



# Deploying Load Balancing

---

This chapter provides information about deploying load balancing and redundancy in a Cisco Instant Connect deployment. It includes these sections:

- [Overview, page 11-1](#)
- [Example, page 11-1](#)

## Overview

Load balancing lets you deploy servers to handle Cisco Instant Connect requests and provide redundancy for high availability. With load balancing, you can enhance the operational efficiency of Cisco Instant Connect and its components.

Load balancing is handled by one or two proxy servers. A proxy server is a dedicated server in a Cisco Instant Connect deployment on which you install and configure load balancing software. You can deploy two proxy servers to provide redundancy for these servers.

You can configure load balancing after you install and Configure Cisco Instant Connect in either a small installation or a cluster installation.

To deploy load balancing, you can use a Cisco Instant Connect VM OVA image, or you can install and configure another external load balancer of your choice. The Cisco Instant Connect OVA solution uses the open source HAProxy to provide load balancing, high availability and proxy server functionality.

## Example

This section provides an example of using the Cisco Instant Connect VM OVA load balancing image to provide load balancing and redundancy. This OVA image includes the haproxy\_conf script that you run to configure load balancing when you use this image.

You can use this example as a model when you deploy load balancing for Cisco Instant Connect.

This example assumes the following:

- Cisco Instant Connect has been deployed and configured as a cluster installation and has two database servers and two services servers.
- Two proxy servers will be deployed and paired for load balancing and redundancy.

Deploying load balancing requires the following information:

- IP address of each database server in the deployment.

- IP address of each services server in the deployment.
- IP address of each proxy server in the deployment.
- Valid virtual IP address. In a load balancing deployment, this virtual IP address directs system operations to the appropriate server.

### Example Procedure

**Step 1** On each proxy server, deploy the VM OVA image for Cisco Instant Connect as described in the [“Obtaining and Deploying a VM OVA Image for the Cisco Instant Connect Operating System”](#) section on page 2-3.

In [Step 1c](#) of the procedure in that section, make sure to download the VM OVA image for the load balancing functionality.

**Step 2** On the proxy server that should be the active proxy server, take the following actions.

The primary proxy server can be either server on which you installed the load balancer OVA image. Information is written to the primary server and that server replicates data to other server.

- Access the server via an SSH client and log in as the root user.
- Enter the following commands:
 

```
[root]# cd /opt/cisco/haproxy
[root]# ./haproxy_conf
```
- When you see the “Continue?” prompt, type **y** and then press **Enter**.
- When you see the prompt for the virtual IP address for heartbeat configuration, enter the appropriate virtual IP address and then press **Enter**.
- When you see the prompt for the IP address of the other proxy server, enter that IP address and then press **Enter**.
- When you see the prompt for the host name of the other proxy server, enter that host name and then press **Enter**.
- When you see the prompt that provides information about the IP address that will be used heartbeat configuration, press **Enter** to continue.
- When you see the prompt for the IP address of the Instant Connect Tomcat server, enter the IP address of one of the services servers and then press **Enter**.
- When you see the prompt for the host name the Instant Connect Tomcat server, enter the host name of the services server for which you just entered the IP address and then press **Enter**.
- When you see the prompt for the IP address of the next Instant Connect Tomcat server, enter the IP address of that server and then press **Enter**.
- When you see the prompt for the host name the next Instant Connect Tomcat server, enter the host name of the services server for which you just entered the IP address and then press **Enter**.
- When you see the prompt for the IP address of the next Instant Connect Tomcat server, type **DONE** and then press **Enter**.
- When you see the prompt that provide information about adding other servers, press **Enter** to continue.
- Make a note of the information for completing the deployment that the system displays. You will need this information later.

**Step 3** Repeat [Step 2](#) on the proxy server that should be the secondary proxy server.

**Step 4** Follow the steps for completing the deployment that you noted as described in [Step 2n](#).

**Step 5** For medium installations only, take the following actions on each proxy server:

- a. Access the server via an SSH client and log in as the root user.
  - b. Enter the following commands:  

```
[root]# cd /opt/cisco/haproxy
```

```
[root]# ./setup_haproxy_redis
```
  - c. Follow the onscreen prompts to complete the redundancy configuration for redis.
-

■ Example