Abstract

This white paper examines deployment and integration of Microsoft® SQL Server® 2012 AlwaysOn Availability Group solutions on EMC® VPLEX™ storage federation systems. Details of how to deploy with VPLEX systems are documented with practical examples for SAN, WAN, and SQL database administrators.

March 2013
Copyright © 2012 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is.” EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

All other trademarks used herein are the property of their respective owners.

Part Number H11622
# Table of Contents

- Executive summary 4
- Introduction 4
- Audience 4
- VPLEX technology overview 5
- VPLEX product offerings 7
- VPLEX Local, VPLEX Metro, VPLEX Metro-IP 7
- Overview of VPLEX features and benefits 8
- Overview of SQL Server High-Availability Solutions 10
- Overview of AlwaysOn Availability Groups 11
- Benefits of using Failover Clustering with VPLEX 12
- Benefits for using AlwaysOn Availability Groups with VPLEX 13
- WSFC Section One: Standard 2-Node Cluster 14
- WSFC Section Two: Standard 4-Node Cluster with AlwaysOn 15
- WSFC Section Three: Standard 4-Node Cluster with AAG and Hyper-V 16
- Implementing WSFC Stretched Clusters using VPLEX 17
- Provisioning Storage for Stretched Clusters on VPLEX 17
- Creating Consistency Groups and Applying Detach Rules 18
- VPLEX Storage Views 19
- Adding Roles and Features for VPLEX Stretched Cluster 20
- Building your Windows Server Failover Cluster 20
- SQL Server Failover Cluster Instances (FCIs) and Availability Groups 20
- SQL Server 2012 Setup for FCI 21
- SQL Server 2012 Setup for AAG 25
- Configuring SQL Server 2012 25
- Configuring SQL Server 2012 AlwaysOn Availability Groups 27
- Synchronous & Asynchronous Mirroring 27
- Creating AlwaysOn Availability Groups 27
- Conclusion 33
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 Server Failover Clustering (WSFC).

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.

Administrators are able to execute dynamic migrations and load-balancing operations with no impact to application availability. In multi-site configurations, VPLEX Metro and Geo provide support for a single federated view of the storage resources across the multiple sites, and will extend the functionality provided by VPLEX Local and Geo to provide multi-site high availability and disaster recovery solutions. VPLEX Metro and Geo also supports Workload Mobility which will allow customers to dynamically load balance SQL storage across different datacenters 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 physical or virtualized 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 Servers, SQL Servers, and Hyper-V-based cloud offerings. 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 of Microsoft SQL Server 2012 AlwaysOn Availability Groups running on Microsoft Windows Server 2012 and its Windows Server Failover Clustering feature solutions on EMC VPLEX storage federation systems. Details of integration with VPLEX systems are documented with practical examples for storage and SQL database administrators.

Audience

This white paper is intended for Microsoft SQL Server 2012 and Microsoft Windows Server 2012 administrators, SAN/WAN administrators, architects, customers, and EMC field personnel who want to understand the implementation of new features and functions of VPLEX with Microsoft SQL Server 2012 and Windows Server Failover Clustering.
VPLEX technology overview

At the highest level, EMC® VPLEX™ has unique capabilities that storage administrator’s value and can use to enhance their data centers. It delivers distributed, dynamic, and smart functionality into existing or new data centers to provide storage virtualization across geographical boundaries.

- VPLEX is distributed because it is a single interface for multivendor storage and delivers dynamic data mobility, which is being able to move applications and data in real time, with no outage required.
- VPLEX is dynamic because it provides data availability and flexibility as well as maintains business through failures that traditionally have required outages of manual restore procedures.
- VPLEX is smart because its unique AccessAnywhere technology can present and keep the same data consistent within and between sites and enable distributed data collaboration.

Because of these capabilities, VPLEX delivers unique and differentiated value to address three distinct requirements within our target customers’ IT environments:

- The ability to dynamically move applications and data across different compute and storage installations, whether they are within the same data center, across a campus, within a geographical region – and now, with VPLEX Metro-IP, across even greater distances.
- The ability to create high-availability storage and a compute infrastructure across these same varied geographies with unmatched resiliency.
- The ability to provide efficient real-time data collaboration over distance for such “big data” applications as video, geographic/oceanographic research, and more.

EMC VPLEX technology is a scalable, distributed-storage federation solution that provides zero disruptions, heterogeneous data movement and volume management functionality.

Insert VPLEX technology between hosts and storage in a storage area network (SAN) and data can be extended over distance within, between, and across data centers.

The VPLEX architecture provides a highly available solution suitable for many deployment strategies including:

- Application and data mobility — defined as the movement of virtual machines (VM) without downtime. An example is shown in Figure 1.

Storage administrators have the ability to automatically balance loads through VPLEX, using storage and compute resources from either cluster’s location. When combined with server virtualization, VPLEX allows users to transparently move and relocate virtual machines and their corresponding applications and data over distance. This provides a unique capability allowing users to relocate, share, and balance infrastructure resources between sites, which can be within a campus or between data centers, up to 10 ms apart with VPLEX Metro, or farther apart (50 ms RTT) across asynchronous distances with VPLEX Metro-IP.
High availability (HA) is a term that several products claim they can deliver. Ultimately, a HA solution is supposed to protect against a failure and keep an application online. Storage administrators plan around HA to provide near-continuous uptime for their critical applications, and automate the restart of an application once a failure has occurred, with as little human intervention as possible. With conventional solutions, customers typically have to choose a recovery point objective (RPO) and a recovery time objective (RTO). But even while some solutions offer small RTOs and RPOs, there can still be downtime, and for most customers, any downtime at all can be costly.

Distributed data collaboration — Increases utilization of passive data recovery (DR) assets and provides simultaneous access to data. An example is shown in Figure 3

This is when a workforce has multiple users at different sites that need to work on the same data, and maintain consistency in the dataset when changes are made. Use cases include co-development of software where the development happens across different teams from separate locations, and collaborative workflows such as engineering, graphic arts, videos, educational programs, designs, research reports, and so forth.

When customers have tried to build collaboration across distance with the traditional solutions, they normally have to save the entire file at one location and then send it to another site using FTP. This is slow, can incur heavy bandwidth costs for large files, or even small files that move regularly, and negatively impacts productivity because the other sites can sit idle while they wait to receive the latest data from another site. If teams decide to do their own work independent of each other, then the dataset quickly becomes inconsistent, as multiple people are working on it at the same time and are unaware of each other’s most recent changes. Bringing all of the changes together in the end is time-consuming and costly, and grows more complicated as the dataset gets larger.
**VPLEX product offerings**

VPLEX first meets high-availability and data-mobility requirements and then scales up to the I/O throughput required for the front-end applications and back-end storage.

A VPLEX cluster consists of one, two, or four engines (each containing two directors), and a management server. A dual-engine or quad-engine cluster also contains a pair of Fibre Channel switches for communication between directors.

Each engine is protected by a standby power supply (SPS), and each Fibre Channel switch gets its power through an uninterruptible power supply (UPS). (In a dual-engine or quad-engine cluster, the management server also gets power from a UPS.)

The management server has a public Ethernet port, which provides cluster management services when connected to the customer network.

**VPLEX Local, VPLEX Metro, VPLEX Metro-IP**

EMC offers VPLEX in three configurations to address customer needs for high-availability and data mobility:

- **VPLEX Local**
- **VPLEX Metro**
- **VPLEX Metro-IP**

The following diagram provides an example of each:

![VPLEX offerings diagram]

**VPLEX Local**

VPLEX Local provides seamless, zero disruption data mobility and the ability to manage multiple heterogeneous arrays from a single interface within a data center.

VPLEX Local allows increased availability, simplified management, and improved utilization across multiple arrays.

**VPLEX Metro with AccessAnywhere**

VPLEX Metro with AccessAnywhere enables active-active, block-level access to data between two sites within synchronous distances. The distance is limited as to what synchronous behavior can withstand as well as consideration to host application stability and MAN traffic. It is recommended that depending on the application that consideration for Metro be less than or equal to 10 ms RTT.

The combination of virtual storage with VPLEX Metro and virtual servers enables the transparent movement of virtual machines and storage across a distance. This technology provides improved utilization across heterogeneous arrays and multiple sites.
VPLEX Metro-IP with AccessAnywhere

VPLEX Metro-IP with AccessAnywhere enables active-active, block-level access to data between two sites within asynchronous distances. VPLEX Metro-IP enables better cost-effective use of resources and power. Geo provides the same distributed virtual volume flexibility as Metro but extends the distance up to and within 50 ms RTT. As with any asynchronous transport media, bandwidth is also important to consider for optimal behavior as well as application sharing on the link.

VPLEX Architecture highlights

VPLEX support is open and heterogeneous, supporting both EMC storage and common arrays from other storage vendors, such as HDS, HP, and IBM. VPLEX conforms to established World Wide Naming guidelines that can be used for zoning.

VPLEX supports a wide variety of operating systems including both physical and virtual server environments with Microsoft Windows Server 2012 and Hyper-V. VPLEX supports network fabrics from Brocade and Cisco including legacy McData SANs. For a complete and up-to-date list of supported platforms, storage, and other equipment please refer to the VPLEX Simple Support Matrix.

The following table lists an overview of VPLEX features along with benefits.

Overview of VPLEX features and benefits

<table>
<thead>
<tr>
<th>Features</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobility</td>
<td>Move data and applications without impact on users.</td>
</tr>
<tr>
<td>Resiliency</td>
<td>Mirror across arrays without host impact, and increase high availability for critical applications.</td>
</tr>
<tr>
<td>Distributed cache coherency</td>
<td>Automate sharing, balancing, and failover of I/O across the cluster and between clusters.</td>
</tr>
<tr>
<td>Advanced data caching</td>
<td>Improve I/O performance and reduce storage array contention.</td>
</tr>
<tr>
<td>Virtual storage federation</td>
<td>Achieve transparent mobility and access in a data center and between data centers.</td>
</tr>
</tbody>
</table>

For all VPLEX products, the appliance-based VPLEX technology:

- Presents SAN volumes from back-end arrays to VPLEX engines
- Packages the SAN volumes into sets of VPLEX virtual volumes with user-defined configuration and protection levels
- Presents virtual volumes to production hosts in the SAN via the VPLEX front end

The VPLEX Metro and VPLEX Metro-IP products present a global, block-level directory for distributed cache and I/O between VPLEX clusters.

Location and distance determine high-availability and data-mobility requirements. For example, if all storage arrays are in a single data center, a VPLEX Local product federates back-end storage arrays within the data center.

When back-end storage arrays span two data centers, the AccessAnywhere feature in a VPLEX Metro or a VPLEX Metro-IP product federates storage in an active-active configuration between VPLEX clusters. Choosing between VPLEX Metro and VPLEX Metro-IP depends on distance and data synchronicity requirements.

Application and back-end storage I/O throughput determine the number of engines in each VPLEX cluster. High-availability features within the VPLEX cluster allow for zero disruptive software upgrades and expansion as I/O throughput increases.
**VPLEX Metro-IP and VPLEX Geo**

EMC VPLEX Metro presents storage in an active-active fashion across synchronous distances. This is achieved by creating VPLEX distributed virtual volumes, which consist of a RAID 1 mirror with extents located across two VPLEX Metro and Geo clusters. Once the distributed virtual volumes have been created, they are then placed into a consistency group for management decisions made by either static bias rules or by a VPLEX Witness.

Being an active-active presentation, hosts can perform read-write operations on the distributed virtual volume while being simultaneously presented to both clusters at both sites. This is fundamentally different from traditional synchronous mirroring solutions in arrays where the volume on the secondary clusters remains Write Disabled until a failover procedure is performed.

With VPLEX distributed virtual volumes, no manual failover processes are required. The Global Cache Coherency layer in VPLEX presents a consistent view of data at any point in time.
Overview of SQL Server High-Availability Solutions

SQL Server provides several options for creating high availability for a server or database. The high-availability options include the following:

**Availability Groups**

As its name suggests, Availability Group is a grouping of related databases. When you were setting up database mirroring before SQL Server 2012, you could set up multiple mirrors, but you were only able to set up to mirror a single database at a time. If you have multiple databases that are reliant on each other for the application to work, there is no simple way of ensuring that all of the databases failed over together. Availability Groups allow you to group up to 10 AAG’s per SQL instance. Each of these 10 AAG’s may have up to 100 replica databases participating.

**Availability Modes**

As previously mentioned, when configuring your SQL Server 2012 AAG’s, there are some considerations that need to be taken into account when determining what type of availability mode you can use.

If you want to use AAGs for a reporting process, you could have your secondary replica located in the same physical data center and implement synchronous-commit mode to give you a read only near time group of databases to report against without impacting the performance of the primary databases with reporting overheads. You probably would not consider this type of availability mode where there are large distances between data centers.

If you have the requirement for a reporting process, that does not require the data to be near real time, you could consider implementing your secondary replica in a separate data center using asynchronous-commits for your AAG. By implementing an asynchronous-commit method, you would reduce the latency of the transactions on the primary site but it would open you up to the possibility of data loss.

In the following example you would have your primary databases in AAG1 residing at the primary datacenter. You would then set up a secondary replica at the primary datacenter configured in a synchronous-commit mode. Now you may offload your testing, reporting, and backup requirements without the reporting overhead impacting on your primary database. This also ensures that you have met HA requirements, by having a secondary environment that is transaction consistent with the ability to failover to in the event of an issue with your primary databases. You could then setup secondary replicas at remote datacenters in asynchronous-commit mode. These asynchronous secondary replicas allow you to meet your DR requirements by having multiple copies in multiple geographical dispersed locations, with the ability to failover to in the event of an issue on the primary site.

**AlwaysOn Failover Cluster Instances**

AlwaysOn Failover Cluster Instances leverages Windows Server Failover Clustering (WSFC) functionality to provide local high availability through redundancy at the server-instance level—a failover cluster instance (FCI). An FCI is a single instance of SQL Server that is installed across Windows Server Failover Clustering (WSFC) nodes and, possibly, across multiple subnets. On the network, an FCI appears to be an instance of SQL Server running on a single computer, but the FCI provides failover from one WSFC node to another if the current node becomes unavailable.

**AlwaysOn Availability Groups**

AlwaysOn Availability Groups is an enterprise-level high-availability and disaster recovery solution used to maximize availability for one or more user databases. AlwaysOn Availability Groups requires that the SQL Server instances reside on all WSFC nodes.

**Database mirroring**

Database mirroring is a solution used to increase database availability by supporting almost instantaneous failover. Database mirroring can be used to maintain a single standby database, or mirror database, for a corresponding production database that is referred to as the principal database. (to be removed in future releases)
Log shipping
Like AlwaysOn Availability Groups and database mirroring, log shipping operates at the database level. You can use log shipping to maintain one or more warm standby databases (referred to as secondary databases) for a single production database that is referred to as the primary database.

Overview of AlwaysOn Availability Groups
As with database mirroring and WSFC, AAG’s provide the ability to failover between the primary and secondary replicas. There are three forms of failover which can be undertaken with AAG’s:

- **Automatic** - Supported by Synchronous-Commit Mode - No Data Loss
- **Manual** - Supported by Synchronous-Commit Mode - No Data Loss
- **Forced** - Supported by Asynchronous-Commit - Possible Data Loss

An availability group supports WSFC for a set of databases (known as availability databases) that will fail over together. An availability group supports a set of primary databases and up to four sets of corresponding secondary databases.

There are two types of availability replicas:

- Single primary replica which hosts the primary databases.
- One to four secondary replicas which hosts a set of secondary databases
  And serves as a potential failover targets

The primary replica makes the primary databases available for read-write connections from clients. Also, in a process known as data synchronization, this occurs at the database level. The primary replica sends transaction log records of each primary database to every secondary database. Every secondary replica caches the transaction log records and then applies them to its corresponding secondary database. Data synchronization occurs between the primary database and each connected secondary database, independently of the other databases. Therefore, a secondary database can be suspended or fail without affecting other secondary databases, and a primary database can be suspended or fail without affecting other primary databases.

Deploying AlwaysOn Availability Groups requires a Windows Server Failover Clustering (WSFC) cluster. Each availability replica of a given availability group must reside on a different node of the same WSFC cluster. A WSFC resource group is created for every availability group that you create. The WSFC cluster monitors this resource group to evaluate the health of the primary replica. The quorum for AlwaysOn Availability Groups is based on all nodes in the WSFC cluster regardless of whether a given cluster node hosts any availability replicas.

The failover behavior for each secondary replica depends on which availability replica is currently the primary replica. The following diagram illustrates an availability group that contains one primary replica and three secondary replicas (4 replicas is the maximum). These replicas may span one, two, or even more sites based on the architectural design and other limitations.
Failover Sets
The forms of failover that are possible for a given availability group can be understood in terms of failover sets. A failover set consists of the primary replica and secondary replicas that support a given form of failover, as follows:

**Automatic Failover Set** (optional): Within a given availability group, a pair of availability replicas (including the current primary replica) that are configured for synchronous-commit mode with automatic failover, if any. An automatic failover set takes effect only if the secondary replica is currently SYNCHRONIZED with the primary replica.

**Synchronous-Commit Failover Set** (optional): Within a given availability group, a set of two or three availability replicas (including the current primary replica) that are configured for synchronous-commit mode, if any. A synchronous-commit failover set takes effect only if the secondary replicas are configured for manual failover mode and at least one secondary replica is currently SYNCHRONIZED with the primary replica.

**Entire Failover Set**: Within a given availability group, the set of all availability replicas whose operational state is currently ONLINE, regardless of availability mode and of failover mode. The entire failover set becomes relevant when no secondary replica is currently SYNCHRONIZED with the primary replica.

Benefits of using Failover Clustering with VPLEX
Windows Server Failover Clustering provides high availability and scalability to many different types of server workloads such as Microsoft Exchange Server, Microsoft SQL Server, and Hyper-V. When adding VPLEX to this solution we are adding the capabilities for each server and/or application to run on either physical servers or virtual machines located within or across data centers. Each accessing the same read/write data from both locations simultaneously in a truly Active/Active configuration. See Standard 4-Node Cluster diagram below.

The following lists some key benefits for using SQL Server 2012 with Failover Clustering and VPLEX:

- Protection at the SQL Instance level through added redundancy
- Automatic failover within or across data centers in the event of a SW, HW, or Application failures
- Heterogeneous support for storage array technologies and Multi-Path software
- Disaster Avoidance and Recovery solution using multi-subnet support
- Zero reconfiguration of applications and clients during failovers
- Flexible failover policy for granular trigger events for automatic failovers
- Reliable planned and unplanned failovers through health detection and monitoring
- Configurable and tunable parameters for predictability in failover events
- Throttled resource usage during failovers
Benefits for using AlwaysOn Availability Groups with VPLEX

Windows Server Failover Clustering (WSFC) combined with SQL Server AlwaysOn provides both a high availability (HA) and disaster recovery (DR) solution for your SQL Database environment. AlwaysOn isn’t a completely new technology, but rather a continuing evolution of the database mirroring features offered in previous releases of SQL Server. AAG’s when coupled with WSFC will provide more granular control of your database replication to achieve the five 9’s of availability that are demanded in today’s data centers.

Now complete your solution by adding the VPLEX and instantly you have virtualized heterogeneous storage within or across your data centers with the ability to transparently move data arrays using workload or application mobility, you may perform tech refreshes and other types of maintenance without incurring outages or disruptions in service, and best of all… You will have simultaneous read/write access to the same data from two separate data centers making this the ultimate solution for continuous availability.

Planning for Manual Failovers

A manual failover causes a synchronized secondary replica to transition to the primary role after a failover command has been issued. To support manual failover, the primary and secondary replicas must be configured for synchronous-commit mode. Every secondary database on the availability replica must be synchronized with its corresponding primary database. This guarantees that every transaction that was committed on a former primary database has also been committed on the new primary database. Therefore, the new primary databases are identical to the old primary databases.

The following figure illustrates the stages of a planned failover:

1. Before the failover, the primary replica is hosted by WIN2012_N1.
2. Failover is initiated and the “sqldb1” instance moves to WIN2012_N2.
3. WIN2012_N2 becomes the new primary replica and WIN2012_N1 switches to the secondary replica.
WSFC Section One: Standard 2-Node Cluster
For the purpose of this white paper we will build our cluster in four separate sections to show how each piece fits together with the previous section. This environment will utilize backend storage from a Symmetrix VMAX at Site-A and then a CLARiiON VNX at Site-B. The storage will then be virtualized with a VPLEX Metro-IP. The two datacenters (DC’s) are named Atlanta & Birmingham and have more than sufficient infrastructure to support the synchronous behavior of a VPLEX Metro-IP environment in combination with WSFC and Microsoft SQL Server 2012. Two physical Microsoft Windows Server 2012 servers were used, (1) located in Atlanta and (1) in Birmingham. Each of the cluster nodes has been setup with the following Roles, Features, and Software:

- **Windows Roles**
  - Hyper-V
- **Windows Features**
  - Failover Clustering
  - Multipath I/O
- **SQL Server 2012 for Failover Cluster Installation (instance = “sqldb1”)**
On this 2-Node cluster, SQL Server 2012 will be installed as a Failover Cluster Instance (FCI) on each clustered node. This is accomplished by first installing SQL Server 2012 as a Clustered Instance ("sqlb1") on WIN2012_N1, then adding an additional node for WIN2012_N2. In our setup we will designate instance "SQLDB1" to run on WIN2012-N1 (local site) with a preference to failover to WIN2012-N2 (remote site). Another consideration for this setup is to build a 4-Node cluster for redundancy into each site so the first failover may be done locally instead of across datacenters. For this scenario, if the "sqlb1" instance was running in Atlanta, then it would failover from N1 to N3 locally, then over to N2 remotely. Likewise if the "sqlb1" instance were running at the Birmingham site N2 would failover to N4 locally, then over to N3 remotely.

Benefits:
- Increases the availability of Microsoft WSFC Cluster resources
- Simplified deployment of WSFC Stretch Clusters by not requiring additional equipment or software
- Eliminating Single Points of Failure (SPOF)
- Providing Active/Active DC’s for next generation Disaster Avoidance (DA)
- Increased Utilization of hardware and software
- Automated Workload and Application Mobility within or across datacenters
- Zero downtime, Zero disruption for planned maintenance and tech refreshes

WSFC Section Two: Standard 4-Node Cluster with AlwaysOn
In the second section of our environment, we will continue to build out our WSFC cluster by adding an AlwaysOn Availability Group (AAG) to the existing 4-Node stretched cluster from section one. The assumption that all four existing node have been configured to run the “sqlb1” instance and all failover preferences are set for local and remote failovers. The AAG will be created for the “sqlb1” instance and will leverage two newly added cluster nodes called WIN2012-N5 and WIN2012-N6. It should be noted that the storage used for AAG’s is not shared between each clustered node and is not replicated by VPLEX, but instead by way of SQL2012’s TCP/IP transport mechanism. We will talk more about this behavior in a later section.
Previously, when we set up the “sqlb1” instance for WSFC it was necessary to install SQL2012 as a Failover Clustered Instance (FCI) on all participating nodes (N1-N4), but in order to leverage AlwaysOn the SQL2012 must be installed as a Stand-Alone Instance on each of the DR nodes (N5 and N6). In our setup we will designate instance “sqlb1_dr1” to run on N5 (local to SQLDB1) with a preference to failover to N6 (which is remote to SQLDB1). This process will create (2) read-only replicas for disaster recovery and off-loading of processes like backups & test/dev.

AlwaysOn Availability Groups (AAGs) offer a high-availability and disaster-recovery solution that helps provide an alternative to database mirroring. AlwaysOn Availability Groups maximize the availability and recoverability of a particular set of databases. An AAG supports failover environments by grouping databases together, into what's known as availability databases. These availability databases will fail over together when needed. An AAG supports a set of read-write primary databases and one to four sets of corresponding secondary databases. Secondary databases can be made available for read-only access for disaster recovery and the off-loading of processes like backups & test/dev.

Each node in a WSFC cluster participates in periodic heartbeat communication to share the node’s health status with all participating cluster nodes. Unresponsive nodes are considered to be in a failed state. A "quorum node set" is a majority of the voting nodes and witnesses in the WSFC cluster. The overall health and status of a WSFC cluster is determined by a periodic quorum vote. The presence of a quorum means that the cluster is healthy and able to provide node-level fault tolerance. It is best practice to always have an odd number of quorum votes in a WSFC cluster. In our configuration we have opted to disable the quorum voting for the two nodes holding the "sqlb1_dr1" replicas and to use a remote file share as a tie-breaker.

Benefits:
- Continuous availability of Servers, Applications, and Storage to achieve (5)9's
- Performing backup operations from the read-only secondary replicas
- Allowing for non-disruptive testing against the read-only secondary replicas
- Database consistency checks (dbcc) may be done with secondary replicas
- Flexible failover policies for controlling AAG failover activities
  - Synchronous commit with automatic failover
  - Synchronous commit with planned manual failover only
  - Asynchronous commit with only forced failover

WSFC Section Three: Standard 4-Node Cluster with AAG and Hyper-V
In this third section we will build on the previous configuration by adding the Hyper-V role to the “sqlb1_dr1” instances. This will allow us to use virtual machines (VMs) to offload various workloads from the production servers to secondary replicas of the production databases. It should be noted that any of the read-only replicas may be used for access to the production data. As is the case for AAG's, the storage used for the AAG’s is non-shared and will be utilized as pass-through devices within each of the Hyper-V instances. However, the VMs will all be installed on Cluster Shared Volumes (CSVs) to ensure failover capabilities for all VM’s running on N5 and N6.
Implementing EMC VPLEX® With Microsoft® SQL Server® 2012

By virtualizing all or a portion of the WSFC Cluster we may build in some additional flexibility for distributing our SQL Server 2012 environment with WSFC and VPLEX. In the above example, Nodes N1-N4 have been setup with "shared" storage for FCI and will automatically handle any failover activity for the "sqldb1" instance. Likewise, nodes N5 and N6 have been added to the WSFC Cluster and setup as stand-alone SQL Server's to host the AAGs with "non-shared" storage. These two nodes will handle the automatic & manual failover of the read-only secondary replicas. It should be noted that in this current configuration that there are currently one primary and two secondary replicas of the database. The primary database will use FCI to failover and the secondary replicas will rely on WSFC Manager or SQL Tools to perform failovers or promotions of a secondary replica to primary.

Additional Benefits:
VPLEX, WSFC, and Hyper-V offer improved scalability and performance
VPLEX, WSFC, and Hyper-V enable workload consolidation and load balancing within or across datacenters
VPLEX, WSFC, and Hyper-V provide a simplified virtualization of servers, applications, and storage
VPLEX, WSFC, and Hyper-V help to reduce operational costs, improve utilization, and provision new servers
VPLEX, WSFC, and Hyper-V support for live migration for "zero downtime" planned migrations

Implementing WSFC Stretched Clusters using VPLEX

Now that we have covered off the basics of what an AlwaysOn solution could possibly look like we are ready to tie it all together with planning a deployment and implementation of this solution to meet your ever increasing High Availability (HA), Disaster Recovery (DR), and your Disaster Avoidance (DA) requirements. In the next sections we will take a look at the steps required to build out this solution and put it to use.

Provisioning Storage for Stretched Clusters on VPLEX

The following steps outline the requirements for provisioning VPLEX Virtual and Distributed volumes that live within and across both data centers. This storage will then be used to create our Windows Server Failover Cluster with SQL Server 2012 AlwaysOn Availability Group Instances.

Once logged into Unisphere for VPLEX Metro-IP webpage, you will need to validate the system health before navigating to the “Provision Storage” tab to claim storage, create extents, and the devices needed for our "shared" and "non-shared" volumes.

Now that the storage has been claimed along with the creation of 1:1 devices, it's time to create our local and distributed volumes for our Windows Server Failover Cluster and SQL Server 2012 AlwaysOn Availability Groups.
The above shows the distributed devices created for the WSFC stretched cluster nodes N1-N4. These “distributed” volumes are “shared” RAID-1 (mirrored) storage that will span across the two datacenters and below we have the “local” non-shared volumes used by the AAG’s on nodes N5 and N6. The replication of the “non-shared” volumes will be handled by TCP/IP and the AAG’s. Ultimately the AAG’s will be managed by WSFC. This brings the total replica count to three; one primary and two secondaries.

Creating Consistency Groups and Applying 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 Storage Views

As the Windows Servers are deployed over 2 geographic sites the Windows servers for each site must belong to a separate VPLEX Storage View. Shown below are the Storage Views from WIN2012-N1 in Atlanta and WIN2012-N2 which is located in Birmingham. Nodes N3 and N4 will be identical by site assignments. The Atlanta site has its storage virtualized from a VMAX array and the Birmingham site uses a VNX array. All writes from Atlanta will be replicated to Birmingham and vice versa through the VPLEX.

Note: When working with Microsoft Windows Server 2012 and WSFC Clusters it should be noted that although it’s no longer necessary to match the Host LUN Numbering (HLN) assignments across nodes, the ability to keep them aligned across each Storage View is shown in the above diagrams for N1 and N2.

Note: LUNs 7 and 8 shown above are the non-shared virtual volumes used by AAG on nodes N5 and N6. These virtual volumes are not replicated by VPLEX; instead TCP/IP keeps the replicas in sync via the AAG.
Adding Roles and Features for VPLEX Stretched Cluster

For our configuration we are going to build a 4-Node SQL Server 2012 AlwaysOn Cluster. The first step will be to ensure that all of the nodes have .NET Framework 3.5.1 and the Failover Clustering feature enabled.

**Required Features**

Building your Windows Server Failover Cluster

There is no difference between the task of building your WSFC for use with SQL Server 2012 AlwaysOn and you’re previously built WSFC for SQL Server 2008. In this white paper, we will not be going to go through the WSFC build, but instead concentrate on SQL Server 2012. It is important to mention that your WSFC must meet the validation tests as defined in the “Validate a Configuration Wizard” within the “Failover Cluster Manager”.

**Validate Cluster**

SQL Server Failover Cluster Instances (FCIs) and Availability Groups

SQL Server 2012 AlwaysOn is a new high availability and disaster recovery solution that takes advantage of WSFC. AlwaysOn provides an integrated, flexible solution that increases application availability, provides better returns on hardware investments, and simplifies high availability deployment and management. An availability replica can be hosted by either a standalone instance of SQL Server or from an FCI instance. The WSFC cluster will proactively monitor your SQL instances for adverse conditions and/or failures and then signal the need to restart or automatically fail that SQL instance over to a different cluster node.
SQL Server 2012 Setup for FCI

Now that we have validated our potential cluster nodes, we are ready to start the building out our SQL Server 2012 AlwaysOn Cluster. On the first node of our cluster (N1), we start the SQL Server Installation Center to begin the installation process. Once started, navigate to the Installation tab and select ‘New SQL Server Failover Cluster installation” or add features to an existing installation’ as shown below.

![SQL Server Installation Center – Installation tab](image)

Next, validate each cluster node (N1-N4) using the “Setup Support Rules” dialogue. This process will identify potential problems before actually launching into the SQL Server installation process. Note: Nodes N2-N4 will use the “Add Node to a SQL Server Failover Cluster” option.

![Validate Setup Support Rules](image)

After validation has completed, you will need to choose the “Installation Type”. From the “Installation Type” screen, select the option to ‘Perform a new Installation of SQL Server 2012’ as shown below.
Next you will select ‘SQL Server Feature Installation.’

Select the Features for your SQL Server instance.
Feature Selection

Enter your SQL Server 2012 Instance Name (sqldb1) for the Instance that you are building. Later we will add nodes N2-N4. This will happen for each cluster node to be configured for use with the FCI.

Instance Configuration

You will need to create your domain service accounts and set the passwords on this Server Configuration screen during the installation. Once you have set the passwords, make sure you click on the Collation Tab so as to configure your Collation for the Instance, click ‘Next’. For simplicity we have used the administrator account.
Service Account Details

On the Database Engine Configuration screen there are three tabs that we need to pay attention to. The first tab is the “Server Configuration” tab. This is where you set your authentication security mode. We will be using “Windows Authentication Mode” (recommended) with the Administrator domain account.

NOTE: Prior to using the New Availability Group Wizard, each server instance will require a database mirroring endpoint. These endpoints are shared by all the availability replicas, database mirrors, and witnesses. If a server instance that you select to host an availability replica is running under a domain user account and does not yet have a database mirroring endpoint, the “New Availability Group Wizard” will create the endpoint and grant the CONNECT permission to the server instance service account. However, if the SQL Server service is running as a built-in account, such as Local System, Local Service, or Network Service (or a non-domain account) you must use certificates for endpoint authentication as the wizard cannot create endpoints on the server instance. In this case, the best practice is to create the database mirroring endpoints manually before you launch the wizard.

The second tab is for setting the “Data Directories.” This tab allows you to specify where you want to locate your database and log files. Traditionally you would have four separate drive locations depending on your storage for data files, log files, TempDB, and backups.

Database Engine Configuration

The “Ready to Install” screen is a summary of features to be installed.
Repeat this process for all WSFC Cluster nodes (N1-N4) that will utilize the FCI for SQL Server (sqldb1).

**SQL Server 2012 Setup for AAG**

Now that we have configured our FCI cluster nodes, we are ready to start building out our two AAG nodes. These two nodes will be used to host the two read-only secondary replicas from the "sqldb1" primary database. This will be implemented on nodes N5 and N6 by starting the SQL Server Installation Center and navigating to the Installation tab and select ‘New SQL Server stand-alone installation” as shown below.

**Configuring SQL Server 2012**

Now that we have installed (4) primary nodes with shared storage for the “sqldb1” FCI and the (2) secondary nodes with non-shared storage for the AAG secondary replicas, we need to undertake some post-installation configuration to put it all together. This is achieved by using the SQL Server Configuration Manager which is available from Start | All Programs | Microsoft SQL Server 2012 | Configuration Tools.

First we need to validate the Log On properties for each SQL2012 instance. Each of these logon’s should be set to use a domain account instead of the default setting that uses the Local NT Authority login. For the purpose of simplicity we will use the Administrator account from the VPLEXLAB domain.
You may also want to validate the database logins using the SQL Server 2012 Configuration Manager.

<table>
<thead>
<tr>
<th>Name</th>
<th>State</th>
<th>Start Mode</th>
<th>Log On As</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Server (SQLB1)</td>
<td>Running</td>
<td>Automatic</td>
<td>sysdb\admin123</td>
</tr>
<tr>
<td>SQL Server (SQLB1 Dir1)</td>
<td>Running</td>
<td>Automatic</td>
<td>VPLEX\Admin123</td>
</tr>
<tr>
<td>SQL Server (SQLB1 Dir2)</td>
<td>Running</td>
<td>Automatic</td>
<td>VPLEX\Admin123</td>
</tr>
</tbody>
</table>

In previous versions of SQL Server we would have installed a Clustered SQL Server Instance across all nodes for HA, but with AAG’s you may have noticed that we have installed stand-alone instances of SQL Server on nodes N5 and N6 so that they are dedicated to AAG for DR and read-only replicas. This change in architecture is to facilitate the many architectural changes made to SQL Server 2012 and the use of AAG’s. At this point we need to enable the AlwaysOn Availability Groups for HA and DR. This can be done from the SQL Server Configuration Manager. Simply select the instance of SQL, then right-click and select Properties. Go to the ‘AlwaysOn High Availability’ tab and check the ‘Enable AlwaysOn Availability Groups’ check box. (repeat on both nodes N5 and N6)
Configuring SQL Server 2012 AlwaysOn Availability Groups

Before SQL Server 2012, one of the options available for you to use to build your High Availability (HA) solution was to utilize Database Mirroring. The Database Mirroring technology is very good at what it was created for. However, it has some limitations when it comes to your HA solution. The limitations include:

- A Single Secondary database
- Mirrored database is accessible via database snapshot (only until failover occurs)
- Lack of support for MSDTC (distributed transactions)
- Related databases are not able to be grouped together

SQL Server 2012 AAG’s resolve most of these issues giving you more flexibility over your environment and more granular control over your environment to meet your ever growing complex HA requirements. SQL Server 2012 AAG’s still utilize the database mirroring technology from previous versions of SQL Server, but allow you to transfer your data via TCP/IP either synchronously or asynchronously to one or more replicas while also granting read access to the replicas.

Synchronous & Asynchronous Mirroring

AlwaysOn uses SQL Server’s existing mirroring technology. Synchronous Mirroring, as its name indicates, requires the transactions to be written at both sites for the transaction to be completed. Whereas this can lead to increased latency in your system, it gives you zero data loss. AlwaysOn will support up to two secondary replicas synchronously replicated per availability group. Asynchronous Mirroring, on the other hand, is faster but increases the risk of data loss as it does not have the requirement to complete the transaction at the secondary site. One advantage that AlwaysOn has over mirroring is that it allows multiple usable secondary’s of the database. With VPLEX hosting the volumes it becomes clear that the ability to combine both synchronous & asynchronous mirroring will offer the best performance and reliability to your overall solution. Depending on your HA/DR requirements you could use VPLEX Distributed Volumes for your shared storage needs on your stretched Failover Cluster between data centers and then use SQL AlwaysOn to synchronously (or asynchronously) replicate your data to a second or third data center.

Creating AlwaysOn Availability Groups

From SQL2012 Server Management Studio, attach to the SQL instances that own the databases to be included in the AlwaysOn Availability Group. Then from the tree view of that instance, right click on the “AlwaysOn High Availability” to launch the “New Availability Group Wizard.”

To create Availability Groups for SQL2012, you will need the following information:

- Specify an AAG name
- Select one or more databases on this instance of SQL2012 server
- Specify one of more instances of SQL Server to host secondary availability replicas
- Specify your availability group listener preference
- Select your initial data synchronization preference
- Check the validation results of the availability group creation
- Review your selections
Specify Availability Group Name
From the New Availability Group dialogue, enter a name for your AAG

![New Availability Group – Specify Name](image1)

Select Databases for Availability Group
Select the “sqlb1” database that you desire to include in your AAG.

![New Availability Group – Select Databases](image2)

Specify Replicas for Availability Group
Your primary replica will automatically be available for you to configure. Choose Automatic Failover, Synchronous Commit, and Readable Secondary mode setting for each SQL instance needed. You may use the ‘Add Replica’ to connect and configure your appropriate secondary servers. (primary and secondary settings should match)
New Availability Group - Specify Replicas

Validate Database Endpoint Values for Availability Group
Select the Endpoint tab, validate the Server Name, Endpoint URL, Port Number, Endpoint Name, Encrypt Data, and the SQL Server Service Account settings to ensure they are correct, and then proceed to the “Listener” tab.

New Availability Group - Validate Endpoints

Choose Backup Strategy for Availability Group
Select the “Backup Preferences” tab, choose where backups will occur. The default is to “Prefer Secondary”, but the options are available to change this behavior as needed.
In order to take advantage of the various solutions that we have stepped through in this article, we need to set up and allow for the applications to maintain connectivity to the SQL Server Databases after a failover. This is where the AlwaysOn Availability Group Listeners (AAGL’s) come into use.

An Availability Group Listener is a Virtual Server Name that applications connect to. From the applications point of view it does not matter where the Availability Database is active and available for use. The AAGL consists of:

- Virtual Network Name (VNN)
- Listener Port (1433)
- One or more Virtual IP Addresses (VIPs)

Selecting the “Listener” tab and enter your AAGL name, port number, and IP Address
Select Database Synchronization Preference
Every replica must have access to a shared network share in order to access the database backups created and then used for synchronizing the secondary databases. Enter your share path, click 'Next'.

New Availability Group - Select Initial Data Synchronization

Select Database Synchronization Preference
Once the AAG configuration steps have been completed it is necessary to process these settings with the “Results” validation process. At the completion of this process you should see a message that reads: “The Wizard has Completed Successfully.” If you did not receive all green ticks for your validation, resolve and re-run process.

New Availability Group - Results Verification
Automatic Failover Requirements
Automatic failover occurs only under the following conditions:

- An automatic failover set exists with a primary replica and a secondary replica that are both configured for synchronous-commit mode and set to AUTOMATIC failover.
- The automatic failover target has a healthy synchronization state.
- The Windows Server Failover Clustering (WSFC) cluster has quorum.
- The primary replica has become unavailable, and the defined failover-conditions have been met.

Note: AlwaysOn Availability Groups monitors the health of both replicas in an automatic failover set. If either replica fails, the availability group's health state is set to CRITICAL. If the secondary replica fails, automatic failover is not possible because the automatic failover target is unavailable. If the primary replica fails, the availability group will fail over to the secondary replica. Until the former primary replica comes online, no automatic failover target exists. In either case, to ensure availability in the unlikely case of a sequential failure, we recommend that you configure a different secondary replica as the automatic failover target.

Manual Failover Requirements
To support a manual failover, the current primary replica must be set to synchronous-commit mode and a secondary replica must be:

- Configured for synchronous-commit mode.
- Currently synchronized with the primary replica.

Forced Failover Requirements (with Possible Data Loss)
Forcing failover of an availability group (with possible data loss) is a disaster recovery method that allows you to use a secondary replica as a warm standby server. Because forcing failover risks possible data loss, it should be used cautiously and sparingly. We recommend forcing failover only if you must restore service to your availability databases immediately and are willing to risk losing data. This method also requires your WSFC to have quorum.

Risks of Forcing Failover
It is essential to understand that forcing failover can cause data loss. Data loss is possible because the target replica cannot communicate with the primary replica which means there is no guarantee that the databases are synchronized. Forcing a failover will start a new recovery fork. Because the original primary databases and secondary databases are on different recovery forks, each of them now contains different versions of the same database. The original primary database contains the changes that were not yet sent from its send queue to the former secondary database… and the former secondary databases contain whatever changes occurred after failover was forced.

If failover is forced because the primary replica has failed, potential data loss is dependent on whether any transaction logs had not been sent to the secondary replica before the failure. Under the asynchronous-commit mode, accumulated unsent log is always a possibility. Under synchronous-commit mode, this is possible only until the secondary database becomes synchronized.

Secondary databases track only two recovery forks, so if you perform multiple forced failovers, any secondary database that did start data synchronization with the previous force failover might not be able to resume. If this occurs, any secondary databases that cannot be resumed will need to be removed from the availability group, restored to the correct point in time, and rejoined to the availability group. A restore will not work across multiple recovery forks, therefore, be sure to perform a log backup after performing more than one forced failover.
Managing the Potential Data Loss

After failover is forced, all secondary databases are suspended. This includes the former primary databases, after the former primary replica comes back online and discovers that it is now a secondary replica. You must manually resume each suspended database individually on each secondary replica.

Once the former primary replica is available, assuming that its databases are undamaged, you can attempt to manage the potential data loss. The available approach for managing potential data loss depends on whether the original primary replica has connected to the new primary replica. Assuming that the original primary replica can access the new primary instance, reconnecting occurs automatically and transparently.

The Original Primary Replica Has Reconnected

Typically, after a failure, when the original primary replica restarts it quickly reconnects to its partner. Upon reconnecting, the original primary replica becomes the secondary replica. Its database becomes the secondary databases and will enter a SUSPENDED state. The new secondary databases must be RESUMED to perform a rollback operation.

One potential drawback is that the suspended databases are inaccessible. This will prevent you from evaluating what data would be lost if you were to resume that particular database. Therefore, the decision on whether to resume or remove a secondary database depends on the following:

- If losing any data would be unacceptable, then remove the databases from the AAG to salvage them.
- If losing data would be acceptable to your business goals, then resume the secondary databases.

Note: Resuming a new secondary database causes it to be rolled back as the first step in synchronizing the database. If any log records were waiting in the send queue at the time of failure, the corresponding transactions are lost, even if they were committed.

Conclusion

This document provides an overview of a zero data loss solution that combines EMC VPLEX Metro-IP or VPLEX Metro-IP with Microsoft Server 2012 and Microsoft SQL Server 2012 using AlwaysOn Availability Groups. It highlights the benefits of leveraging shared EMC VPLEX Metro & VPLEX Metro-IP replicated storage within and across your data center while also leveraging non-shared, non-replicated storage with SQL Server 2012 AAG’s to offload certain processes such as backups, DBCC’s, and Test/Dev validations. EMC VPLEX provides zero downtime, zero disruption data migrations and tech refreshes, plus the ability to transparently move workloads to larger more robust and faster arrays. (and vice-versa)

This white paper documents the value of leveraging SAN and TCP based synchronous and/or asynchronous replication with EMC VPLEX Metro-IP and VPLEX Metro-IP coupled with Microsoft SQL Server 2012 to achieve a zero data loss and lowest possible RPO. This provides tremendous value to an SQL Server 2012 AAG environment. It allows customers to implement AAG resiliency configurations with database-level replication and failover. EMC’s VPLEX Metro-IP and VPLEX Metro-IP will seamlessly integrate with Microsoft Server 2012 Failover Clustering and SQL Server 2012. This provides a highly available solution for your production data, disaster recovery planning, and process offloading for SQL Server processes.

To meet today’s demanding business challenges an organization’s data must be highly available—in the right place, at the right time, and at the right cost to the enterprise. By using EMC VPLEX to virtualize the storage for Microsoft Server 2012 and Microsoft SQL Server 2012 AAG’s, organizations can manage their SQL environments more effectively through its ability to use VPLEX’s data mobility to easily swap between remote data centers with zero downtime and zero disruption in service.