by Michael Noel

The Essential Guide to
Achieving High Availability for SharePoint Data

The growth of SharePoint around the world has exposed one of the major infrastructure challenges of the application: that there is no inherent two-way data replication engine built into the tool. Because of this, organizations face a choice between centralizing their SharePoint data and having users access it across high-latency or low-bandwidth links, or dealing with a decentralized SharePoint content structure, which often results in duplication of data, outdated content, slow end user performance, operations outages, regulatory risks, and discontented users.

This essential guide focuses on understanding the various approaches organizations have taken in dealing with the explosion of SharePoint content across their organizations, and how to best distribute their content so that everyone can work with the most up-to-date content in the most reliable, efficient manner. We'll focus on some of the built-in options such as new ones provided within SQL Server 2012, and also compare those options to third-party replication approaches, including information on how these approaches can be used to achieve high availability and business continuity objectives.

Understanding SharePoint High Availability Challenges with SharePoint

SharePoint products and technologies have fast become a preferred method of providing for advanced document management for many organizations. Critical records and documents are stored within it, and are managed and made easily searchable within the application. As SharePoint envi-

ronments become more widely used, their importance increases exponentially as well. No longer is SharePoint simply a "dumping ground" for old documents. In many organizations SharePoint has become a mission-critical application for full enterprise collaboration.

As such, mission-critical applications typically need to be built to be both fault tolerant and highly available, and SharePoint is no exception. Building high availability into SharePoint has multiple challenges, however, because it is a complex application with various operation tiers that need to be factored into the discussion.

High availability does not just involve building redundancy into the SharePoint hardware. It also involves architecting a SharePoint environment that allows for relevant data to be close to the users that need to access it, and providing for multiple copies of this data to provide for failover in the event of an outage. The native architecture of SharePoint and the fact that a single SharePoint farm can't easily span across multiple sites can make this a challenge, however.

The first challenge with high availability design for SharePoint has to do with the fact that SharePoint itself operates at three very distinct layers, each with their own operational requirements, as illustrated in Figure 1.
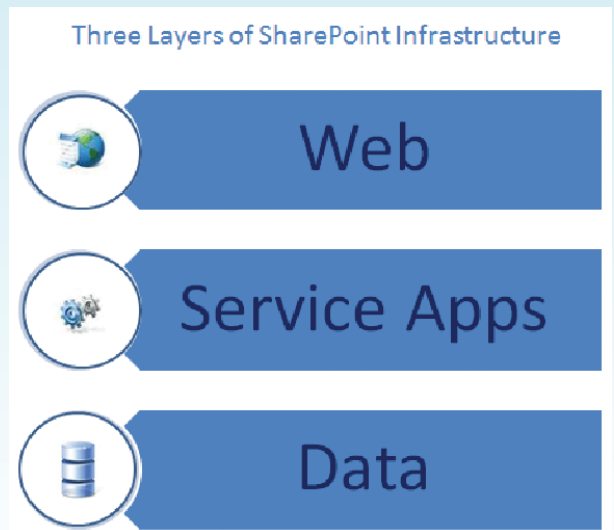

Figure 1 – Three Layers of SharePoint Infrastructure

Exploring Native High Availability Options with SharePoint

The Web Tier handles the traffic from the users, and it runs on Internet Information Services (IIS). It translates all of the user-based traffic directly into the platform, allowing users to browse pages, upload and download documents, and view reports. High availability at the Web Tier typically involves the use of multiple servers in a SharePoint farm all running as Web Front-end Servers. A load balancer distributes user requests to these servers to optimize response times and offer redundancy in the event of a server failure. While in many cases this may be hardware-based load balancers, in some cases, this may also in-

volve the use of a software load balancer such as Windows Network Load Balancing, included in the Windows Server operating system. Windows Network Load Balancing can be a challenge to implement, however, and it is often best to consider the use of a good hardware-based load balancer, many of which are getting considerably cheaper over the years.

The middle tier of SharePoint is known as the Service Application Tier, and it stores shared services such as Search, the Managed Metadata Service, Excel Services, the User Profile Sync Service, and much more. This tier is the most diverse, because each Service Application has its own operational requirements and the majority of them have their own unique high availability concerns. For the most part, high availability at this tier can be achieved simply by running a service application on more than one server in a farm, though the major exceptions to this include the User Profile Sync Service Application and the Search Administration component. But, in general, if you set up at least two servers in a farm, and turn on each service application on both servers, SharePoint will automatically use both servers. In the event of an outage, the second server will then continue to operate, preserving the service application functionality for that farm.

The Search service application is one that requires some additional thought into high availability architecture, however. Simply turning on the service on two servers will not provide for automatic high availability. Instead, you need to make use of the ability to create multiple Index Partitions, query components, and crawlers and split them across the servers running the service applications, similar to what Figure 2 shows.
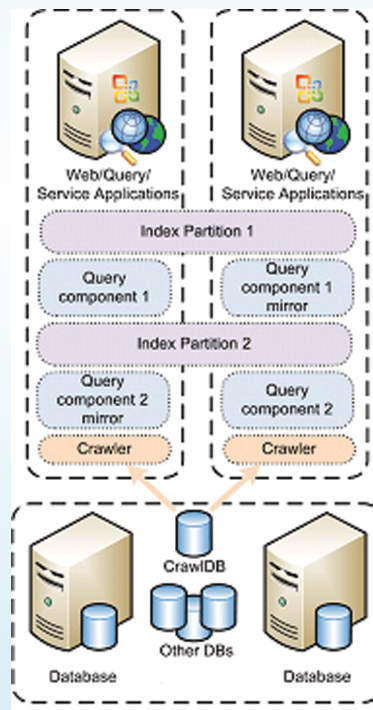

Figure 2 – Search Service Application High Availability

The final tier of SharePoint is the Data Tier, which runs on Microsoft SQL Server. SharePoint itself is highly dependent on this tier, as

all information stored in a SharePoint site is stored in SQL Server databases. Indeed, not only the content itself but also most of the information stored by the various Service Applications is stored on the SQL Server. High Availability at this tier has traditionally taken place via 'traditional' shared storage clustering models that involves two or more SQL Servers that share a common set of databases, through the use of SQL Server 2012.

Using new SQL Server 2012 High Availability Options such as AlwaysOn Availability Groups for SharePoint

The biggest recent change to high availability design for SharePoint farms is related to the recent release of Microsoft SQL Server 2012. The options provided with this version of SQL Server (Table 1) are significantly more sophisticated than some of the prior options, and provide for some fascinating new design options for SharePoint environments.

As indicated in Table 1, vastly different levels of Service Level Agreements (SLAs) are provided by the various Data Tier options. Traditional Backup and Restore options, for example, may end up with data loss that totals hours or days, and may take weeks to bring back online, while some of the newer options provide for zero data loss and failover within seconds. Subsequently,

it is critical to carefully consider what the preferred Data Tier availability options for an organization will be.

The most significant new feature added in Microsoft SQL Server that changes the design paradigm for SharePoint administrators is a feature known as AlwaysOn Availability Groups (AOAGs). AOAG technology is a combination of SQL database mirroring technologies, where exact copies of databases can be made on other servers, together with clustering technologies that allow for automated failover. By combining these technologies, AOAGs allow SharePoint design architects to create multiple redundant copies of SharePoint databases, allowing for up to five total copies of the content.

Key features for AOAGs for a SharePoint environment include the ability
- to create synchronous replicas of SharePoint databases on up to three servers, with automatic failover between two of the replicas,
- to create up to four asynchronous replicas of SharePoint content databases, providing for the ability to direct content to remote sites,
- to set some of the replicas as 'read-only' replicas, creating scenarios where that data is connected to a 'Read Only SharePoint farm' in a remote location,
- to provide for automated failover within five to seven seconds, of all SharePoint content in a farm.

High Availability Options for SQL Server	Potential Data Loss (RPO)	Potential Recovery Time (RTO)	Auto-Failover	Additional Readable Copies	Limitations for SharePoint HA
AlwaysOn Availability Groups-Synchronous (Dual-phase commit, no data loss, can't operate across WAN	None	5-7 Seconds	Yes	0-2	Cannot work across low bandwidth/high latency networks
AlwaysOn Availability Groups- Asynchronous (Latency tolerant, cross WAN option, potential for data loss)	Seconds	Minutes	No	0-4	Cannot be used for SharePoint Service Application and/or Config databased
AlwaysOn Failover Cluster Instance (FCI) - Traditional shared storage clustering	N/A	30 Seconds to several minutes (depending on disk failover)	Yes	N/A	Cannot work across low bandwidth/high latency networks. Only one copy of database files.
Database Mirroring - High-safety (Synchronous)	None	5-10 Seconds	Yes	N/A	Cannot work across low bandwidth/high latency networks
Database Mirroring - High-performance (Asynchronous)	Seconds	Manually initiated, can be a few minutes if automated	No	N/A	Cannot be used for SharePoint Service Application and/or Config databases
SQL Log Shipping	Minutes	Manually imitated, can be a few minutes if automated, by typically hours	No	Not during a restore	Cannot be used for SharePoint Service Application and/or Config databases
Traditional Backup and restore	Hours to Days	Typically multiple hours, days or weeks	No	Not during a restore	Not a high Availability technology

Table 1 – Comparison of High Availability Options for SQL Server

This creates new design options for SharePoint farms, which are illustrated in Figure 3.

Figure 3 illustrates a sample design for AOAGs that takes full advantage of all of the features of the technology. In this ex-

ample, a primary datacenter is configured for automated high availability of the SharePoint Data Tier, with synchronous copies of all databases configured to automatically failover within five to seven seconds between the primary SQL server and
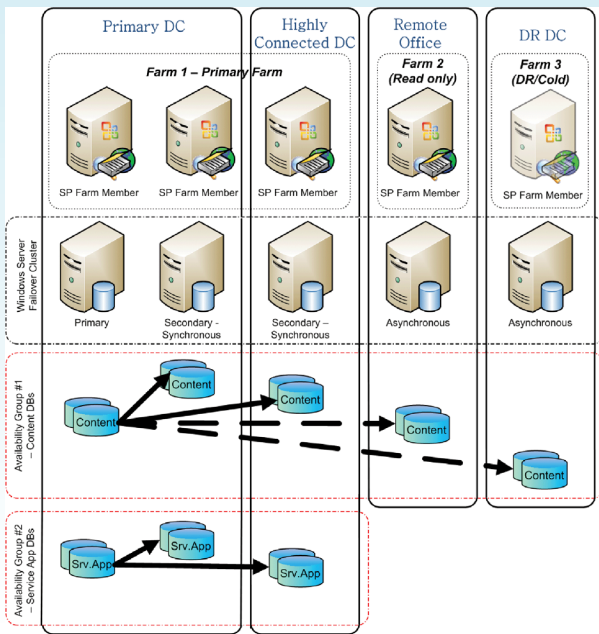
Figure 3 – AOAG Design Options for SharePoint

the secondary replica. In addition, a third synchronous copy of the content is set for manual failover to a highly connected data-center located close by. For synchronous AOAGs to work properly, the connectivity must be very high (1 GB or greater) and the latency must be very low (10ms or less.)

In this example, two additional asynchronous replicas of content database are created in remote datacenters. The reason that only content databases have these fourth and fifth replicas created has to do with the fact that only the Secure Store database and the content databases of SharePoint are supported for asynchronous replicas;

all other SharePoint database types support only Synchronous replicas. In this model, the first asynchronous replica is created in a remote site that is configured with a Read-Only SharePoint farm that is configured to allow users in that location to have a local readable copy of their SharePoint content. The final datacenter has an asynchronous copy put in place to provide for a location to failover the SharePoint content to in the event of a disaster.

This model illustrates an example of how you can use the SQL AOAGs to improve some of the design options, and provide for levels of local high availability that were previously unattainable. However, with every new solution comes a set of limitations, which are important to note for AOAGs. These include the following:

• While the read-only replicas allow for a remote site to have faster access to SharePoint content, a separate URL must be given to the users to access the read-only copy vs. the full read-write copy in the home office. This can lead to confusion, because user's must understand to go to http://readonly-sharepoint.companyabc.com when they want fast read access to SharePoint but go to http://readwritesharepoint.com-panyabc.com when they want to make changes to content.

- Using AOAGs with SharePoint require the most expensive Enterprise editions of both SQL Server and Windows Server 2008 R2 or 2012.
- Service Application functionality can only be replicated synchronously; the asynchronous replicas in remote locations cannot have the critical service application functionality replicated to them.

Examining Third-Party Replication Models

Because of the limitations inherent in SharePoint high availability options, many clients over the years have opted for approaches that use application-layer replication technologies to be able to provide for high availability for SharePoint content. These technologies typically operate at the SharePoint API level, looking for new documents and content within a SharePoint site, and then sending that content to another active farm in a different location. Every time a document is added or modified in SharePoint, a copy of that document is then replicated to the multiple farms within the organization.

In many cases, these multiple farms are configured to take advantage of global load balancing modules that allow for a single URL to be used for all of the farms. That way, if a user clicks on a link to a SharePoint site, the load balancer determines what the closest replicated farm is and sends the user to that farm. Since all content is kept in sync via the third-party replication tools, this allows for a seamless experience for the end users, and keeps content closer to the users, while allowing for multiple live copies of the content to exist in multiple locations. For a high availability perspective, if a local farm is down due to an outage or for maintenance, the load balancers can also send the users to a different farm in a remote location, preserving availability. These options have been successfully implemented in various forms throughout the years, and are especially interesting from an architecture perspective as they provide for concepts unavailable with the out-of the box tools available with SharePoint or SQL Server.

Think Through the Process

Designing for high availability in a SharePoint farm can be a complex process. SharePoint architects need to take into account availability at the Web Tier, the Service Application Tier, and the all-important Data Tier. New options in SQL Server 2012, including AlwaysOn Availability Groups, can improve an architect's design options, while other organizations may choose to look at the flexibility and robust high availability models that are provided by third-party replication models instead. In any

case, because of the importance of Share-Point in most organizations, proper thought and care should be put into the design and implementation of high availability for SharePoint farms.

ABOUT THE AUTHOR

Michael Noel, is an internationally recognized technology expert, best-selling author, and well known public speaker on a broad range of IT topics. He has authored several industry books that have been translated into over a dozen languages, with sales exceeding 500,000 copies worldwide. Significant titles include *SharePoint 2013 Unleashed, Windows Server 2012 Unleashed,* and *Exchange Server 2013 Unleashed*. Currently a partner at Convergent Computing in the San Francisco Bay Area, Michael's writings and extensive public speaking experience across over 100 countries and all seven continents leverage his real-world expertise helping organizations realize business value from Information Technology infrastructure.