports RAID 1, RAID 0, or RAID-C, which allow for mirrored, concatenated, or user-defined striping, respectively. As storage volumes are abstracted in this manner, virtual volumes created on such objects can be mirrored for the purposes of availability or migration.

**Connectivity recommendations**

Symmetrix VPLEX configurations provide a highly available connectivity environment, which allows for the creation of scalable and resilient virtualized environments. Both back-end (array) and front-end (host) connectivity must be configured in a redundant, highly available configuration. Such configurations remove single points of failure and ensure scalable configurations for the most demanding of customer environments.

**Back-end connectivity for storage**

Storage connectivity for array resources is provided by means of Fibre Channel connectivity through a series of ports across the two directors within any given VPLEX Cluster configuration. To protect against single points of failure for storage array resources, multiple discrete connections should be configured on separate Fibre Channel fabrics.

Figure 5 represents a connectivity methodology for providing highly available connectivity across VPLEX director ports. Storage array connectivity will be configured in multiple Fibre Channel switched fabrics, which are not represented in the overview. These redundant fabrics not only provide redundancy against single points of failure but also provide a scalable mechanism to provide connectivity to multiple storage arrays.
To cater for the potential aggregate workload against the provisioned storage resources, connectivity should be established in a scalable manner across back-end resources. The available connectivity ports for any given storage array will depend on the specific characteristics of the array; however, in all cases, sufficient port connectivity should be provided as would normally be necessary to cater for the host workloads.

**Front-end connectivity for hosts**

Figure 6 represents a logical view of a single VPLEX Engine and front-end connectivity to two physical Windows Server environments. The configuration implements a highly available and scalable design where Windows Server hosts are dual-pathed, and each path connects to two separate front-end modules on different directors. Not shown in this graphic is any SAN fabric that should also be configured in a highly available manner.
In instances where a VPLEX Cluster is configured, host connectivity should be configured in a highly available, redundant configuration where Windows Server hosts are connected to front-end ports provided by all engines within a cluster. This style of connectivity provides path redundancy for the specific software implementations on the given servers. Additionally, the multiple paths provide a scalable storage interconnect, allowing for the I/O demands of an active SQL Server instance, or the aggregate workload requirements of a Hyper-V deployment and the subsequent demands of the virtual machines and their application workloads.

It is recommended to configure at least two host bus adapters (HBA) per Windows Server host with the goal of presenting multiple unique paths to the VPLEX Cluster across the multiple directors within a cluster.

To provide the highest levels of availability, all single points of failure need to be addressed. While not a regular occurrence, it may be necessary to occasionally perform director maintenance. These procedures may require the removal of the director and its associated connectivity from the VPLEX system. As a result, each Windows Server host should have redundant paths to multiple front-end directors. Each Windows Server host should be connected to both directors within a single VPLEX Engine, and across all directors within a VPLEX Cluster, as applicable.

For each HBA port, at least one discrete front-end port should be configured. It is recommended that each HBA port be configured to two VPLEX front-end ports on the two directors for a given engine. This methodology for connectivity ensures all front-end directors and processors are utilized, providing maximum potential performance and load balancing for I/O-intensive SQL Server and Windows Hyper-V environments.

Configurations with multiple paths to storage LUNs will require a path management software solution on the Windows host. The recommended solution for multipathing software is EMC PowerPath®, which is the industry-leading path management software with benefits including:

- Enhanced path failover and failure recovery logic
- Improved I/O throughput based on advanced load-balancing algorithms and failover policies
- Ease of management including a Microsoft Management Console (MMC) GUI snap-in and CLI utilities to control all PowerPath features
- Value-added functionality including RAS data encryption technology
• Product maturity with proven reliability over years of development and use in the most demanding enterprise environments.

While PowerPath is recommended, an alternative is the use of the Multipath I/O (MPIO) capabilities native to the Windows operating system. The MPIO framework has been available for Windows for many years; however, it was not until the release of Windows Server 2008 where a generic device specific module (DSM) from Microsoft was included to manage Fibre Channel devices. For more information regarding the Windows MPIO DSM implementation, please see the “Multipath I/O Overview” at http://technet.microsoft.com/en-us/library/cc725907.aspx.

Provisioning storage with VPLEX

EMC VPLEX provides administrators with a simplified, flexible model for storage provisioning. This new storage provisioning model introduces a level of storage virtualization that provides the basis for a dynamic infrastructure.

Historically, administrators were required to provide somewhat static relationships between host visible storage devices and the underlying storage array, and mappings of those devices to front-end directors for host connectivity. They additionally would manage masking operations to ensure that hosts were able to access the requisite storage devices. This methodology served administrators well, and these were often only required to be made once.

Increasingly, administrators need to deal with a dynamic environment, where the introduction of new servers, virtual machines, and storage systems occurs on a regular basis. Data migrations from an existing storage array to a new array are often required for purposes such as technology refresh or lease rotation. These processes are often highly complex, disruptive and infer risk to the ongoing operations of the business applications.

EMC VPLEX provides extensive capabilities within the storage infrastructure to allow for seamless online data migrations within and across storage arrays, even where such capabilities are not a core feature of the storage array itself. Heterogeneous migrations across multiple storage arrays provide administrators not only with the capability to transition to new storage systems but also to dynamically balance workloads across all available resources within the storage infrastructure.

Deployments of clustered instances of both SQL Server database and Windows Hyper-V environments are much more commonplace, as customers consolidate applications and resources. To assist administrators with the ability to create flexible relationships between application resources and their placement within, or across data centers, VPLEX provides a flexible storage provisioning methodology.

Administrators are now able to define relationships between storage objects and host connectivity, and allow the VPLEX storage connectivity to implement the required connectivity. This ability to create logical relationships through views also helps to ensure that appropriate devices are automatically included in changes. For example, in a cluster configuration, only a single pool of storage devices needs to be defined for a given VPLEX Cluster. Views created based on this pool of devices ensures that any hosts (defined by their initiators) included in those views will be able to access the required devices. This is in contrast to many commonly implemented solutions where a manual process is implemented by administrators to ensure that mapping and masking entries have been created.

The following steps outline the requirements for implementing VPLEX storage provisioning functionality:

1. Claim storage presented to the VPLEX Cluster. This assumes that the relevant storage array has been connected to all required back-end ports for the VPLEX Cluster. Thus providing a scalable and highly available back-end storage infrastructure.

2. Define extents on the claimed storage devices. This allows for the allocation of some portion (or all of) the claimed storage device to be represented as an extent. Extents may subsequently be utilized in the definition of virtual volumes.

3. Define virtual volumes that will ultimately be presented to host servers.
4. Define the host view, which includes the WWNs of the HBAs that are used by the host, the VPLEX ports that will be used for providing storage connectivity, and all devices that are to be presented to the hosts.

It will be necessary to have appropriate zoning configurations in place within any fabric to allow the respective HBAs to connect to the director ports.

**VPLEX insertion into a SQL Server environment**

In many instances, customers will implement VPLEX environments within pre-existing environments. This implementation is referred to as VPLEX insertion. Migration of existing Microsoft application environments into a VPLEX environment can be quickly implemented with minimal downtime requirements. Migrations may come in multiple forms, whereby all user and system databases are migrated into a VPLEX environment, or implementations where only the user database locations are migrated. The latter case is discussed in the following section, as this is the most typical use case expected. Other migration scenarios are also possible, including utilizing host-based copy operations to transfer data from the source LUN to the target. In the tested scenario, the data remains resident on the original LUNs, but these devices are federated through a VPLEX.

To demonstrate the manner in which an existing SQL Server database instance may be moved into a VPLEX configuration, a sample environment was configured as shown in Figure 7. An existing SQL Server database on server LICOC211 utilizing three Symmetrix VMAX™ storage devices was to be moved into a VPLEX environment. This VPLEX insertion requires the remapping of the storage resources from the physical server into the target environment via the VPLEX environment.
In the tested configuration, the target environment was a four-node Windows Server 2008 Failover Cluster configuration. This was done as a means to demonstrate the mechanisms required to provision federated volumes to multiple Windows servers forming a single Windows Failover Cluster. Environments where the server environment is not intended to change are similarly supported, and would simply require changing storage device presentation in a similar manner to that described for the target cluster environment.

Additionally, while the example documents a Microsoft SQL Server movement, similar steps would be required for any given application environment; however the nature of the steps would differ between those for a SQL Server environment, which implements SQL Server-specific steps. Hyper-V environments, for example, may require that all virtual machine resources co-resident on a given LUN be taken offline as the storage device is placed under the control of VPLEX.

**Mapping SQL Server storage objects**

For many customer deployments of either Microsoft SQL Server or Windows Hyper-V environments, the environments will invariably require the migration of multiple LUNs. These LUNs represent the various storage areas of a database instances or virtual machine VHD storage or pass-through disks. Figure 8 details a given SQL Server database named “DBtoMigrate”, which is initially resident on Symmetrix.
VMAX devices. Storage utilized by this SQL Server database instance is comprised of three Windows NTFS volumes located on three Symmetrix VMAX LUNs.

**Figure 8. Display of an existing SQL Server database instance prior to migration**

The transition to VPLEX managed storage will require a short outage for applications utilizing the targeted storage devices as they are placed under VPLEX management and provisioned to the host servers. This is a one-time operation required to move the storage into the VPLEX environment. As a result, it will be necessary to place the database into an offline state, or to detach the database, as appropriate. In this example, the database will be migrated to new server hardware in a cluster configuration, and the database was detached after the mapping of database files and disk storage objects was obtained.

It will be necessary to ensure that all existing devices are appropriately identified for migration to ensure that the resulting environment results in a valid database instance. An alternative method to identify all the files (both data files and transaction logs) is to utilize the SQL Server stored procedure “sp_helpdb”. The usage and output of the command are shown in Figure 9, as executed within SQL Server Management Studio. It can be seen that all the data files, as represented by the “filename” value, are located within drives “K:” and “L:” and the transaction log is located on drive “M:”. These represent the storage devices that need to be appropriately provisioned to the target VPLEX Cluster.

**Figure 9. Using sp_helpdb to display all database components**

EMC Solutions Enabler command line utilities provide the ability to map Windows volumes to Windows disk objects and subsequently to storage array devices. Figure 10 demonstrates the mapping of Windows volume information to VPLEX target devices.
disk resources to Symmetrix devices using the SYMDEV command from EMC Solutions Enabler. In the displayed example Symmetrix device 01DA is seen by the server as Physical Drive 1.

![Symmetrix device list](image)

**Figure 10. SYMCLI example of mapping storage devices**

Once the devices have been identified, it will be necessary to execute relevant storage array operations to ensure that these devices can be mapped to the required back-end VPLEX ports. In the tested environment Symmetrix VMAX Auto-Provisioning Groups were utilized to present the storage devices to the VPLEX back-end ports after the database was detached, and the disk resources were placed in an offline state on the source Windows server.

**Claiming storage volumes with VPLEX**

After necessary changes within the storage environment, which facilitate presenting the relevant storage volumes to the VPLEX Cluster, it is necessary to claim the storage volumes. The claim process ensures that only relevant storage objects can be processed in subsequent steps. Claiming storage volumes is accomplished by executing the Claim Storage wizard within the VPLEX Web UI. Once the storage system is selected, and the claim procedure begins, a user-defined storage designation is entered as shown in Figure 11.

![VPLEX storage provisioning - Claim storage](image)

**Figure 11. VPLEX storage provisioning - Claim storage**

The VPLEX interface will subsequently step through a process of selecting appropriate available storage devices. For the purposes of identifying claimed storage volumes, VPLEX implements a mechanism to allow user-defined names to be applied to storage devices. In Figure 12, the claim wizard allows for
naming a storage tier. In this example, the value of “_SQL” is applied to identify these volumes as belonging to the SQL Server environment being processed.

Figure 12. User-defined name for a claimed storage tier

Once the tier has been defined, a subsequent dialog presents the available storage devices for the specific storage array, as shown in Figure 13. As three storage devices that were the storage LUNs for the SQL Server database environment were mapped to the VPLEX Cluster, these are seen as the available storage volumes. The applied name for the storage volumes is also displayed, and is constructed from the Symmetrix VMAX serial number (last four digits of the serial number), the user-defined tier, and the Symmetrix device identifier.

Figure 13. Display of unclaimed storage devices available

At the completion of the storage volume claim process, the three storage devices will be made available for subsequent processing. The three new storage devices can be seen in the storage volume listing for the VMAX storage array in Figure 14. This list includes previously processed storage volumes already processed.
Defining encapsulated devices

As the storage volumes already contain Windows defined volumes, and subsequently Windows NTFS volumes and the required data files and transaction log, the volumes will be configured as encapsulated. These encapsulated volumes will subsequently be surfaced to the required Windows host.

The first step in this process is to define volume extents on the claimed storage volumes, as shown in Figure 15.

The create extent wizard provides the opportunity to define the storage extents. The desire in this case is to simply pass these federated devices directly to the target Windows host, and not to create any segmentation of the underlying storage volume. In Figure 16 the previously selected devices have been added and will be used to define extents.
After the selection of the required devices, the user is presented with the ability to define the size of extents that may be smaller allocations of the entire storage device. Again, the desire in this case is to utilize the entire storage device, and pass through the existing Windows NTFS volume and the SQL Server database files contained within them, and as such, the full storage allocation is utilized as shown in Figure 17.

After the definition of the appropriate storage extents, the wizard will display the newly created extents. In Figure 18, the storage volumes now display a “used” status, and have no additional free storage available. Storage extents will have been created from the storage volumes, and are used for subsequent steps.
In Figure 19, the created storage extents are displayed. The created extents have an automatically generated name that includes the storage volume name with an “extent_” prefix. Storage extents within VPLEX may be defined in various configurations to provide advanced configurations for redundancy and reliability above and beyond those provided by the underlying storage array.

For the purpose of the tested environment a one-to-one relationship between storage volumes, storage extents created on those volumes, and the higher-level virtual volumes was defined. Figure 20 displays the selection of the three defined extents. Also of note is the selection of the Create a Virtual Volume on each device checkbox. This operation will define each extent as a separate fully encapsulated volume, which can be mapped to the appropriate server environment.
Configuring devices into a VPLEX view

After the definition of the encapsulated storage devices, there is a need to include the newly created virtual volumes into a host access view. In this instance an existing host view “PRDCluster” was defined. This host access view included the host initiators, associated VPLEX ports, and pre-existing virtual volumes, as shown in Figure 21. The addition of the new volumes only required the addition of these new volumes into the existing view.

Host access and registration

Access to the volumes will be possible from all hosts defined by inclusion of their initiators in the view. In this instance the hosts all form part of a single Windows Failover Cluster configuration, and would be able to access the defined virtual volume. Specifically because shared disk environments can exhibit this
immediate access, Windows Server environments implement a default disk policy of placing new storage devices into an Offline mode. This mode can be managed through either the Disk Management UI or through the DISKPART command line interface. In this instance, to validate the encapsulation of the storage devices and the import process, the disk devices were brought into an online mode as shown in Figure 22. The encapsulated volumes provided the same NTFS volume labels and data content as when they were presented directly to the original host.

**Figure 22. Windows disk view of new volumes placed into online mode**

However, since the final intent was to present the devices into a Hyper-V virtual machine, and protect the availability of the virtual machines and the SQL Server instance as a cluster resource, the disks were subsequently taken offline to be configured as pass-through resources for a virtualized SQL Server instance. The disk devices were then configured as resources for a Hyper-V virtual machine define within a Windows Failover Cluster configuration as a highly available virtual machine. This process is detailed in subsequent sections of this white paper.

**EMC VPLEX and Microsoft Windows Server Failover Clusters**

Microsoft Windows Server Failover Clustering provides customers with a solution to provide protection for business applications. Failover Clusters may scale up to 16 Windows Server nodes, and support even the largest of customers’ application environments. The clustering components are built on a foundation of shared storage resources. The design principle of Windows Failover Clustering requires that storage utilized by the application is accessible from any of the supported nodes at the time when the application service is started or resumed. Such mechanisms require robust, scalable storage solutions.

Windows Server 2008 Failover Clustering has a primary goal of maintaining availability of the virtual machine in those cases where the virtual machine becomes unavailable due to unforeseen failures. However, this protection does not always infer that the virtual machine state will be maintained through such transitions. As an example of this style of protection, consider the case of a physical node failure where one or more virtual machines were running. Windows Failover Clustering will detect that the virtual machines are not operational and that a node is no longer available and will attempt to restart the virtual machines on a remaining node within the cluster configuration. Such operations require a restart process and will cause any running applications to be completely restarted.

EMC VPLEX is a compliant storage federation offering for Microsoft Windows Failover Cluster configurations. Additionally, as the storage resources are federated by the VPLEX Clusters, it is possible to implement solutions for which the underlying storage may not have been compliant. For example, Windows Server 2008 Failover Clustering requires storage systems to support SCSI-3 Persistent Group Reservations (PGR) utilized to implement device arbitration as a component of the high-availability design. Storage systems that may not inherently possess this functionality can be supported by VPLEX in a Windows 2008 Failover Cluster environment. All SCSI-3 compliant mechanisms are maintained by the VPLEX environment through the virtual volumes defined. The storage system is required only to deal with the host and application I/O workload rather than provide volume arbitration.

The federation solution provided by EMC VPLEX fully supports and meets all requirements of Windows Server 2008 Failover Clustering. Indeed, as VPLEX solutions are designed with redundant and scalable
connectivity requirements for the most aggressive of application environments, they are a natural extension of the highly available design provided by Windows Failover Clustering.

**Windows Failover Clustering and Windows Hyper-V**

Microsoft Windows Server Hyper-V deployments are supported under Windows Server Failover Clusters to provide extremely robust and highly available application solutions. Windows Hyper-V utilizes features of the Windows Failover Clustering environment to enhance and extend certain availability functions, and form the basis for solutions such as Hyper-V Live Migrations and Cluster Shared Volumes (CSV).

A Windows Hyper-V virtual machine instance imported into a high-availability configuration will need to include all related storage disk devices utilized by the virtual machine itself and any applications within the VM, such that the virtual machine can be managed correctly. The ability to import a virtual machine into a Windows Cluster is enabled by running the conversion wizard provided through the Windows Failover Cluster Manager user interface. The high availability wizard will fail if it is unable to include all storage configured for the virtual machine within the cluster environment. EMC VPLEX, through its storage view functionality, significantly simplifies this process and improves reliability. Subsequent addition of disk storage devices will require that the new storage is also configured appropriately as shared storage within the cluster.

Windows Hyper-V virtual machines are able to access storage devices in a number of ways. Most commonly, the storage will be provisioned as virtual hard drives (VHD) on the parent partition, and be assigned to the virtual machine. Utilizing this method, the virtual machine will see the storage as locally attached. This form is typically used for the initial operating system area. Storage may also be presented through the parent system to the virtual machine directly as SCSI targets where those devices are configured as pass-through storage. The final form of storage for virtual machine use is through the implementation of iSCSI-connected storage to the virtual machine via its network infrastructure.

In this example, the storage devices were provisioned both as VHD storage for virtual machine use, and as pass-through storage. The storage utilized by the SQL Server database instance that was migrated to the VPLEX environment was defined as pass-through devices.

**Windows Hyper-V Live Migration**

Movement of virtual machines within a cluster, that is, a proactive move request either by an administrator, or some automated management tool, can take advantage of Windows Hyper-V Live Migration capabilities to mitigate any loss of application available. These proactive requests allow Failover Clustering mechanisms to invoke those processes available to coordinate and protect the state of the virtual machine.

When a Live Migration is executed, Failover Clustering will begin a process to replicate the virtual machine configuration and memory state to the target node of the migration. Multiple cycles of replicating the memory state will begin to take place, in an effort to reduce the amount of changes that need to be replicated on subsequent cycles of memory replication. The execution of this memory replication process can be seen through the Failover Cluster manager console. Given the ability of the network connectivity to allow for the timely transfer of state, the migration process will as a final phase, momentarily suspend the machine instance, and switch all disk resources to the target node. After this process, the virtual machine will immediately resume processing. The transition of the virtual machine is required to complete within a TCP/IP timeout interval such that no loss of connectivity is experienced by client applications.

**Configuring VPLEX storage as Hyper-V pass-through**

In the example outlined in the section “VPLEX insertion into a SQL Server environment,” the database devices were comprised of three storage volumes. These volumes were presented to a VPLEX Cluster, and subsequently configured as encapsulated virtual volumes and presented to the nodes within the target cluster.

In a Windows Failover Cluster configuration supporting Hyper-V virtual machines, it is recommended to not deploy applications such as Microsoft SQL Server in the parent partitions. Therefore, a Hyper-V virtual
machine was defined to implement the SQL Server database environment. This also necessitated the need to configure the storage devices as pass-through storage to the virtual machine.

**Figure 23. Adding storage devices as cluster resources**

In Figure 23, the Add a disk wizard is selected to begin the process of allocating the VPLEX supported devices into the Windows Failover Cluster configuration. The wizard implements a validation check to ensure that only storage devices shared among all member nodes are available to add to the configuration. As previously mentioned, the use of the VPLEX storage volume view mechanism ensures that all nodes are assigned the correct volumes. The three volumes are seen in Figure 24.

**Figure 24. Shared storage devices available to all cluster nodes**

After selecting all applicable storage devices, which in this example require the three volumes utilized for the SQL Server database files, the storage devices are added to the clustered disk resource list. These devices are then required to be assigned to the virtual machine that will run the SQL Server database instance. In Figure 25 the storage devices are assigned to virtual machine “CSV1SRV1”. This assignment ensures that the storage devices are tied to the virtual machine, and in cases where the virtual machine is moved to run on a different physical node within the cluster, the disk storage will be accessible.
In addition to adding the disk resources as clustered resources, the definition of the virtual machine must be altered to include the disk resources as locally attached storage devices in the form of pass-through storage devices. In Figure 26 the three storage devices are defined as pass-through storage by assigning them to a SCSI controller as **Physical hard disk** resources. This process makes the disk storage directly accessible to the virtual machine.
As the original storage devices contained valid NTFS volumes, and the SQL Server database files within the volumes, as validated when the devices were accessed by the cluster nodes, the pass-through storage devices will also display the same attributes. The view of the accessible storage devices from virtual machine CSV1SRV1 is shown in Figure 27. In the view the original storage volume labels can be seen, and normal disk management operations are possible. For example, the disk drive letter assignments were altered for the NTFS volumes within the virtual machine to match those that were originally assigned on the source system.

![Figure 27. Disk management view from the virtual machine](image)

With the storage volumes accessible, and the appropriate SQL Server software installed in the virtual machine, it is possible to mount the SQL Server database instance represented by the files located within the volumes. For the purpose of the tested environment, the stored procedure `sp_attach_db` was utilized, and its execution is shown in Figure 28.

![Figure 28. Attaching a SQL Server database environment](image)
The resulting SQL Server database instance passed all tested scenarios, including DBCC CHECKDB executions that validated all data pages and structures within the database files. The resulting configuration represented a valid highly available solution for a SQL Server environment.

**VPLEX Metro and Cluster Shared Volumes**

Extending on the availability and flexibility provided by VPLEX Local, customers are able to deploy solutions across multiple sites utilizing the VPLEX Metro configuration. This multi-site configuration provides a unique offering in Windows Failover Clustering solutions by supporting robust active/active configurations.

Traditional block storage replication solutions typically only support access to the storage devices on a source site. The constantly changing nature of the data blocks on the target site rarely represented a viable state for any application to utilized. As of the introduction of Windows Server 2008 R2, Microsoft introduced support for functionality referred to as Cluster Shared Volumes, or CSVs. Specifically designed for deployments of Hyper-V highly available solutions, the CSV environment provided the ability for all member nodes within a cluster configuration to maintain direct access to a common namespace representing these specially shared volumes. The namespace is typically of the form “C:\SharedStorage\” appended with a signifier of the specific volume for each CSV.

The CSV solution supports configurations where virtual machine hard disks, or VHDs, could be co-located on the common storage devices. A coordinator node is elected from the member nodes of the cluster and is responsible for ensuring that any given VHD file is only accessed from a single node, specifically the node that is currently supporting the virtual machine. To be clear, the shared volume is fully accessible from all member nodes, and it is only the VHDs that are locked for access to the node that runs the virtual machine. Thus it is possible to consolidate large numbers of VHDs into a smaller number of CSVs, and have the virtual machines distributed across all member nodes accessing respective VHDs from a single storage volume. Each member node generates I/O operations to the locally accessed storage device through its local HBAs.

The implementation of CSVs within a Windows Failover Cluster provides the ability to construct very large deployments of virtual machines, while limiting the management complexity of having large numbers of discrete storage volumes. Additionally, by having direct local access to the storage devices from all member nodes, in the event that a virtual machine is moved or migrated from one member node to another, no disk arbitration is required. This improves failover or Live Migration times considerably, and mitigates issues associated with disk registration.

CSVs are supported in geographically dispersed configurations, but this solution does not cater for asymmetric access to the shared storage as provided by typical block storage devices replicated between systems. CSV deployments assume fully accessible local storage devices irrespective of the nature of differing sites.

VPLEX Metro provides support for active/active storage configurations in multi-site deployments through the ability of defining distributed storage volumes using AccessAnywhere technology. Distributed storage volumes are configured as fully accessible devices to all VPLEX Clusters and therefore to all attached server resources. Utilizing an advanced cache coherency mechanism, VPLEX Metro provides full support for multi-site configurations of Windows Failover Clustering utilizing CSVs with active/active access for all member nodes.

Furthermore, VPLEX Metro configurations support the use of heterogeneous storage arrays at the different site locations. This provides customers with the ability to select the most appropriate storage array at each site, and rely on the VPLEX Clusters to facilitate the replication. Traditionally, the replication was tied to functionality specific to the storage array.

**Defining a geographically dispersed Windows Failover Cluster**

A single Windows Failover Cluster that spans multiple geographically separated sites provides customers with the ability to provide exceptionally high levels of application available and disaster tolerance. Microsoft Windows Clustering can support such configurations but requires that shared storage defined...
within the environment is effectively able to replicate data between sites, and coordinate storage device access as necessary.

EMC VPLEX Metro configurations provide the infrastructure to support distributed Windows Failover Cluster configurations and additionally provide support for active/active shared storage devices. This latter feature allows geographically dispersed cluster support for solutions such as Windows Hyper-V and CSVs.

The tested environment was initially defined to be using CSVs for 20 virtual machines on a VPLEX Local configuration. Initially, four CSV devices were configured for operating system VHD locations and four CSVs were configured for VHDs associated to applications utilized by the virtual machines within the cluster. The operating system VHDs each supported five virtual machine VHDs, for five discrete virtual machines. The virtual machine names, as they appear within the Failover Cluster management interface, were defined by the CSV volume being used to store the VHD (CSV1 thru CSV4), and then a unique identifier for the specific virtual machine instance on the CSV (SRV1 thru SRV5). Thus, it is trivial to identify CSV1SRV1 as being the first server instance that has its VHD file located on the first CSV storage volume.

The initial VPLEX Local configuration was enhanced by adding a remote VPLEX Cluster environment, and thus forming a VPLEX Metro configuration. This targeted environment is shown in Figure 29. Inter-site connectivity is defined in a redundant, scalable manner, similar to the design of both front-end and back-end connectivity as previously discussed.

Figure 29. Geo-cluster configuration using VPLEX Metro
In the defined configuration, the distributed devices were supported by local storage provisioned to each VPLEX Cluster environment. The original storage devices were located on the Symmetrix VMAX array. The data contained on these storage volumes was replicated by the VPLEX distributed volume functionality to storage provisioned from the CLARiiON® CX4 array.

Creating remote storage devices

To provide local storage resources, storage devices were provisioned from the CX4 array to match the source storage volumes. As there were eight storage volumes defined as CSV storage devices, eight matching devices were created.

The CX4 storage volumes are detailed in Figure 30, and were used to create volume extents that match the configuration on the source VMAX configuration.

![Figure 30. CLARiiON provisioned storage volumes](image)

As the storage devices in this instance were also being created as encapsulated devices, where the extent was only created from a single storage volume, the size of the defined extents was defined to match the originating extents. In Figure 31, the extents match the original device size of 250 GB each.

![Figure 31. Creation of storage extents](image)
Once the target extents were created, and the VPLEX Metro connectivity was implemented, the original storage volumes could be implemented as distributed volumes providing local and remote protection in the form of distributed mirrored (RAID 1) devices.

**Defining distributed storage devices**

The creation of distributed storage devices requires that any existing virtual volume on the device to be implemented as a distributed device be removed. This is required to implement the definition of a new device that can be replicated. This process therefore requires that access to the existing virtual volume be terminated to allow for the creation of the distributed storage device.

Due to the implementation of VPLEX storage views, the process is significantly simplified by allowing for virtual volumes to be removed from a view, reimplemented as distributed storage, and reintroduced to the view. Data contents of the existing virtual volumes are retained through the process, and are automatically mirrored to the remote storage device.

As the existing devices were already being utilized within the “PRDCluster” environment, it was necessary to suspend or shut down all virtual machines on a given CSV device, as the storage device was to be reconstructed. Once all virtual machines for the CSV device were suspended or shut down, the volume was placed in a Maintenance Mode state through Failover Cluster Manager. This process effectively places the storage device in an offline state, and suspends all health checks for the device. This allows the device to be manipulated as necessary, and allows for minimal changes to the environment to facilitate the required changes.

In Figure 32 the process of defining a new distributed device is begun. This process guides the definition of the new distributed device.

**Figure 32. Creating a new distributed device**

Prior to the steps outlined here, the existing virtual volume had been removed, leaving the underlying device. In this example the “Dev_MS_VOL_1” device was the original device used by the virtual volume representing the first CSV device.

To reimplement the local device as a member of a new distributed device, it is added as a member of the distributed device, as shown in Figure 33. The distributed device will have a remotely mirrored device represented by “device_CX4_377_LUN1_1” on the remote CX4 storage array.
The newly created distributed storage device will be called “Windows_CSV_1” and will automatically have a virtual volume defined for it.

Once the definition of the distributed storage device is completed, it is possible to display the device, and obtain details on the members of the distributed device. The display in Figure 34 shows the details of the created device, including component extents in the respective clusters comprising the VPLEX Metro environment.

After the creation of the new distributed device, the new virtual volume will need to be added to the appropriate storage view. To begin the process, the view is selected through the VPLEX UI, and the Add/Remove Virtual Volumes option is selected, as shown in Figure 35.
Figure 35. Adding a distributed device to the storage view

In Figure 36 the device is added back to the “PRDCluster” view, which will result in the ability of all cluster nodes to regain access to the storage device.

Figure 36. Adding the new virtual volume to the storage view

To complete the process for the Windows Failover Cluster configuration, it would simply be necessary to take the storage device out of Online Maintenance Mode, and then restart or resume the virtual machines. The transformation of the storage into a distributed storage device was complete.

Enabling remote node access

The implementation of the devices into distributed storage devices is a critical component of defining a geographically dispersed cluster supported by VPLEX Metro. To complete the configuration, it is necessary to implement a storage view on the remote VPLEX Cluster so as to provide access to the distributed storage devices. This process is identical in nature to the definition of the initial storage view, and would only differ with respect to the unique components of the site (host initiator addresses, VPLEX Cluster ports, and so on).

As the distributed storage devices are identified as the same storage devices in local and remote sites, no further changes are required to the Windows Failover Cluster configuration. In fact, the cluster configuration is entirely unaware that the storage devices are distributed. All member nodes will treat the storage devices as they would any local storage device. All Windows Failover Cluster mechanisms will apply to the distributed devices in the same manner as they did to local-only storage.
Additional VPLEX Metro cluster features

In addition to previously outlined functionality provided by VPLEX Metro configurations, a number of additional features provide value add in Windows Server Failover Cluster environments. Beyond providing support for distributed storage devices, VPLEX Metro provides support for a configuration whereby storage devices available in one site can be made accessible to hosts located in a remote site. Unlike the previously discussed distributed storage devices, exported volumes require no remote storage allocations.

Exported volumes

Customers can often find the need to provide remote access to data located on storage in a single site. Often this may come as a need to provide reporting capabilities or to extract data to remote systems. Invariably there is limited need to retain this data in the remote site, and therefore little to no need or desire to provision storage for it. In such cases, exported volumes represent a unique advantage provided by VPLEX Metro configurations.

In the tested scenario for the SQL Server device encapsulation, the storage devices were only presented from the underlying Symmetrix VMAX storage array. Additionally, these devices were not defined as Distributed Storage Devices, so they remain accessible only to the local nodes within the site where the Symmetrix VMAX system is located.

In Figure 37 the local encapsulated devices are being defined as exported volumes. This will effectively provide the remote VPLEX Cluster, which is located in the remote site, to provide access to the exported volumes as if they were local devices. All I/O operations will be serviced through the VPLEX Cluster interconnect between the two VPLEX Clusters.

![Figure 37. Defining virtual volumes as exported devices](image)

This step will confirm that the selected volumes will be exported to the remote VPLEX Cluster as shown in Figure 38. These volumes will then appear within the remote VPLEX Cluster, and become available as storage resources that can be included into storage views.
Once the exported volumes are accessible at the remote VPLEX Cluster, it will be necessary to include these new volumes within the view managing access to the nodes. For the “remote” nodes defined within the configuration, the view managing access was “RMTCluster”. In Figure 39 we begin the process of adding the exported volumes to allow the nodes within the site to gain access.

As the newly exported volumes are the only volumes that have not been defined in the view, and as we need to provide them to all nodes within the cluster, in Figure 40 the three exported volumes are all added to the view.
After the completion of this process, it will be necessary to utilize Windows Disk Management to rescan for the new devices on each node, or alternatively restart the servers. The newly presented devices will become accessible. Windows Failover Clustering will automatically recognize the newly added devices as being cluster disk resources.

Full Windows Failover Clustering functionality will be available to the exported volumes, as they will appear to be fully accessible as local disk resources. This will allow for movement of Cluster Resource Groups that contain these storage objects between the sites, including Hyper-V Live Migrations, if these volumes are utilized by virtual machines.

Clearly, the accessibility of exported volumes is dependent upon the availability of the VPLEX Cluster and the storage that is maintaining the data. In the event that the owning site becomes inaccessible, the exported volumes will also become inaccessible. This differs from distributed volumes that are defined to have local mirror copies.

Use of exported volumes within a Windows Failover Cluster and for Windows servers in general is fully supported. This functionality does provide a viable solution in scenarios where there is a desire not to provide a complete duplicate copy in the remote site, but it does limit high availability to conditions where full access to the owning cluster is available.

Conclusion

The EMC VPLEX Local and VPLEX Metro solutions implement a new strategy in storage federation solutions for applications such as Microsoft SQL Server and Microsoft Hyper-V. They provide significant value to customers by introducing support for dynamic management of storage devices.

Advantages include:

- Support for storage federation across a number of heterogeneous storage systems, allowing administrators to distribute applications across multiple storage devices, and dynamically migrate them.
- Support for heterogeneous storage arrays in a VPLEX Metro system
- Support for exported volumes providing remote site access with zero local storage requirements
- Full support for Windows Failover Clustering
- Enhanced support for multi-site clustering in active/active configurations. Providing advanced solutions for Hyper-V and Cluster Shared Volumes

These new technologies provide an easier and more reliable way to provision storage in Microsoft Windows Server, SQL Server, and Hyper-V environments, while enabling scalable, flexible data mobility between storage tiers across storage systems and across sites.