

Supporting geographically dispersed Windows Server® 2008

Failover Clustering with SteelEye DataKeeper for Windows

Overview

SteelEye DataKeeper is a highly optimized host-based replication solution which integrates seamlessly with Windows Server 2008 Failover Clustering. New features of Windows Server 2008 Failover Clustering such as cross-subnet failover and tunable heartbeat parameters makes it possible for administrators to deploy geographically dispersed clusters. SteelEye DataKeeper provides the data replication mechanism which extends Windows Server 2008 Failover Clustering, allowing administrators to take advantage of these advanced features to support high availability and disaster recovery configurations.

A Brief History of Geographically Dispersed Clusters

In 1997, Microsoft® entered the high availability space with Microsoft Cluster Services (MSCS, aka Wolfpack) which supported Windows NT. Traditionally MSCS relied on shared SCSI, then shared fiber channel and iSCSI SAN as storage options matured. Before Windows Server 2008, the options for deploying geographically dispersed clusters with MSCS were limited. These options involved very specific pre-qualified array based replication technology and were also restricted to certain environments due to the latency requirements of MSCS and the requirement that the cluster nodes reside in the same subnet. This made deploying geographically dispersed clusters challenging.

Some customers facing these challenges turned to 3rd party software vendors with competing technology that included data replication and some level of application and/or system level monitoring and failover. In many cases, these solutions were geared towards the most popular business critical applications such as Microsoft Exchange and SQL Server as well as BlackBerry Enterprise Server and VMware Virtual Center. SteelEye LifeKeeper is one such example.

The release of Microsoft Exchange Server 2007 included "Continuous Cluster Replication" (CCR) which marked Microsoft's first foray into "shared-nothing" clustering - clustering without the use of a shared storage array. CCR is what is known as application-based replication where the application itself provides the replication technology. To date, Exchange 2007 CCR is the only native replication technology provided by Microsoft which works with Microsoft clustering technology.

What's New in Windows Server 2008 Failover Clustering

Windows Server 2008 Failover Clustering (WSFC) is the new name for Microsoft's failover clustering solution. A new name was needed in order to differentiate Windows "failover" clustering from another new "high performance" computing clustering technology offered by Microsoft called Windows HPC Server 2008.

The new features included in WSFC make it the most robust and easy to use failover clustering solution offered by Microsoft to date. These new features, described below, minimize the challenges with deploying geographically dispersed clusters.

Cluster Validation Tool

Perhaps the most significant enhancement made by Microsoft in WSFC involves the elimination of the Cluster Hardware Compatibility List (HCL). Prior to WSFC, all hardware used in a supported MSCS cluster had to be certified and appear on the Cluster HCL. With the release of WSFC, the Cluster Validation Tool eliminates the need to refer to the Cluster HCL when building clusters. Instead, the Cluster Validation Tool performs "tests to determine whether your system, storage, and network configuration is suitable for a cluster".

The Cluster Validation Tool enables users to choose the hardware and software that best meets their needs and allows them to self-certify a cluster configuration without having to wait for an official Microsoft validation test.

New Quorum Models

Leading up to the release of WSFC, Microsoft introduced new quorum options including the majority node set (MNS) and the use of a File Share Witness (FSW) in a MNS. These innovations were the predecessors to the four quorum models introduced in WSFC:

- No Majority – Disk Only
- Node Majority
- Node and Disk Majority
- Node and File Share Majority



Replicate Any Data. Protect Any Application



SteelEye
TECHNOLOGY INC

www.steeleye.com

Specifically, the “Node and File Share Majority” model is what facilitates geographically dispersed clusters. In this model, a two node cluster stretched across a WAN can be supported by identifying a simple File Share as a “witness” to the cluster. The file share and the two cluster nodes all get a vote and the cluster will come online as long as a majority of votes is achieved.

Different Subnets

Prior to WSFC, cluster nodes all had to reside in the same subnet. This complicated deploying geographically dispersed clusters due to the network changes required to span a subnet across WAN links with a VLAN. In Windows Server 2008, the introduction of “OR” logic allows the use of two IP addresses, which means that these IP addresses can reside in different subnets across routed networks, eliminating the need to create a VLAN.

Configurable Heartbeat Timeout

In WSFC, the round-trip latency restriction of <500ms has been lifted. Heartbeats are now tunable so that high latency networks are now supported when deploying geographically dispersed clusters.

What is SteelEye DataKeeper

SteelEye DataKeeper is a highly optimized replication engine with advanced features that ensures your data is replicated quickly and efficiently. Some of the features are as follows:

- Host-based synchronous or asynchronous block level volume replication
- True Continuous Data Protection (CDP) allows any point in time rewind
- Built in WAN optimization enables SteelEye DataKeeper to “pack the pipe” of a high speed/high latency network connection without the need for WAN accelerators
- Compression algorithms make efficient use of available bandwidth
- Bitmap-based intent logging eliminates the need for complete re-syncs
- Bandwidth-throttling provides complete control over replication traffic
- Intuitive wizard-driven MMC 3.0 User Interface
- Supports multiple targets across LAN/WAN for cascading failover configurations
- Extends traditional 2-node shared storage clusters to a 3rd node for disaster recovery

Below we take a deeper look at these features, illustrating why they are important considerations when choosing a data replication solution.

Block Level Replication

SteelEye DataKeeper replicates data at the block level, ensuring that data is replicated as quickly and efficiently as possible. Implemented as a Windows filter driver, SteelEye DataKeeper sits between the file system and the data volume. By sitting below the file system, open files, locked files, NTFS permissions, encrypted files, etc. pose no problem. SteelEye DataKeeper simply sees these files as blocks of data that need to be replicated, just like any other block of data. By removing the need to deal with these issues, SteelEye DataKeeper is able to go about its primary job of real-time replication without adding additional overhead to the host system.

Synchronous or Asynchronous

With two modes of replication, SteelEye DataKeeper is able to offer the most flexibility when choosing the optimal data protection configuration. While choosing synchronous replication guarantees that the target system is always in sync with the source, there is a trade off; the source system must wait for each write to complete on the target volume before it can complete on the source volume. This delay is particularly noticeable on high latency networks and is therefore only recommended on high speed connections with minimal latency, or where absolute data protection exceeds the importance of application performance.

Asynchronous replication, on the other hand, allows writes to complete on the source volume while at the same time sending the write to the target system. SteelEye DataKeeper ensures write order integrity by employing a queue on the source system in time of excessive disk writes. The benefit of asynchronous replication is that it allows replication across high-latency WAN links without impacting the performance of the source server. The downside is that in an unexpected failure of the source system, any data in the queue will not be available on the target system.

WAN Optimization

When pushing data across a high latency network, one must deal with the effect that latency has on the ability to use the available bandwidth. This is especially true once you introduce a combination of high speed and high latency on the same link. SteelEye DataKeeper has been built from the ground up with this requirement in mind – it must not be affected by latency. So, regardless of the latency, SteelEye DataKeeper is able to use up to 90% of the available bandwidth without the use of hardware-

based WAN accelerators. In addition, SteelEye DataKeeper offers nine levels of compression, which in some cases results in a 5:1 compression ratio, allowing even more efficient use of the available bandwidth.

Figure 1 illustrates the effective throughput of SteelEye DataKeeper in asynchronous mode at different WAN speeds with compression turned off (0) and also with a medium compression setting of 3. It also illustrates how compression is affected by different data types, with some data being more compressible than others. The latency on these WAN links has no effect on the throughput, as latency settings of <1 ms, 32ms, 55ms and 128ms all show the same results.

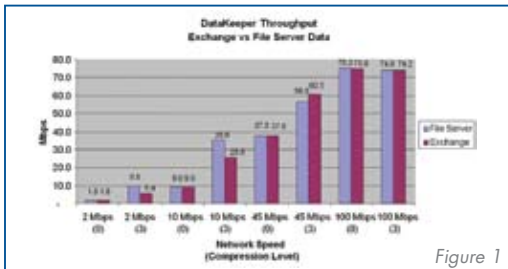


Figure 1

Bitmap File

When replicating data across a WAN connection, it is critical to ensure that once the data is in sync, it is never necessary to re-sync the entire data set again. Large data sets can take hours to resync across slow WAN links. SteelEye DataKeeper implements a persistent bitmap which ensures that a complete re-sync of the data is never necessary, regardless of the type of failure. In the event that the target of a mirror becomes unavailable, the source of the mirror tracks the blocks that change in the bitmap.

The bitmap is created during mirror creation and is a fixed-size file stored in memory and on any non-mirrored volume on the source system. While in a non-mirroring state, the blocks that change on the source are marked as "dirty" in the bitmap. Once communication is re-established to the target(s) of the mirror, only the dirty blocks indicated in the bitmap are transmitted to the target.

Continuous Data Protection (CDP)

Storage Networking Industry Association (SNIA) has this to say about CDP Continuous data protection (CDP) is a methodology that continuously captures or tracks data modifications and stores changes independent of the primary data, enabling recovery points from any point in the past. CDP systems may be block-file or application-based and can provide fine granularities of restorable objects to infinitely variable recovery points. So, according to this definition, all CDP solutions incorporate these three fundamental attributes:

1. Data changes are continuously captured or tracked
2. All data changes are stored in a separate location from the primary storage
3. Recovery point objectives are arbitrary and need not be defined in advance of the actual recovery

A number of recognized technological approaches deliver CDP, including block-file and application-based. Today, many vendors offer varying degrees of support and awareness of specific application and data environments. But regardless of the underlying technological approach utilized, CDP can offer faster data retrieval, enhanced data protection, and increased business continuity with lower overall cost and complexity.

SteelEye DataKeeper implements CDP as defined by SNIA by tracking each write that occurs on the target system in a circular log file. This log file enables the administrator to unlock the target volume and move back and forth through the replicated data stream until the optimal recovery point is found. This provides the highest possible Recovery Point Objective (RPO) and ensures that data can be rolled back to any previous point in time.

DataKeeper and WSFC together

With the new features of WSFC making geographically dispersed clusters easier to implement and more robust, there is still one key ingredient that is needed in order to fully enable this feature - wide-area data replication. With the exception of Exchange 2007 CCR, Microsoft is supporting 3rd party hardware and software vendors to bring this technology to WSFC. SteelEye DataKeeper is one such solution and provides the optimal balance of ease of use, features and price-performance to facilitate geographically dispersed clusters.

Example Configurations

The most basic configuration (see Figure 2) consists of a two node Microsoft failover cluster, with one node in the primary data center and the second node in a different data center at the other side of a WAN link. SteelEye DataKeeper replicates the volume resource(s) defined on the primary server to the secondary server. The volume resource can be a local attached disk, or it can be an iSCSI or fiber channel SAN attached volume.

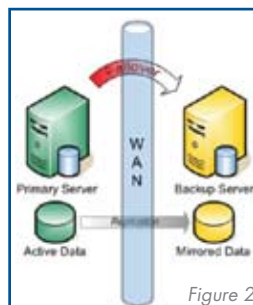


Figure 2

In the event of planned or unplanned activation of the back-up server, the replicated volume(s) will be unlocked and the mirror(s) will automatically be reversed so that any changes written to the volume on the back-up server will be automatically replicated back to the primary server once it becomes available again. Due to the use of a fixed-size bitmap, SteelEye DataKeeper is able to survive both planned and unplanned outages for extended periods of time without requiring a complete re-sync of the data.

Figure 3 shows a hybrid Shared-Storage/Replicated Storage model. In this configuration, the primary server and the back-up server are in the primary data center, attached to a shared storage device. At the same time, SteelEye DataKeeper is replicating the protected volume to the disaster recovery server at the remote site.

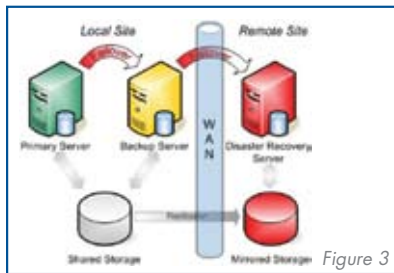


Figure 3

Should the primary server fail, the secondary server comes online and replication continues to the remote site without a lapse. Should the entire local site become unavailable, the disaster recovery server comes online and the mirror is once again reversed. Changes that occur to the volume(s) on the disaster recovery server are tracked in the bitmap and are replicated back to the primary server once it becomes available again.

SteelEye DataKeeper supports replicating to multiple targets. In this configuration, a multi-node Microsoft failover cluster can be configured with the option of having nodes spread across more

About SteelEye Technology

SteelEye is the leading provider of data and application availability management solutions for business continuity and disaster recovery for Linux and Windows and virtual environments.

The SteelEye family of data replication, high availability clustering and disaster recovery solutions are priced and architected to enable enterprises of all sizes to ensure continuous availability of business-critical applications, servers and data.

To complement its software solutions, SteelEye also provides a full range of high availability consulting and professional services to assist organizations with the assessment, design and implementation of solutions for ensuring High Availability within their environments.

SteelEye is a subsidiary of SIOS Technology, Inc. To contact SteelEye, visit www.steeeye.com or call:

US/Canada 866.318.0108

Europe + 44 (0)1223 208701

Int'l + 1.650.843.0655

SteelEye Technology, Inc. 4400 Bohannon Drive, Suite 110, Menlo Park, CA 94025

Replicate Any Data. Protect Any Application

©2008 SteelEye Technology, Inc

than one geographic location. This feature opens up multiple configuration options which allow the administrator to protect against local, regional and national outages. Figure 4 illustrates one possible scenario where a single server in New York has a local node for HA, another node in Chicago for a regional outage and fourth node in London for a national outage.

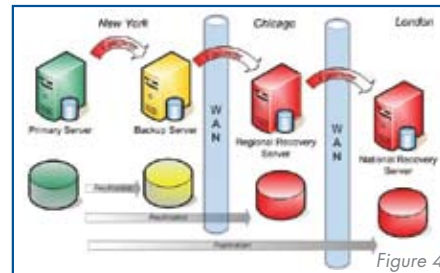


Figure 4

In the example configuration, a failure of the currently active node would cause the next available node to come into service. Recovery cascades across back-up server, regional recovery server and national recovery server until the application can be fully restored. Regardless of the node that comes into service, the in-service node becomes the mirror source and any changes are then replicated to the remaining nodes as they become available once again.

Conclusions

Microsoft has delivered an extremely powerful, flexible and easy to use enhancement to Windows Server Failover Clustering with the release of Windows Server 2008. By providing the framework for geographically dispersed clustering, Microsoft has greatly extended the possible configurations for application availability and business continuity. SteelEye DataKeeper further extends these new configuration possibilities by providing a fast, efficient and easy to use data replication engine with a set of advanced features such as sync/async replication, CDP, multiple targets, data compression and much more.



SteelEye
TECHNOLOGY INC
www.steeeye.com