CHAPTER 4

Configuring Real Servers and Server Farms

This section provides an overview of server load balancing and procedures for configuring real servers and server farms for load balancing on an ACE appliance.

Note When you use the ACE CLI to configure named objects (such as a real server, virtual server, parameter map, class map, health probe, and so on), consider that the Device Manager (DM) supports object names with an alphanumeric string of 1 to 64 characters, which can include the following special characters: underscore (_), hyphen (-), dot (.), and asterisk (*). Spaces are not allowed.

If you use the ACE CLI to configure a named object with special characters that the DM does not support, you may not be able to configure the ACE using DM.

Topics include:

Server Load Balancing Overview

Server load balancing (SLB) is the process of deciding to which server a load-balancing device should send a client request for service. For example, a client request can consist of an HTTP GET for a Web page or an FTP GET to download a file. The job of the load balancer is to select the server that can successfully fulfill the client request and do so in the shortest amount of time without overloading either the server or the server farm as a whole.

Depending on the load-balancing algorithm or predictor that you configure, the ACE appliance performs a series of checks and calculations to determine the server that can best service each client request. The ACE appliance bases server selection on several factors, including the server with the fewest connections with respect to load, source or destination address, cookies, URLs, or HTTP headers.

The ACE Appliance Device Manager allows you to configure load balancing using:

- Virtual servers—See Configuring Virtual Servers, page 3-2.
- Real servers—See Configuring Real Servers, page 4-5.
- Dynamic Workload Scaling—See Configuring Dynamic Workload Scaling, page 4-13.
- Server farms—See Configuring Server Farms, page 4-17.
- Sticky groups—See Configuring Sticky Groups, page 5-6.
- Parameter maps—See Configuring Parameter Maps, page 6-1.

For information about SLB as configured and performed by the ACE appliance, see:

- Configuring Virtual Servers, page 3-2
- Load-Balancing Predictors, page 4-2
- Real Servers, page 4-3
- Dynamic Workload Scaling Overview, page 4-4
- Server Farms, page 4-5
- Configuring Health Monitoring, page 4-35
- TCL Scripts, page 4-36
- Configuring Stickiness, page 5-1

Load-Balancing Predictors

The ACE appliance uses the following predictors to select the best server to satisfy a client request:

- Hash Address—Selects the server using a hash value based on either the source or destination IP address, or both. Use these predictors for firewall load balancing (FWLB).

Note FWLB allows you to scale firewall protection by distributing traffic across multiple firewalls on a per-connection basis. All packets belonging to a particular connection must go through the same firewall. The firewall then allows or denies transmission of individual packets across its interfaces. For more information about configuring FWLB on the ACE appliance, see the *Cisco 4700 Series Application Control Engine Appliance Server Load-Balancing Configuration Guide*.

- Hash Content— Selects the server by using a hash value based on the specified content string of the HTTP packet body
- Hash Cookie—Selects the server using a hash value based on a cookie name.
- Hash Secondary Cookie—The ACE selects the server by using the hash value based on the specified cookie name in the URL query string, not the cookie header.
- Hash Header—Selects the server using a hash value based on the HTTP header name.
- Hash Layer4—Selects the server using a Layer 4 generic protocol load-balancing method.
- Hash URL—Selects the server using a hash value based on the requested URL.You can specify a beginning pattern and an ending pattern to match in the URL. Use this predictor method to load-balance cache servers. Cache servers perform better with the URL hash method because you can divide the contents of the caches evenly if the traffic is random enough. In a redundant

configuration, the cache servers continue to work even if the active ACE appliance switches over to the standby ACE appliance. For information about configuring redundancy, see Configuring High Availability, page 9-1.

- Least Bandwidth—Selects the server with the least amount of network traffic or a specified sampling period. Use this type for server farms with heavy traffic, such as downloading video clips.
- Least Connections—Selects the server with the fewest number of active connections based on server weight. For the least connection predictor, you can configure a slow-start mechanism to avoid sending a high rate of new connections to servers that you have just put into service.
- Least Loaded—Selects the server with the lowest load as determined by information from SNMP probes.
- Response—Selects the server with the lowest response time for a specific response-time measurement.
- Round Robin—Selects the next server in the list of real servers based on server weight (weighted roundrobin). Servers with a higher weight value receive a higher percentage of the connections. This is the default predictor.

Note The different hash predictor methods do not recognize the weight value that you configure for real servers. The ACE uses the weight that you assign to real servers only in the round-robin and least-connections predictor methods.

Related Topic

Configuring Health Monitoring, page 4-35

Real Servers

To provide services to clients, you configure real servers on the ACE appliance. Real servers are dedicated physical servers or VMware virtual machines (VMs) that you configure in groups called server farms.

Note VMs that you define as real servers are VMs that the ACE recognizes when configured for Dynamic Workload Scaling (see "Configuring Dynamic Workload Scaling" section on page 4-13).

These servers provide client services such as HTTP or XML content, Web site hosting, FTP file uploads or downloads, redirection for Web pages that have moved to another location, and so on. You identify real servers with names and characterize them with IP addresses, connection limits, and weight values. The ACE appliance also allows you to configure backup servers in case a server is taken out of service for any reason.

After you create and name a real server on the ACE appliance, you can configure several parameters, including connection limits, health probes, and weight. You can assign a weight to each real server based on its relative importance to other servers in the server farm. The ACE appliance uses the server weight value for the weighted round-robin and the least-connections load-balancing predictors. The load-balancing predictor algorithms (for example, round-robin, least connections, and so on) determine the servers to which the ACE appliance sends connection requests. For a listing and brief description of the load-balancing predictors, see Load-Balancing Predictors, page 4-2.

The ACE appliance uses traffic classification maps (class maps) within policy maps to filter out interesting traffic and to apply specific actions to that traffic based on the SLB configuration. You use class maps to configure a virtual server address and definition.

If a primary real server fails, the ACE appliance takes that server out of service and no longer includes it in load-balancing decisions. If you configured a backup server for the real server that failed, the ACE appliance redirects the primary real server connections to the backup server. For information about configuring a backup server, see the Configuring Virtual Server Layer 7 Load Balancing, page 3-29.

The ACE appliance can take a real server out of service for the following reasons:

- Probe failure
- ARP timeout
- Specifying Out Of Service as the administrative state of a real server
- Specifying In Service Standby as the administrative state of a real server

The Out Of Service and In Service Standby selections both provide the graceful shutdown of a server.

Related Topics

- Configuring Real Servers, page 4-5
- Configuring Health Monitoring for Real Servers, page 4-37

Dynamic Workload Scaling Overview

The ACE Dynamic Workload Scaling feature permits on-demand access to remote resources, such as VMs, that you own or lease from an Internet service provider or cloud service provider. This feature uses Cisco Nexus 7000 series switches with Overlay Transport Virtualization (OTV) technology to create a Data Center Interconnect (DCI) on a Layer 2 link over an existing IP network between geographically distributed data centers. The local data center Nexus 7000 contains an OTV forwarding table that lists the MAC addresses of the Layer 2 extended virtual private network (VPN) and identifies the addresses as either local or remote.

When you configure the ACE to use this feature, the ACE uses an XML query to poll the Nexus 7000 and obtain the OTV forwarding table information to determine the locality of the local or remote VMs. The ACE also uses a health monitor probe that it sends to the local VMware vCenter Server to monitor the load of the local VMs based on CPU usage, memory usage, or both. When the average CPU or memory usage of the local VMs reaches its configured maximum threshold value, the ACE bursts traffic to the remote VMs. The ACE stops bursting traffic to the remote VMs when the average CPU or memory usage of the local VMs drops below its configured minimum threshold value.

To use Dynamic Workload Scaling, you configure the ACE to connect to the Data Center Interconnect device (Cisco Nexus 7000 series switch) and the VMware Controller associated with the local and remote VMs. You also configure the ACE with the probe type VM to monitor a server farm's local VM CPU and memory usage, which determines when the ACE bursts traffic to the remote VMs.

For more details on this feature, see the *Cisco 4700 Series Application Control Engine Appliance Server Load-Balancing Configuration Guide*.

Related Topics

- Configuring Dynamic Workload Scaling, page 4-13

Server Farms

Typically, in data centers, servers are organized into related groups called *server farms*. Servers within server farms often contain identical content (referred to as mirrored content) so that if one server becomes inoperative, another server can take its place immediately. Also, having mirrored content allows several servers to share the load of increased demand during important local or international events, such as the Olympic Games. This phenomenon of a sudden large demand for content is called a *flash crowd*.

After you create and name a server farm, you can add existing real servers to it and configure other server farm parameters, such as the load-balancing predictor, server weight, backup server, health probe, and so on. For a listing and brief description of load-balancing predictors, see Load-Balancing Predictors, page 4-2.

Related Topic

Configuring Server Farms, page 4-17

Configuring Real Servers

Real servers are dedicated physical servers that are typically configured in groups called server farms. These servers provide services to clients, such as HTTP or XML content, streaming media (video or audio), TFTP or FTP services, and so on. When configuring real servers, you assign names to them and specify IP addresses, connection limits, and weight values.

The ACE appliance uses traffic classification maps (class maps) within policy maps to filter specified traffic and to apply specific actions to that traffic based on the load-balancing configuration. A load-balancing predictor algorithm (round-robin or least connections) determines the servers to which the ACE appliance sends connection requests. For information about configuring class maps, see Configuring Virtual Context Class Maps, page 10-8.

Use this procedure to configure load balancing on real servers.

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Real Servers**. The Real Servers table appears.

Step 2 Click **Add** to add a new real server, or select a real server you want to modify, then click **Edit**. The Real Servers configuration screen appears.

Step 3 Configure the server using the information in Table 4-1.

Table 4-1 Real Server Attributes

Field	Description
Name	Either accept the automatically incremented value in this field, or enter a unique name for this server. Valid entries are unquoted text strings with no spaces and a maximum of 64 characters.
Type	Select the type of server: • Host—Indicates that this is a typical real server that provides content and services to clients. • Redirect—Indicates that this server is used to redirect traffic to a new location.
State	Select the state of this real server: • In Service—The real server is in service. • Out Of Service—The real server is out of service.
Description	Enter a brief description for this real server. Valid entries are unquoted alphanumeric text strings with no spaces and a maximum of 240 characters.
IP Address	This field appears for only real servers specified as hosts. Enter a unique IP address in dotted-decimal format (such as 192.168.11.1). The IP address cannot be an existing virtual IP address (VIP).
Fail-On-All	This field appears only for real servers identified as host servers. By default, real servers with multiple probes configured for them have an OR logic associated with them. This means that if one of the real server probes fails, the real server fails and enters the PROBE-FAILED state. Click this checkbox to configure a real server to remain in the OPERATIONAL state unless all probes associated with it fail (AND logic). The Fail-On-All function is applicable to all probe types.
Min. Connections	Enter the minimum number of connections to be allowed on this server before the ACE appliance starts sending connections again after it has exceeded the Max. Connections limit. This value must be less than or equal to the Max. Connections value. By default, this value is equal to the Max. Connections value. Valid entries are integers from 1 to 4000000.
Max. Connections	Enter the maximum number of active connections allowed on this server. When the number of connections exceeds this value, the ACE appliance stops sending connections to this server until the number of connections falls below the Min. Connections value. Valid entries are integers from 1 to 4000000, and the default is 4000000.
Weight	This field appears only for real servers identified as hosts. Enter the weight to be assigned to this real server in a server farm. Valid entries are integers from 1 to 100, and the default is 8.

Table 4-1 Real Server Attributes (continued)

Field	Description
Probes	In the Probes field, select the probes that are to be used for health monitoring in the list on the left, then click **Add**. The selected probes appear in the list on the right. The redirect real server probe list contains only configured probes of the type Is Routed, which means that the ACE routes the probe address according to the ACE internal routing table (see the "Configuring Health Monitoring for Real Servers" section on page 4-37). **Note** The Probes field list on the left does not display the VM probe type. To remove probes that you do not want to use for health monitoring, select them in the list on the right, then click **Remove**. The selected probes appear in the list on the left.
Web Host Redirection	URL string used to redirect requests to another server. This field appears only for real servers identified as redirect servers. Enter the URL and port used to redirect requests to another server. Valid entries are in the form http://host.com:port where host is the name of the server and *port* is the port to be used. Valid host entries are unquoted text strings with no spaces and a maximum of 255 characters. Valid port numbers are from 1 to 65535. The relocation string supports the following special characters: • %h—Inserts the hostname from the request Host header • %p—Inserts the URL path string from the request
Redirection Code	This field appears only for real servers identified as redirect servers. Select the appropriate redirection code: • N/A—Indicates that the webhost redirection code is not defined. • 301—Indicates that the requested resource has been moved permanently. For future references to this resource, the client should use one of the returned URIs. • 302—Indicates that the requested resource has been found, but has been moved temporarily to another location. For future references to this resource, the client should use the request URI because the resource may be moved to other locations from time to time.
Rate Bandwidth	The bandwidth rate is the number of bytes per second and applies to the network traffic exchanged between the ACE and the real server in both directions. Specify the real server bandwidth limit in bytes per second. Valid entries are integers from 1 to 300000000.
Rate Connection	The connection rate is the number of connections per second received by the ACE and applies only to new connections destined to a real server. Specify the limit for connections per second. Valid entries are integers from 1 to 350000.

Step 4 Click:

- **Deploy Now** to deploy this configuration on the ACE appliance.
- **Cancel** to exit the procedure without saving your entries and to return to the Real Servers table.
- **Next** to save your entries and to configure another real server.

Related Topics

- Configuring Health Monitoring for Real Servers, page 4-37
- Configuring Server Farms, page 4-17
- Configuring Sticky Groups, page 5-6

Managing Real Servers

The Real Servers table (**Config > Operations > Real Servers**) provides the following information by default for each server:

- Server name
- IP address
- Port
- Admin State (In Service, Out Of Service, or In Service Standby)
- Operational state (See Table 4-3 for descriptions of real server operational states.)
- Number of current connections
- Current server weight
- Locality
- Associated server farm
- Associated virtual server
- Owner, such as the associated virtual context

In the table, Disabled indicates that either the information is not available from the database or that it is not being collected via SNMP. To identify any SNMP-related issues, select the real server's virtual context in the object selector. If there are problems with SNMP, SNMP status will appear in the upper right above the content pane.

The following options are available from the Real Servers table:

- Activating Real Servers, page 4-9
- Suspending Real Servers, page 4-9
- Modifying Real Servers, page 4-10
- Viewing All Real Servers, page 4-11

Activating Real Servers

Use this procedure to activate a real server.

Procedure

Step 1 Select **Config > Operations > Real Servers**. The Real Servers table appears.

Step 2 Select the servers that you want to activate, then click **Activate**. The Activate Server screen appears.

Step 3 In the Task field, confirm that this is the server that you want to activate.

Step 4 In the Reason field, enter a reason for this action. You might enter a trouble ticket, an order ticket, or a user message. **Do not enter a password in this field.**

Step 5 Click:

- **Deploy Now** to deploy this configuration and to return to the Real Servers table. The server appears in the table with the status Inservice.
- **Cancel** to exit this procedure without activating the server and to return to the Real Servers table.

Related Topics

- Managing Real Servers, page 4-8
- Suspending Real Servers, page 4-9
- Viewing All Real Servers, page 4-11

Suspending Real Servers

Use this procedure to suspend a real server.

Procedure

Step 1 Select **Config > Operations > Real Servers**. The Real Servers table appears.

Step 2 Select the server that you want to suspend, then click **Suspend**. The Suspend Server screen appears.

Step 3 In the Reason field, enter the reason for this action. You might enter a trouble ticket, an order ticket, or a user message. **Do not enter a password in this field.**

Step 4 Select one of the following from the Type pulldown menu:

- Graceful
- Suspend
- Suspend and Clear Connections to clear the existing connections to this server as part of the shutdown process

Step 5 Click:

- **Deploy Now** to deploy this configuration and to return to the Real Servers table. The server appears in the table with the status Out Of Service.
- **Cancel** to exit this procedure without suspending the server and to return to the Real Servers table.

Related Topics

- Managing Real Servers, page 4-8
- Activating Real Servers, page 4-9
- Viewing All Real Servers, page 4-11

Modifying Real Servers

Use this procedure to modify weight and connection limits for real servers.

Procedure

Step 1 Select the servers whose configuration you want to modify, then click **Change Weight** below the table to the right of **Activate** and **Suspend**. The **Change Weight Real Servers** window appears.

Step 2 Enter the following information for the selected server:

- Reason for change—Such as trouble ticket, order ticket or user message. **Do not enter a password in this field.**
- Weight—Select a value from 1 to 100.

Step 3 Click:

- **Deploy Now** to accept your entries and to return to the Real Servers table. The server appears in the table with the updated information.
- **Cancel** to exit this procedure without saving your entries and to return to the Real Servers table.

Related Topics

- Managing Real Servers, page 4-8
- Activating Real Servers, page 4-9
- Viewing All Real Servers, page 4-11

Viewing All Real Servers

To view all real servers, select **Config > Operations > Real Servers**. The Real Servers table displays the following information in Table 4-2 by default:

Table 4-2 Real Server Table Fields

Item	Description
Name	Real server name.
IP address	Real server IP address.
Port	Port used to by the real server for communications.
Admin	Administrative state of the real server: In Service, Out Of Service, or In Service Standby.
Oper	Operational state of the real server (see Table 4-3 for descriptions of real server operational states).
Conn	Number of current connections.
Wt	Current server weight.
Locality	Locality requires that you configure the Dynamic Workload Scaling on the ACE (see the "Configuring Dynamic Workload Scaling" section on page 4-13). Location of the real server, which must be a VM and not a physical server. Possible locality states are as follows: • N/A—he ACE cannot determine the real server location (local or remote). A possible cause for this issue is that Dynamic Workload Scaling is not configured correctly. • Local—The real server is located in the local network. • Remote—The real server is located in the remote network. The ACE bursts traffic to this server when the CPU or memory usage of the local real server reaches the specified maximum threshold value.
Server Farm	Associated server farm.
Virtual Servers	Associated virtual server.
Virtual Context	Associated virtual context.

In the table, Disabled indicates that either the information is not available from the database or that it is not being collected via SNMP. To identify any SNMP-related issues, select the real server's virtual context in the object selector. If there are problems with SNMP, SNMP status will appear in the upper right above the content pane.

Table 4-3 Real Server Operational States

State	Description
ARP Failed	An ARP request to this server has failed.
Failed	The server has failed and will not be retried for the amount of time specified by its retry timer.
Inactive	The server is disabled as it has become inactive such as in the case when the real server is not associated to any server farm.
Inband probe failed	The server has failed the inband Health Probe agent.

Table 4-3 Real Server Operational States (continued)

State	Description
In service	The server is in use as a destination for server load balancing client connections.
Max. Load	The server is under maximum load and cannot receive any additional connections.
Operation wait	The server is ready to become operational but is waiting for the associated redirect virtual server to be in service.
Out of service	The server is not in use by a server load balancer as a destination for client connections.
Probe failed	The server load-balancing probe to this server has failed. No new connections will be assigned to this server until a probe to this server succeeds.
Probe testing	The server has received a test probe from the server load balancer.
Ready to test	The server has failed and its retry timer has expired; test connections will begin flowing to it soon.
Return code failed	The server has been disabled because it returned an HTTP code that matched a configured value.
Standby	The server is in standby state. No connections will be assigned to it unless the primary server fails.
Test wait	The server is ready to be tested. This state is applicable only when the server is used for HTTP redirect load balancing.
Testing	The server has failed and has been given another test connection. The success of this connection is not known.
Throttle: DFP	DFP has lowered the weight of the server to throttle level; no new connections will be assigned to the server until DFP raises its weight.
Throttle: max clients	The server has reached its maximum number of allowed clients.
Throttle: max connections	The server has reached its maximum number of connections and is no longer being given connections.
Unknown	The state of the server is not known.

Related Topics

- Activating Real Servers, page 4-9
- Suspending Real Servers, page 4-9
- Modifying Real Servers, page 4-10

Configuring Dynamic Workload Scaling

This section describes how to configure the ACE Dynamic Workload Scaling (DWS) feature. DWS enables an ACE to burst traffic to a remote pool of VMs when the average CPU or memory usage of the local VMs has reached a specified maximum threshold value. When the usage drops to a specified minimum threshold value, the ACE stops bursting traffic to the remote VMs. For more information about the Dynamic Workload Scaling feature, see the "Dynamic Workload Scaling Overview" section on page 4-4.

DWS requires configuring an ACE with the following:

- Nexus 7000 switch—XML interface IP address of the local Cisco Nexus 7000 series switch that the ACE polls to obtain VM location information (local or remote).
- VM Controller—IP address of the VM Controller (also known as VMware vCenter Server) that the ACE sends a health probe to monitor local VM load.
- VM probe—Probe that the ACE sends to the VM Controller to monitor local VM load based on CPU usage, memory usage, or both (see the "Configuring Health Monitoring" section on page 4-35).
- Server Farms—Groups of networked real servers (physical servers and VMs) that provide content delivery. See the "Configuring Server Farms" section on page 4-17.

Note To enable the ACE to use the VMs associated with DWS for load balancing, you must configure them as real servers on the ACE (see the "Configuring Real Servers" section on page 4-5).

Prerequisites

Dynamic Workload Scaling requires the following configuration elements:

- A Nexus 7000 series switch configured for DCI/OTV in the local data center and in the remote data center. For details about configuring a Nexus 7000 for DCI/OTV, see the *Cisco Nexus 7000 NX-OS OTV Configuration Guide, Release 5.x*.
- VMware vCenter Server 4.0 or later.
- Multiple local and remote VMs configured as real servers and associated with server farms configured on the ACE.
- ACE backend interface MTU set to 1430 or less to accommodate DCI encapsulation and the Don't Fragment (DF) bit is automatically set on the DCI link. For details about setting the ACE MTU, see the *Cisco 4700 Series Application Control Engine Appliance Routing and Bridging Configuration Guide*.

This section includes the following topics:

Configuring and Verifying a Nexus 7000 Connection

This procedure describes how to configure an ACE with the Nexus 7000 series switch attributes required to allow the ACE to communicate with the Nexus 7000 using SSH. The ACE uses the Nexus 7000 to obtain VM location information (local or remote).

Guidelines and Restrictions

Configure only one Nexus 7000 per ACE in the Admin context.

Procedure

Step 1 Choose **Config > Virtual Contexts > Load Balancing > Dynamic Workload Scaling > Nexus 7000 Setup**.

The Nexus 7000 Setup pane appears.

Step 2 From the Nexus 7000e Setup pane, define the Nexus 7000 using the information in Table 4-4.

Table 4-4 Nexus 7000 Setup Attributes

Field	Description
Name	Nexus 7000 name (see the Note at the beginning of this chapter for ACE object naming specifications) with a maximum of 64 characters.
Primary IP	Nexus 7000 XML interface IP address in dotted-decimal format (such as 192.168.11.1).
User Name	Username that the ACE uses for access and authentication on the Nexus 7000. Valid entries are unquoted text strings with a maximum of 64 characters with no spaces. **Note** The user must have either the vdc-admin or network-admin role to receive the Nexus 7000 output for the VM location information in XML format.
Password	Password that the ACE uses for authentication on the Nexus 7000. Valid entries are unquoted text strings with a maximum of 64 characters with no spaces. Reenter the password in the Confirm field.

Step 3 Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.

Note Configuring the ACE for Dynamic Workload Scaling also requires configuring the ACE with the VM Controller information (see "Configuring and Verifying a VM Controller Connection" section on page 4-15) and configuring a VM health probe (see the "Configuring Health Monitoring" section on page 4-35).

Step 4 (Optional) Click **Details** to verify connectivity between the ACE and the Nexus 7000.

The ACE **show nexus-device** *device_name* **detail** CLI command output displays in a pop-up window and includes the device name, IP address, and connection information. For more information about the command output, see the *Cisco 4700 Series Application Control Engine Appliance Server Load-Balancing Configuration Guide*.

(Optional) Click **Delete** to delete the currently configured Nexus 7000.

Caution If the ACE is currently configured for Dynamic Workload Scaling, deleting the Nexus 7000 disables the feature.

Related Topics

- Configuring and Verifying a VM Controller Connection, page 4-15
- Configuring Health Monitoring, page 4-35
- Configuring Dynamic Workload Scaling, page 4-13
- Dynamic Workload Scaling Overview, page 4-4
- Configuring Real Servers, page 4-5
- Configuring Server Farms, page 4-17

Configuring and Verifying a VM Controller Connection

This procedure describes how to configure an ACE with the VM Controller (VMware vCenter Server) attributes required to allow the ACE to communicate with the VM Controller to obtain local VM load information.

Guidelines and Restrictions

Configure only one VM Controller per ACE Admin context.

Prerequisites

The ACE is configured to communicate with the local Nexus 7000 that enables the ACE to discover the locality of the VM Controller VMs (see the "Configuring and Verifying a Nexus 7000 Connection" section on page 4-14).

Procedure

Step 1 Choose **Config > Virtual Contexts > Load Balancing > Dynamic Workload Scaling > VM Controller Setup**.

The VM Controller Setup pane appears.

Step 2 From the VM Controller Setup pane, define the VM Controller using the information in Table 4-5.

Table 4-5 VM Controller Setup

Field	Description
Name	VM Controller name (see the Note at the beginning of this chapter for ACE object naming specifications).
URL	IP address or URL for the VM Controller web services API agent. The URL must point to the VM Controller software development kit (SDK), for example, https://1.2.3.4/sdk). Enter a maximum of 255 characters.

Table 4-5 VM Controller Setup (continued)

Field	Description
User Name	Username that the ACE uses for access and authentication on the VM Controller. The user must have a read-only role at least or a role with a read privilege. Valid entries are unquoted text strings with a maximum of 64 characters and no spaces.
Password	Password to be used for authentication on the VM Controller. Valid entries are unquoted text strings with a maximum of 64 characters and no spaces. Reenter the password in the Confirm field.

Step 3 Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.

Note Configuring the ACE for Dynamic Workload Scaling also requires configuring the ACE with the Nexus 7000 information (see "Configuring and Verifying a Nexus 7000 Connection" section on page 4-14) and configuring a VM health probe (see the "Configuring Health Monitoring" section on page 4-35).

Step 4 (Optional) Click **Details** to verify connectivity between the ACE and the remote VM Controller.

The ACE **show vm-controller** *device_name* **detail** CLI command output displays in a pop-up window and includes VM Controller status, IP address, and connection information.

Step 5 (Optional) Click **Delete** to delete the currently configured VM Controller.

Note If the ACE is currently configured to use the Dynamic Workload Scaling, before you can delete the VM controller, you must delete the associated VM health probe (see the "Configuring Health Monitoring" section on page 4-35).

Related Topics

- Configuring and Verifying a Nexus 7000 Connection, page 4-14
- Configuring Health Monitoring, page 4-35
- Configuring Dynamic Workload Scaling, page 4-13
- Dynamic Workload Scaling Overview, page 4-4
- Configuring Real Servers, page 4-5
- Configuring Server Farms, page 4-17

Configuring Server Farms

Server farms are groups of networked real servers (physical servers and VMs) that contain the same content and that typically reside in the same physical location in a data center.

Note With Dynamic Workload Scaling configured on the ACE, the real servers that are VMs can also reside in a remote datacenter (see the "Configuring Dynamic Workload Scaling" section on page 4-13).

Web sites often comprise groups of servers configured in a server farm. Load-balancing software distributes client requests for content or services among the real servers based on the configured policy and traffic classification, server availability and load, and other factors. If one server goes down, another server can take its place and continue to provide the same content to the clients who requested it.

Use this procedure to configure load balancing on server farms.

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Server Farms**. The Server Farms table appears.

Step 2 Click **Add** to add a new server farm, or select an existing server farm, then click **Edit**. The Server Farms configuration screen appears.

Step 3 Enter the server farm attributes (see Table 4-6).

Table 4-6 Server Farm Attributes

Field	Description
Name	Either accept the automatically incremented value in this field, or enter a unique name for this server farm. Valid entries are unquoted text strings with no spaces and a maximum of 64 characters.
Type	Select the type of server farm: • Host—Indicates that this is a typical server farm that consists of real servers that provide content and services to clients • Redirect—Indicates that this server farm consists only of real servers that redirect client requests to alternate locations specified in the real server configuration. (See Configuring Real Servers, page 4-5.)
Description	Enter a brief description for this server farm. Valid entries are unquoted alphanumeric text strings with no spaces and a maximum of 240 characters.
Fail Action	Select the action the ACE appliance is to take with respect to connections if any real server in the server farm fails: • N/A—Indicates that the ACE appliance is to take no action if any server in the server farm fails. • Purge—Indicates that the ACE appliance is to remove connections to a real server if that real server in the server farm fails. The ACE appliance sends a reset command to both the client and the server that failed. • Reassign—The ACE is to reassign the existing server connections to the backup real server (if configured) if the real server fails after you enter this command. If a backup real server has not been configured for the failing server, this selection leaves the existing connections untouched in the failing real server.

Table 4-6 Server Farm Attributes (continued)

Field	Description
Failaction Reassign Across Vlans	This field appears only when the Fail Action is set to Reassign. Check the check box to specify that the ACE reassigns the existing server connections to the backup real server on a different VLAN interface (commonly referred to as a bypass VLAN) if the real server fails. If a backup real server has not been configured for the failing server, this option has no effect and leaves the existing connections untouched in the failing real server. Note the following configuration requirements and restrictions when you enable this option: • Enable the Transparent option (see the next Field) to instruct the ACE not to use NAT to translate the ACE VIP address to the server IP address. The Failaction Reassign Across Vlans option is intended for use in stateful firewall load balancing (FWLB) on your ACE, where the destination IP address for the connection coming in to the ACE is for the end-point real server, and the ACE reassigns the connection so that it is transmitted through a different next hop. • Enable the MAC Sticky option on all server-side interfaces to ensure that packets that are going to and coming from the same server in a flow will traverse the same firewalls or stateful devices (see the "Configuring Virtual Context VLAN Interfaces" section on page 8-8). • Configure the Predictor Hash Address option. See the "Configuring the Predictor Method for Server Farms" section on page 4-26 for the supported predictor methods and configurable attributes for each predictor method. • You must configure identical policies on the primary interface and the backup-server interface. The backup interface must have the same feature configurations as the primary interface. • If you configure a policy on the backup-server interface that is different from the policies on the primary-server interface, that policy will be effective only for new connections. The reassigned connection will always have only the primary-server interface policies. • Interface-specific features (for example, NAT, application protocol inspection, outbound ACLs, or SYN cookie) are not supported. • You cannot reassign connections to the failed real server after it comes back up. This restriction also applies to same-VLAN backup servers. • Real servers must be directly connected to the ACE. This requirement also applies to same-VLAN backup server. • You must disable sequence number randomization on the firewall (see the "Configuring Connection Parameter Maps" section on page 6-2). • Probe configurations should be similar on both ACEs and the interval values should be low. For example, if you configure a high interval value on ACE-1 and a low interval value on ACE-2, the reassigned connections may become stuck because of the probe configuration mismatch. ACE-2 with the low interval value will detect the primary server failure first and will reassign all its incoming connections to the backup-server interface VLAN. ACE-1 with the high interval value may not detect the failure before the primary server comes back up and will still point to the primary server. To minimize packet loss, we recommend the following probe parameter values on both ACEs: Interval: 2, Faildetect: 2, Passdetect interval: 2, and Passdetect count: 5.

Table 4-6 Server Farm Attributes (continued)

Field	Description
Dynamic Workload Scaling	This field appears only for host server farms. Allows the ACE to burst traffic to remote VMs when the average CPU or memory usage of the local VMs has reached it's specified maximum threshold value. The ACE stops bursting traffic to the remote VMs when the average CPU or memory usage of the local VMs has dropped below its specified minimum threshold value. This option requires that you configure the ACE for Dynamic Workload Scaling using a Nexus 7000, VM Controller, and VM probe (see the "Configuring Dynamic Workload Scaling" section on page 4-13). Click one of the following radio button options: • N/A—Not applicable (default). • Local—Restricts the ACE to use of local VMs only for server load balancing. • Burst—Enables the ACE to burst traffic to remote VMs when needed. When you choose Burst, the VM Probe Name field appears along with a list of available VM probes. Choose an available VM probe or click **Add** to display the Health Monitoring pop-up window and create a new VM probe or edit an existing one (see the "Configuring Health Monitoring" section on page 4-35).
Fail-On-All	This field appears only for host server farms. By default, real servers that you configure in a server farm inherit the probes that you configure directly on that server farm. When you configure multiple probes on a server farm, the real servers in the server farm use an OR logic with respect to the probes, which means that if one of the probes configured on the server farm fails, all the real servers in that server farm fail and enter the PROBE-FAILED state. With AND logic, if one server farm probe fails, the real servers in the server farm remain in the operational state. If all the probes associated with the server farm fail, then all the real servers in that server farm fail and enter the PROBE-FAILED state. Click this checkbox to configure the real servers in a server farm to use AND logic with respect to multiple server farm probes. The Fail-On-All function is applicable to all probe types.

Table 4-6 Server Farm Attributes (continued)

Field	Description
Inband-Health Check	This field appears only for host server farms. By default, the ACE monitors the health of all real servers in a configuration through the use of ARPs and health probes. However, there is latency period between when the real server goes down and when the ACE becomes aware of the state. The inband health monitoring feature allows the ACE to monitor the health of the real servers in the server farm through the following connection failures: • For TCP, resets (RSTs) from the server or SYN timeouts. • For UDP, ICMP Host, Network, Port, Protocol, and Source Route unreachable messages. When you configure the failure-count threshold and the number of these failures exceeds the threshold within the reset-time interval, the ACE immediately marks the server as failed, takes it out of service, and removes it from load balancing. The server is not considered for load balancing until the optional resume-service interval expires. Choose one of the following: • Count—Tracks the total number of TCP or UDP failures, and increments the counters as displayed by the s**how serverfarm** *name* **inband** CLI command. • Log—Logs a syslog error message when the number of events reaches the configured connection failure threshold. • Remove—Logs a syslog error message when the number of events reaches the threshold and removes the server from service. **Note** You can configure this feature and health probes to monitor a server. When you do, both are required to keep a real server in service within a server farm. If either feature detects a server is out of service, the ACE does not select the server for load balancing.
Connection Failure Threshold Count	This field appears only when the Inband-Health Check is set to Log or Remove. Enter the maximum number of connection failures that a real server can exhibit in the reset-time interval before ACE marks the real server as failed. Valid entries are integers from 1 to 4294967295.
Reset Timeout (Milliseconds)	This field appears only when the Inband-Health Check is set to Log or Remove. Enter the number of milliseconds for the reset-time interval. Valid entries are integers from 100 to 300000. The default interval is 100. This interval starts when the ACE detects a connection failure. If the connection failure threshold is reached during this interval, the ACE generates a syslog message.When Inband-Health Check is set to Remove, the ACE also removes the real server from service. Changing the setting of this option affects the behavior of the real server, as follows: • When the real server is in the OPERATIONAL state, even if several connection failures have occurred, the new reset-time interval takes effect the next time that a connection error occurs. • When the real server in the INBAND-HM-FAILED state, the new reset-time interval takes effect the next time that a connection error occurs after the server transitions to the OPERATIONAL state.

Table 4-6 Server Farm Attributes (continued)

Field	Description
Resume Service (Seconds)	This field appears only when the Inband-Health Check is set to Remove. Enter the number of seconds after a server has been marked as failed to reconsider it for sending live connections. Valid entries are integers from 30 to 3600. The default setting is 0. The setting of this option affects the behavior of the real server in the inband failed state, as follows: • When this field is not configured and has the default setting of 0, the real server remains in the failed state until you manually suspend and then reactivate it. • When this field is not configured and has the default setting of 0 and then you configure this option with an integer between 30 and 3,600, the failed real server immediately transitions to the Operational state. • When you configure this field and then increase the value, the real server remains in the failed state for the duration of the previously-configured value. The new value takes effect the next time the real server transitions to the failed state. • When you configure this field and then decrease the value, the failed real server immediately transitions to the Operational state. • When you configure this field with an integer between 30 and 3,600 and then reset it to the default of 0, the real server remains in the failed state for the duration of the previously-configured value. The default setting takes effect the next time the real server transitions to the failed state. Then the real server remains in the failed state until you manually suspend and then reactivate it. • When you change this field within the reset-time interval and the real server is in the OPERATIONAL state with several connection failures, the new threshold interval takes effect the next time that a connection error occurs, even if it occurs within the current reset-time interval.
Transparent	This field appears only for real servers identified as host servers. Check the check box to specify that network address translation from the VIP address to the server IP is to occur. Clear the check box to indicates that network address translation from the VIP address to the server IP address is not to occur (default).
Partial-Threshold Percentage	This field appears only for host server farms. Enter the minimum percentage of real servers in the primary server farm that must remain active for the server farm to stay up. If the percentage of active real servers falls below this threshold, the ACE takes the server farm out of service. Valid entries are integers from 0 to 99. The default is 0.

Table 4-6 Server Farm Attributes (continued)

Field	Description
Back Inservice	This field appears only for host server farms. Enter the percentage of real servers in the primary server farm that must be active again for the ACE to place the server farm back into service. Valid entries are integers from 0 to 99. The value in this field should be larger than the value in the Partial Threshold Percentage field. The default is 0.
Probes	In the Available list, choose the probes to use for health monitoring, and click **Add**. The selected probes appear in the Selected list. The redirect server farm probe list contains only configured probes of the type Is Routed, which means that the ACE routes the probe address according to the ACE internal routing table (see the "Configuring Health Monitoring for Real Servers" section on page 4-37). **Note** The list of Available probes does not display the VM probe type. To choose a VM probe for monitoring local VM usage, see the Dynamic Workload Scaling field. To remove probes that you do not want to use for health monitoring, select them in the Selected list, then click **Remove**. The selected probes appear in the Available list.

Step 4 Click:

- **Deploy Now** to deploy this configuration on the ACE appliance. To add real servers to the farm and to configure server farm attributes, see:
  - Adding Real Servers to a Server Farm, page 4-23
  - Configuring Health Monitoring, page 4-35
  - Configuring Server Farm HTTP Return Error-Code Checking, page 4-33
- **Cancel** to exit the procedure without saving your entries and to return to the Server Farms table.
- **Next** to save your entries and to configure another server farm.

Related Topics

- Configuring Health Monitoring for Real Servers, page 4-37
- Configuring Real Servers, page 4-5
- Configuring Sticky Groups, page 5-6
- Configuring Health Monitoring, page 4-35
- Configuring Server Farm HTTP Return Error-Code Checking, page 4-33
- Configuring Dynamic Workload Scaling, page 4-13

Adding Real Servers to a Server Farm

After adding a server farm, (see Configuring Server Farms, page 4-17), you can associate real servers with it and configure predictors and retcode maps. The configuration screens for these attributes appear beneath the Server Farms table or after you have successfully added a new server farm.

Note If you do not see these tabs beneath the Server Farms table, click the **Switch between Configure and Browse Modes** button.

When creating or editing a server farm, if the real server to be added has the same name as an existing global real server but contains a different IP address (or no IP address), the Device Manager displays the following error message:

```
IP address of pre-existing real sever cannot be changed: "<rs-name>" (ip-addr>).
```

If this error message appears, ensure that you specify an existing real server with the matching IP address.

Use this procedure to add real servers to a server farm.

Assumptions

- A server farm has been added to the ACE Appliance Device Manager. (See Configuring Server Farms, page 4-17.)
- At least one real server exists.

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Server Farms**. The Server Farms table appears.

Step 2 Select the server farm you want to associate with real servers, then select the Real Servers tab. The Real Servers table appears.

Step 3 Click **Add** to add a new entry to the Real Servers table, or select an existing server, then click **Edit** to modify it. The Real Servers configuration screen appears.

Step 4 Configure the real server using the information in Table 4-7.

Table 4-7 Real Server Configuration Attributes

Field	Description
Name	Select the server that you want to associate with the server farm.
Port	Enter the port number to be used for server port address translation (PAT). Valid entries are integers from 1 to 65535.
Backup Server Name	Select the server that is to act as the backup server for the server farm. Leave this field blank to indicate that there is no designated backup server for the server farm.
Backup Server Port	If you select a backup server, enter the backup server port number. Valid entries are integers from 1 to 65535.

Table 4-7 Real Server Configuration Attributes (continued)

Field	Description
State	Select the state of this server: • In Service—Indicates that this server is in service. • In Service Standby—Indicates that this server is a backup server and is to remain inactive unless the primary server fails. If the primary server fails, the backup server becomes active and starts accepting connections. • Out Of Service—Indicates that this server is out of service.
Fail-On-All	This field appears only for real servers identified as host servers. By default, real servers with multiple probes configured for them have an OR logic associated with them. This means that if one of the real server probes fails, the real server fails and enters the PROBE-FAILED state. Click this checkbox to configure a real server to remain in the OPERATIONAL state unless all probes associated with it fail (AND logic). The Fail On All function is applicable to all probe types.
Min. Connections	Enter the minimum number of connections that the number of connections must fall below before the ACE appliance resumes sending connections to the server after it has exceeded the number in the Max. Connections field. The number in this field must be less than or equal to the number in the Max. Connections field. 1 to 4000000. The default value is 4000000.
Max. Connections	Enter the maximum number of active connections that can be sent to the server. When the number of connections exceeds this number, the ACE appliance stops sending connections to the server until the number of connections falls below the number specified in the Min. Connections field. Valid entries are integers from 1 to 4000000. The default is 4000000.
Weight	Enter the weight to assign to the server. Valid entries are integers from 1 to 100, and the default is 8.
Cookie String	This field appears only for real servers identified as hosts. Enter a cookie string value of the real server, which is to be used for HTTP cookie insertion when establishing a sticky connection. Valid entries are text strings with a maximum of 32 alphanumeric characters. You can include spaces and special characters in a cookie string value. Use cookie insertion when you want to use a session cookie for persistence if the server is not currently setting the appropriate cookie. With this feature enabled, the ACE inserts the cookie in the Set-Cookie header of the response from the server to the client. See Chapter 5, "Configuring Stickiness" for details on HTTP cookie sticky connections.
Probes	Select the probes in the Available list that you want to apply to this server, then click **Add**. The selected probes appear in the Selected list. To remove probes you do not want to apply to this server, select the probes in the Selected list, then click **Remove**. **Note** The Available list does not display the VM probe type.

Table 4-7 Real Server Configuration Attributes (continued)

Field	Description
Rate Bandwidth	The bandwidth rate is the number of bytes per second and applies to the network traffic exchanged between the ACE and the real server in both directions. Specify the bandwidth limit in bytes per second. Valid entries are integers from 1 to 300000000.
Rate Connection	The connection rate is the number of connections per second received by the ACE and applies only to new connections destined to a real server. Specify the limit for connections per second. Valid entries are integers from 1 to 350000.

Step 5 When you finish configuring this server for this server farm, click:

- **Deploy Now** to deploy this configuration on the ACE appliance.
- **Cancel** to exit this procedure without saving your entries and to return to the Real Servers table.
- **Next** to save your entries and to add another real server for this server farm.

Related Topics

- Configuring Health Monitoring for Real Servers, page 4-37
- Configuring Real Servers, page 4-5
- Configuring Sticky Groups, page 5-6
- Configuring Health Monitoring, page 4-35
- Configuring Server Farm HTTP Return Error-Code Checking, page 4-33
- Configuring Dynamic Workload Scaling, page 4-13

Configuring the Predictor Method for Server Farms

After adding a server farm, (Configuring Server Farms, page 4-17), you can associate real servers with it and configure the predictor method and retcode maps. The configuration screens for these attributes appear beneath the Server Farms table or after you have successfully added a new server farm.

Note If you do not see these tabs beneath the Server Farms table, click the **Switch between Configure and Browse Modes** button.

Use this procedure to configure the predictor method for a server farm. The predictor method specifies how the ACE appliance is to select a server in the server farm when it receives a client request for a service.

Note You can configure only one predictor method per server farm.

Assumptions

- A server farm has been added to the ACE Appliance Device Manager. (See Configuring Server Farms, page 4-17.)
- At least one real server exists.

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Server Farms**. The Server Farms table appears.

Step 2 Select the server farm you want to configure the predictor method for, then select the Predictor tab. The Predictor configuration screen appears.

Step 3 In the Type field, select the method that the ACE appliance is to use to select a server in this server farm when it receives a client request. Table 4-8 lists the available options and describes them.

Step 4 Enter the required information for the selected predictor method. Round Robin is the default predictor method. See Table 4-8.

Table 4-8 Predictor Method Attributes

Predictor Method	Description / Action
Hash Address	The ACE selects the server using a hash value based on the source or destination IP address. To configure the hash address predictor method: 1. In the Mask Type field, indicate whether server selection is based on source IP address or the destination IP address: – N/A—This option is not defined. – Destination—The server is selected based on the destination IP address. – Source—The server is selected based on the source IP address. 2. In the IP Netmask field, select the subnet mask to apply to the address. If none is specified, the default is 255.255.255.255.
Hash Content	The ACE selects the server by using a hash value based on the specified content string of the HTTP packet body. 1. In the Begin Pattern field, enter the beginning pattern of the content string and the pattern string to match before hashing. If you do not specify a beginning pattern, the ACE starts parsing the HTTP body immediate following the offset byte. You cannot configure different beginning and ending patterns for different server farms that are part of the same traffic classification. Valid entries are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters. The ACE supports regular expressions for matching string expressions. Table 10-31 lists the supported characters that you can use for matching string expressions. 2. In the End Pattern field, enter the pattern that marks the end of hashing. If you do not specify either a length or an end pattern, the ACE continues to parse the data until it reaches the end of the field or the end of the packet, or until it reaches the maximum body parse length. You cannot configure different beginning and ending patterns for different server farms that are part of the same traffic classification. Valid entries are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters. The ACE supports regular expressions for matching string expressions. Table 10-31 lists the supported characters that you can use for matching string expressions. 3. In the Length field, enter the length in bytes of the portion of the content (starting with the byte after the offset value) that the ACE uses for sticking the client to the server. Valid entries are integers from 1 to 1000 bytes. The offset and length can vary from 0 to 1000 bytes. If the payload is longer than the offset but shorter than the offset plus the length of the payload, the ACE sticks the connection based on that portion of the payload starting with the byte after the offset value and ending with the byte specified by the offset plus the length. The total of the offset and the length cannot exceed 1000. You cannot specify both the length and the end-pattern options for a Hash Content predictor. 4. In the HTTP Content Offset field, enter the portion of the content that the ACE uses to stick the client on a particular server by indicating the bytes to ignore starting with the first byte of the payload. Valid entries are integers from 0 to 999 bytes. The default is 0, which indicates that the ACE does not exclude any portion of the content.

Table 4-8 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Hash Cookie	The ACE selects the server by using a hash value based on the cookie name. In the Cookie Name field, enter a cookie name in the form of an unquoted text string with no spaces and a maximum of 64 characters.
Hash Secondary Cookie	The ACE selects the server by using the hash value based on the specified cookie name in the URL query string, not the cookie header. In the Cookie Name field, enter a cookie name in the form of an unquoted text string with no spaces and a maximum of 64 characters.
Hash Header	The ACE selects the server by using a hash value based on the header name. In the Header Name field, select the HTTP header to be used for server selection: • To specify an HTTP header that is not one of the standard HTTP headers, select the first radio button and enter the HTTP header name in the Header Name field. Valid entries are unquoted text strings with no spaces and a maximum of 64 characters. • To specify one of the standard HTTP headers, select the second radio button, then select one of the HTTP headers from the list.

Table 4-8 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Hash Layer4	The ACE selects the server by using a Layer 4 generic protocol load-balancing method. Use this predictor to load balance packets from protocols that are not explicitly supported by the ACE. 1. In the Begin Pattern field, enter the beginning pattern of the Layer 4 payload and the pattern string to match before hashing. If you do not specify a beginning pattern, the ACE starts parsing the HTTP body immediate following the offset byte. You cannot configure different beginning and ending patterns for different server farms that are part of the same traffic classification. Valid entries are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters. The ACE supports regular expressions for matching string expressions. Table 10-31 lists the supported characters that you can use for matching string expressions. 2. In the End Pattern field, enter the pattern that marks the end of hashing. If you do not specify either a length or an end pattern, the ACE continues to parse the data until it reaches the end of the field or the end of the packet, or until it reaches the maximum body parse length. You cannot configure different beginning and ending patterns for different server farms that are part of the same traffic classification. Valid entries are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters. The ACE supports regular expressions for matching string expressions. Table 10-31 lists the supported characters that you can use for matching string expressions. 3. In the Length field, enter the length in bytes of the portion of the payload (starting with the byte after the offset value) that the ACE uses for sticking the client to the server. Valid entries are integers from 1 to 1000 bytes. The offset and length can vary from 0 to 1000 bytes. If the payload is longer than the offset but shorter than the offset plus the length of the payload, the ACE sticks the connection based on that portion of the payload starting with the byte after the offset value and ending with the byte specified by the offset plus the length. The total of the offset and the length cannot exceed 1000. You cannot specify both the length and end-pattern options for a Hash Layer 4 predictor. 4. In the HTTP Content Offset field, enter the portion of the content that the ACE uses to stick the client on a particular server by indicating the bytes to ignore starting with the first byte of the payload. Valid entries are integers from 0 to 999 bytes. The default is 0, which indicates that the ACE does not exclude any portion of the content.
Hash URL	The ACE selects the server using a hash value based on the URL. Use this method to load balance firewalls. Enter values in one or both of the pattern fields: • In the URL Begin Pattern field, enter the beginning pattern of the URL and the pattern string to parse. • In the URL End Pattern field, enter the ending pattern of the URL and the pattern string to parse. Valid entries for these fields are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters for each pattern you configure. The following special characters are also allowed: @ # $

Table 4-8 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Least Bandwidth	The ACE selects the server with the least amount of network traffic over a specified sampling period. 1. In the Assess Time field, enter the number of seconds for which the ACE is to collect traffic information. Valid entries are integers from 1 to 10 seconds. 2. In the Least Bandwidth Samples field, enter the number of samples over which you want to weight and average the results of the probe query to calculate the final load value. Valid entries are 1, 2, 4, 8, and 16 (integers from 1 to 16 that are also a power of 2).
Least Connections	The ACE selects the server with the fewest number of connections. In the Slow Start Duration field, enter the slow-start value to be applied to this predictor method. Valid entries are integers from 1 to 65535, where 1 is the slowest ramp-up value. The slow-start mechanism is used to avoid sending a high rate of new connections to servers that you have just put into service.

Table 4-8 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Least Loaded	The ACE selects the server with the lowest load based on information from SNMP probes. 1. In the SNMP Probe Name field, select the name of the SNMP probe to use. 2. In the Auto Adjust field, configure the autoadjust feature to instruct the ACE to apply the maximum load of 16000 to a real server whose load reaches zero or override the default behavior. By default, the ACE applies the average load of the server farm to a real server whose load is zero. The ACE periodically adjusts this load value based on feedback from the server SNMP probe and other configured options. Options include: – Average—Applies the average load of the server farm to a real server whose load is zero. This setting allows the server to participate in load balancing, while preventing it from being flooded by new connections. This is the default setting. – Maxload—Instruct the ACE to apply the maximum load of 16000 to a real server whose load reaches zero. – Off—Instruct the ACE to send all new connections to the server that has a load of zero until the next load update arrives from the SNMP probe for this server. If two servers have the same lowest load (either zero or nonzero), the ACE load balances the connections between the two servers in a round-robin manner. 3. In the Weight Connection field, check the check box to instruct the ACE to use the current connection count in the final load calculation for a real server. When you configure this option, the ACE includes the current connection count in the total load calculation for each real server in a server farm. Clear the check box to reset the behavior of the ACE to the default of excluding the current connection count from the load calculation. To instruct the ACE to select the server with the lowest load, use the predictor least-loaded command in server farm host or redirect configuration mode. With this predictor, the ACE uses SNMP probes to query the real servers for load parameter values (for example, CPU utilization or memory utilization). This predictor is considered adaptive because the ACE continuously provides feedback to the load-balancing algorithm based on the behavior of the real server. To use this predictor, you must associate an SNMP probe with it. The ACE queries user-specified OIDs periodically based on a configurable time interval. The ACE uses the retrieved SNMP load value to determine the server with the lowest load. The syntax of this predictor command is as follows: **predictor least-loaded probe** *name* The name argument specifies the identifier of the existing SNMP probe that you want the ACE to use to query the server. Enter an unquoted text string with no spaces and a maximum of 64 alphanumeric characters. For example, to configure the ACE to select the real server with the lowest load based on feedback from an SNMP probe called PROBE_SNMP, enter: `host1/Admin(config)# serverfarm SF1` `host1/Admin(config-sfarm-host)# predictor least-loaded probe PROBE_SNMP` `host1/Admin(config-sfarm-host-predictor)#` To reset the predictor method to the default of Round Robin, enter: `host1/Admin(config-sfarm-host)# no predictor`

Table 4-8 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Response	The ACE selects the server with the lowest response time for a requested response-time measurement. 1. In the Response Type field, select the type of measurement to use: – App-Req-To-Resp—The response time from when the ACE sends an HTTP request to a server to the time that the ACE receives a response from the server for that request. – Syn-To-Close—The response time from when the ACE sends a TCP SYN to a server to the time that the ACE receives a CLOSE from the server. – Syn-To-Synack—The response time from when the ACE sends a TCP SYN to a server to the time that the ACE receives a SYN-ACK from the server. 2. In the Response Samples field, enter the number of samples over which you want to average the results of the response-time measurement. Valid entries are 1, 2, 4, 8, and 16 (integers from 1 to 16 that are also a power of 2). 3. In the Weight Connection field, check the check box to instruct the ACE to use the current connection count in the final load calculation for a real server. When you configure this option, the ACE includes the current connection count in the total load calculation for each real server in a server farm. Clear the check box to reset the behavior of the ACE to the default of excluding the current connection count from the load calculation.
Round Robin	The ACE selects the next server in the list of servers based on server weight. This is the default predictor method.

Step 5 Click:

- **Deploy Now** to deploy this configuration on the ACE appliance.
- **Cancel** to exit this procedure without saving your entries and to return to the t Connection field table.

Related Topics

- Configuring Health Monitoring for Real Servers, page 4-37
- Configuring Real Servers, page 4-5
- Configuring Sticky Groups, page 5-6
- Adding Real Servers to a Server Farm, page 4-23
- Configuring Server Farm HTTP Return Error-Code Checking, page 4-33
- Configuring Dynamic Workload Scaling, page 4-13

Configuring Server Farm HTTP Return Error-Code Checking

After adding a server farm, (Configuring Server Farms, page 4-17), you can associate real servers with it and configure the predictor method and retcode maps. The configuration screens for these attributes appear beneath the Server Farms table or after you have successfully added a new server farm.

Use this procedure to configure HTTP return error-code checking (retcode map) for a server farm.

Note This feature is available only for server farms configured as hosts. It is not available for server farms configured with the type Redirect.

Assumption

A host type server farm has been added to the ACE Appliance Device Manager. (See Configuring Server Farms, page 4-17.)

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Server Farms**. The Server Farms table appears.

Step 2 Select the server farm you want to configure return error-code checking for, then select the Retcode Map tab. The Retcode Map table appears. If you do not see tabs beneath the Server Farms table, click the **Switch Between Configure And Browse Modes** button.

Step 3 Click **Add** to add a new entry to the table. The Retcode Map configuration screen appears.

Note You cannot modify an entry in the Retcode Map table. Instead, delete the existing entry, then add a new one.

Step 4 In the Lowest Retcode field, enter the minimum value for an HTTP return error code. Valid entries are integers from 100 to 599. This number must be less than or equal to the number in the Highest Retcode field.

Step 5 In the Highest Retcode field, enter the maximum number for an HTTP return error code. Valid entries are integers from 100 to 599. This number must be greater than or equal to the number in the Lowest Retcode field.

Step 6 In the Type field, specify the action to be taken and related options using the information in Table 4-9.

Table 4-9 Return-Code Type Configuration Options

Option	Description
Count	The ACE tracks the total number of return codes received for each return code number that you specify.
Log	The ACE generates a syslog error message when the number of events reaches a specified threshold. 1. In the Threshold field, enter the number of events that the ACE is to receive before generating a syslog error message. Valid entries are integers from 1 to 4294967295. 2. In the Reset field, enter the time interval in seconds for which the ACE checks for the return code. Valid entries are integers from 1 to 2147483647 seconds.
Remove	The ACE generates a syslog error message when the number of events reaches a specified threshold and then removes the server from service. 1. In the Threshold field, enter the number of events that the ACE is to receive before generating a syslog error message and removing the server from service. Valid entries are integers from 1 to 4294967295. 2. In the Reset field, enter the time interval in seconds for which the ACE checks for the return code. Valid entries are integers from 1 to 2147483647 seconds. 3. In the Resume Service field, enter the number of seconds that the ACE waits before it resumes service for the real server automatically after taking the real server out of service. Valid entries are 30 to 3600 seconds. The default setting is 0. The setting of this field affects the behavior of the real server in the failed state, as follows: – When this field is not configured has the default setting of 0, the real server remains in the failed state until you manually remove it from service and readd it. – When this field is not configured has the default setting of 0 and then you configure it with an integer between 30 and 3,600, the failed real server immediately transitions to the Operational state. – When you configure this field and then increase the value, the real server remains in the failed state for the duration of the previously-configured value. The new value takes effect the next time the real server transitions to the failed state. – When you configure this field and then decrease the value, the failed real server immediately transitions to the Operational state. – When you configure this field with an integer between 30 and 3,600 and then reset it to the default of 0, the real server remains in the failed state for the duration of the previously-configured value. The default setting takes effect the next time the real server transitions to the failed state. Then the real server remains in the failed state until you manually remove it from service and readd it.

Step 7 Click:

- **Deploy Now** to deploy this configuration on the ACE appliance.
- **Cancel** to exit this procedure without saving your entries and to return to the Retcode Map table.
- **Next** to save your entries and to add another retcode map.

Related Topics

- Using Virtual Contexts, page 2-2
- Configuring Virtual Context Class Maps, page 10-8
- Configuring Virtual Context Policy Maps, page 10-33

- Configuring Real Servers, page 4-5
- Configuring Sticky Groups, page 5-6
- Configuring Dynamic Workload Scaling, page 4-13

Viewing All Server Farms

Use this procedure to view all server farms associated with a virtual context.

Procedure

Step 1 Select **Config > Virtual Contexts**. The All Virtual Contexts table appears.

Step 2 Select the virtual context with the server farms you want to view, then click **Load Balancing > Server Farms**. The Server Farms table appears with the following information:

- Server farm name
- Server farm type (either host or redirect)
- Description

Depending on the server farms selected, additional tables appear below the Server Farms table. These tables include:

- Real Servers—This table identifies the real servers associated with the selected server farm.
- Predictor—This configuration screen displays the selected predictor method for the selected server farm.
- Retcode Map—This table displays the HTTP return error-code checking that has been configured for the selected server farm.

Related Topics

- Configuring Server Farms, page 4-17
- Adding Real Servers to a Server Farm, page 4-23
- Configuring Health Monitoring, page 4-35
- Configuring Server Farm HTTP Return Error-Code Checking, page 4-33
- Configuring Dynamic Workload Scaling, page 4-13

Configuring Health Monitoring

You can instruct the ACE appliance to check the health of servers and server farms by configuring health probes (sometimes referred to as *keepalives*). After you create a probe, you assign it to a real server or a server farm. A probe can be one of many types, including TCP, ICMP, Telnet, HTTP, and so on. You can also configure scripted probes using the TCL scripting language (see TCL Scripts, page 4-36).

The ACE appliance sends out probes periodically to determine the status of a server, verifies the server response, and checks for other network problems that may prevent a client from reaching a server. Based on the server response, the ACE appliance can place the server in or out of service, and, based on the status of the servers in the server farm, can make reliable load-balancing decisions.

Health monitoring on the ACE appliance tracks the state of a server by sending out probes. Also referred to as out-of-band health monitoring, the ACE appliance verifies the server response or checks for any network problems that can prevent a client to reach a server. Based on the server response, the ACE appliance can place the server in or out of service, and can make reliable load balancing decisions.

Note You can configure the inband health monitoring feature and health probes to monitor the health of the real servers in a server farm. For more information on inband health monitoring, see the "Configuring Server Farms" section on page 4-17.

The ACE appliance identifies the health of a server in the following categories:

- Passed—The server returns a valid response.
- Failed—The server fails to provide a valid response to the ACE appliance is unable to reach a server for a specified number of retries.

By configuring the ACE appliance for health monitoring, the ACE appliance sends active probes periodically to determine the server state.

The ACE appliance supports 4000 unique probe configurations which includes ICMP, TCP, HTTP, and other predefined health probes. The ACE appliance also allows the opening of 1000 sockets simultaneously.

Related Topics

- Configuring Health Monitoring for Real Servers, page 4-37
- TCL Scripts, page 4-36

TCL Scripts

The ACE appliance supports several specific types of health probes (for example HTTP, TCP, or ICMP health probes) when you need to use a diverse set of applications and health probes to administer your network. The basic health probe types supported in the current ACE appliance software release may not support the specific probing behavior that your network requires. To support a more flexible health-probing functionality, the ACE appliance allows you to upload and execute TCL scripts on the ACE appliance.

The TCL interpreter code in the ACE appliance is based on Release 8.44 of the standard TCL distribution. You can create a script to configure health probes. Script probes operate similar to other health probes available in the ACE appliance software. As part of a script probe, the ACE appliance executes the script periodically, and the exit code that is returned by the executing script indicates the relative health and availability of specific real servers. For information on health probes, see Configuring Health Monitoring for Real Servers, page 4-37.

For your convenience, the following sample scripts for the ACE appliance are available to support the TCL feature and are supported by Cisco TAC:

- CHECKPORT_STD_SCRIPT
- ECHO_PROBE_SCRIPT
- FINGER_PROBE_SCRIPT
- FTP_PROBE_SCRIPT
- HTTP_PROBE_SCRIPT
- HTTPCONTENT_PROBE

- HTTPHEADER_PROBE
- HTTPPROXY_PROBE
- IMAP_PROBE
- LDAP_PROBE
- MAIL_PROBE
- POP3_PROBE
- PROBENOTICE_PROBE
- RTSP_PROBE
- SSL_PROBE_SCRIPT
- TFTP_PROBE

These scripts are located in the probe: directory and are accessible in both the Admin and user contexts. Note that the script files in the probe: directory are read-only, so you cannot copy or modify them. However, you can copy files from the probe: directory. For more information, see the *Cisco 4700 Series Application Control Engine Appliance Administration Guide*.

To load a script into memory on the ACE appliance and enable it for use, use the **script file** command. For detailed information on uploading and executing Toolkit Command Language (TCL) scripts on the ACE appliance, refer to the *Cisco 4700 Series Application Control Engine Appliance Server Load-Balancing Configuration Guide*.

Configuring Health Monitoring for Real Servers

To check the health and availability of a real server, the ACE appliance periodically sends a probe to the real server. Depending on the server response, the ACE appliance determines whether to include the server in its load-balancing decision.

Note You can configure the inband health monitoring feature and health probes to monitor the health of the real servers in a server farm. When you do, both are required to keep a real server in service within a server farm. If either feature detects a server is out of service, the ACE does not select the server for load balancing. For more information on inband health monitoring, see the "Configuring Server Farms" section on page 4-17.

Use this procedure to establish monitoring of real servers to determine their viability in load-balancing decisions.

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Health Monitoring**. The Health Monitoring table appears.

Step 2 Click **Add** to add a new health monitoring probe, or select an existing entry, then click **Edit** to modify it. The Health Monitoring screen appears.

Step 3 In the Name field, enter a name that identifies the probe and that associates the probe with the real server. Valid entries are unquoted text strings with no spaces and a maximum of 64 characters.

Step 4 In the Type field, select the type of probe you want to use. The probe type determines what the probe sends to the real server. See Table 4-10 for the types of probes and their descriptions.

Table 4-10 Probe Types

Probe Type	Description
DNS	Sends a request to a DNS server giving it a configured domain. To determine if the server is up, the ACE appliance must receive the configured IP address for that domain.
ECHO-TCP	Sends a string to the server and compares the response with the original string. If the response string matches the original, the server is marked as passed. If not, the ACE appliance retries as configured before the server is marked as failed.
ECHO-UDP	Sends a string to the server and compares the response with the original string. If the response string matches the original, the server is marked as passed. If not, the ACE appliance retries as configured before the server is marked as failed.
FINGER	Sends a probe to the server to verify that a defined username is a username on the server.
FTP	Initiates an FTP session. By default, this probe is for an anonymous login with the option of configuring a user ID and password. The ACE appliance performs an FTP GET or LS to determine the outcome of the problem. This probe supports only active connections.
HTTP	Sets up a TCP connection and issues an HTTP request. Any valid HTTP response causes the probe to mark the real server as passed.
HTTPS	Similar to an HTTP probe, but this probe uses SSL to generate encrypted data.
ICMP	Sends an ICMP request and listens for a response. If the server returns a response, the ACE appliance marks the real server as passed. If there is no response and times out, or an ICMP standard error occurs, such as DESTINATION_UNREACHABLE, the ACE appliance marks the real server as failed.
IMAP	Initiates an IMAP session, using a configured user ID and password. Then, the probe attempts to retrieve e-mail from the server and validates the result of the probe based on the return codes received from the server.
POP	Initiates a POP session, using a configured user ID and password. Then, the probe attempts to retrieve e-mail from the server and validates the result of the probe based on the return codes received from the server.
RADIUS	Connects to a RADIUS server and logs into it to determine if the server is up.
RTSP	Establishes a TCP connection and sends a request packet to the server. The ACE compares the response with the configured response code to determine whether the probe succeeded.
Scripted	Executes probes from a configured script to perform health probing. This method allows you to author specific scripts with features not present in standard probes.
SIP-TCP	Establishes a TCP connection and sends an OPTIONS request packet to the user agent on the server. The ACE compares the response with the configured response code or expected string, or both, to determine whether the probe has succeeded. If you do not configure an expected status code, any response from the server is marked as failed.

Table 4-10 Probe Types (continued)

Probe Type	Description
SIP-UDP	Establishes a UDP connection and sends an OPTIONS request packet to the user agent on the server. The ACE compares the response with the configured response code or expected string, or both, to determine whether the probe has succeeded. If you do not configure an expected status code, any response from the server is marked as failed.
SMTP	Initiates an SMTP session by logging into the server.
SNMP	Establishes a UDP connection and sends a maximum of eight SMNP OID queries to probe the server. The ACE weighs and averages the load information that is retrieved and uses it as input to the least-loaded algorithm for load-balancing decisions. If the retrieved value is within the configured threshold, the server is marked as passed. If the threshold is exceeded, the server is marked as failed.
TCP	Initiates a TCP handshake and expects a response. By default, a successful response causes the probe to mark the server as passed. The probe then sends a FIN to end the session. If the response is not valid, or if there is no response, the probe marks the real server as failed.
TELNET	Establishes a connection to the real server and verifies that a greeting from the application was received.
UDP	Sends a UDP packet to a real server. The probe marks the server as failed only if an ICMP Port Unreachable messages is returned.
VM	Sends a probe to the VMware VM Controller to determine the average amount of both CPU and memory usage of its associated local VMs. The probe response determines whether the ACE load-balances traffic to the local VMs only or bursts traffic to the remote VMs due to high usage of the local VMs. **Note** Use a VM probe when you configure the ACE for Dynamic Workload Scaling (see the "Configuring Dynamic Workload Scaling" section on page 4-13).

Step 5 Enter health monitoring general attributes (see Table 4-11).

Note Click **More Settings** to access the additional general attributes for the selected probe type. By default, the Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-11 Health Monitoring General Attributes

Field	Action
Description	Enter a description for this probe. Valid entries are unquoted alphanumeric text strings with no spaces and a maximum of 240 characters.
Probe Interval (Seconds)	Enter the number of seconds that the ACE is to wait before sending another probe to a server marked as passed. Valid entries are from 2 to 65535 for all probe types except the VM probe, which has a range from 300 to 65535. The default is 15 for all probe types except the VM probe, which has a default of 300 seconds.
Pass Detect Interval (Seconds)	Enter the number of seconds that the ACE is to wait before sending another probe to a server marked as failed. Valid entries are integers from 2 to 65535 with a default of 60. **Note** This field is not applicable for the VM probe type.
Fail Detect	Enter the consecutive number of times that an ACE must detect that probes have failed to contact a server before marking the server as failed. Valid entries are integers from 1 to 65535 with a default of 3. **Note** This field is not applicable for the VM probe type.

Table 4-11 Health Monitoring General Attributes (continued)

Field	Action
More Settings (Not applicable for the VM probe type)	
Pass Detect Count	Enter the number of successful probe responses from the server before the server is marked as passed. Valid entries are integers from 1 to 65535 with a default of 3.
Receive Timeout (Seconds)	Enter the number of seconds the ACE is to wait for a response from a server that has been probed before marking the server as failed. Valid entries are integers from 1 to 65535 with a default of 10.
Dest IP Address[1]	By default, the probe uses the IP address from the real or virtual server configuration for the destination IP address. To override the destination address that the probe uses, enter the preferred destination IP address in this field using dotted-decimal notation, such as 192.168.11.1.
Is Routed[2]	Check the check box to indicate that the destination IP address is routed according to the ACE internal routing table. Clear the check box to indicate that the destination IP address is not routed according to the ACE internal routing table.
Port	By default, the precedence in which the probe inherits the port number is as follows: • The port number that you configure for the probe. • The configured port number from the real server in server farm. • The configured port number from the VIP in a Layer 3 and Layer 4 class map. • The default port number. Table 4-12 lists the default port number for each probe type. If you explicitly configure a default port, the ACE always sends the probe to the default port. The probe does not dynamically inherit the port number from the real server in a server farm or from the VIP specified in the class map.

1. The Dest IP Address field is not applicable to the Scripted probe type.
2. The Is Routed field is not applicable to the RTSP, Scripted, SIP-TCP, and SIP-UDP probe types.

Table 4-12 Default Port Numbers for Probe Types

Probe Type	Default Port Number
DNS	53
Echo	7
Finger	79
FTP	21
HTTP	80
HTTPS	443
ICMP	Not applicable
IMAP	143
POP3	110
RADIUS	1812
RTSP	554

Table 4-12 Default Port Numbers for Probe Types (continued)

Probe Type	Default Port Number
Scripted	1
SIP (both TCP and UDP)	5060
SMTP	25
SNMP	161
Telnet	23
TCP	80
UDP	53
VM	443

Step 6 Enter the attributes for the specific probe type selected:

- For DNS probes, see Table 4-13.
- For Echo-TCP probes, see Table 4-14.
- For Echo-UDP probes, see Table 4-15.
- For Finger probes, see Table 4-16.
- For FTP probes, see Table 4-17.
- For HTTP probes, see Table 4-18.
- For HTTPS probes, see Table 4-19.
- There are no specific attributes for ICMP probes.
- For IMAP probes, see Table 4-20.
- For POP probes, see Table 4-21.
- For RADIUS probes, see Table 4-22.
- For RTSP probes, see Table 4-23.
- For Scripted probes, see Table 4-24.
- For SIP-TCP probes, see Table 4-25.
- For SIP-UDP probes, see Table 4-26.
- For SMTP probes, see Table 4-27.
- For SNMP probes, see Table 4-28.
- For TCP probes, see Table 4-29.
- For Telnet probes, see Table 4-30.
- For UDP probes, see Table 4-31.
- For VM probes, see Table 4-32.

Step 7 Click:

- **Deploy Now** to deploy this configuration on the ACE appliance.
- **Cancel** to exit this procedure without saving your entries and to return to the Health Monitoring table.
- **Next** to save your entries and to configure another probe.

Related Topics

- Configuring DNS Probe Expect Addresses, page 4-60
- Configuring Headers for HTTP and HTTPS Probes, page 4-60
- Configuring Health Monitoring Expect Status, page 4-61
- Configuring Real Servers, page 4-5
- Configuring Server Farms, page 4-17
- Configuring Sticky Groups, page 5-6

Probe Attribute Tables

Refer to the following topics to configure health monitoring probe-specific attributes:

- DNS Probe Attributes, page 4-44
- Echo-TCP Probe Attributes, page 4-44
- Echo-UDP Probe Attributes, page 4-45
- Finger Probe Attributes, page 4-45
- FTP Probe Attributes, page 4-46
- HTTP Probe Attributes, page 4-46
- HTTPS Probe Attributes, page 4-48
- IMAP Probe Attributes, page 4-50
- POP Probe Attributes, page 4-50
- RADIUS Probe Attributes, page 4-51
- RTSP Probe Attributes, page 4-52
- Scripted Probe Attributes, page 4-53
- SIP-TCP Probe Attributes, page 4-54
- SIP-UDP Probe Attributes, page 4-55
- SMTP Probe Attributes, page 4-56
- SNMP Probe Attributes, page 4-56
- TCP Probe Attributes, page 4-57
- Telnet Probe Attributes, page 4-57
- UDP Probe Attributes, page 4-58
- VM Probe Attributes, page 4-59

Refer to the following topics for additional configuration options for health monitoring probes:

- Configuring DNS Probe Expect Addresses, page 4-60
- Configuring Headers for HTTP and HTTPS Probes, page 4-60
- Configuring Health Monitoring Expect Status, page 4-61
- Configuring an OID for SNMP Probes, page 4-62

DNS Probe Attributes

Note Click **More Settings** to access the additional attributes for the DNS probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-13 DNS Probe Attributes

Field	Action
Domain Name	Enter the domain name that the probe is to send to the DNS server. Valid entries are unquoted text strings with a maximum of 255 characters.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.

To configure expect addresses for DNS probes, see Configuring DNS Probe Expect Addresses, page 4-60.

Echo-TCP Probe Attributes

Note Click **More Settings** to access the additional attributes for the Echo-TCP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-14 Echo-TCP Probe Attributes

Field	Action
Send Data	Enter the ASCII data that the probe is to send to the server. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
More Settings	
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.

Echo-UDP Probe Attributes

Note Click **More Settings** to access the additional attributes for the Echo-UDP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-15 Echo-UDP Probe Attributes

Field	Action
Send Data	Enter the ASCII data that the probe is to send to the server. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.

Finger Probe Attributes

Note Click **More Settings** to access the additional attributes for the Finger probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-16 Finger Probe Attributes

Field	Action
Send Data	Enter the ASCII data that the probe is to send to the server. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.

FTP Probe Attributes

Note Click **More Settings** to access the additional attributes for the FTP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-17 FTP Probe Attributes

Field	Action
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.

To configure probe expect statuses for FTP probes, see Configuring Health Monitoring Expect Status, page 4-61.

HTTP Probe Attributes

Note Click **More Settings** to access the additional attributes for the HTTP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-18 HTTP Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Request Method Type	Select the type of HTTP request method that is to be used for this probe: • N/A—This option is not defined. • Get—The HTTP request method is a GET with a URL of "/". This request method directs the server to get the page, and the ACE calculates a hash value for the content of the page. If the page content information changes, the hash value no longer matches the original hash value and the ACE assumes the service is down. This is the default request method. • Head—The server is to only get the header for the page. Using this method can prevent the ACE from assuming that the service is down due to changed content and therefore changed hash values.

Table 4-18 HTTP Probe Attributes (continued)

Field	Action
Request HTTP URL	This field appears if you select Head or Get in the Request Method Type field. Enter the URL path on the remote server. Valid entries are strings of up to 255 characters specifying the URL path. The default path is "/".
More Settings	
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.
User Name	Enter the user identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	Enter the password to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters. Reenter the password in the Confirm field.
Expect Regular Expression	Enter the expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Valid entries are integers from 1 to 4000.
Hash	Check the Hash check box to indicate that the ACE is to use an MD5 hash for an HTTP GET probe. Clear the Hash check box to indicate that the ACE should not use an MD5 hash for an HTTP GET probe.
Hash String	This field appears if the Hash check box is selected. Enter the 32-bit hash value that the ACE is to compare with the hash that is generated from the HTTP page sent by the server. If you do not provide this value, the ACE generates a value the first time it queries the server, stores this value, and matches this value with other responses from the server. A successful comparison causes the probe to maintain an Alive state. Enter the MD5 hash value as a quoted or unquoted hexadecimal string with 16 characters.

To configure probe headers and expect statuses for HTTP probes, see:

- Configuring Headers for HTTP and HTTPS Probes, page 4-60
- Configuring Health Monitoring Expect Status, page 4-61

HTTPS Probe Attributes

Note Click **More Settings** to access the additional attributes for the HTTPS probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-19 HTTPS Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Request Method Type	Select the type of HTTP request method that is to be used for this probe: • N/A—This option is not defined. • Get—The HTTP request method is a GET with a URL of "/". This request method directs the server to get the page, and the ACE calculates a hash value for the content of the page. If the page content information changes, the hash value no longer matches the original hash value and the ACE assumes the service is down. This is the default request method. • Head—The server is to only get the header for the page. Using this method can prevent the ACE from assuming that the service is down due to changed content and therefore changed hash values.
Request HTTP URL	This field appears if you select Head or Get in the Request Method Type field. Enter the URL path on the remote server. Valid entries are strings of up to 255 characters specifying the URL path. The default path is "/'.
Cipher	Select the cipher suite to be used with this HTTPS probe: • RSA_ANY—The HTTPS probe accepts all RSA-configured cipher suites and that no specific suite is configured. This is the default action. • RSA_EXPORT1024_WITH_DES_CBC_SHA • RSA_EXPORT1024_WITH_RC4_56_MD5 • RSA_EXPORT1024_WITH_RC4_56_SHA • RSA_EXPORT_WITH_DES40_CBC_SHA • RSA_EXPORT_WITH_RC4_40_MD5 • RSA_WITH_3DES_EDE_CBC_SHA • RSA_WITH_AES_128_CBC_SHA • RSA_WITH_AES_256_CBC_SHA • RSA_WITH_DES_CBC_SHA • RSA_WITH_RC4_128_MD5 • RSA_WITH_RC4_128_SHA

Table 4-19 HTTPS Probe Attributes (continued)

Field	Action
SSL Version	Select the version of SSL or TLS to be used in ClientHello messages sent to the server: • All—The probe is to use all SSL versions. • SSLv3—The probe is to use SSL version 3. • TLSv1—The probe is to use TLS version 1. By default, the probe sends ClientHello messages with an SSL version 3 header and a TLS version 1 message.
More Settings	
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.
User Name	Enter the user identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	Enter the password to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters. Reenter the password in the Confirm field.
Expect Regular Expression	Enter the expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Value entries are integers from 1 to 4000.
Hash	Check the Hash check box to indicate that the ACE is to use an MD5 hash for an HTTP GET probe. Clear this check box to indicate that the ACE is not to use an MD5 hash for an HTTP GET probe.
Hash String	This field appears if the Hash check box is selected. Enter the 32-bit hash value that the ACE is to compare with the hash that is generated from the HTTP page sent by the server. If you do not provide this value, the ACE generates a value the first time it queries the server, stores this value, and matches this value with other responses from the server. A successful comparison causes the probe to maintain an Alive state. Enter the MD5 hash value as a quoted or unquoted hexadecimal string with 16 characters.

To configure probe headers and expect statuses for HTTPS probes, see:

- Configuring Headers for HTTP and HTTPS Probes, page 4-60
- Configuring Health Monitoring Expect Status, page 4-61

IMAP Probe Attributes

Note Click **More Settings** to access the additional attributes for the IMAP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-20 IMAP Probe Attributes

Field	Action
User Name	Enter the user identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	Enter the password to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters. Reenter the password in the Confirm field.
Mailbox Name	Enter the user mailbox name from which to retrieve e-mail for this IMAP probe. Valid entries are unquoted text strings with a maximum of 64 characters.
Request Command	Enter the request method command for this probe. Valid entries are text strings with a maximum of 32 characters and no spaces.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.

POP Probe Attributes

Note Click **More Settings** to access the additional attributes for the POP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-21 POP Probe Attributes

Field	Action
User Name	Enter the user identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	Enter the password to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters. Reenter the password in the Confirm field.
Request Command	Enter the request method command for this probe. Valid entries are text strings with a maximum of 32 characters and no spaces.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.

RADIUS Probe Attributes

Note Click **More Settings** to access the additional attributes for the RADIUS probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-22 RADIUS Probe Attributes

Field	Action
User Secret	Enter the shared secret to be used to allow probe access to the RADIUS server. Valid entries are case-sensitive strings with no spaces and a maximum of 64 characters.
User Name	Enter the user identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	Enter the password to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters. Reenter the password in the Confirm field.
More Settings	

Table 4-22 RADIUS Probe Attributes (continued)

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
NAS IP Address	Enter the IP address of the Network Access Server (NAS) in dotted-decimal format, such as 192.168.11.1.

RTSP Probe Attributes

Note Click **More Settings** to access the additional attributes for the RTSP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-23 RTSP Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
RTSP Require Header Value	Enter the Require header for this probe.
RTSP Proxy Require Header Value	Enter the Proxy-Require header for this probe.
RTSP Request Method Type	Select the request method type: • N/A—No request method is selected. • Describe—This probe is to use the Describe request type.
More Settings	
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.

To configure probe expect statuses for RTSP probes, see Configuring Health Monitoring Expect Status, page 4-61.

Scripted Probe Attributes

Note Click **More Settings** to access the additional attributes for the Scripted probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-24 Scripted Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Script Name	Enter the local name that you want to assign to this file on the ACE. This file can reside in the disk0: directory or the probe: directory (if the probe: directory exists). **Note** The script file must first be established on the ACE device and the name must be entered exactly as is appears on the device. Please refer to your ACE documentation for more details. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
Script Arguments	Valid arguments are unquoted text strings with no spaces; separate multiple arguments with a space. The field limit is 255 characters.
More Settings	
Script Needs To Be Copied From Remote Location?	Check this check box to indicate that the file needs to be copied from a remote server. Clear this check box to indicate that the script resides locally.
Protocol	This field appears if the script is to be copied from a remote server. Select the protocol to be used for copying the script: • FTP—The script is to be copied using FTP. • TFTP—The script is to be copied using TFTP.
User Name	This field appears if FTP is selected in the Protocol field. Enter the name of the user account on the remote server.

Table 4-24 Scripted Probe Attributes (continued)

Field	Action
Password	This field appears if FTP is selected in the Protocol field. Enter the password for the user account on the remote server. Reenter the password in the Confirm field.
Source File Name	This field appears if the script is to be copied from a remote server. Enter the host IP address, path, and filename of the file on the remote server in the format *host-ip/path/filename* where: • *host-ip* represents the IP address of the remote server. • *path* represents the directory path of the file on the remote server. • *filename* represents the filename of the file on the remote server. For example, your entry might resemble `192.168.11.2/usr/bin/my-script.ext`.

SIP-TCP Probe Attributes

Note Click **More Settings** to access the additional attributes for the SIP-TCP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-25 SIP-TCP Probe Attributes

Field	Action
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.
Expect Regular Expression	Enter the expected response data from the probe destination. Valid entries are text strings with a maximum of 255 characters. This field accepts both single and double quotes. Double quotes are considered delimiters so they don't appear on the device. Single quotes will appear on the device.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Value entries are integers from 1 to 4000.

To configure probe expect statuses for SIP-TCP probes, see Configuring Health Monitoring Expect Status, page 4-61.

SIP-UDP Probe Attributes

Note Click **More Settings** to access the additional attributes for the SIP-UDP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-26 SIP-UDP Probe Attributes

Field	Action
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Enable Rport	Check the check box to indicate that the server will be forced to send a reply from the same port on which the request was received. Clear the check box to indicate that the server can send the reply from a different port than the port from which the request was received.
Expect Regular Expression	Enter the expected response data from the probe destination. Valid entries are text strings with a maximum of 255 characters. This field accepts both single and double quotes. Double quotes are considered delimiters so they don't appear on the device. Single quotes will appear on the device.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Value entries are integers from 1 to 4000.

To configure probe expect statuses for SIP-UDP probes, see Configuring Health Monitoring Expect Status, page 4-61.

SMTP Probe Attributes

Note Click **More Settings** to access the additional attributes for the SMTP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-27 SMTP Probe Attributes

Field	Action
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.

To configure probe expect statuses for SMTP probes, see Configuring Health Monitoring Expect Status, page 4-61.

SNMP Probe Attributes

Note Click **More Settings** to access the additional attributes for the SNMP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-28 SNMP Probe Attributes

Field	Action
SNMP Community	Enter the SNMP community string. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
SNMP Version	Select the SNMP version for this probe: • N/A—No version is selected. • SNMPv1—This probe is to use SNMP version 1. • SNMPv2c—This probe is to use SNMP version 2c.

To configure the SNMP OID for SNMP probes, see Configuring an OID for SNMP Probes, page 4-62.

TCP Probe Attributes

Note Click **More Settings** to access the additional attributes for the TCP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-29 TCP Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Send Data	Enter the ASCII data that the probe is to send to the server. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
More Settings	
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.
Expect Regular Expression	Enter the expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Value entries are integers from 1 to 4000.

Telnet Probe Attributes

Note Click **More Settings** to access the additional attributes for the Telnet probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-30 Telnet Probe Attributes

Field	Action
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.

Table 4-30 Telnet Probe Attributes (continued)

Field	Action
Is Connection	Check the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 1.

UDP Probe Attributes

Note Click **More Settings** to access the additional attributes for the UDP probe type. By default, ACE appliance Device Manager hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 4-31 UDP Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Send Data	Enter the ASCII data that the probe is to send to the server. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
More Settings	
Expect Regular Expression	Enter the expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Value entries are integers from 1 to 4000.

VM Probe Attributes

Note Use a VM probe when you configure the ACE for Dynamic Workload Scaling (see the "Configuring Dynamic Workload Scaling" section on page 4-13).

Configure the VM probe attributes to control when the ACE bursts traffic to remote VMs based on an average of local VM CPU usage, memory usage, or both. The ACE obtains the usage information by sending the VM probe to the specified VM Controller associated with the local VMs It calculates the average aggregate load information for all local VMs as a percentage of CPU usage or memory usage and uses either or both percentages to determine when to burst traffic to the remote data center. If the server farm consists of both physical servers and VMs, the ACE considers load information only from the VMs.

By default, the VM probe checks the percentage of usage for either the CPU or memory against the maximum threshold value. Whichever percentage reaches its maximum threshold value first causes the ACE to burst traffic to the remote data center. The default maximum burst threshold value of 99 percent instructs the ACE to always load balance traffic to the local VMs unless the load value is equal to 100 percent or the VMs are not in the OPERATIONAL state. If you configure the maximum burst threshold value to1 percent, the ACE always bursts traffic to the remote data center.

When the usage percentage is less than the minimum threshold value, the ACE stops bursting traffic to the remote data center and continues to load balance traffic to the local VMs. Any active connections to the remote data center are allowed to complete.

Table 4-32 lists the VM probe attributes, which allow you to control when the ACE bursts traffic to remote VMs.

Table 4-32 VM Probe Attributes

Field	Action
Probe Interval (seconds)	Frequency in seconds with which the ACE sends probes to the VM controller. Enter an integer from 300 to 65535. The default is 300 (5 minutes).
Max CPU Burst Threshold	Threshold for the maximum percentage of the CPU usage based on the average load information for all local VMs. When the CPU usage percentage reaches or exceeds this threshold, the ACE starts bursting traffic to the remote VMs. Enter a value from 1 to 99. The default is 99.
Min CPU Burst Threshold	Threshold for the minimum percentage of the CPU usage based on the average load information for all local VMs. When the CPU usage percentage drops below this threshold, the ACE stops bursting traffic to the remote VMs. Enter a value from 1 to 99 percent. The default is 99.
Max Memory Burst Threshold	Threshold for the maximum percentage of the memory usage based on the average load information for all local VMs. When the memory usage percentage reaches or exceeds this threshold, the ACE starts bursting traffic to the remote VMs. Enter a value from 1 to 99 percent. The default is 99
Min Memory Burst Threshold	Threshold for the minimum percentage of the memory usage based on the average load information for all local VMs. When the memory usage percentage drops below this threshold, the ACE stops bursting traffic to the remote VMs. Enter a value from 1 to 99 percent. The default is 99.
VM Controller Name	Identifier of the VM controller that you configured in the "Configuring and Verifying a VM Controller Connection" section on page 4-15. Click the radio button for the VM controller.

Related Topics

- Configuring Dynamic Workload Scaling, page 4-13

Configuring DNS Probe Expect Addresses

When a DNS probe sends a domain name resolve request to the server, it verifies the returned IP address by matching the received IP address with the configured addresses.

Use this procedure to specify the IP address that the ACE appliance expects to receive in response to a DNS request.

Assumption

A DNS probe has been configured. See Configuring Health Monitoring for Real Servers, page 4-37 for more information.

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Health Monitoring**. The Health Monitoring table appears.

Step 2 Select the DNS probe that you want to configure with an expected IP address. The Expect Addresses subtable appears.

Step 3 Click **Add** to add an entry to the Expect Addresses table. The Expect Address configuration screen appears.

Note You cannot modify an entry in the Expect Addresses table. Instead, delete the existing entry, then add a new one.

Step 4 In the IP Address field, enter the IP address that the ACE appliance is to expect as a server response to a DNS request. Valid entries are unique IP addresses in dotted-decimal notation, such as 192.168.11.1.

Step 5 Click:

- **Deploy Now** to deploy this configuration on the ACE appliance.
- **Cancel** to exit this procedure without saving your entry and to return to the Expect Addresses table.
- **Next** to save your entry and to add another IP Address to the Expect Addresses table.

Related Topics

- Configuring Health Monitoring for Real Servers, page 4-37
- DNS Probe Attributes, page 4-44

Configuring Headers for HTTP and HTTPS Probes

Use this procedure to specify header fields for HTTP and HTTPS probes.

Assumption

An HTTP or HTTPS probe has been configured. See Configuring Health Monitoring for Real Servers, page 4-37 for more information.

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Health Monitoring**. The Health Monitoring table appears.

Step 2 Select the HTTP or HTTPS probe that you want to configure with header. The Probe Headers subtable appears.

Step 3 Click **Add** to add an entry, or select an existing entry, then click **Edit** to modify it. The Probe Headers configuration screen appears.

Step 4 In the Header Name field, select the HTTP header the probe is to use.

Step 5 In the Header Value field, enter the string to assign to the header field. Valid entries are text strings with a maximum of 255 characters. If the string includes spaces, enclose the string with quotes.

Step 6 Click:

- **Deploy Now** to deploy this configuration on the ACE appliance.
- **Cancel** to exit this procedure without saving your entry and to return to the Probe Headers table.
- **Next** to save your entry and to add another header entry to the Probe Headers table.

Related Topics

- Configuring Health Monitoring for Real Servers, page 4-37
- HTTP Probe Attributes, page 4-46
- HTTPS Probe Attributes, page 4-48

Configuring Health Monitoring Expect Status

When the ACE appliance receives a response from the server, it expects a status code to mark a server as passed. By default, there are no status codes configured on the ACE appliance. If you do not configure a status code, any response code from the server is marked as failed.

Expect status codes can be configured for FTP, HTTP, HTTPS, RTSP, SIP-TCP, SIP-UDP, and SMTP probes.

Use this procedure to configure a single or range of code responses that the ACE appliance expects from the probe destination.

Assumption

An FTP, HTTP, HTTPS, RTSP, SIP-TCP, SIP-UDP, or SNMP probe has been configured. See Configuring Health Monitoring for Real Servers, page 4-37 for more information.

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Health Monitoring**. The Health Monitoring table appears.

Step 2 Select the FTP, HTTP, HTTPS, or SMTP probe that you want to configure for expect status codes. The Expect Status subtable appears.

Step 3 Click **Add** to add an entry, or select an existing entry, then click **Edit** to modify it. The Expect Status configuration screen appears.

Step 4 To configure a single expect status code:

a. In the Min. Expect Status Code field, enter the expect status code for this probe. Valid entries are integers from 0 to 999.

b. In the Max. Expect Status code, enter the same expect status code that you entered in the Min. Expect Status Code field.

Step 5 To configure a range of expect status codes:

a. In the Min. Expect Status Code, enter the lower limit of the range of status codes. Valid entries are integers from 0 to 999.

b. In the Max. Expect Status Code, enter the upper limit of a range of status codes. Valid entries are integers from 0 to 999. The value in this field must be greater than or equal to the value in the Min. Expect Status Code field.

Step 6 Click:

- **Deploy Now** to deploy this configuration on the ACE appliance.
- **Cancel** to exit this procedure without saving your entries and to return to the Expect Status table.
- **Next** to save your entries and to add another expect status code to the Expect Status table.

Related Topics

- Configuring Health Monitoring for Real Servers, page 4-37
- FTP Probe Attributes, page 4-46
- HTTP Probe Attributes, page 4-46
- SNMP Probe Attributes, page 4-56

Configuring an OID for SNMP Probes

When the ACE sends a probe with an SNMP OID query, the ACE uses the retrieved value as input to the least-loaded algorithm for load-balancing decisions. Least-loaded load balancing bases the server selection on the server with the lowest load value. If the retrieved value is within the configured threshold, the server is marked as passed. If the threshold is exceeded, the server is marked as failed.

The ACE allows a maximum of eight OID queries to probe the server.

Assumption

An SNMP probe has been configured. See Configuring Health Monitoring for Real Servers, page 4-37 for more information.

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Health Monitoring**. The Health Monitoring table appears.

Step 2 Select the SNMP probe that you want to specify an OID for. The SNMP OID for Server Load Query table appears.

Step 3 Click **Add** to add an entry, or select an existing entry, then click **Edit** to modify it. The SNMP OID configuration pane appears.

Step 4 In the SNMP OID field, enter the OID that the probe is to use to query the server for a value. Valid entries are unquoted strings with a maximum of 255 alphanumeric characters in dotted-decimal notation, such as .1.3.6.1.4.2021.10.1.3.1. The OID string is based on the server type.

Step 5 In the Maximum Absolute Server Load Value field, enter the OID value in the form of an integer and to indicate that the retrieved OID value is an absolute value instead of a percent. Valid entries are integers from 1 to 4294967295.

When the ACE sends a probe with an SNMP OID query, the ACE uses the retrieved value as input to the least-loaded algorithm for load-balancing decisions. By default, the ACE assumes that the retrieved OID value is a percentile value. Use this option to specify that the retrieved OID value is an absolute value.

Step 6 In the Server Load Threshold Value field, specify the threshold at which the server is to be taken out of service:

- When the OID value is based on a percent, valid entries are integers from 1 to 100.
- When the OID is based on an absolute value, valid entries are from 1 to the value specified in the Maximum Absolute Server Load Value field.

Step 7 In the Server Load Weighting field, enter the weight to assign to this OID for the SNMP probe. Valid entries are integers from 0 to 16000.

Step 8 Click:

- **Deploy Now** to deploy this configuration.
- **Cancel** to exit this procedure without saving your entries and to return to the SNMP OID table.
- **Next** to deploy your entries and to add another item to the SNMP OID table.

Related Topics

- Configuring Health Monitoring for Real Servers, page 4-37
- SNMP Probe Attributes, page 4-56

Configuring Secure KAL-AP

A keepalive-appliance protocol (KAL-AP) on the ACE allows communication between the ACE and the Global Site Selector (GSS), which send KAL-AP requests, to report the server states and loads for global-server load-balancing (GSLB) decisions. The ACE uses KAL-AP through a UDP connection to calculate weights and provide information for server availability to the KAL-AP device. The ACE acts as a server and listens for KAL-AP requests. When KAL-AP is initialized on the ACE, the ACE listens on the standard 5002 port for any KAL-AP requests. You cannot configure any other port.

The ACE supports secure KAL-AP for MD5 encryption of data between it and the GSS. For encryption, you must configure a shared secret as a key for authentication between the GSS and the ACE context.

Use this procedure to configure secure KAL-AP associated with a virtual context.

Assumptions

- You have created a virtual context that specifies the Keepalive Appliance Protocol over UDP.
- You have enabled KAL-AP on the ACE by configuring a management class map and policy map, and apply it to the appropriate interface.

Procedure

Step 1 Select **Config > Virtual Contexts >** ***context*** **> Load Balancing > Secure KAL-AP**. The Secure KAL-AP table appears.

Step 2 Click **Add** to configure secure KAL-AP for MD5 encryption of data. The Secure KAL-AP configuration screen appears.

Step 3 In the IP Address field, enable secure KAL-AP by configuring the VIP address for the GSS. Enter the IP address in dotted-decimal notation (for example, 192.168.11.1).

In the Hash Key field, enter the MD5 encryption method shared secret between the KAL-AP device and the ACE. Enter the shared secret as a case-sensitive string with no spaces and a maximum of 31 alphanumeric characters. The ACE supports the following special characters in a shared secret:

, . / = + - ^ @ ! % ~ # $ * ( )

Step 4 Click:

- **Deploy Now** to save your entries. The ACE appliance validates the secure KAL-AP configuration and deploys it.
- **Cancel** to exit this procedure without accepting your entries and to return to the Secure KAL-AP table.
- **Next** to accept your entries.

Related Topics

- Creating Virtual Contexts, page 2-2
- Setting Match Conditions for Layer 3/Layer 4 Management Traffic Class Maps, page 10-13