Dell High Availability and Disaster Recovery Solutions Using Microsoft SQL Server 2012 AlwaysOn Availability Groups

Dell servers and storage options available for AlwaysOn Availability Groups deployment.

Gopu Narasimha Reddy

Dell Database Solutions Engineering

August 2012| Rev 1.0

Contents

Figures

Executive summary

Organizations typically require that their critical applications are available to end users on a continuous basis, and thus the need for reliable solutions that provide High Availability (HA) and Disaster Recovery (DR) continues to increase. Most HA and DR solutions available today use multiple redundant components to extend these vital capabilities to hosted data. Several factors are important when choosing the best and most appropriate HA and DR solutions for a given infrastructure, including reliability, robustness, cost, efficiency, automation, ease-of-use, and related performance impacts to the hosted application.

In this technical paper, we will discuss an important HA and DR feature, Microsoft SQL Server 2012 AlwaysOn Availability Groups (AVG), and how it addresses the issues mentioned above when deployed on Dell™ 12th Generation PowerEdge™ Servers. The guidelines and recommendations for implementing AVG will be illustrated by a review of some of the possible user scenarios. We will also discuss how PowerEdge Servers can be combined in a variety of server and storage configurations that make Dell the right platform for deploying Microsoft AlwaysOn HA and DR solutions.

Introduction

Achieving the best possible level of availability and recovery of critical applications is a big challenge. Many problems—software failures, hardware failures, planned and unplanned downtimes to name just a few—can make the critical data unavailable.

High Availability features provide continuous access to the applications/infrastructure, and are generally achieved by having redundant hardware and software combination located within the same room or a datacenter. Disaster Recovery, on the other hand, is a recovery plan which will be used when the primary setup is not accessible. Usually a DR setup is located at the remote site rather than the same site of primary setup.

Recovery Time Objective (RTO) and Recovery Point Objective (RPO) are the important metrics that help customers in choosing the best suitable HA and DR solution for a given application. RTO refers to how long a customer can afford to go without a given application being available for use. The appropriate HA solution has to be implemented so that the application recovers from failure within an acceptable time period.

RPO refers to the amount of data that a customer determines it can afford to lose in the case of a disaster/outage. This dictates the kind of recovery solution that needs to be in place for a given application. Achieving the required RPO and RTO goals involves ensuring continuous uptime of critical applications and protection of critical data from unplanned and planned downtime.

Choosing the right hardware components such as servers, storage, and switches for implementing a HA and DR solution can significantly benefit the customers in terms cost, performance, and availability of the application. The servers and different storage options chosen for implementing the AlwaysOn AVG have a great impact on performance and failback/failover times. In the following sections, we will

discuss various server and storage options available from Del that are most conducive to optimal AVG deployments.

Objectives

- Understanding how AlwaysOn Availability Groups may be the right choice for implementing High Availability and Disaster Recovery solutions for business critical SQL databases based on customer RTO and RPO requirements.
- Exploring the various hardware combinations possible using the latest Dell PowerEdge servers and available internal/external database storage options for HA and DR implementations.
- Defining useful HA and DR implementation scenarios with AlwaysOn Availability Groups.
- Discussing guidelines and recommendations for implementing AlwaysOn Availability Groups.

Why Dell server and storage hardware for AlwaysOn setup?

Dell recently launched its 12th Generation PowerEdge servers which bring to the market many new features and enhancements that benefit the AlwaysOn Availability Group deployments. These features and enhancements are detailed below.

Processors

By supporting Intel® E5-series processors, Dell's 12th Generation PowerEdge servers provide the heavy-duty computing power required by high performing AVG databases. Together with High Quick Path Interconnect (QPI) links, 20MB L3 cache, and support for many 3rd Gen PCIe slots, this makes the Dell 12th generation PowerEdge servers the right platform for hosting high performing AVG databases.

Memory

AVG deployments benefit from the high speed memory DIMMs. Dell's 12th Generation PowerEdge servers support 1333 and 1600 MHz high speed memory DIMMS with ECC. Single, Dual and Quad rank DIMMs are supported. With the high speed memory DIMMS, Dell's solutions can achieve noticeable improvement in the overall database throughput.

Internal storage (HDD and SSD)

Dell's 12th Generation PowerEdge rack servers have a flexible storage chassis that supports 2.5- and 3.5-inch SAS HDDs and PCIe-based SSD drives with various capacities and speeds. This flexibility allows customers to choose drive type and number of drives based on their application requirements.

For example, the PowerEdge R720xd can handle the storage-intensive databases needs of the entry-level customer without demanding external storage. It supports 24 front-accessible 2.5-inch SAS drives, and provides a single server and storage platform for deploying AlwaysOn AVG databases.

The Dell PowerEdge R720/R820 supports up to 4 front loading 2.5-inch PCIe SSD drives. This innovative server architecture allows the deployment of high performing AVG databases on PCIe SSD drives of the server, and achieves better throughputs and failover/failback times at lower latencies.

The PowerEdge R720 supports up to 7 PCIe slots and offers an interface for several available storage categories such as fiber channel, iSCSI and Direct Attached Storage (DAS), making possible the installation of applications that require high-bandwidth on the same system.

By combining the processing power of the multi-core Intel Sandy Bridge processors, high performing internal storage, large levels of RAM and enough PCIe Cards, the PowerEdge 12th Generation rack servers provide an excellent platform for hosting AVG database that require high processing power, memory and storage.

Introduction to Microsoft SQL Server 2012 AlwaysOn Availability Groups (AVG)

Microsoft® SQL Server 2012 introduces the AVG feature and a number of other high-availability enhancements to provide more robust and reliable HA and DR solutions to SQL databases. AVG is the one of the important features included in Microsoft SQL Server 2012, allowing customers to implement HA and DR solutions for their business critical databases. This feature is available only with SQL Server 2012 Enterprise Edition; Standard Edition supports only SQL Failover Cluster Instances (FCI) for high availability.

AlwaysOn Availability Groups Architecture

SQL Server 2012 AVG provides flexible design choices for selecting an appropriate high availability and disaster recovery solution for specific database applications. AVG is a set of databases that can failover together to an available secondary server. It supports a set of primary databases running on a primary server, and one to four sets of corresponding secondary databases running on secondary servers.

The following are the important elements that are used in AVG implementation:

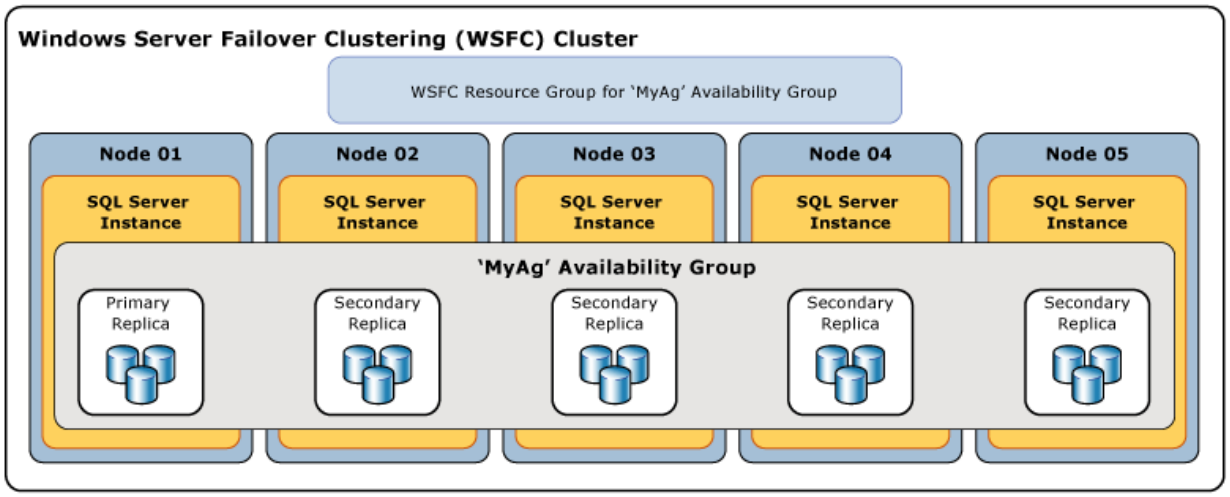
- **Availability Databases**: Group of databases that are configured as an availability group and that can failover together from a primary replica to any other secondary replica.
- **Availability Replicas:**
    - **Primary Availability Replica**: An SQL instance that hosts the availability group for read-write access. For a given Availability Group, only one primary replica is allowed.
    - **Secondary Availability Replica(s)**: An SQL instance that hosts the secondary sets of databases for the primary set of databases. For a given availability group, up to four secondary availability replicas are allowed. Secondary sets of databases are available for either read-only access or no access at all.
- **Availability Mode**: Two availability modes are supported.
    - **Synchronous-Commit Mode:** In this mode, the primary replica commits the transaction only when it receives an acknowledgement of writing (hardening) the transaction to T-Log of the secondary replicas configured for Synchronous-Commit Mode. This mode makes sure that primary and Synchronous-Commit secondary replicas are synchronized with each other. Since the primary replica has to wait for arrival of acknowledgement from the secondary replicas, this mode comes at the cost of increased transaction latency. Up to two Synchronous-Commit mode replicas (one with automatic failover and other with manual failover) are supported.

- **Asynchronous-Commit Mode:** The primary replica commits the transaction without waiting for an acknowledgement from the secondary replicas that are configured for Asynchronous-Commit mode. This mode minimizes the transaction latency, but might cause Asynchronous-Commit mode replicas to lag behind the primary replicas. This may, in turn, lead to some data loss in case of failover. Up to four Asynchronous-Commit Mode replicas (with manual failover) are supported.

- **Windows Server Failover Cluster (WSFC):** All availability replicas (primary and secondary) participating in the AVG deployment must be installed on the different nodes of the same WFSC cluster. However, it is not required that all the cluster nodes be part of AVG. Support for Multi-Subnet clusters is introduced in SQL server 2012, and allows the SQL databases to span long distances across datacenters.

Figure 1 shows an implementation of availability group that contains the maximum possible number of availability replicas, i.e. one primary replica and four secondary replicas.

Figure 1. Availability Groups on WSFC



At any point in time, AVG can have only one primary replica serving read-write requests. It is possible to configure up to four secondary replicas exclusively for read-only access.

Most common HA and DR deployments for SQL databases

Previous versions of SQL Server required that the customer combine one or more features together to create a robust HA and DR solution for their critical SQL server databases. The most typical solutions implemented with previous versions of SQL Server are:

- **Non-shared Storage Implementations**: Implementing database mirroring (with witness and automatic failover) for High Availability, and combining it with log shipping for Disaster Recovery. This method uses local storage (non-shared) in the primary and DR sites. This setup can easily be replaced by AlwaysOn Availability Groups and provides both HA and DR. This setup is a right choice for SMB customers who cannot invest more in SAN implementations.

- **Shared Storage Implementations**: Implementing FCI for High Availability and combining it with Database Mirroring (Asynchronous) or Log Shipping for Disaster recovery. This method uses shared storage in the primary site for HA and either shared or non-shared storage in the DR

site. This setup can be replaced by FCI for HA and AlwaysOn Availability Groups for DR. This setup is a right choice for SMB and Enterprise customers who can invest to some extent in SAN implementations.

- **Multi-site Implementations**: Implementing Multi-site FCI across datacenters for both HA and DR. Mostly this setup uses shared storage in both the sites. This method can be replaced by the Multi-site FCI and Availability Groups for HA and DR. This setup is a right choice for large Enterprise customers who have built a robust SAN implementation across multiple data centers.

For more information on the design patterns of AVG, refer to "AlwaysOn Availability Design Patterns".

Legacy HA and DR features have their own pros and cons in terms of setup complexity, transaction latencies, maximum number of secondary instances possible, readable secondary instances, automatic failover/failback times, and other maintenance activities. SQL Server 2012 AVG combines the best of legacy HA and DR features as well as providing a single feature with many enhancements in many areas as pointed out above. Some of the benefits that AVG can offer are:

- HA and DR solutions with AlwaysOn Availability Groups (with non-shared storage)
- HA and DR solutions with FCI (with shared-storage) and multi-site cluster
- Multiple secondary replicas and hence reduced overhead on primary replicas
- Automatic routing of read-only connections to secondary replicas
- Offloading backups to the secondary replicas and prioritizing secondary replicas for backup operations
- Automatic page repair and robust failure detection and correction when compared to previous versions of SQL Server
- Improved flexible failover polices and diagnostics
- Flexible, integrated, easy and automated implementation with Power shell

Dell reference configuration for AVG

In the following sections, we will discuss the different AVG implementation scenarios that are possible using Dell PowerEdge servers, incorporating internal/external storage options. We will explore the AlwaysOn AVG deployments both on non-shared and shared storage configurations. The concepts of AVG multi-site deployments are similar apart from the multi-site WSFC cluster setup. A detailed description of the multi-site configuration is outside the scope of this paper.

AVG deployments on non-shared storage

In addition to SAN storage deployments, SQL Server 2012 enables us to use non-shared storage to host databases. We can deploy AVG on local storage, or on the server's directly attached. This feature benefits the customer who cannot invest much on SAN deployments.

As discussed in the above sections, with its high performing internal storage and huge processing power, Dell's 12th Generation PowerEdge R720xd server is an excellent single platform choice for hosting AlwaysOn databases that have large processing power, memory, and high-performing storage requirements. R720xd supports a maximum of 38TB of internal storage.

The Dell H710 mini-integrated PowerEdge RAID Controller (PERC) card can be used efficiently to manage internal storage of the server. The H710P mini-card is designed for enhanced performance, increased reliability, and simplified management. For more details on the H710 PERC card refer dell.com/perc.

In our test configuration, we used fully populated PowerEdge R720xd servers in the primary site and DR site. Two RAID-10 (10 disks for each) virtual disks are carved out from 20 disks to store the database files, and one RAID 1 virtual disk from the two drives for storing T-Log files. The last two hard drives might be used for Global Hot spares as shown in Figure 2.

Figure 2. Simple AVG deployment on PE R720XD local storage

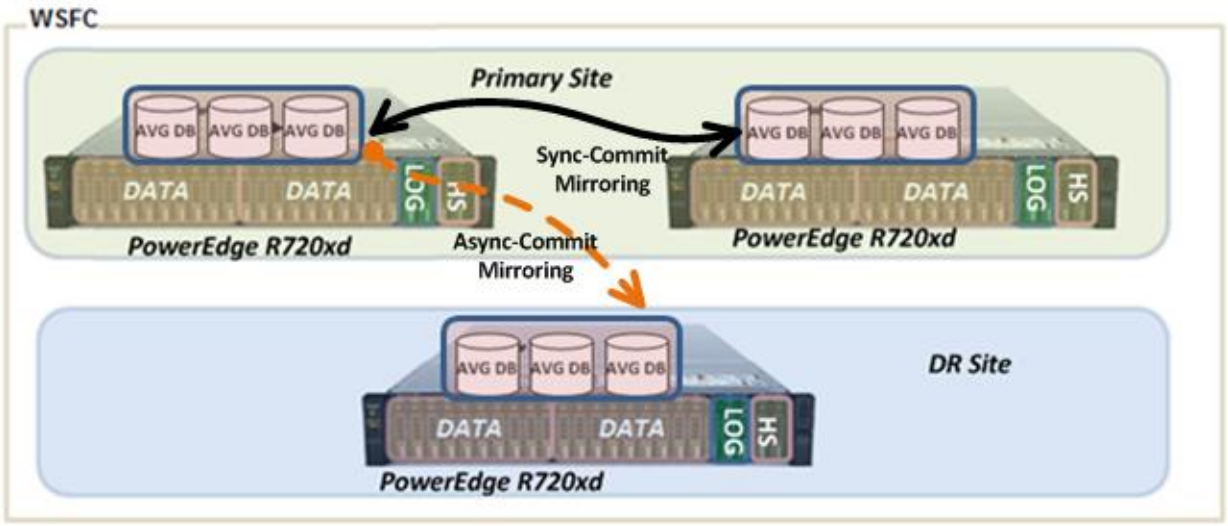


Figure 2 shows the basic AlwaysOn Availability Group deployment with two replicas configured for Synchronous-Commit mode with automatic failover to provide High Availability. Another replica was configured with Asynchronous-Commit mode to provide DR capabilities to the AVG databases. Customers can add two more replica nodes (one in the primary and one in the secondary site) for further HA and DR capabilities.

The quorum model and NodeWeight settings of the configuration should be based on the customer's specific application requirements. For more information on deploying AVGs, including prerequisites, restrictions, and recommendations for host computers, Windows Server Failover Clustering (WSFC) clusters, server instances, and availability groups refer http://msdn.microsoft.com/en-us/library/ff878487.aspx.

Major advantages and limitations of implementing AVG on non-shared storage

Advantages:

- A simple and cost effective solution for HA and DR
- Perfectly suits SMB customer requirements
- Quicker failover and failback operations because storage is locally attached
- Reduced datacenter footprint and less power consumption when compared to shared storage solutions

Limitations:

- Limits on storage scalability.
- No enhanced SAN capabilities.

AVG deployments on shared storage

In addition to the failover capability provided by AlwaysOn Availability Group, we may also set up a secondary layer of availability at the server instance level. This is done by implementing SQL Server FCI together with the underlying WSFC cluster. FCI is configured with shared storage to host the SQL server instances.

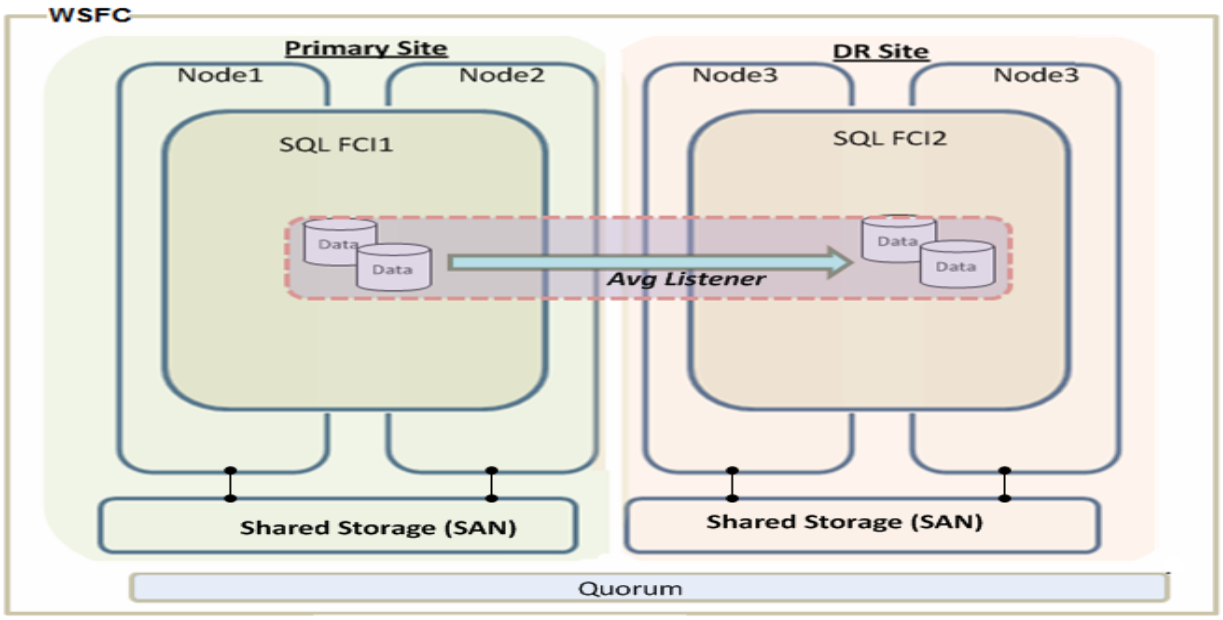
FCIs are basically designed to provide High Availability at instance level whereas AlwaysOn Availability Group (AVG) is designed to provide HA and DR capabilities at database level.

- Consider the following when you plan to combine the FCI with AVG.

- AVG alone does not require shared storage, but when we use FCI to host one or more availability replicas, each of the nodes which are part of FCIs will require shared storage as per standard SQL Server failover instance installation.
- An FCI may be used together with an availability group to enhance the availability of an availability replica. However, to prevent potential race conditions in the WSFC cluster, automatic failover of the availability group is not supported to or from an availability replica that is hosted on a FCI.
- Since SQL Server FCIs do not support automatic failover by availability groups, any availability replica that is hosted by an FCI can only be configured for manual failover.
- All the prerequisites and restrictions that need to be addressed before proceeding with AVG deployment may be found at http://msdn.microsoft.com/en-us/library/ff878487.aspx#FciArLimitations. Care should be taken to make sure all the specified hotfixes are installed on the appropriate nodes.

  Figure 3 shows the typical AVG deployment along with SQL Server FCI.

Figure 3. Scenario 1 : SQL Server FCI with AVG on shared storage



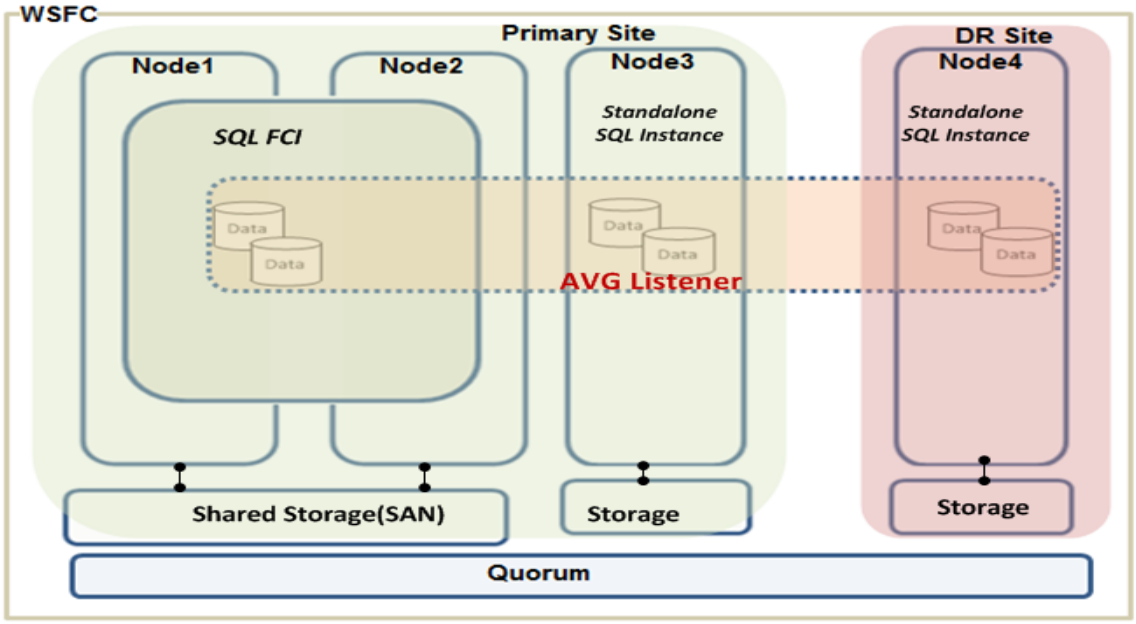
In Figure 3, Node1 and Node2 are deployed in the primary site providing HA for the databases with the FCI (namely, SQL FCI1). Node3 and Node4 are implemented in the DR site. An FCI (namely, SQL FCI2) is installed on Node3 and Node4 in the DR site in order to provide High Availability at the DR site too. So we now have one FCI in the primary site serving read-write access to the clients, and another FCI at the DR site for disaster recovery purpose. Now AVG can be configured over two FCIs (i.e. over "SQL FC1" and "SQL FC2") and may be accessed through the AVG listener.

The major aspects of the configuration described in Figure 3 are:

- HA is achieved at both the primary and DR sites.
- Should the primary site fail completely (both nodes in primary site), a manual failover has to be triggered at the DR site (this is because Automatic failover is not supported when FCIs are deployed with AVG).
- Only a single SQL Server instance is active at either of the sites (primary and DR). With FCI, only one node will be actively servicing the user load; the other node will be idle (passive), and we will be unable even to access the databases on the passive node.

In order to further enhance the resource utilization of this setup, we can configure Node3 to be part of the primary site and use it as a secondary replica with Synchronous-Commit mode. Node4 will act as another secondary replica with Asynchronous-Commit mode in the DR site as shown in Figure 4.

Figure 4. Scenario 2 : SQL Server FCI with AVG on shared storage



The advantage of the setup described within Figure 4 is, with the help of the read-only routing list, we may redirect or offload all read-only requests to Node3 and thereby reduce the overhead on the primary replica. Clients can see the latest data without any delay on Node3 as it is configured as Synchronous-Commit replica with primary replica. This is achieved at the cost of compromising the High Availability at the DR site. On the other hand, in the scenario described in Figure 3, we do not have a secondary node to serve the read-only requests in the primary site. However, we do have High Availability at the DR site. So it is up to the customer to choose which scenario best suits their requirements.

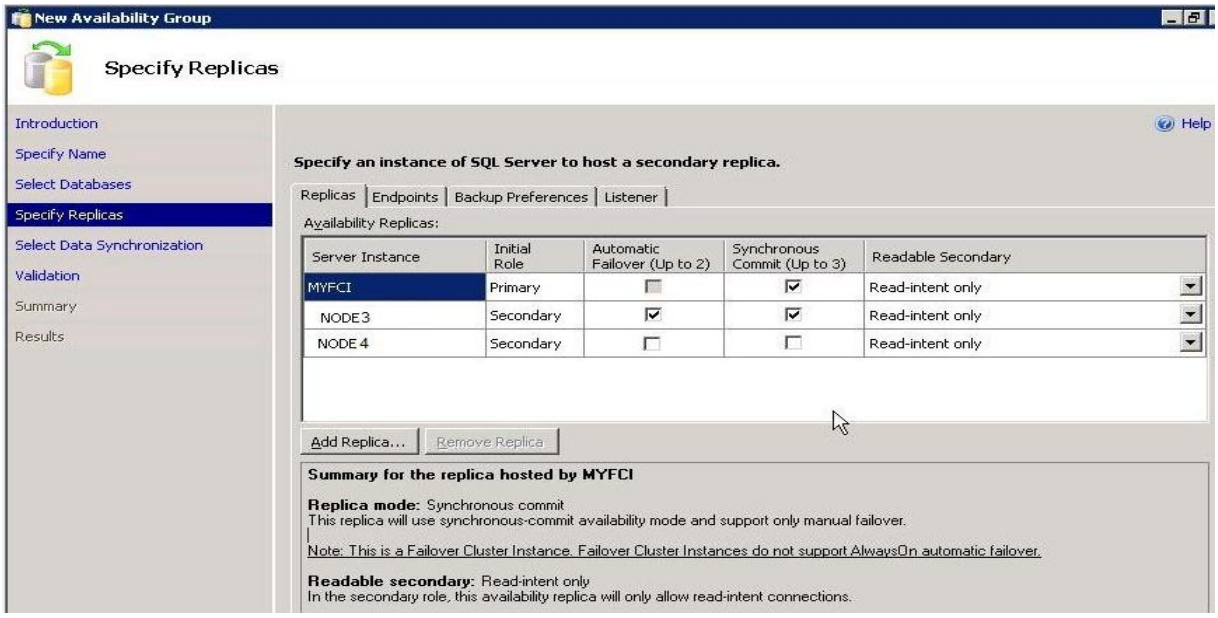
In the scenario illustrated in Figure 4, we have installed the SQL FCI on Node1 and Node2 in the primary site. The main purpose of this FCI instance is to provide High Availability to the databases with automatic failover capability, and this is available for access as long as any of the two physical nodes is up and running. Now the Availability Groups can be configured over the FCI (SQL FCI running on Node1 and 2) and the standalone instances of other two nodes (Node3 and Node4).

Since this is a 4-node cluster, we can configure this cluster with the “Node and Disk Majority” quorum setting. When disk witness is online, this quorum setting can sustain a maximum of any two node failures. The advantage of using the “Node and Disk Majority” quorum setting is that even when the FCI instance (if Node1 and Node2 goes down) is not available, we can still access the AVG databases through AVG listener on Node3 without needing to failover to DR site. Node4 is still available for manual failover for disaster recovery purposes.

> NOTE: In this configuration, automatic failover is achieved with the FCI instance which might be running on Node1 or Node2 at any point in time. Since the FCI instance does not support automatic failover/failback when we combined it with Availability Groups, we may have to do a manual failover to Node3. Since Node3 is Synchronous-Commit replica, failover can happen without any data loss.

We validated the configuration outlined in Figure 4 using four PowerEdge R720 servers. Figure 5 shows the replica configuration window of the New Availability Group wizard.

Figure 5. Creating AVG with FCI



In Figure 5, MYFCI is name of the FCI (running on either Node1 or Node2) and is configured as the primary replica with Synchronous-Commit mode. Node3 is configured as secondary replica with Synchronous-Commit mode in the primary site. Node4 is configured as another secondary replica with Asynchronous-Commit mode at the DR site. The readable secondary option has been set to "Read-Intent Only" for all the nodes. The Automatic Failover option has been disabled for MYFCI instance as automatic failover is not supported when FCIs is implemented with AVG.

Another important setting to be considered is NodeWeight. In order to avoid the impact of failures in the DR site, we may set the "NodeWeight" parameter of Node4 to zero so that Node4 will not participate in cluster quorum voting. When we set the "NodeWeight" of Node4 to 0, we may have to change the quorum configuration to "Node Majority". In this case only first three nodes (on the primary site) will take part in quorum voting. But this configuration can sustain only 1 node failure.

More configuration details of this setup are shown in Figure 6.

Figure 6. FCI with AVG deployment setting

Node Name	SQL Instance Name	AVG Name for access when the given node is Primary	AVG Name for access when the given node is Secondary	Cluster NodeWeight	Site	AVG Commit Mode	SQL instance Auto Failover/FailBack	Data Loss
Node1	MyFCI (FCI)	MY_FCI_AVG_Lis	MyFCI	1	Primary	Synchronous	Only Between Node1 & Node2	No
Node2				1	Primary	Synchronous	Only Between Node1 & Node2	No
Node3	Node3 (standalone)	MY_FCI_AVG_Lis	Node3	1	Primary	Synchronous	Manual Failover to/from Node1/2/4	No
Node4	Node4 (Standalone)	MY_FCI_AVG_Lis	Node4	1/0	Secondary (DR)	Asynchronous	Manual Failover to/from Node1/2/3	Yes

As shown in Figure 6, there will not be any data loss as long as the AVG databases are failover/failback among the Node1, 2 and 3. There may be a data loss, however, when AVG databases are failover to Node4 in DR site as Node4 is configured as Asynchronous-Commit mode.

Advantages and limitations of implementing AVG on shared storage

Advantages:

- Cost effective HA and DR solution for Enterprise Customers
- Instance level protection using FCIs
- Support of multi-subnet FCI clustering enables a DR solution using AlwaysOn with multi-subnet FCI
- Flexible failover policy for granular trigger events for automatic failovers
- Support for variety of storage types and configurations at each site (iSCSI, FC, SAS etc.)
- May benefit from enhanced SAN features

Limitations:

- No Automatic failover/failback support for FCI when it is combined with AVG functionality

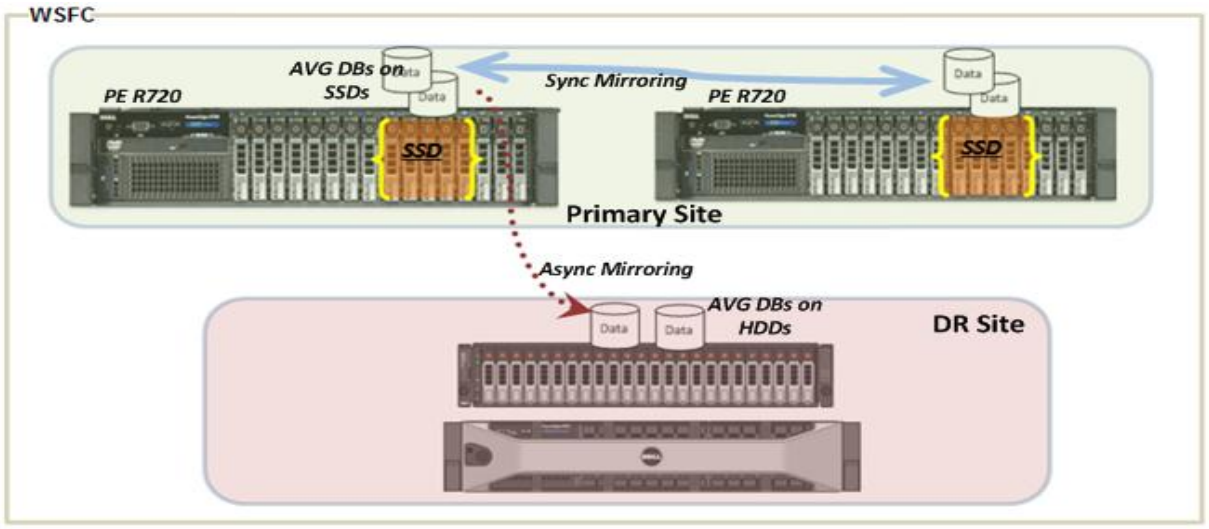
AVG deployments on hybrid (HDD and SSD) storage

In HA and DR solutions, it is always a good practice to use high performing drives such as SSDs in the primary site to get faster failover benefits and high performing databases within the primary site. Dell 12th generation servers (such as the R720 and R820) come with support for front-loading PCIe SSD drives.

Benefits

Dell™ PowerEdge R720 and R820 servers support up to four 2.5-inch 350G front-access PCIe SSDs drives. Figure 7 shows a possible deployment of AlwaysOn Availability Groups on the PCIe SSDs drives of Dell PowerEdge servers.

Figure 7. AVG deployments on SSD and HDD



In the above sample implementation, two PowerEdge R720 servers in the primary site were deployed with PCIe SSD drives. The databases that are part of the AVG are stored on the SSD drives. Remaining databases may be stored on the traditional hard drives. These two servers should be configured in synchronous commit mode with automatic failover. In order to avoid network bottleneck between the hosts, make sure that the enough network bandwidth is deployed between the hosts. This will help to achieve high performing AVG databases, with fast failover/failback capabilities.

As per the sample implementation shown in Figure 7, the DR site may have less costly hard drives (and/or storage) to host the secondary replica. The DR replicas should be configured in Asynchronous commit mode. This implementation will help the customer reduce the overall cost of the entire HADR configuration.

Advantages and limitations of implementing AVG on hybrid (SSDs and HDDs) storage

Advantages:

- Simple and cost effective solution for SMB customers who cannot invest in SAN
- Improved DB performance with PCIe SSDs
- Quicker failover and failback times compared to other solutions

Limitations:

- Limited storage scalability
- No enhanced SAN capabilities

Flexible failover policies and enhanced diagnostics for detecting failover conditions

SQL Server 2012 introduces more flexible failover polices and a rich set of diagnostics that enable WSFC to detect failures and take required actions as soon as possible. The SQL Server setup periodically reports a set of component diagnostics to the WSFC resource group. The WSFC resource group maintains the failover policy which defines the failure conditions that trigger restarts and failovers.

SQL Server setup uses sp_server_diagnostics to collect the health status of the SQL Server components. The diagnostic information that is collected by sp_server_diagnostics includes System, Resource, Query process, io_subsystem and Events. The first three components' information is used for failover detection, while the last two components' information is used for diagnostic purpose only.

Failure conditions are set on an increasing scale. For levels 1-5, each level includes all the conditions from the previous levels in addition to its own conditions. This means that with each level, there is an increased probability of a failover or restart. 3 is the default setting for AVG and FCI. We use either Cluster Management tools or T-SQL scripts to set the failover condition for FCI and AVG. For more information about the SQL server flexible failover polices and diagnostics, refer http://msdn.microsoft.com/en-us/library/ff878664.aspx

Dell conducted a variety of tests to determine how effectively WSFC and SQL Server work together to identify the failover condition and take appropriate action. We observed that an AVG took a maximum

of 5-10 seconds based on a different range of tests to successfully failover to the other nodes and re-establish availability.

Summary

AlwaysOn Availability Groups is a new HADR feature designed to meet ever-increasing RTO and RPO requirements. It is aimed to provide more granular control to achieve HA and DR of the SQL databases.

Dell 12th Generation PowerEdge servers provide the right platform for implementing the AVGs by supporting the tremendous processing power with E5 series processors, large memory (with 1600Mhz DIMMS) and high performing local storage. Dell PowerEdge servers and their close integration with a variety of storage platforms make an outstanding platform for implementing AVGs with FCIs.

Suited for deployment in a wide variety of environments, SQL Server 2012 AVG provides the customer with a high level of flexibility. The robust and reliable nature of Dell PowerEdge servers makes them a good choice for AVG implementations. The major common types of AVG implementations discussed are as below.

- **Non-shared Storage implementations**: The PowerEdge R720xd provides a unique single server platform for deploying AVG, with great internal capacity and high processing power. This configuration can benefit the SMB customers who cannot invest more on SAN.
- **Shared Storage Implementations**: Together with the wide variety of external storage solutions available from Dell (such as Compellent, EqualLogic, and PowerVault), the PowerEdge servers provide an enterprise level server and storage combination for deploying AVG on shared storage.
- **PCIe SSD Advantage**: With PCIe based SSDs on Dell PowerEdge servers (eg. R720, R820), customers can enjoy high performing and cost effective HA and DR solutions. These configurations achieve fast failover capabilities for the AVG databases.