IBM Netfinity Availability Extensions for Microsoft Cluster Server

Continuing leadership in clustering technology

Executive Summary

In today's business environment, highly available and scalable server-based computer systems and applications are a priority. One method of configuring server systems in a network so that they can provide backup for each other based on the needs of client/server applications is known as clustering. Clustering software allows a server application, service (such as file/print) or resource (for example a shared drive) to migrate smoothly from one server to another at the choosing of the system administrator or in the event of failure. The more servers in the cluster (scaling), the more potential candidates to which the application can be migrated, thus reducing the chance of an outage. Clients accessing the application should notice little or no impact to their support during this migration, thus ensuring the high availability of the application.

Microsoft's enhanced version of their Windows NT® operating system—Windows NT Server 4.0 Enterprise Edition—introduced a clustering technology for the Windows NT operating system, which they named Microsoft Cluster Server (MSCS). MSCS software provides a clustering environment that allows server-based applications to become highly available by linking two servers running Microsoft® Windows NT 4.0 Enterprise Edition. MSCS's primary function is to allow resources (for example files or disk drives) that are dedicated to one server to move to another, backup server, in case of failure or being taken offline. In this event, clients using the server resources experience little or no interruption in service because the resource functions are moved from one server to another.

IBM has developed a new technology that provides the ability to extend Microsoft's implementation of MSCS across multiple MSCS clusters. This allows for the expansion of support from three to eight nodes, each acting as a single cluster. This extension confirms IBM's intention to maintain our strategic alliance with Microsoft for API selection and validation for industry users of Windows clustering products. This paper offers an overview of the IBM Netfinity® technology that makes this extension possible.

Introduction to High-availability Clustering

System and application availability is a critical factor for almost every industry in today's electronic world. Because a company's operations might be based on one computer system staying up and running, its failure to perform even for a short time can result in significant losses. Productivity of an entire enterprise can be interrupted if one server goes down, especially if it runs supply chain or other equally demanding applications. According to an estimate made by Standish Group, the average cost per minute of downtime is $10,000 in revenue, productivity or profit.[1] Because of this potential threat to businesses' competitive edge in the marketplace, there is a growing demand for application solutions with increased reliability and availability. And these applications may require businesses to operate 24 hours a day, 7 days a week and 365 days a year.

In such environments, choosing the most available computer system becomes a vital business decision. And more and more businesses are deciding to buy Intel processor-based servers. They are less expensive than large-system servers, conform to industry standards and, usually, help lower businesses' total cost of ownership.

Many businesses today are connecting servers together into clusters, which are rapidly becoming a preferred configuration in demanding environments. In these environments, having the right clustering software is as important as having the right hardware. Today's Netfinity clustering technology offers the reliability, availability, scalability and manageability that enterprises need to help achieve the following benefits:

- High availability through Predictive Failure Analysis® by notifying you of a failing component and initiating automatic recovery mechanisms
- Access to data and shared devices
- Improved performance and the ability to manage future growth
- Workload balancing
- Single point of control and management
- Elimination of single points of failure

Single points of failure can lead to costly unplanned downtime, and even planned downtime for upgrades or maintenance can put a cluster out of action. IBM's Netfinity products have been designed to leverage the high-availability technology of our large servers to help customers greatly reduce or even eliminate both planned and unplanned downtime.

IBM has been a leader in providing high-availability solutions in the high-end server market. IBM's High-availability Cluster Multiprocessing (HACMP) for AIX® has been rated the number-one high-availability solution on the UNIX® platform for a number of years.[2] Now IBM has developed a new technology that provides the ability to expand on Microsoft's implementation of MSCS across multiple MSCS clusters to allow for the expansion of support from three to eight clusters.

The technology that allows this is X-architecture—IBM's blueprint for Netfinity servers. Netfinity X-architecture takes the best management capabilities from larger IBM systems and adapts them into a framework that will integrate with a wide range of industry-standard, customer-chosen management and operating system environments.

---

[1] Standish Group Research Note: Pound Foolishness, 1998 High Availability Forecast.

[2] D.H. Brown Associates, Inc.: Competitive Analysis of Reliability, Availability, Serviceability and Cluster Features, 1998.

IBM Netfinity X-architecture

The following is a summary of the key elements of Netfinity X-architecture that have been incorporated into selected Netfinity servers.[3] They include powerful processors, core logic, Chipkill memory (which protects against the failure of a single memory chip), reliable and highly available memory systems, scalable I/O, advanced caching software and world-class silicon and module technology. Netfinity X-architecture also includes clustered systems featuring technology from IBM's industry-leading S/390® and RS/6000® SP™ product lines, as well as interoperability with existing large and midrange systems.

Netfinity X-architecture is evident in these features in selected Netfinity servers:

- Fibre Channel-attached storage options for scalable, highly available, cluster-enabled storage, improved security and disaster protection
- Hot-plug hard disk drives, power supplies, fans and PCI slots for cluster availability and reliability
- Clustering solutions for higher system availability and performance scalability
- Light-path diagnostics to improve availability and serviceability
- Integration with enterprise systems management software such as Tivoli™ Management Software, Microsoft SMS and Intel® LANDesk® for management flexibility

IBM has focused on these features in developing IBM Netfinity Availability Extensions for MSCS.

And because clusters can run diverse application solutions, it is also vital that applications work well across systems. To that end IBM recently announced the IBM ClusterProven™ Program on Netfinity.

IBM ClusterProven Program on Netfinity Servers

IBM is the leader in clustering technology. That leadership has been reinforced by the IBM ClusterProven Program on Netfinity, which is focused on providing robust and effective support to qualified Netfinity solution developers to join the high-availability trend through delivery of a wide range of proven, highly available software applications. These clustering solutions for Netfinity server products will be tested so that they meet IBM's strict standards for high availability, and will be identified in the IBM Global Software Solutions Guide, accessible online by customers.

In the first half of 1999 the ClusterProven Program on Netfinity will focus on solutions for Microsoft Windows NT and MSCS. In the second half of 1999 and beyond, IBM intends to expand the program to include additional clustering platforms.

---

[3] *To learn more about Netfinity servers, see "Additional Information" at the end of this paper and visit the Netfinity Web site at **www.ibm.com/netfinity**.

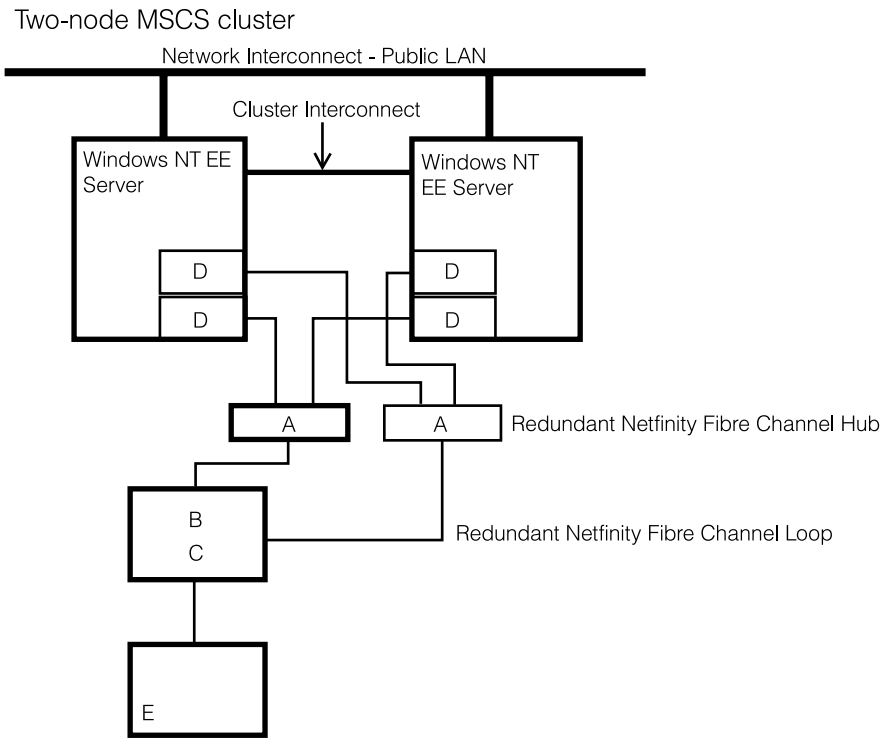
Microsoft Cluster Server

Microsoft Corporation released an enhanced version of their Windows NT operating system, called NT Server 4.0 Enterprise Edition (EE), in late 1997. As part of this release, Microsoft introduced a clustering technology for the Windows NT operating system, which they named Microsoft Cluster Server (MSCS). MSCS software provides a clustering environment that allows server-based applications to become highly available by linking two servers or nodes running Microsoft Windows NT 4.0 EE. And Microsoft has announced its intention to achieve greater than two-node support with Windows 2000 Data Center.

MSCS offers many potential benefits:

- High availability
- Dramatically reduced planned and unplanned downtime
- Failover and failback
- Graphical management console

IBM has certified MSCS on select Netfinity server hardware. It works as shown in the following sample figure containing two nodes and IBM Fibre Channel storage subsystems.



Product names:

A - IBM Netfinity Fibre Channel Hubs
B - IBM Netfinity Fibre Channel RAID Controller Unit A
C - IBM Netfinity Fibre Channel Failsafe RAID Controller B
D - IBM Netfinity Fibre Channel PCI Host Bus Adapters
E - IBM Netfinity EXP15 Rack Storage Enclosure

For each application running on a system in the cluster, there is a set of resources defined to support that application when running on the server. These resources are under the control of MSCS, and MSCS-enabled applications have resource dynamic link libraries (DLLs) written so that they communicate with these resources through the MSCS control layer. Resource DLLs allow applications to be cluster-aware. MSCS provides a resource monitor that interacts with Windows NT Cluster Server to provide status and to maintain the resources for the system. MSCS provides a set of default resource types and associated resource DLLs.

A more detailed description of how MSCS works can be found in the *Microsoft Cluster Server Administrator's Guide* from Microsoft Corporation.

MSCS's primary function is to allow resources (which can include physical hardware devices such as disk drives and network adapters, or logical items such as logical disk volumes, TCP/IP addresses, applications and databases) that are dedicated to one server to move to another, backup server, in case of failure or planned downtime. In the event of failure or downtime, clients using the server resources experience little or no interruption in service because the resource functions are moved from one node to another.

One of the first widely accepted clustering solutions for industry-standard servers, MSCS provides excellent, high-availability operations in a two-node cluster. For businesses requiring higher availability and reliability provided by larger clusters, and a single point of control, IBM has developed Netfinity Availability Extensions for MSCS.

IBM Netfinity Availability Extensions for MSCS

Enterprises that have standardized on Microsoft Windows or have established MSCS clusters are now planning to expand their clustered systems. The new IBM Netfinity Availability Extensions complements and extends MSCS capabilities. Delivering a virtual multinode cluster that manages a collection of MSCS clusters as if they were one, Availability Extensions facilitates failover between clusters with support for three to eight nodes in an odd or even configuration. Binary compatibility is maintained with the MSCS Cluster Services API and with MSCS Resource DLLs. This means that applications that are MSCS cluster aware are supported with minimum to no modification. The Netfinity Availability Extensions enables customers to exploit the combined advantages of Windows NT EE and reliable Netfinity servers.

In order that this can be accomplished, IBM and Microsoft have jointly defined and developed a test suite for clustering extensions to validate API compliance for industry users of these standard Windows NT cluster APIs.

The IBM Netfinity Availability Extensions package integrates software and services, including pre-installation planning, installation and setup by trained technicians from IBM Global Services, and post-installation technical support.

This clustering solution is based on the award-winning[4] Netfinity 7000 M10 enterprise server. Powerful, versatile and reliable, the 7000 M10 is well matched to the requirements of clustering. And it comes with its own extensive package of service and support.

Based on IBM Cluster Systems Management with an easy-to-use GUI, IBM Cluster Server Manager has been enhanced to manage clusters of more than two nodes. The ability to specify

[4] Netfinity 7000 M10 was awarded the "Best New Product–Enterprise Server or System" award for innovation, specification, usability and value at Federal Office Systems Exposition (FOSE) in March 1999. FOSE is the largest U.S. Government information technology show and conference.

failover policy and prioritized sequence of failover and failback sequence allows an administrator to remain within a system's capacity in the event of a failure. If Netfinity Manager™ 5.2 or higher is running (recommended), an administrator can also have improved alert management for a cluster.

Netfinity Availability Extensions Components

There are three major components in a Netfinity Availability Extensions cluster: Cluster Services, Recovery Services and Trace and Logging Services.

Cluster Services. Cluster Services manages cluster nodes and keeps track of resources on nodes. Cluster Services take recovery actions for failure, configuration and administration events. For example when a node fails, Cluster Services brings the resource groups that were running on the failed node back online on one or more of the remaining nodes.

Recovery Services. Recovery Services deals with events caused by administrative operations on groups, nodes and resources. Examples are events that bring a group online or offline, or that move it from one node to another. Such administrative operations can be initiated from IBM Cluster Manager. They export status information about nodes, adapters, network interface, groups and resources, and also deal with cluster partitioning, where a cluster is split into two or more separate clusters.

Trace and Logging Services. Trace and Logging Services provides tracing and logging APIs for use by other cluster components. You can view the log entries by viewing the log files in the log directory.

Advantages of IBM Netfinity Availability Extensions for MSCS

IBM's Netfinity Availability Extensions for MSCS provides the following functions and benefits:

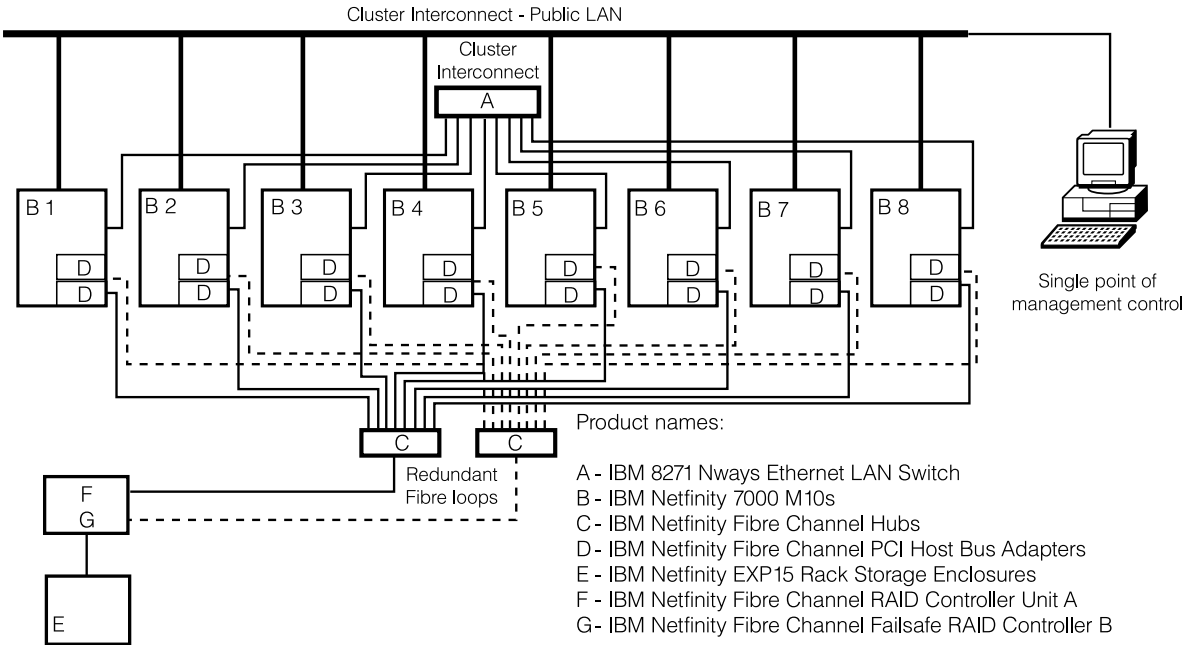
- The ability to manage, from a single console, an increasing number of nodes as a larger virtual cluster
- Clustering support for applications to expand up to 8 nodes, where each node is set up as a one-node MSCS cluster
- Clustering support in either even or odd node configurations
- MSCS-compatible APIs with minimal deviations, and some extensions to support more than two nodes

IBM's Netfinity Availability Extensions for MSCS provides advantages to customers using MSCS-certified IBM Netfinity hardware. Customers have greater flexibility in their hardware resources and can use them more effectively:

- Configuration to minimize any single point of failure
- Because three or more nodes are included in a cluster, availability is improved through cascading failover (B1 to B2, B2 to B3...).
- One server (or as many as are defined by the cluster administrator) can be used as hot-standby in a many-to-one configuration (B8 in the example in the figure).
- Any-node-to-any-node failover services

- Lower system cost implementation for a given level of availability
- Higher level of availability for a given level of cost
- Any node can be set up as a primary node and as a backup node for other applications running on other nodes.

8-Node Netfinity Availability Extensions cluster



Netfinity Availability Extensions for MSCS failover and failback. The failover policy implemented in Netfinity Availability Extensions for MSCS is a cascading failover policy. That is, in the event of a node failure, an application (resource group) on the failed node will be restarted on the node next to the one on which it was running, in the best owner list for this group (with a wraparound when the end of the list is reached). The best owner list is obtained by taking the list of possible nodes for this group and prioritizing it according to the list of preferred owners for this group. The list of possible owners for a group is derived from the list of possible owners for the resources contained in the group. The other MSCS failover properties will also be honored by this policy.

Two types of failover are supported in this release of Netfinity Availability Extensions for MSCS: *N+1* and *Nway* failover. N+1 failover allows recovery to a node in the cluster that is an online standby node (a node that does not run any resource groups when all the nodes are up). Nway failover allows recovery to any surviving node in the cluster that is defined in the best owner list for each group.

The failback policy implemented in Netfinity Availability Extensions for MSCS is an auto-homing failback policy. For example, consider a resource group which has a preferred owner list (1, 2, 3, 4). If node 1 fails, the group is failed over to node 2. If node 2 fails, the group is failed over to node 3. If node 2 comes back up, the group is not failed back to node 2, but if node 1 comes back up the group is failed back to node 1. So the auto-homing failback policy fails back only when the first node in the preferred owner list comes back up. The other MSCS failback properties will also be honored by this policy.

Disk pooling. Disk pooling with Netfinity Availability Extensions is the ability for up to eight Netfinity 7000 M10 servers to connect to and share (via a Fibre Channel based Netfinity Storage Area Network [SAN]) one or more Netfinity Fibre Channel RAID storage subsystems. Because each Fibre Channel RAID subsystem can have multiple RAID arrays (LUNs) defined, one way to share the storage is to define multiple LUNs, which can be allocated to individual servers in the Availability Extensions cluster. Using the MSCS capabilities, these resources can be allowed to fail over between cluster nodes in the event of node failure or can be dedicated to one particular server. By using the Availability Extensions, it is possible to easily change the allocation of resources among the servers as storage needs change or for performance balancing.

The potential benefits of this disk pooling capability to customers are:

- Lower total solution cost by allowing sharing Netfinity Fibre Channel subsystems by several servers
- Improved manageability and reduced administration costs by allowing resources to be managed as a common pool that can be reallocated on demand without needing to physically move or reconfigure the hardware components
- Once the Netfinity SAN is in place to support the shared disk storage, the ability to add Fibre Channel-to-SCSI bridges to support the connection of shared tape libraries to allow tape pooling (sharing a tape backup unit among multiple servers) is an attractive addition to the solution.

Netfinity Availability Extensions for MSCS Cluster Management

Netfinity Availability Extensions for MSCS enables management of an eight-node cluster from a central management console and is easily accessed through Netfinity Manager.

Netfinity Availability Extensions for MSCS software consists of two major components: Services and IBM Cluster Manager. Services uses Core Cluster Services (CCS) to extend the scalability of MSCS to eight nodes for certain IBM Netfinity servers. Each node in a cluster is configured as a one-node MSCS cluster.

IBM Cluster Manager is responsible for configuration and administration of the Netfinity Availability Extensions for MSCS product.

Netfinity Availability Extensions for MSCS Services. Services has three major, integrated components (discussed previously) that provide many functions in a cluster. Together the components are responsible for managing the resources and resource groups in a cluster. They manage cluster nodes and keep track of resources on nodes. They take recovery actions for failure, configuration and administration events. For example, when a node fails, they bring the resource groups that were running on the failed node back online on one or more of the remaining nodes.

They also deal with events caused by administrative operations on groups, nodes and resources. Examples are events that bring a group online or offline, or that move it from one node to another. Such administrative operations can be initiated from IBM Cluster Manager. They export status information about nodes, adapters, network interface, groups and resources, and also deal with cluster partitioning, where a cluster is split into two or more separate clusters.

These services provide tracing and logging APIs for use by other cluster components. You can view the log entries either in the NT Event log or by viewing the log files located in the log directory.

IBM Cluster Systems Manager. Management and middleware technologies are two components critical to horizontal scalability of clusters. IBM Netfinity servers offer IBM Cluster Systems Manager, which builds management and control features on top of the MSCS feature of Windows NT Server Enterprise Edition. IBM Cluster Systems Manager gives MSCS administrators improved control of clustered installations including Netfinity Availability Extensions. IBM's offering simplifies cluster administration by providing single-console control of multiple clusters and their respective cluster resources. It also can increase management control by providing resource alerting capabilities to IBM Netfinity Manager, Microsoft SMS and Intel LANDesk management software.

Conclusion

High availability is a critical factor for almost every industry in today's electronic world. The growing popularity of Enterprise Resource Planning systems and core computing support enterprise development and growth, but at the same time they increase dependency on information technology. If a server goes down or must be taken offline for maintenance, the losses can be staggering.

Clustering enables the grouping of servers so that the resources on one server can be moved to another in the event of failure or being taken offline, with little or no impact on end users. Microsoft's MSCS allows the formation of two-node clusters, which means increased availability.

The IBM Netfinity Availability Extensions for MSCS is a validation of the unique capabilities of IBM in bringing management services previously available only on its SP cluster products to the Windows NT customer. This helps Windows NT meet increasing business-critical requirements on Windows NT, and IBM is working with Microsoft to enhance Windows NT enterprise readiness. IBM extensions provide today's MSCS customer with the ability to manage an increasing number of nodes—up to eight— at a lower potential cost for a given level of availability. It also offers an any-to-any or any-to-one cluster implementation.

In summary, Netfinity Availability Extensions for MSCS provides an additional management layer to manage multiple MSCS clusters and to monitor and manipulate the application resources defined to those MSCS clusters. It can provide Netfinity customers with an MSCS-compatible, flexible, lower cost method of providing highly available, and highly scalable, client/server applications.

Additional Information

For more information on IBM Netfinity directions, products and services, refer to the following white papers, available from our Web site at **www.ibm.com/netfinity**.

Management

Implementing IBM Netfinity Server Management
Integrating IBM Netfinity Manager with Microsoft System Management Server
Integrating IBM Netfinity Manager with Intel LANDesk Server Manager
IBM Netfinity Manager 5.2
IBM Netfinity Manager Plus for Tivoli Enterprise Overview
IBM Netfinity Advanced Systems Management
IBM Netfinity Advanced Systems Management for Servers
IBM ServerGuide for Netfinity and PC Server Systems

Other Topics

Enterprise Storage Solutions
IBM Chipkill Memory
IBM Netfinity X-architecture
IBM ClusterProven Program on Netfinity
IBM Netfinity Predictive Failure Analysis
IBM Netfinity Cluster Directions
IBM Netfinity Web Server Accelerator
Lotus Domino Clusters Overview
Lotus Domino Clusters Installation Primer
Implementing Microsoft IIS on Netfinity 5500 M10
IBM Netfinity ESCON Adapter
IBM Netfinity Hot-Plug Solutions
IBM Netfinity Storage Management Solutions Using Tape Subsystems
IBM Netfinity 8-Way SMP Directions
IBM Netfinity Fibre Channel Directions
IBM Netfinity Server Ultra2 SCSI Directions
IBM Netfinity Server Quality
IBM Netfinity 5000 Server
IBM Netfinity 5500 Server Family
IBM Netfinity 7000 M10 Server
Achieving Remote Access Using Microsoft Virtual Private Networking
At Your Service...Differentiation beyond technology

IBM Personal Computer Company
3039 Cornwallis Road
Dept. LO6A
Research Triangle Park, NC 27709

Printed in the United States of America
5-99

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

IBM reserves the right to change specifications or other product information without notice.

IBM Netfinity systems are assembled in the U.S., Great Britain, Japan, Australia and Brazil and are comprised of U.S. and non-U.S. components.

Are you Year 2000 ready? Visit **www.ibm.com/pc/year2000** or call 1 800 426-3395 (and request document number 10020 from our faxback database) for the latest information.