Feature description

The Network Load Balancing (NLB) feature distributes traffic across several servers by using the TCP/IP networking protocol. By combining two or more computers that are running applications into a single virtual cluster, NLB provides reliability and performance for web servers and other mission-critical servers. The servers in an NLB cluster are called *hosts*, and each host runs a separate copy of the server applications. NLB distributes incoming client requests across the hosts in the cluster. You can configure the load that is to be handled by each host. You can also add hosts dynamically to the cluster to handle increased load. NLB can also direct all traffic to a designated single host, which is called the *default host*. NLB allows all of the computers in the cluster to be addressed by the same set of IP addresses, and it maintains a set of unique, dedicated IP addresses for each host. For load-balanced applications, when a host fails or goes offline, the load is automatically redistributed among the computers that are still operating. When it is ready, the offline computer can transparently rejoin the cluster and regain its share of the workload, which allows the other computers in the cluster to handle less traffic.

Practical applications

NLB is useful for ensuring that stateless applications, such as web servers running Internet Information Services (IIS), are available with minimal downtime, and that they are scalable (by adding additional servers as the load increases). The following sections describe how NLB supports high availability, scalability, and manageability of the clustered servers that run these applications.

High availability

A high availability system reliably provides an acceptable level of service with minimal downtime. To provide high availability, NLB includes built-in features that can automatically:

- Detect a cluster host that fails or goes offline, and then recover.

- Balance the network load when hosts are added or removed.

- Recover and redistribute the workload within ten seconds.

Scalability

Scalability is the measure of how well a computer, service, or application can grow to meet increasing performance demands. For NLB clusters, scalability is the ability to incrementally add one or more systems to an existing cluster when the overall load of the cluster exceeds its capabilities. To support scalability, you can do the following with NLB:

- Balance load requests across the NLB cluster for individual TCP/IP services.

- Support up to 32 computers in a single cluster.

- Balance multiple server load requests (from the same client or from several clients) across multiple hosts in the cluster.

- Add hosts to the NLB cluster as the load increases, without causing the cluster to fail.

- Remove hosts from the cluster when the load decreases.

- Enable high performance and low overhead through a fully pipelined implementation. Pipelining allows requests to be sent to the NLB cluster without waiting for a response to a previous request.

Manageability

To support manageability, you can do the following with NLB:

- Manage and configure multiple NLB clusters and the cluster hosts from a single computer by using NLB Manager or the NLB Windows PowerShell cmdlets.

- Specify the load balancing behavior for a single IP port or group of ports by using port management rules.

- Define different port rules for each website. If you use the same set of load-balanced servers for multiple applications or websites, port rules are based on the destination virtual IP address (using virtual clusters).

- Direct all client requests to a single host by using optional, single-host rules. NLB routes client requests to a particular host that is running specific applications.

- Block undesired network access to certain IP ports.

- Enable Internet Group Management Protocol (IGMP) support on the cluster hosts to control switch port flooding (where incoming network packets are sent to all ports on the switch) when operating in multicast mode.

- Start, stop, and control NLB actions remotely by using Windows PowerShell commands or scripts.

- View the Windows Event Log to check NLB events. NLB logs all actions and cluster changes in the event log.

Important functionality

Hardware requirements

To run an NLB cluster, the following are hardware requirements:

- All hosts in the cluster must reside on the same subnet.

- There is no restriction on the number of network adapters on each host, and different hosts can have a different number of adapters.

- Within each cluster, all network adapters must be either multicast or unicast. NLB does not support a mixed environment of multicast and unicast within a single cluster.

- If you use the unicast mode, the network adapter that is used to handle client-to-cluster traffic must support changing its media access control (MAC) address.

Software requirements

To run an NLB cluster, the following are software requirements:

- Only TCP/IP can be used on the adapter for which NLB is enabled on each host. Do not add any other protocols (for example, IPX) to this adapter.

- The IP addresses of the servers in the cluster must be static.

> **Note**
>
> NLB does not support Dynamic Host Configuration Protocol (DHCP). NLB disables DHCP on each interface t

Server Manager information

In Server Manager, use the Add Roles and Features Wizard to add the Network Load Balancing feature. Optionally you can install the Network Load Balancing Tools to manage a local or remote NLB cluster. The tools include Network Load Balancing Manager and the NLB Windows PowerShell cmdlets. For more information about installing features, see Install or Uninstall Roles, Role Services, or Features.