Technical Overview of Windows HPC Server 2008

Published: June, 2008, Revised September 2008

Abstract

Windows® HPC Server 2008 brings the power, performance, and scale of high performance computing (HPC) to mainstream computing. The centralized management and deployment interface helps to simplify deployment for both large and small compute clusters and provide a simple and effective management experience to increase cluster administrator productivity. Microsoft® HPC Pack 2008 includes a highly scalable Job Scheduler that provides support for interactive service-oriented architecture (SOA) applications using high performance computing for Windows Communication Foundation (WCF) and parallel jobs using Microsoft® Message Passing Interface (MS-MPI). Essential applications from key independent software providers (ISVs) can be run on the cluster to help you meet your business needs in a timely, cost-effective, and highly productive manner. Windows Server® 2008 helps to provide seamless security, storage, and desktop access to cluster resources and management. Microsoft® Visual Studio® 2005 provides parallel debugging capabilities for use with Windows HPC Server 2008, and MS-MPI is now integrated with the Event Tracing for Windows infrastructure. This consolidates application, networking, and operating system events from many compute nodes into a single, time-correlated record to speed debugging.

This document was developed prior to the product's release to manufacturing, and as such, we cannot guarantee that all details included herein will be exactly as what is found in the shipping product.

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, IMPLIED, OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT.

Microsoft

Contents

What Is Windows HPC Server 2008?

Microsoft® Windows® HPC Server 2008 (HPCS), the next generation of high performance computing (HPC), provides enterprise-class tools, performance, and scalability for a highly productive HPC environment. HPCS provides a complete and integrated cluster environment including the operating system, a Job Scheduler, Message Passing Interface v2 (MPI2) support, and cluster management and monitoring components. Built on Windows Server® 2008 64-bit technology, HPCS can efficiently scale to thousands of processing cores and includes a management console that helps proactively monitor and maintain system health and stability. Job scheduling interoperability and flexibility enables integration between Windows and Linux-based HPC platforms, and supports batch and service-oriented architecture (SOA) workloads. Enhanced productivity, scalable performance, and ease of use are some of the features that make Windows HPC Server 2008 best-of-breed for Windows environments.

Windows HPC Server 2008 can help shorten user time-to-insight for HPC workloads through easier deployment and management. By using the existing Windows-based information technology (IT) infrastructure, HPCS simplifies management, security, and storage for the cluster, and provides seamless access from the desktop.

HPCS includes improved provisioning based on the Windows Server 2008 Windows Deployment Services technology, a faster Microsoft Message Passing Interface (MS-MPI) that includes new NetworkDirect support, an advanced Job Scheduler, and a new management interface built on the Microsoft® System Center 2007 user interface (UI), which has support for Windows PowerShell™ as a preferred scripting interface.

Windows HPC Server 2008 integrates with other Microsoft products to help increase HPC productivity and improve the overall end-user experience. This includes collaboration through Microsoft® Office SharePoint® Server 2007 and the Windows Workflow Foundation, as well as improved management and efficiency by integrating with System Center 2007. Through integration with the Windows Communication Foundation (WCF), Windows HPC Server 2008 allows developers working with SOA applications to harness the power of parallel computing offered by HPC solutions.

HPCS uses the failover clustering capabilities of Windows Server 2008 Enterprise and of Microsoft® SQL Server® to provide high availability and redundancy of the head node in your HPC cluster.

What's New in Windows HPC Server 2008?

When developing Windows HPC Server 2008, Microsoft focused on four key areas: systems management, job scheduling, networking and MPI, and storage. The new features include:

Systems Management

- The new Administration Console based on System Center UI framework integrates every aspect of cluster management.
- Node grouping allows administrators to categorize and batch operations on compute nodes.
- Monitoring heat map allows viewing cluster status at-a-glance.
- Scriptable control of cluster using Windows PowerShell and other command-line tools.
- High availability for multiple head nodes.

- Improved compute node provisioning using Windows Deployment Services.
- Built-in support for cluster-wide updating.
- Built-in system diagnostics.
- Built-in cluster reporting.
- Extensible reporting, allowing for job scheduling analysis using external database tools like SQL Server Analysis Services.
- Microsoft System Center Operations Manager 2007 Management Pack.

Job Scheduling

- Integration with WCF, allowing SOA application developers to harness the power of parallel computing offered by HPC solutions.
- Job scheduling granularity at processor core, processor socket, and compute node levels.
- Performance improvements for large clusters.
- Support for external databases for Job Scheduling.
- New job scheduling policies.
- Support for Open Grid Forum's HPC Basic Profile interface for Job Scheduler interoperability.

Networking and MPI

- NetworkDirect, a new Remote Direct Memory Access (RDMA) network interface providing dramatic performance improvements for MPI applications running over high-speed fabrics.
- Improved Network Configuration Wizard.
- New shared memory MS-MPI implementation for multicore servers.
- MS-MPI integrated with Event Tracing for Windows.

Storage

- Improved iSCSI storage area network (SAN) support in Windows Server 2008.
- Improved Server Message Block v2 (SMB v2) in Windows Server 2008.
- New parallel file system support and vendor partnerships for clusters with high performance storage needs.

Table 1 provides a quick glance at some of the main Windows HPC Server 2008 features.

Table 1. Windows HPC Server 2008 Features

Feature	Implementation	Benefit
Operating system	Windows Server 2008	Inherits security and stability features from Windows Server 2008.
Processor type	x64 (AMD64 or Intel EM64T)	The large memory model and processor efficiencies of x64 architecture.
Node deployment	Windows Deployment Services	Image-based deployment, with full support for multicasting.

Head node redundancy	Windows Failover Clustering and SQL Server Failover Clustering	Requires Windows Server 2008 Enterprise and SQL Server Standard, but provides a fully redundant head node and scheduler.
Management	New Integrated Administration Console	Integrates all aspects of node and job management, grouping, monitoring at-a-glance, diagnostics, and reporting in a single application.
Network topology	Network Configuration Wizard	Fully automated, with a more intuitive interface, the Network Configuration Wizard facilitates configuring the cluster's network topology.
Application network	MS-MPI	New, high-speed application network stack using NetworkDirect drivers. New shared memory implementation for multicore processors. Highly compatible with existing MPICH2 implementations.
Scheduler	Job Manager Console	GUI console is integrated into Administration Console, or is stand-alone. Command line supports Windows PowerShell scripting and legacy command-line scripts from Windows Compute Cluster Server (the first version of Windows HPC offering). Greatly improved speed and scalability. Support for SOA applications.
Reporting	Integrated into Administration Console	Standard, prebuilt reports. Extensibility features allow using SQL Server Analysis Services for additional analysis. Historical performance charts.
Monitoring	Integrated into Administration Console	Heat map on cluster or node group. Per node charts. Cluster-wide performance overview.
Diagnostics	Integrated into Administration Console	In-the-box verification tests and performance tests. Store, filter, and view test results and history.

What's in the Box?

Windows HPC Server 2008 is a two-DVD package. The first DVD contains the setup for a 64-bit version of Windows Server 2008 that is restricted to an HPC workload, and the second DVD contains Microsoft HPC Pack 2008, which provides the additional interfaces, tools, and management infrastructure.

Hardware Requirements

The minimum hardware requirements for Windows HPC Server 2008 are similar to the hardware requirements for the x64-based version of the Windows Server 2008 Standard operating system. Windows HPC Server 2008 supports up to 128 GB of RAM.

Supported processors include:

- AMD Opteron
- AMD Athlon 64
- AMD Phenom
- Intel Xeon with Intel EM64T
- Intel Pentium with Intel EM64T
- Intel Core 2 Duo with Intel EM64T

Table 2. Minimum Hardware Requirements

Hardware	Requirements
CPU	x64 architecture computer with Intel Pentium or Xeon family processors with Intel Extended Memory 64 Technology (EM64T) processor architecture; AMD Opteron family processors; AMD Phenom family processors; AMD Athlon 64 family processors; compatible processor(s)
RAM	512 MB
Multiprocessor support	Windows Server 2008 HPC Edition and Windows Server 2008 Standard Edition support up to four processors per server. Windows Server 2008 Enterprise Edition supports up to eight processors per server.
Minimum disk space for setup	50 GB
Disk volumes	A single system volume is required for the head and compute nodes. Redundant array of independent disks (RAID) is supported but not required. The system volume must be Master Boot Record (MBR). Additional volumes can be MBR or GUID Partition Table (GPT).
Network interface card	At least one network interface card (NIC) is required. If a private network is used, the head node requires at least two NICs, and compute nodes at least one. Each node may also require a high-speed NIC for an application network.

Software Requirements

The head and compute nodes for Windows HPC Server 2008 can be any of the following:

- x64-based version of Windows Server 2008 Standard operating system
- Windows Server 2008 Enterprise x64 Edition

The Job Scheduler uses SQL Server as its repository. An existing SQL Server 2005 or SQL Server 2008 installation can be used, or the HPC Pack 2008 installer will install SQL Express. SQL Server Standard Edition or Enterprise Edition is required for head node failover clustering.

The Administration Console and job scheduling user console components are automatically installed on the head node of the compute cluster. These components can also be installed on other computers allowing either remote management of the cluster or job submission from client computers. The supported operating systems for installation of the remote components are:

- Windows Server 2003 Service Pack 2 (SP2) or Release 2 (R2) (32-bit or x64 versions)
- Windows Server 2008 (32-bit or x64 versions)
- Windows® XP Professional SP3
- Windows XP Professional, x64 Edition SP2
- Windows Vista® Business, Enterprise, and Ultimate Editions SP1

Note Windows HPC Server 2008 is designed solely for use with high performance computing applications and does not support use as a general-purpose infrastructure server.

Solution Architecture

Windows HPC Server 2008 is composed of a cluster of servers that includes a single head node and one or more compute nodes (see Figure 1). The head node, which now provides failover via Windows Server 2008 Enterprise high availability services and SQL Server clustering, controls and mediates all access to the cluster resources and is the single point of management, deployment, and job scheduling for the compute cluster. Windows HPC Server 2008 can use an existing Active Directory® directory service-based infrastructure for security, account management, and overall operations management using tools such as System Center Operations Manager.

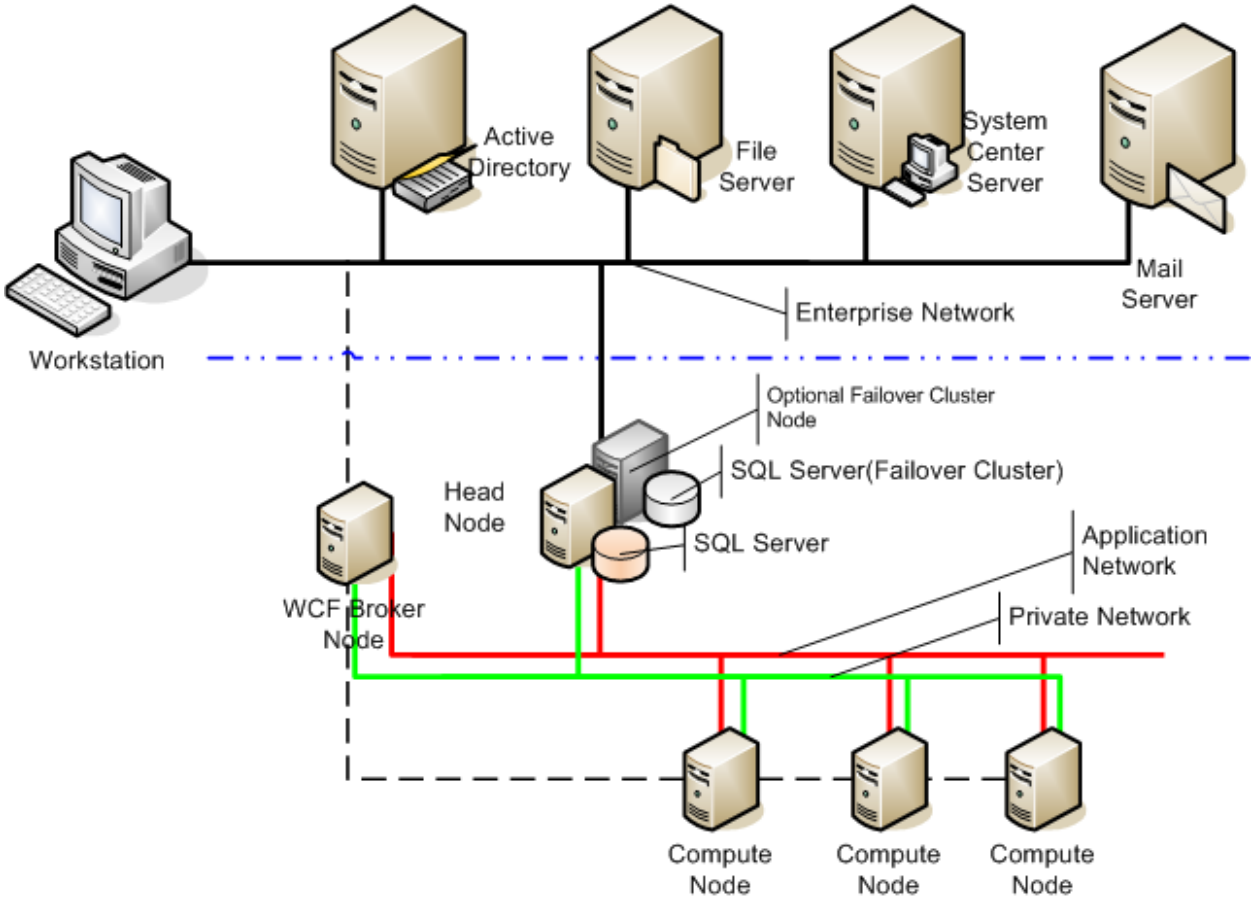

Figure 1. A typical Windows HPC Server 2008 network

The Windows HPC Server 2008 installation involves installing the operating system on the head node, joining it to an Active Directory domain, and then installing HPC Pack 2008.

A To Do List page that shows you the steps necessary to complete configuration of your compute cluster is displayed when first starting the Administration Console. These steps include defining the network topology, configuring automatic deployment, and adding compute nodes to the cluster. Windows HPC Server 2008 provides Node Templates, an easy way to define the desired configuration of the compute nodes, and a simple interface that leverages the underlying Windows Deployment Services of Windows Server 2008 that allows you to deploy to all nodes, or to a group of the nodes in the cluster, with

support for tracking the deployment progress. The Windows HPC Server 2008 management interface includes diagnostic tests that allow administrators to detect connectivity problems, node loading, and job status across the cluster.

Network Topology

Windows HPC Server 2008 supports five different network topologies to allow users to configure networking to suit their environment, infrastructure, and clustering needs. The five topologies have one to three NICs on each node. The Network Configuration Wizard, shown in Figure 2, simplifies configuring your network topology.

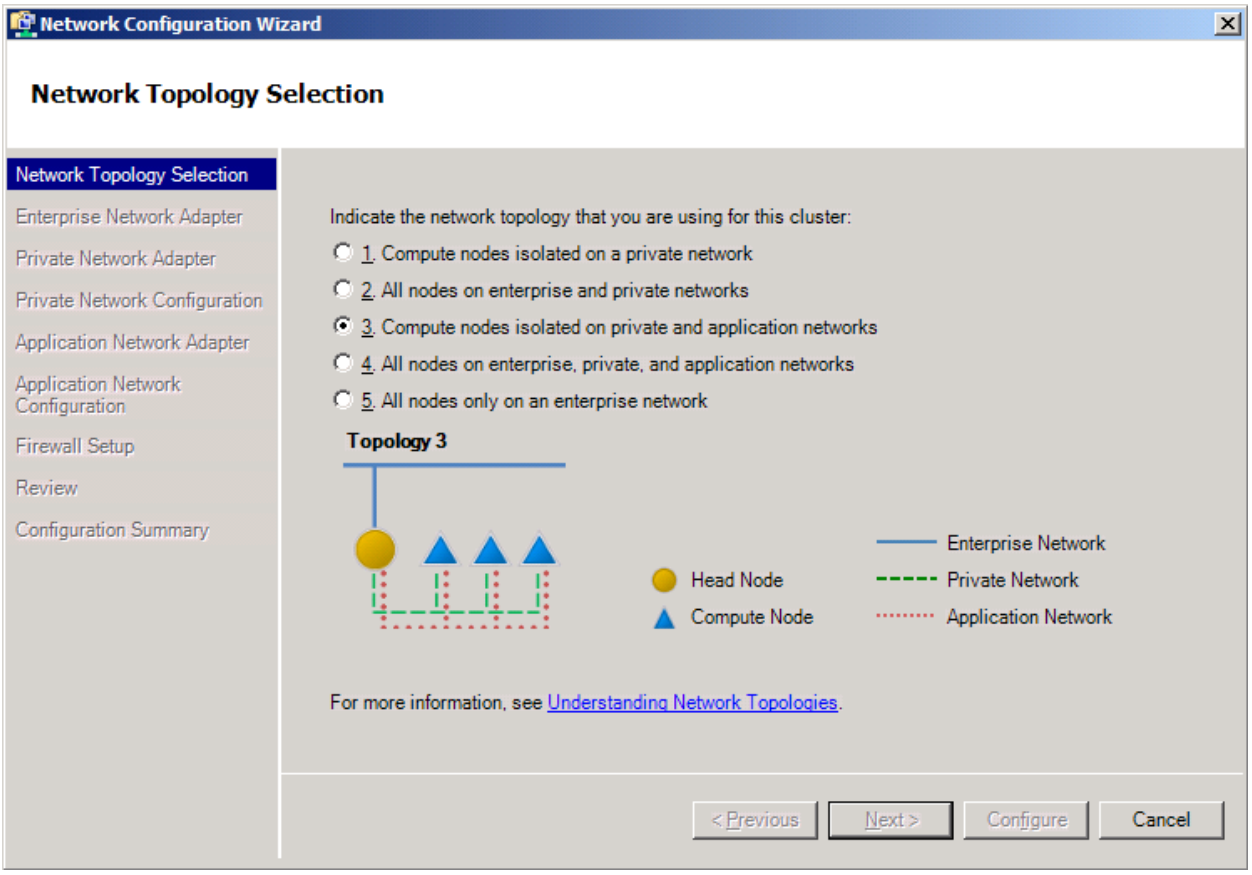

Figure 2. Cluster topology scenarios supported by Windows HPC Server 2008

1. **Compute nodes isolated on a private network.** The head node has two NICs and may provide network address translation (NAT) between the compute nodes, which each having a single NIC connected to a private network, and to the enterprise network.
2. **All nodes on both public and private network.** One NIC is connected to the enterprise network and one is connected to the private, dedicated cluster network.
3. **Compute nodes isolated on private and application network.** The head node has three NICs, one connected to the enterprise network, one to the private network, and one to the application network. The head node may perform NAT between the compute nodes and the enterprise network, with each compute node having a connection to the private network and a connection to a high-speed protocol such as MPI for the application network.

4. **All nodes on enterprise, private, and application networks.** One NIC is connected to the enterprise network, one is connected to a private, dedicated cluster-management network, and one is connected to a high-speed, dedicated application network.
5. **All nodes only on enterprise network.** In this limited networking scenario, where each node has only a single NIC, the use of Windows Deployment Services to deploy compute nodes is not supported, and each compute node must be manually installed and activated.

Note In topologies 1 and 3, where the compute nodes are isolated on a private network, NAT between the private and enterprise networks can be performed by the head node, or by a separate server.

MS-MPI is a high-speed networking interface that runs over Gigabit Ethernet, InfiniBand, Myrinet, or any network that provides a Winsock Direct, NetworkDirect, or TCP/IP interface.

NetworkDirect support is new in Windows HPC Server 2008, and provides a high-speed, hardware-independent, RDMA network architecture that provides speed parity with custom, hardware-native interfaces available from some manufacturers without the disadvantages of reduced flexibility and choice.

Benefits and Elements of the Microsoft HPC Solution

Windows HPC Server 2008 combines the underlying stability and security of Windows Server 2008 with the features of HPC Pack 2008 to provide a robust and productive high performance computing solution. The key elements of the HPCS solution include:

- Deployment
- Management
- Microsoft Message Passing Interface (MS-MPI)
- Job Scheduler
- Security

Deployment

One of the biggest challenges that customers face today in adopting HPC solutions lies in the deployment of clusters and nodes. It's not hard to deploy a four-node cluster—someone just has to insert the DVD and press ENTER as required at each node and then add the nodes to the cluster. Configuration is easy, and the time involved isn't significant. But when you're deploying clusters of a few hundred nodes or a few thousand nodes, manual installation is tedious and prone to errors.

Windows Compute Cluster Server 2003—the predecessor to Windows HPC Server 2008—originally used Remote Installation Service (RIS) for automatic deployment of compute nodes, and this was a major improvement over manually adding nodes one at a time. With the introduction of Windows Compute Cluster Server 2003 SP1, Windows Deployment Services became available for Windows Compute Cluster Server 2003. Windows Deployment Services offers many advantages in deploying larger or heterogeneous clusters. In Windows HPC Server 2008, Microsoft continues to use Windows Deployment Services to deploy compute nodes. Deployment is integrated into the Administration Console, allowing the administrator to select nodes for deployment and easily monitor progress. Figure 3 shows the deployment architecture on the head node of the HPC cluster.

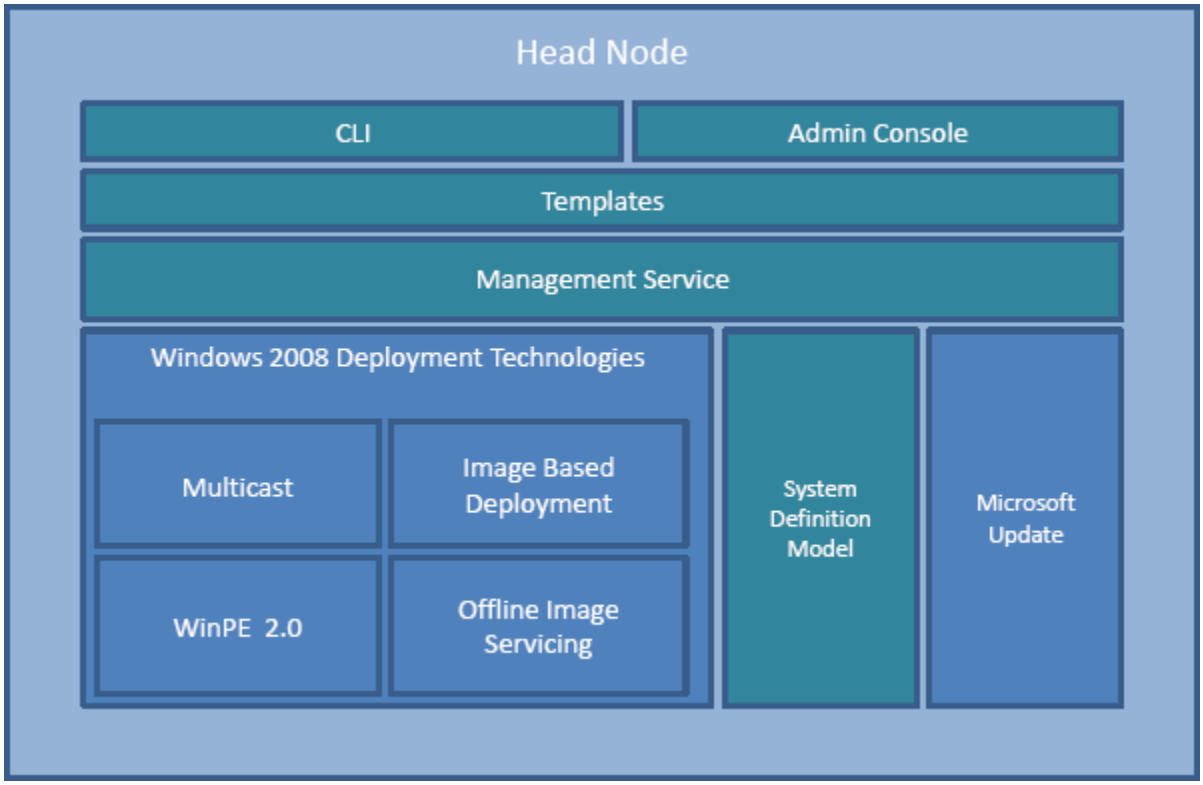

Figure 3. Deployment architecture on head node

Windows HPC Server 2008 includes a Node Template Generation Wizard that guides the administrator through creating a template for compute node configuration. Where advanced configuration is needed, a Template Editor provides additional capabilities, including configuring the template for automatic application deployment. The Windows HPC Server 2008 Template Generation Wizard includes support for injecting drivers into images.

Windows Deployment Services uses Windows Imaging Format (WIM) files and multicasting to rapidly deploy nodes in parallel. Administrators, using the Template Generation Wizard, create Node Templates for the types of node they want to deploy. A Note Template includes the base operating system image, drivers, configuration, and additional software.

Windows Deployment Services running on the head node discovers nodes as they are turned on. Preinstalled compute nodes can be added to the cluster by importing a node list, or the administrator can choose to select nodes interactively as the Administration Console displays the discovered nodes. During either unattended or interactive deployment, the node is imaged and configured, and applications are deployed according to the node template.

The biggest advantages of Windows HPC Server 2008 integrated deployment using Windows Deployment Services are:

- Scalability due to image multicasting
- Direct integration into the Administration Console
- Support for Node Templates
- Driver integration (and updating)

- Support for non-uniform compute node deployment

Windows HPC Server 2008 takes one of the major challenges of building and managing an HPC cluster, deploying and updating nodes, and simplifies it. The ability to have templates for each type of deployment ensures that you can easily deploy or redeploy large numbers of nodes. One of the design goals for HPCS was to enable deployments of up to 256 nodes in an hour.

Management

Another major challenge that customers face today in adopting HPC solutions involves the management and routine administration of clusters and nodes. This problem has traditionally been a departmental- or organizational-level problem, requiring a dedicated IT professional staff to manage and deploy nodes, and users to submit batch jobs, all competing for limited resources. Windows HPC Server 2008 is designed to provide support for ease of deployment and ease of management. Microsoft set the following goals:

- To give clear, prescriptive guidance for deploying clusters and monitoring their health.
- To provide authentication and authorization mechanisms.
- To build scriptable management solutions.
- To provide simple, but effective management that allows at-a-glance view of the cluster with the ability to drill down to node details such as metrics, logs, and configuration status.

New User Interface

Windows HPC Server 2008 builds on a new integrated Administration Console based on the System Center (http://www.microsoft.com/systemcenter) user interface. Windows HPC Server 2008 leverages the Windows PowerShell as the new scripting language and shell for managing HPCS clusters and jobs. Active Directory integration enables easy, role-based cluster management with Cluster Administrator and Cluster User roles. The new Administration Console has five major navigation panes:

- **Configuration.** This navigation pane, which includes the To Do List, is used to set up the cluster configuration. It contains a Network Configuration Wizard for easy network configuration, and a Node Template Generation Wizard to guide the administrator in setting up system templates for consistent compute node deployment, including updating, and application deployment.
- **Node Management.** This navigation pane is used to initiate any node-specific actions such as deployment, monitoring at-a-glance node status, bringing nodes offline or online, or adding and removing nodes.
- **Job Management.** This navigation pane gives you control of job scheduling and status in the cluster.
- **Diagnostics.** This navigation pane allows the administrator to select a node or group of nodes and run diagnostic tests that validate key cluster functions or settings such as network connectivity, job execution, configurations, and performance. Administrators can view the progress of tests and view past test results in support of overall cluster troubleshooting.
- **Charts and Reports.** This navigation pane displays standard reports on jobs and nodes in the cluster and allows for both scheduled and on-demand reporting.

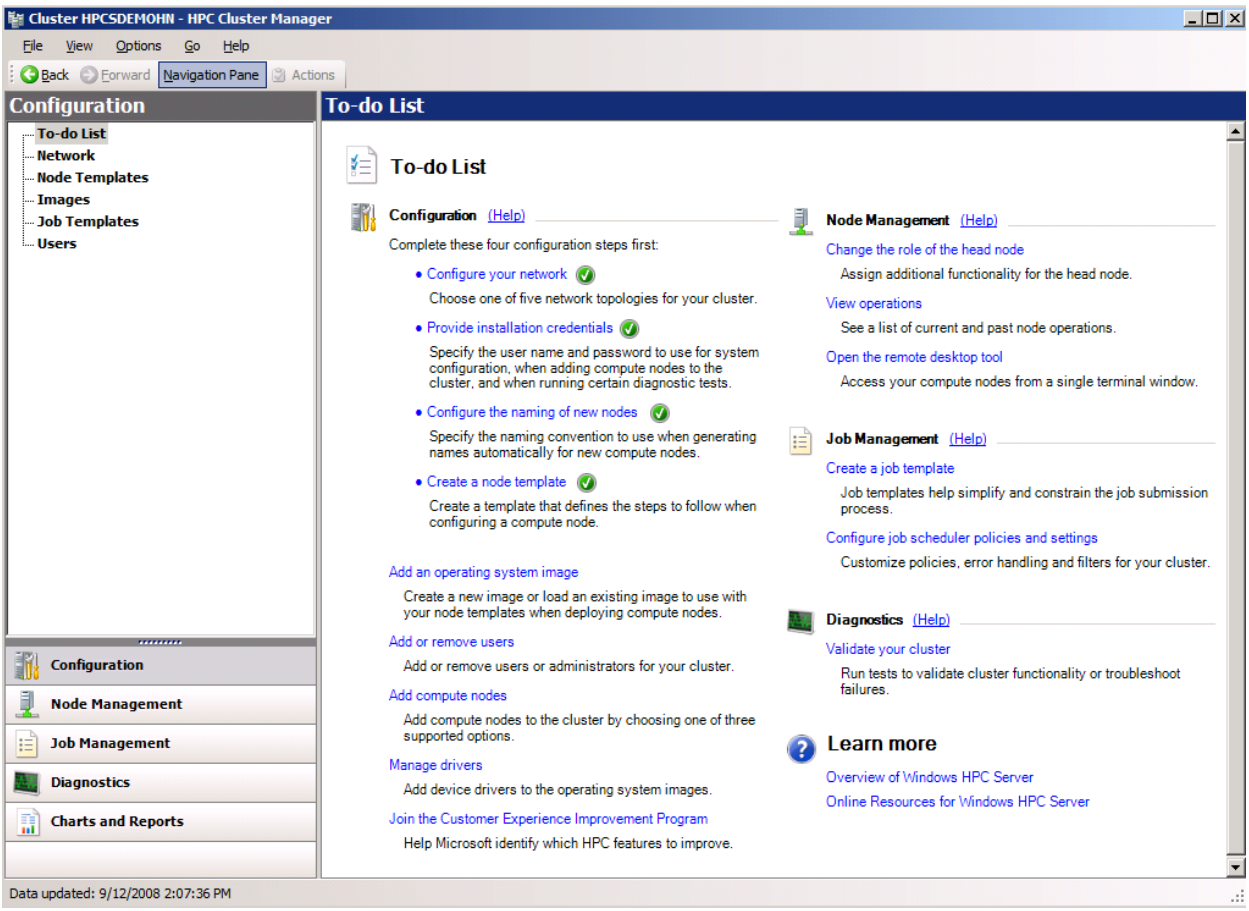

Figure 4. The Administration Console

The new Administration Console integrates management and deployment operations into a single console. Operations that are performed from the graphical console can also be performed from the command line using Windows PowerShell, the new Windows command-line shell and scripting language.

Windows PowerShell Support

Windows HPC Server 2008 maintains compatibility with existing Windows Compute Cluster Server 2003 command-line scripts, but extends this to add support for Windows PowerShell. With more than 130 standard command-line tools, a new admin-focused scripting language, and consistent syntax and utilities, Windows PowerShell accelerates automation of system administration tasks and helps to improve your organization's ability to address the unique system management problems of your environment. Windows PowerShell works with Microsoft .NET objects, giving users and administrators the power of an object-oriented language with the ease of use of an interactive shell.

For those not familiar with Windows PowerShell, a simple example script or two will help to demonstrate the power and simplicity of the language. In the first script, the administrator wants to group all nodes that have at least eight processors on the node as a single group of nodes called FastNodes. He then tests the new grouping by submitting a test job to the group of nodes.

```
Get-HpcNode | where {$_.NumProcessors –gt 8} | Add-HpcGroup “FastNodes”
New-HpcJob –Name “TestJob” –NodeGroup “FastNodes” | Submit-HpcJob
```

This simple script shows the power of the language. The cmdlet **Get-HpcNode** returns nodes as objects, which are then piped to *where* (an alias for the *where* object), which checks the properties of the *node* object to see how many processors the node has. The nodes that have eight or more processors are then piped to the **Add-HpcGroup** cmdlet, which creates a new group of nodes called FastNodes. Finally, a job is created and assigned to the FastNodes group and then submitted using the **Submit-HpcJob** cmdlet.

Another example shows how Windows PowerShell can work with Windows Management Instrumentation (WMI) for more complex tasks. In this example, the administrator wants to get a list of all nodes with more than 1 GB of free disk space.

```
$nodes=Get-HpcNode

foreach ($item in $nodes) {
$disks=get-wmiobject -computername $item.instancename -class win32_logicaldisk |\    where
   {$_.freespace/$gb -gt 1gb}

if ($disks.count -gt 0) {
   write-host "name:" $item.instancename
  }
}
```

The Windows HPC Server 2008 Ecosystem

As shown in Figure 5, Windows HPC Server 2008 takes advantage of the overall Microsoft ecosystem to improve the functionality, usability, and manageability of the cluster.

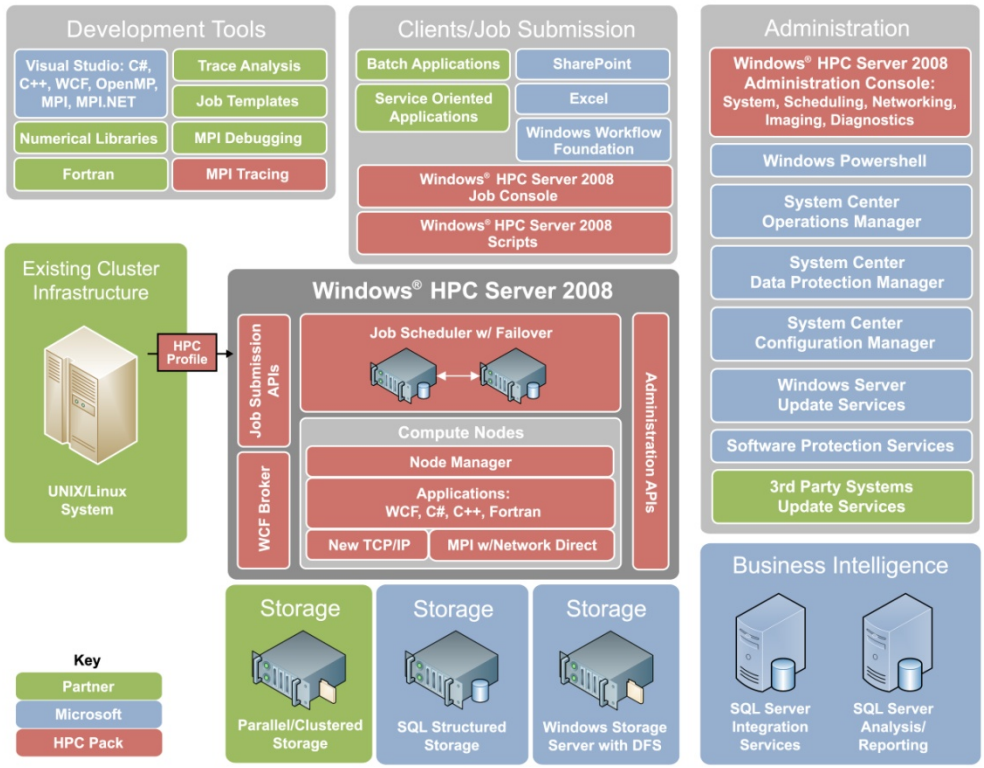

Figure 5. The HPC ecosystem

Windows HPC Server 2008 uses SQL Server (or SQL Express) for the Job Scheduler, greatly improving the scalability and redundancy of the scheduler. Windows HPC Server 2008 also supports integration with the System Center solutions for enterprise management and configuration management, including System Center Operations Manager. Windows HPC Server 2008 will include a custom System Center Operations Manager Pack to allow for advanced management and alerting through System Center Operations Manager 2007.

To improve reporting on jobs and the overall health of the cluster, Windows HPC Server 2008 uses SQL Server Reporting Services to provide standard reports. These reports can be customized, or completely new reports can be created, to meet the specific needs of your environment.

The new Heat Map view, shown in Figure 6, is part of the Cluster Management Console and provides an instant overview of the health of the cluster.

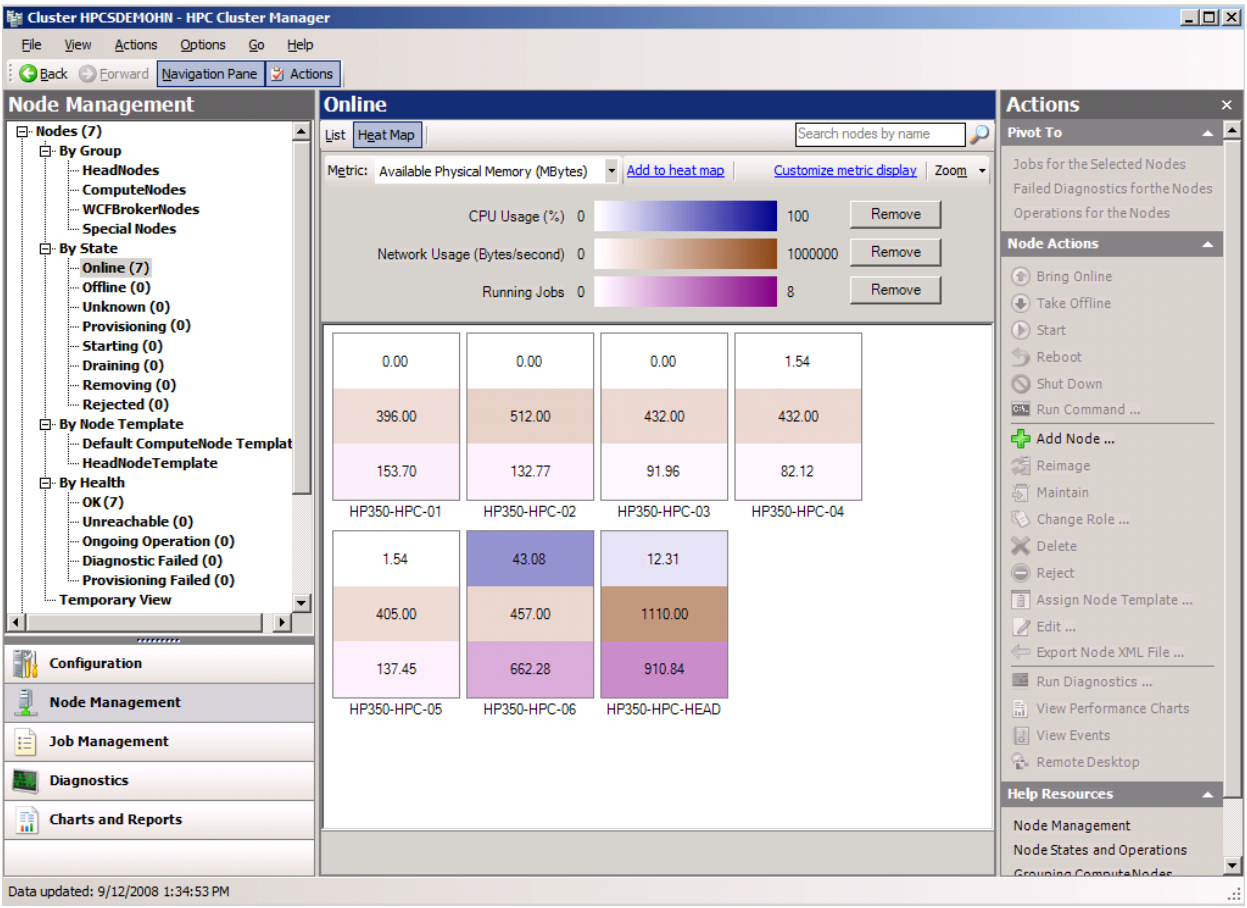

Figure 6. The Heat Map view gives instant feedback on the health of the cluster

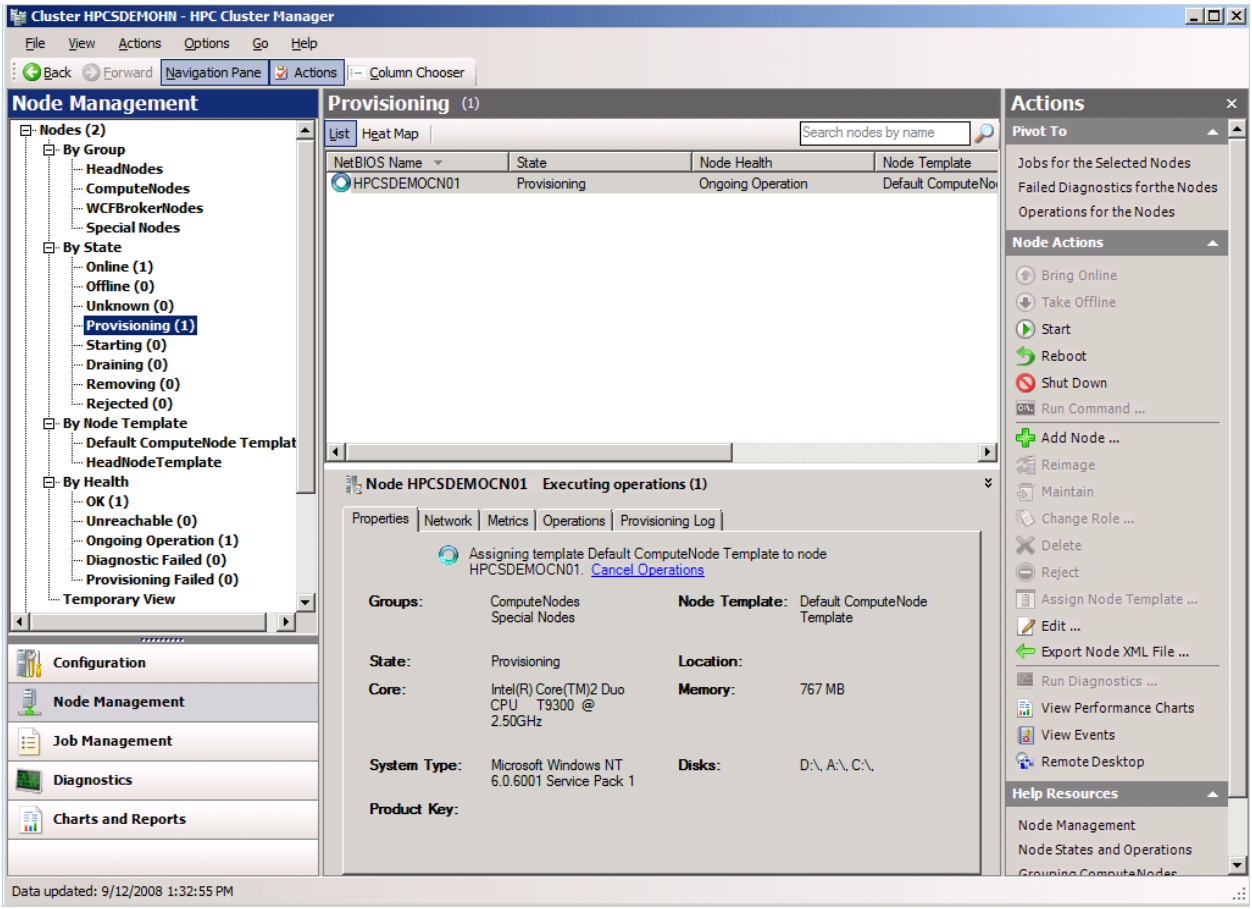

Figure 7. Node management

Failover Clustering of Head Node

An important new feature in Windows HPC Server 2008 is support for Windows Server 2008 Failover Clustering. Windows HPC Server 2008 integrates Failover Clustering in Windows Server 2008 Enterprise to ensure easy setup of the head node and failover node with SQL Server failover cluster. In the event of a failure of the head node, failover clustering will automatically (or manually, if desired) fail over the Job Scheduler. Job Scheduler clients see no change in the head node during the failover (and fallback) process, ensuring uninterrupted cluster operation.

Microsoft Message Passing Interface and NetworkDirect

The Microsoft Message Passing Interface (MS-MPI) standard is a portable, flexible, vendor- and platform-independent standard for messaging within and between HPC nodes. MPI is the specification, and MS-MPI, MPICH2, and others are the implementations of that standard. MPI2 is an extension of the original MPI specification. MS-MPI is a messaging interface based on the Argonne National Laboratory open-source MPI2 implementation and is compatible with the MPICH2 reference implementation and other MPI implementations.

Fundamentally, MPI is the application interconnection between nodes on an HPC cluster, tying nodes together. MPI provides a portable and powerful interprocess communication mechanism that simplifies some of the complexities of communication between hundreds or even thousands of processors working in parallel.

The MS-MPI in Windows HPC Server 2008 uses a new RDMA-based NetworkDirect interface for best performance and CPU efficiency. As shown in Figure 8, NetworkDirect uses a more direct path to support networking hardware, providing extremely fast and efficient networking. Speeds and latencies are of the same order as custom, hardware-specific, high-speed MPI drivers from hardware providers.

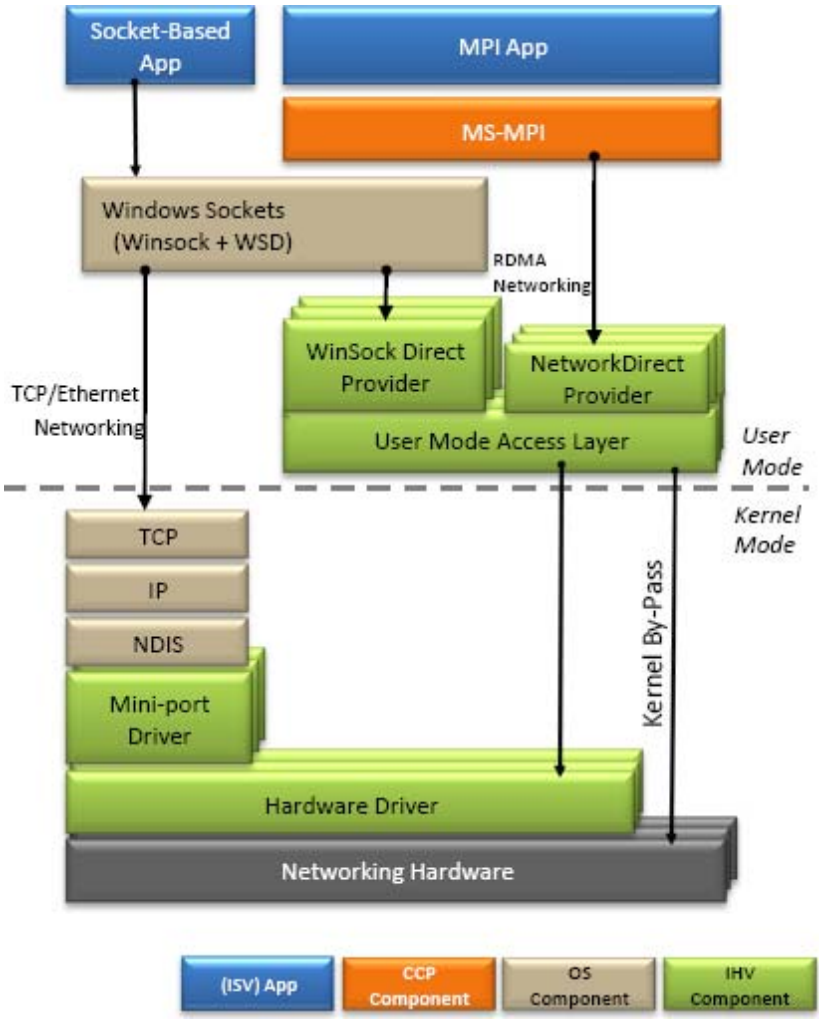

Figure 8. Networking architecture

MS-MPI can use any Ethernet interconnect that is supported by Windows Server 2008 as well as low-latency and high-bandwidth interconnects, such as InfiniBand, Myrinet, and 10GigE, through NetworkDirect drivers provided by the hardware manufacturers. Gigabit Ethernet provides a high-speed and cost-effective interconnect fabric, and interconnects, such as InfiniBand or Myrinet are ideal for latency-sensitive and high-bandwidth applications.

An important new feature in MS-MPI for Windows HPC Server 2008 is integration with Event Tracing for Windows, which allows better tuning for maximum performance and provides a time-synchronized log for debugging MPI system and application events across multiple computers running in parallel.

MS-MPI for Windows HPC Server 2008 includes significant improvements in MPI for shared memory communication. These improvements benefit the multicore systems prevalent in high performance computing.

MS-MPI includes support (bindings) for the C, Fortran77, and Fortran90 programming languages. Microsoft Visual Studio® 2005 includes a parallel debugger that works with MS-MPI. Developers can launch their MPI applications on multiple compute nodes from within the Visual Studio environment, and Visual Studio will then automatically connect the processes on each node, enabling the developer to individually pause and examine program variables on each node.

For more detailed information on MS-MPI see *Using Microsoft Message Passing Interface* (on the Microsoft TechNet Web site at http://go.microsoft.com/fwlink/?LinkId=55930).

Job Scheduler

The Job Scheduler queues jobs and their associated tasks. It allocates resources to these jobs; initiates the tasks on the compute nodes of the cluster; and monitors the status of jobs, tasks, and compute nodes. The Windows HPC Server 2008 Job Scheduler has been enhanced to scale and support much larger cluster installations and a greater number of simultaneous jobs. It now includes new policies for greater flexibility and resource utilization, and is built to address both traditional batch jobs as well as newer service-oriented applications. It is faster, more flexible, and provides support for heterogeneous clusters—clusters whose nodes have different hardware and different software.

The Windows HPC Server 2008 Job Scheduler supports both command-line and graphical interfaces. The GUI-based Job Scheduler is directly integrated into the Administration Console, and the command-line interface uses Windows PowerShell for flexibility and ease of use and supports Common Object Model (COM) access for use with Microsoft Visual Basic® Scripting Edition (VBScript) and other scripting languages. The Windows HPC Server 2008 Job Scheduler is backward compatible with existing command-line interface scripts from Windows Compute Cluster Server and can interoperate with third-party job schedulers for maximum compatibility with existing environments. The command-line interface supports a variety of languages, including Perl, Fortran, C/C++, C#, and Java.

The Windows HPC Server 2008 Job Scheduler enables users to schedule jobs, allocate resources needed for the job, and change the tasks and properties associated with the job. It includes built-in parametric support and custom job filtering, and supports heterogeneous and multicore clusters.

Jobs can be single tasks or multiple tasks and can specify the number of processors required for the job and whether those processors are needed exclusively or can be shared with other jobs and tasks.

The important distinguishing features of the Windows HPC Server 2008 Job Scheduler include:

- **NUMA-aware.** The Windows HPC Server 2008 Job Scheduler is Non-Uniform Memory Architecture (NUMA)-aware, and multicore-aware, allowing intelligent scheduling of jobs on large clusters of multicore nodes.
- **SOA support.** The Windows HPC Server 2008 Job Scheduler supports interactive SOA workloads. SOA workloads run on WCF hosts, and communicate with the submitting systems through WCF Brokers. Additional WCF Brokers can be added as necessary for additional scalability.
- **Scheduling policies**. The Windows HPC Server 2008 Job Scheduler supports five new policies: resource matchmaking, job admission policies through the job templates, multilevel processor allocation, adaptive allocation (grow/shrink), and preemption.

Security

Because HPC clusters are being adopted by a broad range of mainstream users for mission-critical applications, security and integration with the existing infrastructure are essential. Windows HPC Server 2008 uses Active Directory to enable role-based security for all cluster jobs and administration. The scheduler runs each job under the context and credentials of the submitting user, not as a super user or administrator. All credentials are encrypted and stored with the job only until the completion of the job. This behavior enables the compute jobs to access network resources, such as file or database servers, in the context of the user, while enabling systems administrators to apply and audit security policies using the existing and familiar mechanisms available in Active Directory.

Job management communications are done over encrypted and authenticated channels, and the user's credentials are known only to the scheduler and the node manager process on the compute nodes—not to processes or applications started on the user's behalf—further isolating credentials and protecting their integrity.

This additional security is an important addition to the MS-MPI implementation that is not part of the reference MPICH2 implementation. HPCS inherits all the additional security features that are a part of Windows Server 2008, including the new security features such as Network Access Protection, Role Management, Network Policy Management, and an integrated, bidirectional Windows Firewall for enterprise-facing networks.

Another important security and stability feature is the integrated patch management built in to HPC 2008 Pack. This allows the cluster administrator to schedule and deploy updates to groups of cluster nodes with the assurance that patches won't interfere with running jobs.

Summary

For those seeking highly productive solutions for high performance computing, Windows HPC Server 2008 provides a comprehensive platform built on Windows Server 2008, which helps to simplify deployment, management, and integration with existing infrastructure, thus helping to improve the productivity of your system administrators, application developers, and end users. Windows HPC Server 2008 unites the power of commodity x64-based computers, the security of Active Directory, and the Windows Server 2008 operating system to provide an affordable, easy-to-use, and scalable HPC solution while delivering on performance. Windows HPC Server 2008 uses Node Templates to help simplify and speed deployment of compute nodes using standard Windows Server 2008 deployment technologies. Additional compute nodes can be added to a cluster by simply connecting computers to the network. The Microsoft Message Passing Interface implementation is compatible with the reference MPICH2 and uses high-speed NetworkDirect drivers. Integration with Active Directory helps enable role-based security for administration and users, and the use of the System Center user interface model provides a familiar administrative and scheduling interface. The Windows HPC Server 2008 Job Scheduler supports heterogeneous clusters and enables the use of Service Oriented Architecture applications on the cluster.

Resources

For the latest information about Microsoft Windows HPC Server 2008, visit the following Web sites:

- Windows HPC at http://www.microsoft.com/hpc/
- Microsoft HPC Community at http://windowshpc.net

For more information about Microsoft System Center, Windows Server 2008, and the Windows Server System, visit the following Web sites:

- Windows Server 2008 at http://www.microsoft.com/windowsserver2008/
- Microsoft System Center at http://www.microsoft.com/systemcenter/
- Windows Server System at http://www.microsoft.com/servers/