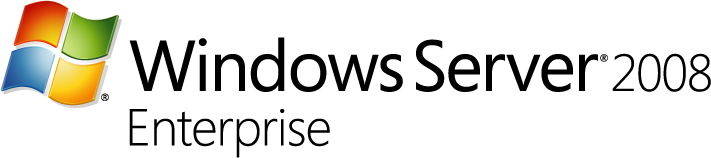
ms-logo_cgray6



Windows Server 2008 Failover Clustering Architecture Overview

New Features and Capabilities

Published: November 2007

For the latest information, please see   
<www.microsoft.com/windowsserver2008/failover-clusters.mspx>

Contents

[Introduction 1](#_Toc183600586)

[Overview of Windows Server 2008 Failover Clustering 1](#_Toc183600587)

[What’s New in Failover Clustering 2](#_Toc183600588)

[Improved Failover Cluster Management Interfaces 3](#_Toc183600589)

[The Validate Tool 4](#_Toc183600590)

[The Validate Inventory 4](#_Toc183600591)

[The Validate Verification 5](#_Toc183600592)

[A New Way to Create Clusters 8](#_Toc183600593)

[Migrating Legacy Clusters 9](#_Toc183600594)

[Support for Windows Server 2008 Server Core 10](#_Toc183600595)

[Improvements in Scoping and Managing Shares 11](#_Toc183600596)

[Better Storage and Backup Support 12](#_Toc183600597)

[Enhanced Maintenance Mode 13](#_Toc183600598)

[Superior Scalability 14](#_Toc183600599)

[A New Quorum Model 15](#_Toc183600600)

[An Improved Security Model 16](#_Toc183600601)

[CSA No Longer Requires a Domain User Account 16](#_Toc183600602)

[Improved Logging and Event Tracing 16](#_Toc183600603)

[More Secure TCP-based Cluster Communications 16](#_Toc183600604)

[Kerberos Authentication By Default 17](#_Toc183600605)

[Improved Auditing of Cluster Services Tools Usage 17](#_Toc183600606)

[Securing Intra-cluster Communications 17](#_Toc183600607)

[New Networking Capabilities and More Flexible Dependencies 18](#_Toc183600608)

[Better Support for Geographically Distributed Clusters 18](#_Toc183600609)

[Support for DHCP 19](#_Toc183600610)

[Improvements in the Cluster Heartbeat Mechanism 19](#_Toc183600611)

[Comprehensive Support for IPv6 19](#_Toc183600612)

[Conclusion 20](#_Toc183600613)

# Introduction

Organizations put a lot of value on mission-critical servers and rely on them heavily to run their businesses. As a result, server downtime can be very costly. A heavily used e-mail or database server can easily cost a business thousands or tens of thousands of dollars in lost productivity or lost business for every hour that it is unavailable. For every benefit and advantage an organization gains by an IT solution, technology and business decision-makers should also think about how to deal with the inevitable downtime of these solutions.

Server availability is a trade-off between availability and cost. The 24/7/365 pace of global commerce makes uninterrupted IT operations vital to an increasing number of industries, from financial services and logistics to manufacturing and travel and tourism. However, achieving the degree of reliability and availability demanded by mission-critical business requirements is expensive to create and support, both in terms of hardware and software costs. In addition, there is the employee time required to manage the solution. The challenge for organizations is to learn what level of IT service availability is justified by their own price of downtime.

*High availability* refers to redundancies built into an IT infrastructure that makes it available to users even in the event of a service disruption. Disruptions can be unexpected and range from anything as localized as the failure of a network card on a single server to something as dramatic (and improbable) as the physical destruction of an entire data center. Service disruptions can also be routine and predictable, such as planned downtime for server maintenance.

Failover clustering can be used as a way to achieve high-service availability. A failover cluster is a group of computers working together to run a common set of applications that presents a single logical system to client applications. Computers in the cluster are physically connected by either local area network (LAN) or wide area network (WAN) and are programmatically connected by the cluster software. These connections let services to fail over to another computer in the cluster in the event of a resource failure on one computer or its connections to the network.

Failover clusters in Windows Server® 2008 provide high availability and scalability for mission-critical applications such as databases, messaging systems, file and print services, and virtualized workloads. If a node in a cluster becomes unavailable (such as a result of failure or having been taken down for maintenance), another node in the cluster will provide service. Users accessing the service continue their work and are unaware of any service disruption.

This document describes these new features and improvements for the failover clustering in Windows Server 2008.

## Overview of Windows Server 2008 Failover Clustering

Clustering in Windows Server 2008 has been radically redesigned to simplify and streamline cluster creation and administration. Rather than worrying about groups and dependencies, administrators can create an entire cluster in one seamless step via a wizard interface. All you have to do is supply a name for the cluster and the servers to be included in the cluster and the wizard takes care of the rest. You do not have to be a cluster specialist or have in-depth knowledge of failover clusters to successfully create and administer Windows Server 2008 failover clusters. This means a far better total cost of cluster ownership for you.

The goal of Windows Server 2008 failover clustering is to make it possible for the non-specialist to create a failover cluster that works. Organizations using previous versions of failover clustering often had staff dedicated to installation and management of failover clusters. This significantly increased the total cost of ownership (TCO) for failover cluster services. With the introduction of Windows Server 2008 failover clusters, even an IT generalist without any special training in failover cluster services will be able to create a server cluster and configure the cluster to host redundant services, and the configuration will work. This means a lower total cost of ownership for you.

You will not need an advanced degree to get failover clusters working. The main reason for this change is that the new administrative interface does the heavy lifting for you. In previous versions of failover clustering, you had to learn an unintuitive, cluster-centric vocabulary and then try to figure out what those words really meant. There is no need to learn the intricacies of cluster vocabulary with Windows Server 2008 failover clustering. Instead, configuration is task based. You are asked if you want to create a highly available file server, Dynamic Host Configuration Protocol (DHCP) server, Windows Internet Name Service (WINS) server, or other type of server and then the wizard walks you through the process.

Ease of use was the number one consideration for Windows Server 2008 clustering. However, there are a number of new features and technical improvements too. The remainder of this paper discusses these new features and improvements.

## What’s New in Failover Clustering

The first thing you might notice is that Windows® clustering has a new name in Windows Server 2008—*failover clustering*. The first version of server clustering, code-named “Wolfpack,” was released for Microsoft® Windows® NT 4.0 under the official name of Microsoft® Cluster Services (MSCS). The name changed to Server Clustering in Microsoft® Windows® 2000 Server and Windows Server® 2003. The name changed again in Windows Server 2008 because of some confusion with another type of cluster: Windows® Compute Cluster Server.

A number of new services and capabilities are included in the new Windows Server 2008 failover cluster service. These include:

* Improved Failover Cluster Management Interfaces.
* The Validate Tool.
* A New Way to Create Clusters.
* Migration of Legacy Clusters.
* Support for Windows Server 2008 Server Core.
* Improvements in Share Scoping and Management.
* Better Storage and Backup Support.
* Enhanced Maintenance Mode.
* Superior Scalability.
* A New Quorum Model.
* An Improved Security Model.
* New Networking Capabilities and More Flexible Dependencies.

# Improved Failover Cluster Management Interfaces

Windows Server 2008 includes a new, easy-to-use management interface. The previous cluster administration interface has been replaced with a Microsoft Management Console (MMC) 3.0 snap-in, **CluAdmin.msc**. This new interface is accessible from within Administrative Tools. It is also possible to open a blank MMC and then add this snap-in along with any others.

The Failover Cluster Administration Console is designed to be task oriented instead of cluster resource oriented, as it was in previous versions of failover clustering. Instead of playing with “knobs and dials,” failover cluster administrators now select the clustering task that they want to undertake (such as making a file share highly available) and supply the necessary information via the wizard. Administrators can even manage Windows Server 2008 clusters from Windows Vista® client computers by installing the Remote Server Administration Tools (RSAT).

In previous versions of cluster administration, the procedure for creating a highly availability file share was a complex process. The administrator had to create a group, a disk resource, an IP address resource, resource name (Network Name), configure IsAlive/LookAlive, configure preferred services and set dependencies. There were a lot of opportunities to get something wrong.

In contrast, Windows Server 2008 failover clustering asks if you want to create a highly available file share and does all the work for you. You never have to deal with resources or dependencies. Instead, you launch the **File Share High Availability** wizard. You are then asked for a client access point name (the Network Name). You do not even need to assign an IP address, as Windows Server 2008 failover clustering supports DHCP (and DHCP addressing for resources is the default in the wizard).

What happens under the hood is that the wizard creates a group, takes a disk from available storage (which is automatically detected), and moves it to that group. It then creates an IP address resource and creates a network name resource (based on client access point entry). Then it defines resources owners, preferred owner, and dependencies automatically.

In addition, you can manage multiple clusters throughout the organization from a single MMC. And since the Windows Server 2008 MMC is a true MMC (unlike the interface available in previous versions), you can create custom management consoles that include the failover cluster snap-in in addition to other management snap-ins.

In addition to the powerful and easy to use failover cluster management console, experienced cluster server administrators may want to get full access to all the commands they had available in the command-line interface. Administrators can access all the knobs and dials that failover clustering has to offer to fine-tune their clusters by using the **cluster.exe** command-line interface. Moreover, Windows Server 2008 Failover Clusters are fully scriptable with Windows Management Instrumentation (WMI).

# The Validate Tool

Currently, clusters too often fail because of configuration complexity. To help solve this problem, Windows Server 2008 clustering comes with the built-in cluster Validate Tool. The Validate Tool is an expansion and integration of the ClusPrep tool that was released for Windows Server 2003 server clustering.

Validate runs a focused set of tests for both functionality and best practices on the servers intended to be in a given cluster as a part of the cluster configuration process. Validate performs a software inventory, tests the network, and validates system configuration.

## The Validate Inventory

The Validate inventory includes:

* **List BIOS Information**. Lists the BIOS information for all the nodes in the cluster.
* **List Environment Variables**. Examples of environment variables are the number of processors, the operating system path, and the location of temporary folders.
* **List Fibre Channel Host Bus Adapters**. Typically, a host bus adapter (HBA) is a PCI card that connects the server to the storage. These support Fibre Channel.
* **List iSCSI Host Bus Adapters**. Typically, an iSCSI host bus adapter is a PCI card that connects the server to the storage.
* **List Memory Information**. Lists the memory configuration for each node in the cluster.
* **List Operating System Information**. Gathers information about the operating system configuration on the node. Items included are operating system version, service pack level, installation date, Windows directory location, page file size and location, boot device, and other operating system information.
* **List Plug and Play Devices**. Lists all the Plug and Play (PnP) devices on each node.
* **List Running Processes**. Lists all the processes running on each node in the cluster.
* **List SAS Host Bus Adapters**. Typically, a host bus adapter is a PCI card that connects the server to the storage. These support Serial Attached SCSI (SAS).
* **List Services Information**. Lists all services installed on all nodes in the failover cluster.
* **List Software Updates**. Lists any updates that have been installed on the servers such as hot fixes.
* **List System Drivers**. Lists all the drivers that are installed on each node of the cluster.
* **List System Information**. The system information includes the following:
  + Computer name.
  + Manufacturer, model, and type.
  + Account Name of the person who ran the validation tests.
  + Domain that the computer is in.
  + Time zone and daylight-saving setting (determines whether the clock is adjusted for daylight-saving changes).
  + Number of processors.
  + List Unsigned Drivers: You can use this to help correct issues uncovered by "Validate All Drivers Signed.”

## The Validate Verification

The Validate verification includes both network validation and storage validation.

### Network Validation

#### Validate Cluster Network Configuration

This test:

* Lists the cluster networks, that is, the network topology as seen from the perspective of the cluster.
* Validates that, for a particular cluster network, all network adapters are provided with IP addresses in the same way, that is, all use static IP addresses or all use DHCP.
* Validates that, for a particular cluster network, all network adapters use the same version of IP, that is, all use IP version 4 (IPv4), all use IP version 6 (IPv6), or all use both IPv4 and IPv6.

#### Validate IP Configuration

This test:

* Lists the IP configuration details.
* Validates that IP addresses are unique in the cluster (no duplication).
* Validates that all tested servers have at least two network adapters.
* Validates that no tested servers have multiple adapters on the same IP subnet.
* Validates that all tested servers use the same version of IP; in other words, all tested servers use IPv4, all use IPv6, or all use both IPv4 and IPv6.

#### Validate Network Communication

This test validates that tested servers can communicate, with acceptable latency, on all networks.

#### Validate Windows Firewall Configuration

This test validates that the current Windows® Firewall configuration on the tested servers is compatible with failover clustering.

### Storage Validation

#### List All Disks

This test lists all disks that are visible to one or more tested servers. The test lists both the disks that can be accessed by all the servers (disks that can support clustering) and the disks on an individual server.

#### List Potential Cluster Disks

This test lists disks that can support clustering and are visible to all tested servers. To support clustering, the disk must be on a SAS, iSCSI, or Fibre Channel bus. The test also provides information about whether each disk that can support clustering has a unique ID (required for iSCSI, and recommended for SAS and Fibre Channel). In addition, the test validates that multi-path I/O (MPIO) is working correctly.

#### Validate Disk Arbitration

This test validates that:

* Each of the clustered servers can use the arbitration process to become the owner of each of the cluster disks.
* For a clustered server that owns a disk, when one or more other clustered servers arbitrate for that disk, the original owner retains ownership.

#### Validate Disk Access Latency

This test validates that for the cluster storage, the latency for disk read and write operations is within an acceptable limit.

#### Validate Disk Failover

This test validates that disk failover works correctly in the cluster. Specifically, the test validates that when a disk owned by one clustered server is failed over, the server that takes ownership of the disk can read it. The test also validates that information written to the disk before the failover is still the same after the disk failover occurs. If the server that takes ownership of a disk after failover cannot read it, the cluster cannot maintain availability of the disk.

#### Validate File System

This test ensures that the file system configured on the shared storage is supported by failover clustering.

#### Validate Microsoft-MPIO Based Disks

This test validates storage that is configured for access via multiple paths. This ensures that the multi-path software that is installed complies with the Microsoft MPIO standard and that it is configured correctly.

#### Validate Multiple Arbitration

This test validates that when multiple clustered servers arbitrate for a cluster disk, only one server obtains ownership.

#### Validate SCSI Device Vital Product Data (VPD)

This test checks to see whether the cluster storage supports unique identifiers (unique IDs) for the disks. In addition, if the storage supports unique IDs, the test validates that the ID for each disk actually is unique.

#### Validate SCSI-3 Persistent Reservation

This test validates that the cluster storage uses the newer (SCSI-3 standard) Persistent Reservation commands, not the older (SCSI-2 standard) reserve/release commands. These newer commands avoid SCSI bus resets, which means they are much less disruptive than the commands used in earlier versions.

#### Validate Simultaneous Failover

This test validates that simultaneous disk failovers work correctly in the cluster. Specifically, the test validates that even when multiple disk failovers occur at the same time, any clustered server that takes ownership of a disk can read it. The test also validates that information written to each disk before a failover is still the same after the failover.

Validate test results are HTML-based for easy collection and remote analysis. The time Validate takes to run can be just a few minutes, though this is a function of how many nodes are in the cluster and how many LUNs are exposed to the servers and may take longer as a result. The minimum number of nodes in a given cluster configuration to run Validate is two.

Running Validate is a required part of cluster creation.

**Note:** When you run Validate, some tests may not pass, but clustering may still install and function. For example, not conforming to a cluster configuration best practice (such as having only one network interface card [NIC] in each node) will raise a warning rather than an error, but the cluster would function. However, passing Validate is the standard for support for clusters in Windows Server 2008: If a cluster does not pass Validate, it is not supported by Microsoft. In addition, running validate does not release the customer from responsibility of using only hardware and software certified under the Windows Server Logo Program for Windows Server 2008.

There should be at least two servers in the cluster before running the Validate Tool. If there is only one server, storage tests that require two servers will not be run and these will be reflected in the report. Validate can also be run after the clusters are created and configured. However, storage tests will not be run on online disk resources.

# A New Way to Create Clusters

The installation process in Windows Server 2008 is fundamentally different from previous versions of Windows Server.. Roles and features (and the distinction between them) are now more important in Windows Server 2008 than before. Failover clustering is a *feature* because it makes other server *roles* highly available. You can install the failover clustering feature through the Initial Configurations Task (ICT) interface or with the Server Manager snap-in in Administrative Tools. You can uninstall clustering by the same means.

The procedure to install cluster functionality in servers has changed dramatically with Windows Server 2008. Windows Server 2008 is far more compartmentalized than Windows Server 2003. Failover cluster services are no longer installed by default as with Windows Server 2003; in Windows Server 2008, you must use the Add Feature Wizard to install the Failover Clustering feature.

Windows Server 2008 uses a *componentization* model wherein components are not added until you need them. Be aware that there may be some roles and features that will be added by default on product installation. Also note that some roles and features may be needed prior to configuring a cluster resource. For example, the DHCP server role must be installed prior to clustering the DHCP service. The uninstall procedure also uses the same model—you remove features and/or roles. The new install model is also reflected in a new directory structure, seen under **Windows\Cluster**.

Creating a cluster can be done from the failover cluster snap-in or from the command-line interface using the **cluster.exe** utility. Using the failover cluster snap-in is the preferred approach. Note that you are strongly encouraged to run Validate before creating the cluster.

When you run the Create Cluster Wizard, you’ll find that you now can enter all the members of the failover cluster at the same time. In contrast, with previous versions of failover clustering you had to create the cluster on one server and then add servers later. While you can still add servers, it was not possible to include all servers in the initial configuration in Windows Server 2003 server clustering.

After adding the servers, you will be asked for an Access Point name. This is the name of the cluster. You will also need to include an IP address for administering the cluster. This can be a static address or an address obtained via DHCP.

You click Finish in the wizard and the cluster is created for you without any more input.

# Migrating Legacy Clusters

To provide significantly enhanced security, Windows Server 2008 failover clusters sacrifice backwards compatibility with earlier clusters (and, by extension, rolling upgrade migrations). This means that Windows Server 2003 server cluster nodes and Windows Server 2008 failover cluster nodes cannot be on the same cluster. In addition, failover cluster nodes must be joined to an Active Directory®–based domain (not a Windows NT 4.0–based domain).

The process in moving from Windows Server 2003 clusters to Windows Server 2008 failover clusters will be a migration. The migration functionality can be accessed from a wizard in the Windows Server 2008 in the cluster management snap-in named Migrate Services and Applications. After the tool is run, a report is created that provides information on the migration tasks.

The migration tool will import critical resource settings into the new cluster registry. The migration process migrates clustered resource configuration information. This involves reading the Windows Server 2003 cluster database information for the resources being migrated and then importing that information into the Windows Server 2008 cluster database, realizing that the location of this information may have changed. The primary examples here are the dependency, crypto checkpoint, and registry checkpoint information has all been relocated within the Windows Server 2008 cluster registry structure.

Windows Server 2008 Failover Clustering will migrate, with some restrictions, specific resource types from Windows 2003 server clusters. The resources that can be migrated are:

* Physical Disk Resource.
* Network Name Resource.
* IP Address Resource.
* DHCP Resource.
* File Share Resource (including Distributed File System [DFS] Root).
* WINS Resource.
* Generic Application Resource.
* Generic Service Resource.
* Generic Script Resource.

After resource information is migrated, you can then move the data.

# Support for Windows Server 2008 Server Core

The Server Core installation of the Windows Server 2008 operating system is a new option for installing Windows Server 2008. A Server Core installation provides a minimal environment for running specific server roles that reduces the maintenance and management requirements and the attack surface for those server roles.

Another major advantage of installing failover clustering in a Server Core environment is that it reduces the serviceability of the system. This means that you’ll be able to significantly increase your uptime since you’ll need to apply fewer updates to the Server Core based failover cluster. This makes Server Core a great enablers for failover cluster high availability.

Server Core supports the following server roles:

* DHCP server.
* File Services.
* Print Services.
* DNS server.
* Active Directory® Domain Services (AD DS).
* Active Directory® Lightweight Directory Services (AD LDS).
* Streaming Media Services.
* Windows Server 2008 Virtualization.

To provide this minimal environment, a Server Core installation installs only the subset of the binaries required by the supported server roles. For example, the Explorer shell is not installed as part of a Server Core installation. Instead, the default user interface for a Server Core installation is the command prompt. Once you have installed and configured the server, you can manage it either locally at the command prompt or remotely by using Remote Desktop. You can also manage the server remotely by using a Microsoft Management Console (MMC) or command-line tools that support remote use.

Server Core supports the failover cluster feature. You can manage failover clusters on Server Core using the **cluster.exe** command line tool, or remotely from the failover cluster MMC. Installation syntax for installing failover cluster services on server core is:

**Start /w ocsetup FailoverCluster –core**

# Improvements in Scoping and Managing Shares

A problem with previous versions of failover clustering was that users accessing a file server could accidentally lose access to shared resources on the cluster. For example, suppose you have an active-active file server setup and you have two groups with file shares in each group. Each group has its own network name resource. If both of those groups were owned by the same node at the same time, users who browsed to one of the network names would see shares not only from that group, but also shares from the other group and even shares on the local computer.

This could confuse users. Suppose a user browses to a server and sees a share. The user then right-clicks on the share and maps that share to a network drive. This works fine for about a month, at which time you decide to fail over that group owns the share for the mapped network drive to another server. Now the user will get an error that the path is no longer valid.

In an active-active file server cluster, users could see shares from both groups, and even local shares on the failover cluster server. Users would then map a network drive to that share. Some time after mapping the drive, the cluster would fail over and the mapped drive would no longer be accessible.

This won’t happen with Windows Server 2008 failover clustering. Users will only see shares accessible by the node that they’re connected to and won’t see shares owned by other groups. This prevents users from being confused and incorrectly mapping a network drives.

There are some significant improvements in how to create a highly available share with Windows Server 2008 failover clustering. In previous versions of failover clustering, you had to go through a moderately complex process of creating a file share resource. You no longer need to do this with Windows Server 2008 failover clustering. Now you can use the **Add a Shared Folder** wizard to do all the under the covers work for you. In fact, you no longer need to type in the UNC path to the share, which can introduce typing errors. Now you can use a **Browse** button to quickly and reliably identify the folder you want to use for the highly available file share.

Windows Server 2008 failover cluster services make it even easier than using a **Browse** button. IT generalists can create shares on a cluster server using the Windows Explorer, just as they would on any other file server. Failover cluster services will automatically see the new file share and create the file share resource in the background. The IT generalist never needs to drop into the failover cluster services manager to create a highly available share—failover cluster services hooks into the shell and does all the work for you.

# Better Storage and Backup Support

Overall, the storage changes made for failover clustering in Windows Server 2008 are designed to improve stability and keep with requirements for future growth. The changes result in the addition of some new features, in addition to some stricter guidelines for cluster storage. Some examples of improved storage and backup support include:

* **Support for both Master Boot Record (MBR) and GUID Partition Table (GPT) disks**. This allows the failover cluster feature to be more consistent with the features available within the core operating system. Now you can create disks greater than 2 terabytes without the need of a third-party solution.
* **Built-in self-healing logic**. Disk identification is based on either the disk signature in the Master Boot Record or the SCSI inquiry page 0x83 data (VPD), which is an attribute of the LUN and self healing will automatically be initiated as long as one attribute can be found.
* **Improved support for modern SAN storage solutions.** Storage hardware must support SCSI-3 SPC3–compliant SCSI commands for persistent reservation/release. Storage vendors can provide more information about what solutions are compliant. The cluster validation tool included with Windows Server 2008 can also be used to test the hardware.
* **All HBAs use the Storport/mini-port driver model to be listed on the Windows Server Catalog**. HBA vendors can provide more information about what drivers are compliant. This change should increase the overall stability of the storage driver stack.
* **All multi-path solutions are MPIO**. Storage vendors can provide more information about which solutions are compliant. The Device Specific Module (DSM) provided by the vendor is responsible for sending PR registration information down each path to the storage.
* **Windows Server 2008 features a closer integration with Volume Shadow Copy Service** **(VSS), for easier backups**. Failover clustering in Windows Server 2008 has its own VSS writer, which enables VSS backup applications to more easily support clusters.

These improvements in storage and backup will provide for increase stability and uptime for failover clusters.

# Enhanced Maintenance Mode

Windows Server 2008 failover cluster includes new Maintenance Mode functionality. This mode now shuts off health monitoring on a disk for a period of time so that it does not fail while you work on it.

Failover clustering in Windows Server 2008 builds on the maintenance mode feature first introduced with Windows Server 2003. This mode lets you perform certain maintenance or administrative tasks on clustered disk resources (for example, volume snapshots, ChkDsk).

There are three of methods for moving a disk in and out of maintenance mode:

* Using the **cluster.exe** command-line tool.
* Through the use of APIs available to third parties for developing solutions requiring the use of maintenance mode. The platform software development kit (SDK) provides more information about the use of these APIs. This is the method through which Maintenance Mode will typically be invoked.
* From inside the Failover Cluster Management snap-in is to right-click on the disk and select **Enable Maintenance Mode for this disk** from the **More actions** dialog box. Once the action is executed, the disk display inside the management snap-in is modified.

To put a disk into maintenance mode, all nodes need to be able to communicate with each other. Initially, the disk will be fenced from all nodes that do not own the resource. In this configuration, other nodes will still see the disk, but they will not be able to access it. The owning node will then remove its persistent reservation from the disk.

# Superior Scalability

Windows Server 2008 clusters can support more nodes than in Windows Server 2003. Specifically, x64-based failover clusters support up to 16 nodes in a single cluster, as opposed to the maximum of 8 nodes in Windows Server 2003, providing even greater scalability.

In addition to support for more cluster nodes, Windows Server 2008 failover clusters now support GPT disks. A GPT disk uses the GUID partition table (GPT) disk partitioning system. A GPT disk offers these benefits:

* Allows up to 128 primary partitions. (MBR disks can support up to four primary partitions and an infinite number of partitions inside an extended partition.)
* Allows a much larger volume size—greater than 2 terabytes (the limit for MBR disks).
* Provides greater reliability due to replication and cyclical redundancy check (CRC) protection of the partition table.

The combination of increased number of nodes and support for GPT disks greatly enhances the scalability of larger volumes in your failover cluster deployments.

# A New Quorum Model

The Windows Server 2008 failover clustering quorum model is entirely new and represents a blend of the earlier shared disk and majority node set models. In Windows Server 2008 failover clustering there are now four ways to establish a quorum:

* No Majority – Disk Only (same as the Windows Server 2003 shared disk quorum)
* Node Majority (same as the Windows Server 2003 majority node set)
* Node and Disk Majority
* Node and File Share Majority

The concept of quorum in Windows Server 2008 moves away from the requirement for a shared storage resource. The concept of quorum now refers to a number of *votes* which must equate to a majority of nodes. All nodes and disk resources get a vote.

This helps eliminate failure points in the old model, where it was assumed that the disk would always be available. If the disk failed, the cluster would fail.

In Windows Server 2008 failover clustering the disk resource that gets a vote is no longer referred to as a quorum disk; now it is called the *witness disk*. With the new quorum models, the cluster can come online even if the witness disk resource is not available.

The No Majority model behaves similarly to the old quorum disk model. If the quorum disk failed, the cluster would not come online, thus representing a single point of failure.

The Node Majority model behaves similarly to the Majority Node Set model. This model requires there or more nodes and there is no dependence on witness-disk availability. The disadvantage of this model is that you cannot run two server clusters, because a majority of nodes is not possible in a two-node cluster scenario.

The Node and Disk Majority and the Node and File Share Majority models are similar. In each case, both the nodes and the disk resource are allowed to vote. The cluster will come online as long as a majority of votes are reached, regardless of the status of the disk resource. In the Node and Disk Majority quorum model, the disk resource is a shared disk, the witness disk. In the Node and File Share Majority model, a file share replaces the disk as a disk-based vote. The Node and File Share Majority model is an excellent solution for geographically dispersed multi-site clusters.

Failover cluster administrators can select the quorum model of choice, depending on the requirements of the clustered resource. The quorum model should be selected after the cluster is first created and prior to putting the cluster into production.

# An Improved Security Model

There have been several changes made to Windows Server 2008 failover clustering that make it a more secure and reliable product. Some of these changes include:

* Removal of the requirement for a domain user account for the Cluster Service Account (CSA).
* Improved logging and event tracing.
* Transition from unsecure datagram remote procedure call (RPC) communications to TCP-based RPC communications.
* Enabling Kerberos authentication by default on all cluster network name resources.
* The ability to audit access to the cluster service (Clussvc.exe) by way of either the Failover Cluster Management snap-in (cluadmin.msc) or the cluster command-line interface (cluster.exe).
* Ability to secure Inter-Cluster communications.

## CSA No Longer Requires a Domain User Account

The cluster service no longer runs under the context of a domain user account, also known as the CSA. This helps solve problems related to CSA privileges being changed by Group Policy and account expiration limits. Now the failover cluster service runs in the context of a local administrator account that has the same privileges as CSA.

Because the cluster service now runs in the context of a local system account, the cluster’s common identity has transitioned to the CNO, which is the computer object created in Active Directory during the Create Cluster process. The CNO represents the Cluster Network Name core resource.

The CNO is different from a Virtual Computer Object (VCO). The CNO is associated with the Cluster Name core resource, while a VCO is a computer object created in Active Directory for all other Network Name resources created in a cluster as part of configuring a Client Access Point (CAP).

## Improved Logging and Event Tracing

The text-file-based cluster log is also gone. Event trace logging (.etl) is now enabled via Event Tracing for Windows (ETW). There is new functionality built into the command line. The **cluster.exe** tool allows you to dump the trace log into a text file. This file looks similar to the cluster log used in previous versions of failover clustering. Use the **cluster.exe Log /Generate** command to see this log. Also, you can create diagnostic views inside the improved Windows Server 2008 Event Viewer.

## More Secure TCP-based Cluster Communications

Windows Server 2008 failover cluster communications are TCP-based, as opposed to User Datagram Protocol (UDP) datagrams as in previous versions of clustering. This provides for more reliable and secure communications among cluster nodes. Additionally, TCP is required for Kerberos authentication. TCP will also be used for Windows NT LAN Manager (NTLM) authentication if required.

## Kerberos Authentication By Default

For backward compatibility, the cluster service (in conjunction with **clusapi.dll**) will support the following combinations of authentication and RPC transport:

* Kerberos over TCP (Default and most secure).
* NTLM (version 2 or version 1) over TCP.
* NTLM (version 2 or version 1) over UDP.

Legacy NTLM is no longer required and Kerberos is the primary authentication protocol. All Network Name resources will have Kerberos enabled. The default authentication mechanism in previous versions of clustering was NTLM authentication via the NTLM Security Support Provider (NTLMSSP). In Windows Server 2008 failover clustering, authentication will primarily use Negotiate via Kerberos, but can fallback to NTLM authentication (version 2 or version 1) if enabled on your network. These changes in the failover clustering authentication architecture provide for a much more secure clustering environment.

Kerberos authentication is supported by the fact that computer objects are created in Active Directory by default for every clustered Network Name resource. Using the “Negotiate” package provides the following advantages:

* Allows the system to use the strongest (most secure) available protocol.
* Ensures forward compatibility for applications.
* Ensures applications exhibit behaviors that are in accordance with your security policy.
* Provides backward compatibility with NTLM authentication.

## Improved Auditing of Cluster Services Tools Usage

New auditing feature for tracking access to cluster enables you to see who has accessed the **cluadmin.msc** and **cluster.exe** tools. If an unexpected event takes place on the cluster, you will be able to go back to check who was using the cluster administration tools to make any changes that might have led to that event.

## Securing Intra-cluster Communications

Windows Server 2008 failover clustering provides for securing communications between nodes in the cluster exclusive of heartbeat communications. By default, the user-mode communications (clussvc-to-clussvc, examples include GUM updates, regroups, form\join, etc.) occurring between nodes in the cluster are signed. A new feature in Windows Server 2008 failover clustering provides the capability for further securing these communications by encrypting them. Inter-node communications are signed by default. If you want to encrypt inter-node communications, you will need to configure the cluster to do so.

# New Networking Capabilities and More Flexible Dependencies

Windows Server 2008 failover clustering includes a new networking model. Major improvements have been to failover cluster networking, including the following:

* Improved support for geographically distributed networks.
* The ability to place cluster nodes on different networks.
* The ability to use DHCP server to assign IP addresses to cluster interfaces.
* Improvements in the cluster heartbeat mechanism.
* New support for IPv6.

## Better Support for Geographically Distributed Clusters

With Windows Server 2008 failover clustering, individual cluster nodes can be placed on separate, routed networks. This requires that resources depending on IP Address resources, (for example, Network Name resource), implement an *OR* logic, since it is unlikely that every cluster node will have a local connection to every network for which the cluster is aware. Support for OR logical for cluster resources is new in Windows Server 2008 failover cluster. This facilitates IP Address and Network Name resources coming online when services and applications fail over to remote nodes.

Additionally, the default behavior for determining if a cluster member is unavailable has been changed in terms of how many replies can be missed before the node is considered unreachable and a Regroup is conducted to obtain a new view of the cluster membership. As an added feature, the cluster administrator is able to modify properties associated with this process. There are properties of the cluster that address the heartbeat mechanism, these are:

* SameSubnetDelay
* CrossSubnetDelay
* SameSubnetThreshold
* CrossSubnetThreshold

The default configuration (shown here) is such that we will wait 5.0 seconds before we consider a cluster node to be “unreachable” and have to regroup to update the “view” of the cluster. The limits on these settings are seen in Table 1.

Table 1: Parameters determining whether a node is unreachable

| Parameter | Default | Range |
| --- | --- | --- |
| SameSubnetDelay | 1000 milliseconds | 250-2000 milliseconds |
| CrossSubnetDelay | 1000 milliseconds | 250-4000 milliseconds |
| SameSubnetThreshold | 5 | 3-10 |
| CrossSubnetThreshold | 5 | 3-10 |

These changes afford much more flexibility in implementing geographically dispersed clusters. Administrators no longer have to stretch virtual local area networks (VLANs) across the WAN to accommodate geographically distant servers that are on different subnets. Failover cluster nodes can now reside on completely different subnets.

Moreover, the network latency requirements in Windows Server 2003 server clustering have been removed from Windows Server 2008 failover clustering. The failover clustering heartbeat requirement is now fully configurable. Geographically dispersed multi-site clusters are easier to deploy and more technically feasible with Windows Server 2008 failover clustering when compared to Windows Server 2003.

## Support for DHCP

Windows Server 2008 failover clustering now includes the capability whereby cluster IP address resources can obtain their addressing from DHCP servers in addition to via static entries. If the cluster nodes are configured to obtain IP addresses from a DHCP server, then the default behavior will be to obtain an IP address automatically for all cluster IP address resources. If the cluster node has statically assigned IP addresses, the cluster IP address resources will have to be configured with static IP addresses as well. Therefore, cluster IP address resource IP assignment follows the configuration of the physical node and each specific interface on the node.

## Improvements in the Cluster Heartbeat Mechanism

The cluster heartbeat mechanism has changed in Windows Server 2008 failover clustering. While still using port 3343, it has transitioned from a UDP broadcast health checking mechanism to a UDP unicast communication that is similar to ping in that it uses a Request-Reply-type process. It provides for higher security and more reliable packet sequence numbering.

## Comprehensive Support for IPv6

Since Windows Server 2008 supports IPv6, the cluster service will support this functionality as well. This includes being able to support IPv6 IP Address resources and IPv4 IP Address resources either alone or in combination in a cluster.

Clustering also supports 6-4 and Intra-site Automatic Tunneling Addressing Protocol (ISATAP). Additionally, clustering supports only IPv6 addresses that allow for dynamic registration in DNS (AAAA host records and the IP6.ARPA reverse look-up zone). Currently there are three types of IPv6 address types—global, site local, and link local. Dynamic DNS registrations will not occur for link local addresses and therefore cannot be used in a cluster.

# Conclusion

A major obstacle to traditional clusters was the complexity in building, configuring, and managing the clusters. Too often this resulted in higher-than-necessary costs for organizations, both in lost opportunities to make more applications highly available and from retaining expensive administrative resources exclusively for maintaining the clusters that they did deploy.

A major facet of the improvements to failover clusters in Windows Server 2008 is aimed squarely at these challenges, producing a clustering paradigm that radically simplifies and streamlines the cluster creation, configuration, and management process.

Changes to clustering in the next iteration of the Windows Server operating system also improve cluster performance and flexibility. Clusters are x64-based and now support up to 16 nodes. Cluster nodes can have their IP addresses assigned by DHCP and geographically dispersed clusters can span subnets. Windows Server 2008 failover clusters are designed to work well with storage area networks (SANs), natively supporting the most commonly used SAN bus types. And Windows Server 2008 failover clusters are built around a more resilient and customizable quorum model.

This is a preliminary document and may be changed substantially prior to final commercial release of the software described herein.

The information contained in this document represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of publication.

This white paper is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS   
DOCUMENT.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in, or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

© 2007 Microsoft Corporation. All rights reserved.

Microsoft, Active Directory, Windows, Windows NT, Windows Server, Windows Vista, and Windows logo are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.