Lesson 1: Configuring Network Load Balancing

MOAC 70-412: Configuring Advanced Windows Server 2012 Services

Microsoft Official Academic Course

WILEY

Overview

- Exam Objective 1.1 – Configure Network Load Balancing (NLB).
- Install NLB nodes
- Configure NLB prerequisites
- Configure affinity
- Configure port rules
- Configure cluster operation mode
- Upgrade an NLB cluster

Understanding Fault Tolerance

Lesson 1: Configuring Network Load Balancing

Availability

- When a server goes down, it most likely causes your company to lose money.
  - If your network contains an external website or database that controls your sales, ordering, inventory, or production, server downtime can be detrimental to these business needs.
  - If it is an internal server, it might not allow your users to perform their jobs.
  - In either case, your company loses money either through lost revenue or lost productivity.

Availability

- When designing servers and the services they provide, servers are often assigned service level agreements (SLA), which specify the level of availability those servers or services must maintain.
- To have a server design that can support an availability of 99.999% is much more expensive than supporting an availability of 99%.

Fault Tolerance

- To make a server more fault tolerant, you should first look at what components are the most likely to fail and implement technology to make a system less likely to fail.
- Redundant components could include:
  - **Disks**: Use some form of RAID and hot spares.
  - **Power supplies**: Use redundant power supplies.
  - **Network cards**: Use redundant network cards.

Fault Tolerance

- A ***cluster*** is a group of linked computers that work together as one computer.
- Common uses may include:
  - A load-balancing cluster for the front end can provide the web interface to the back-end servers.
  - A failover cluster for back-end servers such as a database (e.g., SQL Server) or mail server (e.g., Exchange Server).

Configuring Network Load Balancing

Lesson 1: Configuring Network Load Balancing

Network Load Balancing (NLB)

- ***Network Load Balancing (NLB)*** transparently distributes traffic across multiple servers by using virtual IP addresses and a shared name.
- With NLB, you gain fault tolerance and enhanced performance. It is often used with mission-critical web servers but can also be used with other types of servers.

Network Load Balancing (NLB)

- A cluster has two or more servers known as nodes.
- Each ***node*** runs a separate copy of the desired service application such as a web server, an FTP server, or a Secure Shell (SSH)/Remote Desktop Server.
- NLB is a scalable, high-availability feature found in Windows Server 2012.
- Windows Server 2012 NLB clusters can have between 2 and 32 nodes.
- When you create an NLB cluster, you create a virtual network address and adapter that is assigned to the entire cluster.
- As network requests are sent to the virtual network address, the requests are distributed across the nodes in the cluster.

Heartbeats

- NLB can detect the failure of cluster nodes by sending packets known as ***heartbeats***.
- NLB cluster heartbeats are transmitted every second between nodes in the cluster.
- If a node misses five consecutive heartbeats, the node is automatically removed from the NLB cluster.

Convergence

- When a node is added or removed from a cluster, a process known as ***convergence*** occurs, where the cluster determines its current configuration by building a membership of nodes and mapping client requests based on the available nodes.
- Convergence can occur only if each node is configured with the same port rules.

NLB Prerequisites

- To support NLB, your systems must meet the following requirements:
  - All hosts in the cluster must reside on the same subnet.
  - Within each cluster, all network adapters must be either multicast or unicast. You cannot have some nodes configured as multicast while other nodes are configured as unicast within a single cluster. We discuss multicast and unicast configuration later in the lesson.
  - If unicast mode is used, the network adapter that is used to handle client-to-cluster traffic must support changing its media access control (MAC) address.
  - The IP addresses assigned to the nodes must be static.

Installing NLB Nodes

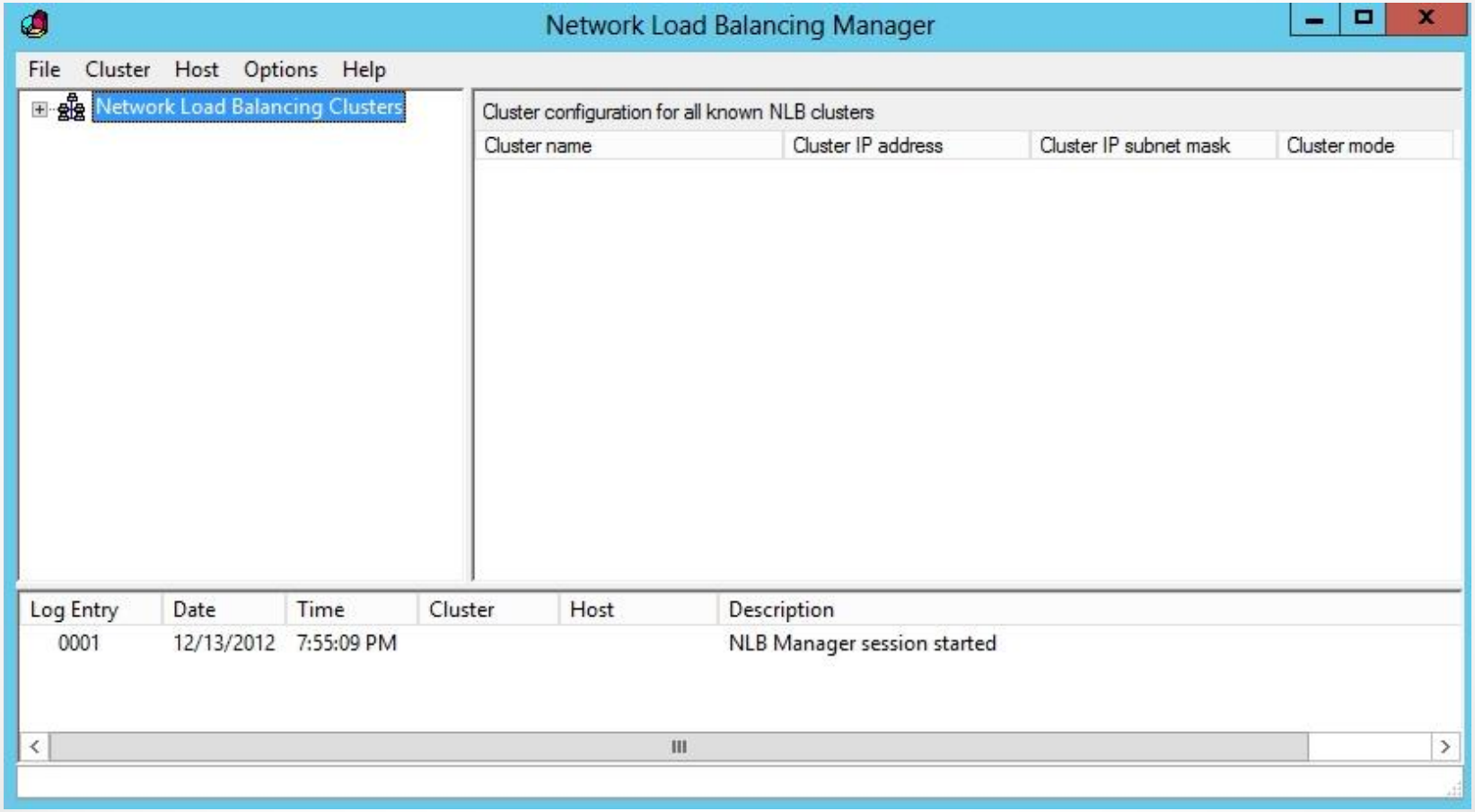
- To install and configure an NLB node, you must first install NLB.
- The NLB is a feature, not a role.
- It is used to enhance other roles such as web services or Remote Desktop Services.
- After NLB is installed on each machine, you then have to create the node and add each host to the cluster.

NLB Cluster

To configure the NLB cluster, you must configure three types of parameters:

- **Host parameters**: Defines what each node can do in an NLB cluster.
- **Cluster parameters**: Configures the NLB cluster as a whole.
- **Port rules**: Controls what ports the NLB cluster services and how requests are balanced across all servers.

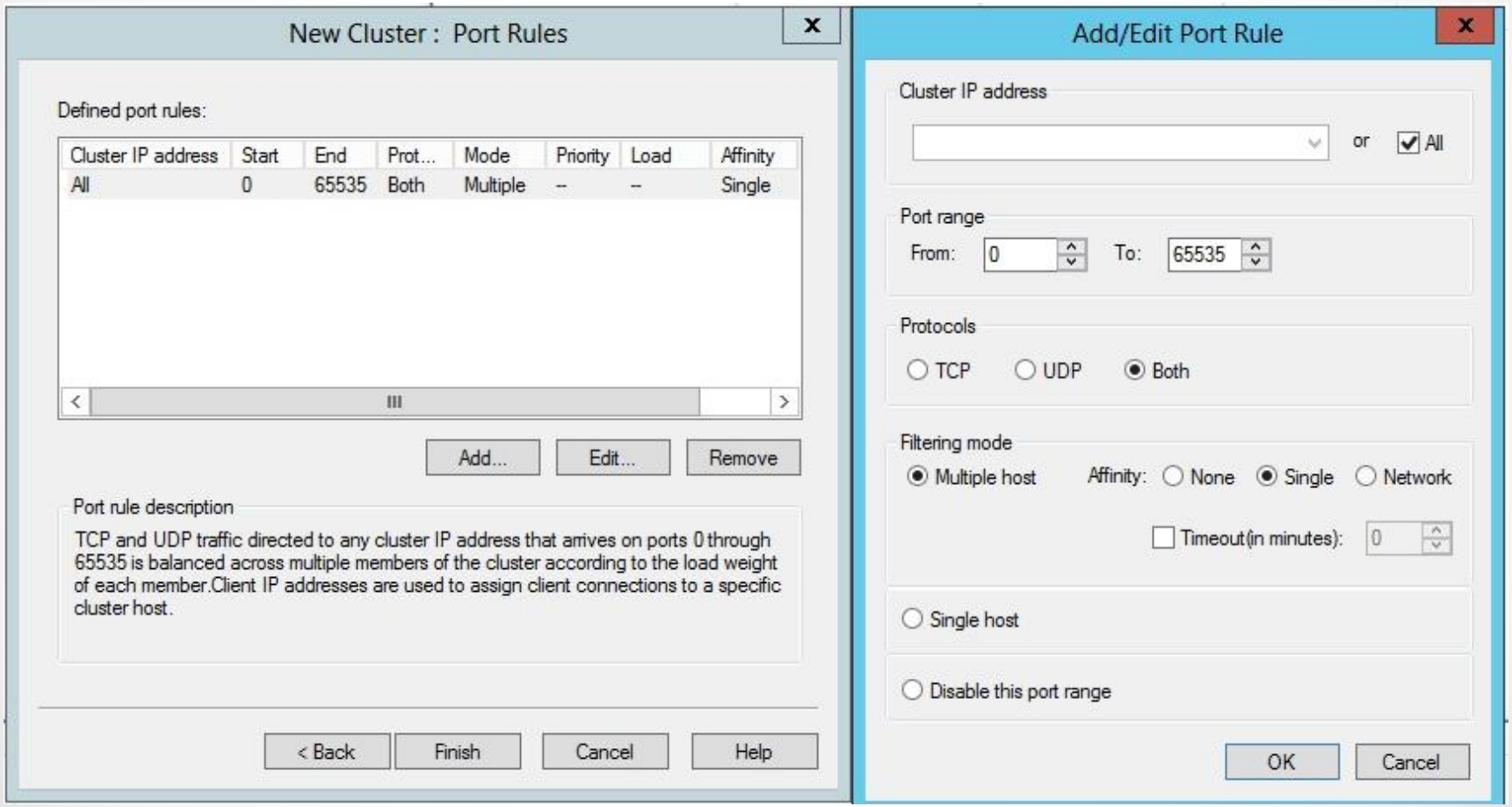
Network Load Balancing Manager



Port Rules

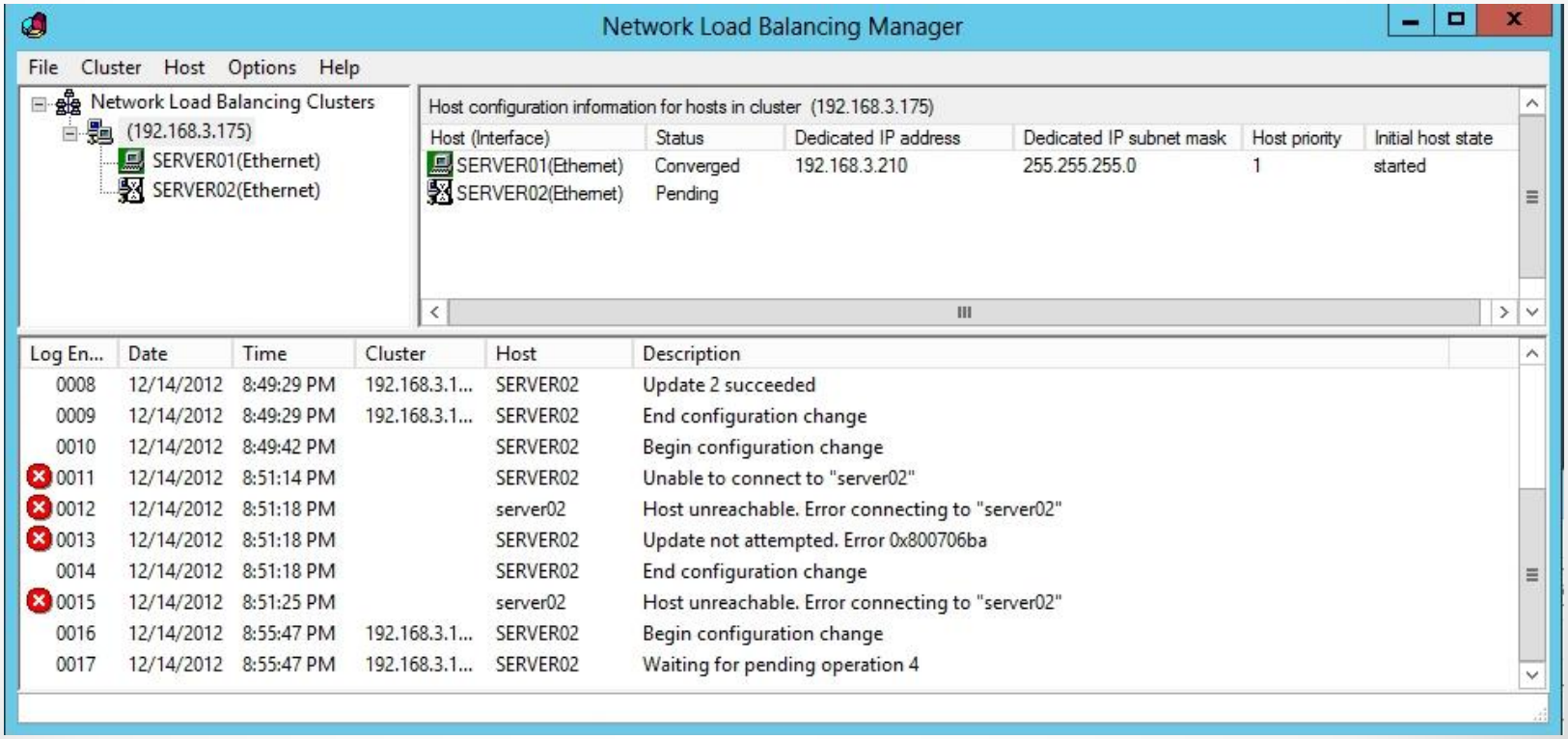
- ***Port rules***
  - Specify how NLB directs traffic based on the port and protocol.
  - Are used to configure how requests to specific IP addresses and ports are directed by the NLB cluster.
- Use port rule configuration to specify:
  - The virtual IP address that the rule should apply to
  - The TCP or UDP port range that this rule should apply to
  - The protocols that this rule should apply to, including TCP, UDP, or both
  - The filtering mode, described by the port range and the protocols, specifies how the cluster handles traffic.

Port Rules



Convergence

- After a host is added, convergence will occur.
- When convergence is complete, the host will participate in the cluster.



Affinity

- ***Affinity*** determines how the servers are going to balance the load.
- You use affinity settings when you use multiple host filter mode.

Filter Mode

- The ***filter mode*** specifies which hosts can respond to requests.
- The filter mode includes:
  - Multiple host
  - Single host
  - Disable

Multiple Host Filtering Mode

- If you choose the multiple host filtering mode, you can then configure the affinity.
- When you configure affinity, you can choose from three options:
  - None
  - Single
  - Class C

Cluster Operation Mode

- On the Cluster Parameters tab, you configure the virtual IP address, subnet mask, and DNS name that the cluster will use.
- You also can configure the cluster operation mode, which specifies whether a multicast MAC address should be used for cluster operations.

Unicast Mode

- When you configure an NLB cluster to use ***unicast mode***, NLB replaces the network card's original MAC address and all cluster hosts use the same unicast MAC address.
- When you use unicast mode with a single network adapter on each node, the computer can communicate only with other computers within the same subnet.

Multicast Mode

When an NLB host is in ***multicast mode***, each NLB network adapter has two MAC addresses (the original MAC address and the virtual MAC address).

Internet Group Management Protocol Multicast Mode

- The ***Internet Group Management Protocol (IGMP) Multicast mode*** is a special form of multicast mode that prevents the network switch from flooding with traffic.
- When you use IGMP multicast mode, traffic is forwarded only through the switch ports that are part of the NLB cluster.
- However, to use IGMP multicast mode, you need switch hardware that supports IGMP multicast mode.

Controlling Hosts in NLB

- To remove a node, you can perform a stop or a drainstop action:
  - ***Stop action***: Terminates all existing connections to the cluster node and stops the NLB service.
  - ***Drainstop action***: Blocks all new connections without terminating existing sessions.
- To control the host, right-click the node, click *Control Host,* and select the appropriate option (Start, Stop, Drainstop, Suspend, or Resume).

Upgrading an NLB Cluster

- There are two ways to upgrade a Windows Server 2008 R2 NLB cluster to Windows Server 2012:
  - Upgrade all the hosts at one time.
  - Upgrade each host, one at a time.

Lesson Summary

- High availability is a system design protocol and associated implementation that ensures a certain degree of operational continuity during a given measurement period.
- A cluster is a group of linked computers that work together as one computer. Based on the technology used, clusters can provide fault tolerance (often referred to as availability), load balancing, or both.
- The two most popular forms of clusters are failover clusters and load-balancing clusters.
- A load-balancing cluster for the front end provides the web interface to the back-end servers.
- A failover cluster for back-end servers such as a database (e.g., SQL Server) or mail server (e.g., Exchange Server).

Lesson Summary

- Network Load Balancing (NLB) transparently distributes traffic across multiple servers by using virtual IP addresses and a shared name. By using NLB, you gain fault tolerance and enhanced performance.
- A cluster has two or more servers known as nodes.
- Each node runs a separate copy of the desired service application such as a web server, an FTP server, or an SSH/Remote Desktop Server.
- NLB is able to detect the failure of cluster nodes by sending packets known as *heartbeats*.
- When a node is added or removed from a cluster, a process known as *convergence* occurs, where the cluster determines its current configuration by building a membership of nodes and mapping client requests based on the available nodes.

Lesson Summary

- To configure the NLB cluster, you must configure three types of parameters: host parameters, cluster parameters, and port rules.
- Port rules specify how NLB directs traffic based on the port and protocol.
- Affinity determines how the servers balance the load. You use affinity settings when you use multiple host filter mode.
- For a system with two network cards use unicast. If a server has only a single network card use multicast mode.
- The drainstop action blocks all new connections without terminating existing sessions.
- To upgrade an NLB cluster to Windows Server 2012, you can upgrade all the hosts at one time or upgrade each host, one at a time.