White Paper

EMC VPLEX with Quantum Stornext

Application Enabled Collaboration

Abstract

The EMC VPLEX storage federation solution together with Quantum StorNext file system enables a stretched cluster solution where hosts has simultaneous read and write access to a file system in a heterogeneous environment.

September 2011

EMC²

Part Number h8922

Table of Contents

Executive summary

EMC® VPLEX™ is a federation solution that can be stretched across two geographically dispersed data centers separated by synchronous distances (maximum round trip latency = 5 ms). VPLEX provides simultaneous access to storage devices at two sites through creation of VPLEX distributed virtual volumes, supported on each side by a VPLEX Cluster. Each VPLEX Cluster is itself highly available, scaling from two directors per VPLEX Cluster up to eight directors per VPLEX Cluster. Furthermore, each director is supported by independent power supplies, fans, and interconnects. Each VPLEX Cluster has no single point of failure.

The EMC VPLEX storage federation solution together with Quantum StorNext file system enables a stretched cluster solution where hosts has simultaneous read and write access to a file system in a heterogeneous environment. This architecture provides new options for the following.

- Disaster recovery and disaster avoidance.
- Supporting organizations in remote offices.
- Deploying private or public cloud computing environments that can span multiple data centers.

While the metadata controllers (MDC) are an active-passive clustering provided by StorNext with robust failover and management capabilities. The SAN clients are active independent of the site location. The MDC workload is failed over seamlessly to the secondary site in case of a primary site failure, thus providing DR capability in the event of site failures with the help of VPLEX. The StorNext MDC cluster is a two-node system with a dedicated private network for cluster data traffic. The SAN clients directly access the storage for read and write and the metadata is managed by the metadata controllers.

Key value propositions

SAN clients can concurrently read and write to a common block storage volume across locations, supporting remote offices and collaboration by making shared data available locally on each site. The VPLEX solution provides immediate benefits such as increased resiliency for unplanned outages and centralized storage management with the ability to dynamically scale out and load balance your applications, servers, and other business needs between local and geographically dispersed sites.

For this white paper we have demonstrated key functionality in a lab environment using VPLEX Metro, which supports geographic clusters up to 5ms apart. VPLEX Metro is generally available and SNFS 4.1/4.1.1 is fully qualified on this platform.

VPLEX overview

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 collected in single data center or multiple data centers that are geographically separated by metropolitan area network (MAN) distances. EMC VPLEX Metro provides non-disruptive, heterogeneous data movement and volume management functionality within synchronous distances both within and between data centers. With a unique scale-up and scale-out architecture, advanced data caching, and distributed cache coherency, VPLEX provides workload resiliency, automatic sharing, balancing, failover of storage domains, enables both local and remote data access with predictable service levels.
EMC Access Anywhere, available with VPLEX, is a break-through technology from EMC that enables a set of data to be shared accessed and relocated over distance.EMC GeoSynchrony™ is the VPLEX operating system.

VPLEX Local™

VPLEX Local provides simplified management and non-disruptive data mobility across heterogeneous arrays.

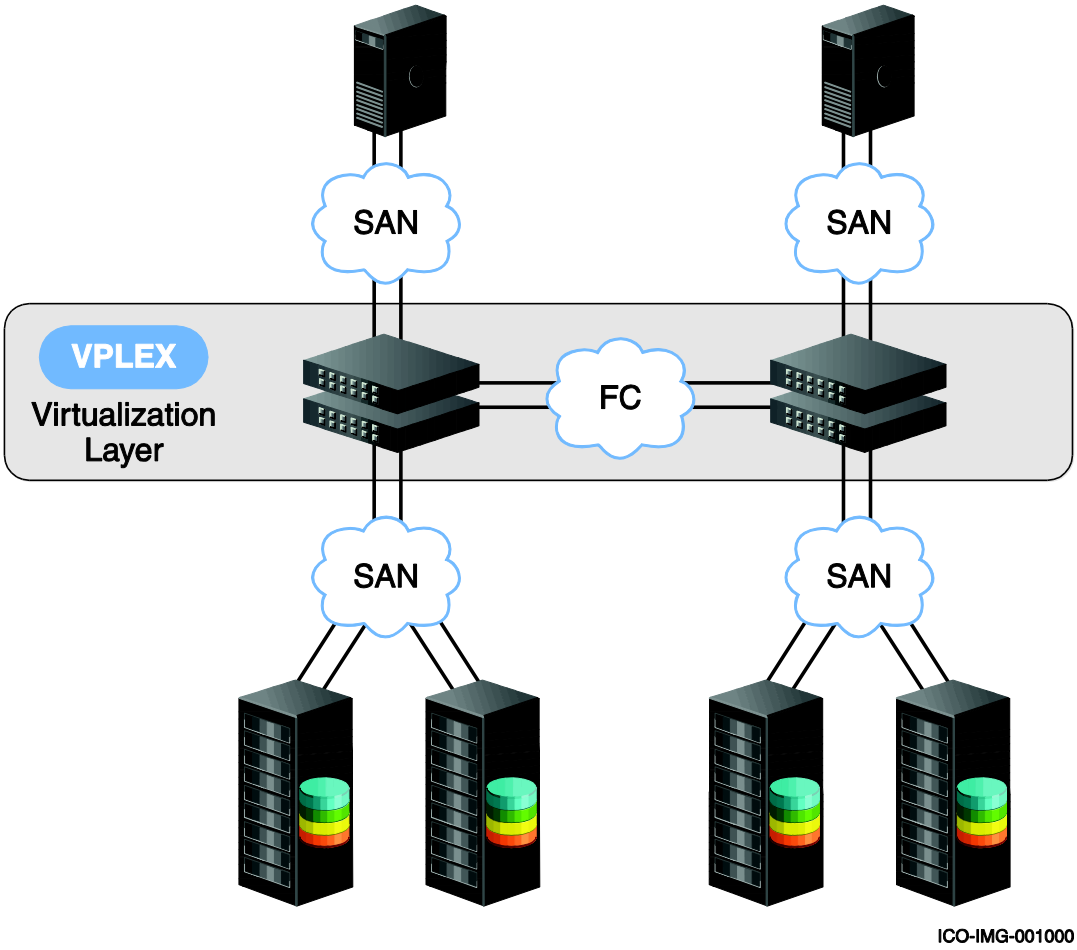
VPLEX Metro™

VPLEX Metro provides data access and mobility between two VPLEX Clusters within synchronous distances. VPLEX Metro supports two forms of devices:

1. Metro-distributed virtual volumes and remote virtual volumes.

   Metro-distributed virtual volumes provide synchronized copies (mirrors) of the volume's data in each cluster. The mirrored volume appears and behaves as a single volume and acts in a similar manner to a virtual volume whose data resides in one cluster. Remote virtual volumes, metro-distributed volumes are able to take advantage of VPLEX distributed coherent cache and its prefetch algorithms to provide better performance than a SAN extension solution.

   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 next, can scale up through the addition of more engines and scale out by connecting multiple clusters to form a VPLEX Metro configuration. The initial release a VPLEX Metro system supports up to two clusters, which can be in the same data centers and optimize resource utilization across data centers. In addition, VPLEX clusters provide non-disruptive data mobility, heterogeneous storage management and improved application availability.



2. Application availability during data center outages.

   A VPLEX Metro distributed between two data centers can be used to protect against data unavailability in the presence of a data center outage by mirroring data between two data centers. Data access will remain available for metro-distributed virtual volumes whose preferred cluster (static bias) is in the surviving data center. For those volumes whose preferred cluster is in the data center with the outage, access to the data for these volumes can be resumed at the surviving cluster with invocation of a manual command to resume the suspended I/O. When combined with failover logic for host clusters or active-passive environments using SNFS, this provides infrastructure that is able to restore service operations quickly, even in the presence of an unplanned data center outage.

VPLEX witness

VPLEX Metro and VPLEX Geo rely on a new component called VPLEX Witness. The VPLEX Witness is an optional component to be deployed in customer environments where the current "static bias" solution described above is unacceptable.

The VPLEX Witness is installed in a VM operating on a customer's host deployed in failure domain separate from either of the VPLEX clusters (to eliminate the possibility of a single fault affecting both a cluster and the VPLEX Witness). VPLEX Witness has IP-based connectivity with both distributed clusters in VPLEX Metro or Geo, and has knowledge about each cluster's view of intercluster connectivity. By reconciling its own observations with the information reported by the clusters, the VPLEX Witness improves the ability to distinguish between intercluster network partition failures and cluster failures.

Stornext overview

StorNext is data management software that enables customers to complete projects faster and confidently store more data at a lower cost. Used in the world's most demanding environments, StorNext is the standard for high performance shared workflow operations and multitier archives. With StorNext data management software, you get high-speed content sharing combined with cost-effective data archiving. It's all about helping you build an infrastructure that consolidates your resources, so workflow runs faster and operations cost less. StorNext offers data sharing and retention in a single solution, so you do not have to piece together multiple products that may not integrate well. Even in heterogeneous environments, all data is easily accessible to all hosts.

StorNext File System streamlines processes and facilitates faster job completion by enabling multiple business applications to work from a single, consolidated data set. Using SNFS, applications running on different operating systems (Windows, Linux, Solaris, HP-UX, AIX, and Mac OS X) can simultaneously access and modify files on a common, high-speed SAN storage pool. This centralized storage solution eliminates slow LAN-based file transfers between workstations and dramatically reduces delays caused by single-server failures. The StorNext FS configuration that we have tested consists of two main components, briefly described next.

1. Metadata Controller (MDC)
   The server on which the StorNext Storage Manager software is running (the metadata controller host.) Also known as the local host, or the primary server on HA systems.
2. SAN clients
   The server/desktop on which StorNext client software is running, they directly access the storage for read and write functions.

In high availability (HA) MDC configurations, which is an active-passive configuration, a redundant server is available to access files and pick up processing requirements of a failed system, and carry on processing.

The primary advantage of an HA system is file system availability, because an HA configuration has redundant servers. During operation, if one server fails, failover occurs automatically and operations are resumed on its peer server.

At any point in time, only one of the two servers is allowed to control and update StorNext metadata and databases. The HA feature enforces this rule by monitoring for conditions that might allow conflicts of control that could lead to data corruption.

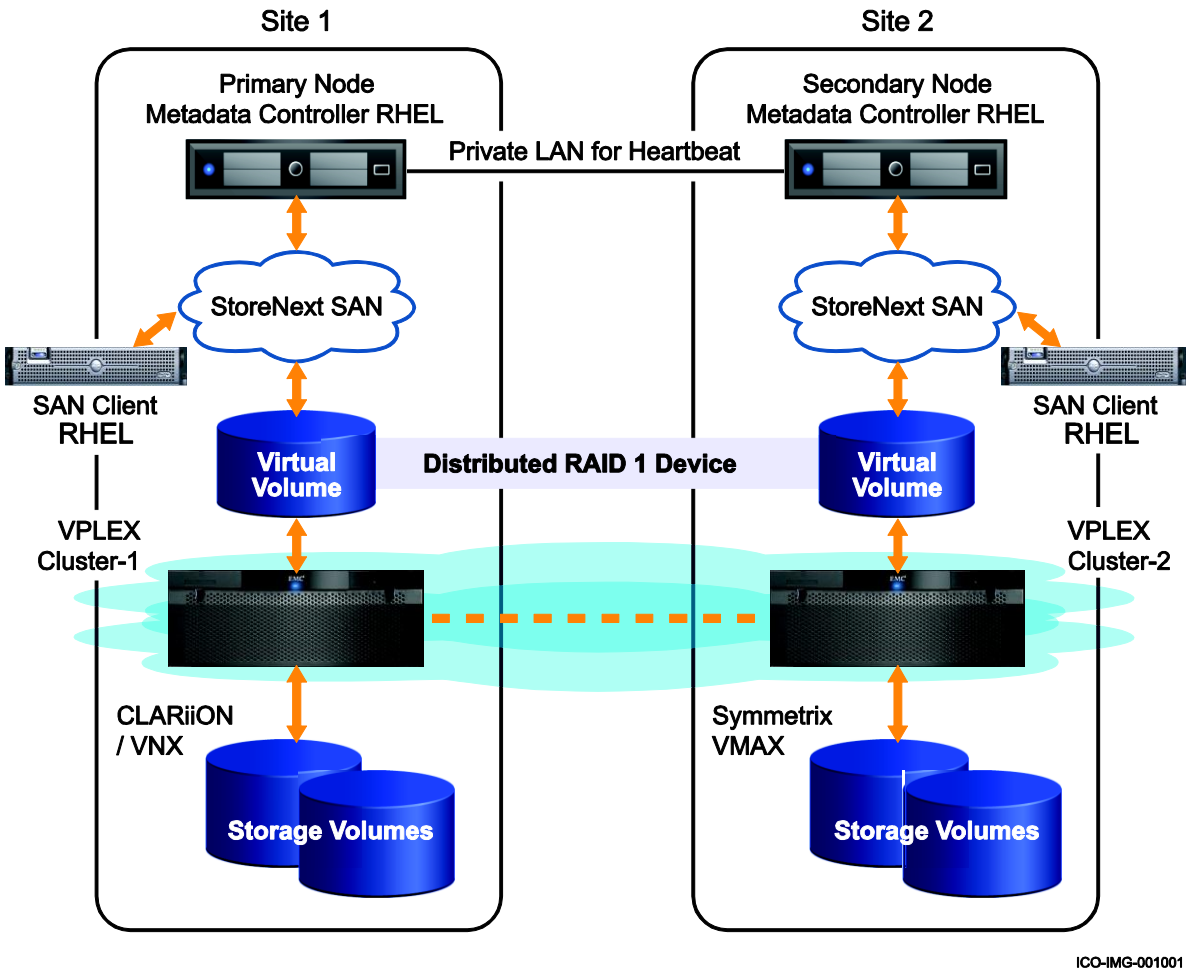
Before this so-called *Split Brain Scenario* would occur, the failing server is reset at the hardware level, which causes it to immediately relinquish all control. The redundant server is able to take control without any risk of split-brain data corruption. The HA feature provides this protection without requiring special hardware, and HA resets occur only when necessary according to HA protection rules.

Technology integration

It has been proven over time that the joint solution by EMC's and StorNext is extremely effective at supporting large infrastructures. SNFS is designed to fully utilize the capabilities of the EMC CLARiiON®, Symmetrix®, and VPLEX and it supports key software applications such as PowerPath®.

Lab configuration

The demonstration configuration includes two VPLEX locations named site1 and site2, each with CLARiiON and EMC VNX™ on site1 and Symmetrix at site2. VPLEX presents shared luns to the metadata controller as well as to the SAN clients. Metadata controllers service the metadata and journal part and SAN clients have the luxury to write directly to the storage allocated by the metadata controller. All the SAN clients have simultaneous read and write access to the file system.



Site -1 consists of:

- An RHEL MDC – Primary
- An RHEL SAN Client
- Storage – CLARiiON and VNX

Site -2 consists of:

- An RHEL MDC – Primary
- An RHEL SAN Client
- Storage – VMAX

5ms latency was tested by placing ANUE in between the two sites.

Configuration

SNFS 4.1.1/SNFS 4.1

MDC (both primary and secondary) OS - RHEL 5.5 Power path - Version 5.3 SP1 OS Kernel - 2.6.18-194.el5	Client 's OS - RHEL 5.5 Power path - Version 5.3 SP1 OS Kernel - 2.6.18-194.el5	VPLEX Config- Metro/Local Version - 5.0.0.00.00.22	Switches - Used in test bed. Brocade -300B- Fabric OS - v6.4.0b Brocade 24k- Fabric OS- 5.3.1b	Array - CX4-480-FLARE® - 4.30.0.5.506 VNX 5500- 05.31.000.3.545 VMAX - 5785.135.91

The lab configuration was used to check following functionalities.

- Creating SNFS file system over VPLEX-based federated storage in a RHEL-based environment.
  - This test makes sure that we can create file system when the devices are presented by the VPLEX to the hosts.
- File system expansion and basic I/O tests.
  - To test FS can be expanded over time whenever there is a need for space, and also make sure different file sizes with different record length can be created.
- Compatibility with PowerPath.
  - To test the path failover features at host level.
- Physicals devices removal and restoration.
  - To test how StorNext behaves when devices are not available and that data is consistent when the devices are brought back.
- SNFS failover capabilities
  - To test the various HA failover scenarios like heart beat failure, manual failover, device unavailability, and network down.
- VPLEX Witness failure
  - To test the robustness when witness server is down.
- WAN link failure
  - To test the failover of VPLEX /SNFS and its recovery.
- Latency between the two sites is less than or equal to 5ms latency RTT (Metro)
- Inducing the shutdown of one VPLEX site down and rebuilding of data.
  - To test how robust the VPLEX/SNFS will be when one complete site is powered down. Also to test the ability for VPLEX/SNS to provide recovery and data consistency.
- Array power down shut down.
- NDU and online upgrades for the storage arrays.
- VPLEX online (non-disruptive) code upgrades.

- VPLEX cluster shutdown and restart.
- VPN in between the two VPLEX sites failure and recovery.

Conclusion

VPLEX storage federation technology provides new capability to distribute workloads across data centers improve collaboration and efficiently provide DR protection using active-active rather than active-passive secondary data center architecture. A clustered file system enables customers to fully utilize this capability providing concurrent shared read-write access to shared data in two or more geographic locations.

SNFS clustered file system easily integrate with VPLEX , EMC storage and EMC PowerPath with robust functionalities and easy management.

The joint solution provides us with:

- High degree of HA at storage level with the help of VPLEX and host level with the help of SNFS.
- Provides accessibility to users over the distance who may share the file system for variety of needs.
- Provides highly scalable file system.
- Provides a clustered file system that will leverage a federated volume across distance.

References

http://www.quantum.com/Products/Software/Index.aspx