# Implementing Load Balancing

This chapter contains the following sections:



# F5 Load Balancing

Cisco UCS Director supports the creation and monitoring of F5 load balancers.

Although load balancing may be prevalent in the routing environment, it is also of growing importance in the virtual networking and VM environment. Server load balancing is a mechanism for distributing traffic across multiple virtual servers, offering high application and server resource utilization.

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 F5 BIG-IP performs a series of checks and calculations to determine the server that can best service each client request. F5 BIG-IP bases server selection on several factors, including the server with the fewest connections regarding load, source or destination address, cookies, URLs, or HTTP headers.

A high-level process flow of load balancing is as follows:

1. A client attempts to connect with a service on the load balancer.
2. The load balancer accepts the connection.
3. The load balancer decides which host should receive the connection and changes the destination IP address (or port) in order to match the service of the selected host.
4. The host accepts the load balancer's connection and responds to the original source, to the client (through its default route), and to the load balancer.
5. The load balancer acquires the return packet from the host and changes the source IP address (or port) to correspond to the virtual server IP address and port, and forwards the packet back to the client.

6. The client receives the return packet, assuming it came from the virtual server, and continues the rest of the process.

Cisco UCS Director enables the management, orchestration, and monitoring of the F5 load balancer. Following is a summary of the crucial processes:

1. Add the F5 load balancer. To add the F5 load balancer, choose **Administration** > **Physical Accounts**. On the **Physical Accounts** page, click **Managed Network Elements** and then click **Add Network Element**.
2. On adding the F5 load balancer as a managed element, Cisco UCS Director triggers Cisco UCS Director task inventory collection. The polling interval configured on the **System Tasks** specifies the frequency of inventory collection.
3. After the F5 load balancer is added to the Pod, it is listed with all other components of the pod environment at the account level. To see the F5 component information, choose **Physical** > **Network**. On the **Network** page, choose the Pod and click **Managed Network Elements**.

There are two ways to implement load balancing on an F5 device using Cisco UCS Director:

1. Use an iApps (BIG-IP) application service.

   iApps application templates let you configure the BIG-IP system for your HTTP applications, by functioning as an interface to consistently deploy, manage, and monitor your servers. You can use default iApps templates or create and customize a template to implement load balancing on the F5 device.
2. Use Cisco UCS Director to:
   - Set up a managed element
   - Create a Pool
   - Add pool members
   - Create a virtual server

# About the Workflow Task for F5 Application Container Setup

Cisco UCS Director includes an F5 BIG-IP workflow task to aid in connecting to the Load Balancer using the Workflow Designer. The crucial workflow tasks are:

- Allocate Container VM Resources
- Provision Container - Network
- Provision Container - VM
- Re-synch Container - VMs
- Setup Container Gateway
- Setup Container F5 Load Balancer
- Send Container Email

# F5 Load Balancing Application Container Prerequisites

You must complete the following tasks before you can create and implement an F5 Load Balancing Application Container within Cisco UCS Director.

- Fenced Container Setup
- Fenced Container Setup - ASA Gateway

**Tip** The Setup Container Load Balancer task is provided for manual creation of the application service. This task is integrated with the Fenced Container Setup-ASA Gateway task to create an F5 load balancing application container.

# Requirements for Setup of a F5 Load Balancing Application Container

Cisco UCS Director can create an application container that provides F5 load balancing properties to the contained VMs. The Cisco UCS Director process workflow is summarized below:

1. Create a load balancing policy
2. Add a network element
3. Create a virtual infrastructure policy
4. Create a tiered application gateway policy (optional)
5. Create a container template
6. Create a container

## F5 Big IP Network Settings Limitations

You must configure the required network settings in the gateway as well as in F5 BIG-IP device manually.

**Note** Configuring the VLAN and NAT settings in the gateway, as well as related network settings in the F5 device, cannot be performed using Cisco UCS Director as part of F5 application container support. This particular automation process will be addressed in an upcoming release of Cisco UCS Director.

## Adding a Network Element

In order to create a virtual server that supports load balancing, first add a network element in Cisco UCS Director. After a Load Balancer is added as a network element in Cisco UCS Director, it appears on the **Managed Network Element** screen.

### Before you begin

You must be logged in to the appliance to complete this task.

**Step 1** Choose **Administration** > **Physical Accounts**.

**Step 2** On the **Physical Accounts** page, click **Managed Network Elements**.

**Step 3** Click **Add Network Element**.

**Step 4** On the **Add Network Element** screen, complete the following fields:

| Name | Description |
|---|---|
| **Pod** drop-down list | Choose the pod to which the network element belongs. |
| **Device Category** drop-down list | Choose the device category for this network element. For example: **F5 Load Balancer**. |
| **Device IP** field | The IP address for this device. |
| **Protocol** drop-down list | Choose the protocol to be used. The list may include the following:<br>• Telnet<br>• SSH<br>• HTTP<br>• HTTPS<br>**Note** When working with an F5 load balancer device, HTTP and HTTPS are the only valid selections. |
| **Port** field | The port to use. |
| **Login** field | The login name. |
| **Password** field | The password associated with the login name. |

**Step 5** Click **Submit**.

Adding the F5 Load Balancer triggers the system task inventory collection. The polling interval configured on the **System Tasks** screen specifies the frequency of inventory collection.

### What to do next

To modify or edit a virtual server, choose the server, and then click **Modify**. To remove a virtual server, choose the server, and then click **Delete**.

# Adding an F5 Load Balancing Policy

**Step 1** Choose **Policies** > **Application Containers**.

**Step 2** On the **Application Containers** page, click **F5 Load Balancer Policies**.

**Step 3** Click **(+) Add Policy**.

**Step 4** On the **Add F5 Load Balancer Policy** screen, complete the following fields:

| Name | Description |
|---|---|
| **Policy Name** field | The name you assign to an F5 load balancer application policy. |
| **Policy Description** field | A description of this policy. |
| **Load Balancer Account Type** drop-down list | Choose **Physical**. |
| **Select F5 Account** field | Click **Select** to choose an **F5 load balancer** account from the available list. |

**Step 5** Click **Next**.

**Step 6** Click **Submit**.

### What to do next

Create a virtual infrastructure policy.

# Adding an F5 Load Balancing Virtual Infrastructure Policy

**Step 1** Choose **Policies** > **Application Containers**.

**Step 2** On the **Application Containers** page, click **Virtual Infrastructure Policies**.

**Step 3** Click **Add Policy (+)** .

**Step 4** On the **Virtual Infrastructure Policy Specification** screen, complete the following fields:

| Name | Description |
|---|---|
| **Template Name** field | A unique name for the policy. |
| **Template Description** field | A description of this policy. |
| **Container Type** drop-down list | Choose **Fenced Virtual** as the container type. |
| **Select Virtual Account** drop-down list | Choose a virtual account to which you want to apply the virtual infrastructure policy. |

**Step 5** Click **Next**.

**Step 6** On the **Virtual Infrastructure Policy - Fencing Gateway** screen, complete the following fields:

| Name | Description |
|---|---|
| **Gateway Required** check box | Check this check box to select a gateway policy, otherwise click **Next**. |
| **Select Gateway Policy** drop-down list | If the **Gateway Required** check box is checked, this field allows you to choose a gateway policy for the virtual infrastructure policy. |

**Step 7** Click **Next**.

**Step 8** On the **Virtual Infrastructure Policy - Fencing Load Balancing** screen, complete the following fields:

| Name | Description |
|---|---|
| **F5 Load Balancer Required** check box | Check this check box to choose an F5 load balancing for the virtual infrastructure policy. |
| **Select F5 Load Balancer Policy** drop-down list | If the **F5 Load Balancer Required** check box is checked, this field allows you to choose an F5 load balancing policy. |

**Step 9** Click **Next** to view the summary of the configuration.

**Step 10** Click **Submit**.

### What to do next

Configure the Tiered Application Gateway Policies.

# Creating a Tiered Application Gateway Policy

**Step 1** Choose **Policies** > **Application Containers**.

**Step 2** On the **Application Containers** page, click **Tiered Application Gateway Policies**.

**Step 3** Click **(+) Add Policy**.

**Step 4** On the **Policy Specification** screen, complete the following fields:

| Name | Description |
|---|---|
| **Policy Name** field | The name you assign to an F5 load balancer tiered application gateway policy. |
| **Policy Description** field | A description of this policy. |
| **Gateway Type** drop-down list | Choose a gateway type. |
| **Select Virtual Account** drop-down list | Choose a cloud account to deploy the container. |

**Step 5** Click **Next**.

**Step 6** On the **Gateway - Linux** screen, complete the following fields for the Linux gateway type (if applicable):

| Name | Description |
|---|---|
| **Provision gateway VM using Content Library template** checkbox | Check to view and choose a VM template from the content library VM templates. If unchecked, you have to choose VM template from VM image templates. |
| **Content Library VM Template** field | This field appears only when the **Provision gateway VM using Content Library template** check box is checked. Click **Select** to view a list of available VM template. Check the check box of the VM template from the content library and click **Select**. |
| **VM Image for the Gateway** drop-down list | This field appears only when the **Provision gateway VM using Content Library template** checkbox is unchecked. Choose a VM image for the gateway from the list. |
| **Number of Virtual CPUs** field | The number of virtual CPUs allowed as per policy. |
| **Memory** drop-down list | Choose the size of the memory. |
| **CPU Reservation in MHz** field | The number of CPU reserved as per the policy. |
| **Memory Reservation in MB** field | The maximum limit of memory reserved as per the policy in MB. |
| **Root Login for the Template** field | The root login name for accessing the template. |
| **Root password for the Template** field | The root password for accessing the template. |
| **Gateway Password Sharing Option** drop-down list | If and how to share the root password for the gateway VM with end users. |

**Step 7** On the **Policy Specification** screen, complete the following fields for the Cisco ASA gateway type:

| Name | Description |
|---|---|
| **Use Unified Fabric** checkbox | Check the box to use unified fabric. |
| **Select Device** drop-down list | Choose a Cisco ASA device. |
| **Outside Interface** drop-down list | Choose the outside interface for the Cisco ASA device. |
| **Outside Interface IP Address** field | This fields appears only when the **Use Unified Fabric** checkbox is unchecked. The IP address of the outside interface. |
| **Outside Interface VLAN ID** field | This fields appears only when the **Use Unified Fabric** checkbox is unchecked. The VLAN ID associated to the outside interface. |
| **Inside Interfaces** field | Choose the inside interfaces from the list. |

**Step 8** On the **Policy Specification** screen, complete the following fields for the Cisco ASAv gateway type:

| Name | Description |
|---|---|
| **Use Unified Fabric** checkbox | Check the box to use unified fabric. |
| **ASAv OVF** field | Check the checkbox to choose an OVF file for the Cisco ASAv device from the table. |
| **ASAv Policy** field | Check the checkbox to choose the ASAv deployment policy from the table. This deployment policy is created earlier by choosing **Policies > Virtual / Hypervisor Policies > Service Delivery > ASAv Deployment Policy**. |
| **Outside Interface** field | Check the checkbox to choose an outside interface from the table. |
| **Inside Interfaces** field | Check the checkbox to choose the inside interfaces from the table. If the **Use Unified Fabric** checkbox is checked, you will be able to select an inside interface from the drop-down list. |

**Step 9** Click **Next**.

**Step 10** Click **Submit** after viewing the summary of the configuration.

# Creating an Application Container Template

**Note** This procedure does not capture the steps to update a template. If you change the templates, the template is applied only to the newly created containers from that template. With this template you can create application containers for use in a variety of networks (including DFA Networks).

### Before you begin

Create a virtual infrastructure policy.

**Step 1** Choose **Policies** > **Application Containers**.

**Step 2** On the **Application Containers** page, click **Application Container Templates**.

**Step 3** Click **Add Template**. The **Add Application Container Template** screen appears. Complete the following fields:

| Name | Description |
|---|---|
| **Template Name** field | The name of the new template. |
| **Template Description** field | The description of the template. |

**Step 4** Click **Next**.

**Step 5** The **Application Container Template - Select a Virtual infrastructure policy** screen appears. Complete the following selection:

| Name | Description |
|---|---|
| **Select Virtual Infrastructure Policy** drop-down list | Choose a policy to deploy the container.<br>**Note** Select a policy that supports load balancing (future wizard screens are then populated with applicable load balancing information). |

**Step 6** Click **Next**. The **Application Container Template - Internal Networks** screen appears.

You can add and configure multiple networks for a container. These networks are applicable to the VM that is provisioned using this template.

**Step 7** Expand **Networks** and click **Add (+)** icon to add a network. The **Add Entry to Networks** screen appears. Complete the following fields:

| Name | Description |
|---|---|
| **Dynamic Fabric Network** check box | If checked, enables the application container for use in Digital Fabric Automation (DFA) Networks. |
| **Network Name** field | The network name. The name should be unique within the container. |
| **Fabric Account** drop-down list | Choose a fabric account. |
| **Network IP Address** field | The network IP address for the container. |
| **Network Mask** field | The network mask address for the container. |
| **Gateway IP Address** field | The IP address of the default gateway for the network. A NIC with this IP address is created on the GW VM. |

**Step 8** Click **Submit**.

Next, you can add and configure the VM that will be provisioned in the application container.

**Step 9** Click **Next**.

**Step 10** Click **Add (+)** icon to add a VM. The **Add Entry to Virtual Machines** screen appears. Complete the following fields:

| Name | Description |
|---|---|
| **VM Name** field | The VM name. |
| **Description** field | The description of the VM. |
| **Provision VM using Content Library Template** check box | Check to view and choose a VM template from the content library VM templates. If unchecked, you have to choose VM template from VM image templates. |
| **Content Library VM Template** field | This field appears only when the **Provision VM using Content Library VM Template** check box is checked. Click **Select** to view a list of available VM template. Check the check box of the VM template that you want to use from the content library and click **Select**. |
| **VM Image** drop-down list | This field appears only when the **Provision VM using Content Library Template** check box is unchecked. Choose the VM image to be deployed. |

| Name | Description |
|---|---|
| **Number of Virtual CPUs** drop-down list | Choose the number of virtual CPUs that can be allowed in the container. |
| **Memory** drop-down list | Choose the size of the memory. |
| **CPU Reservation (MHz)** field | The CPU reservation for the VM. |
| **Memory Reservation (MB)** field | The memory reservation for the VM. |
| **Disk Size (GB)** field | The custom disk size for the VM. To use the template disk size specify the value of 0. The specified disk size overrides the disk size of the selected image. |
| **VM Password Sharing Option** drop-down list | Choose an option to share the VM's username and password with the end users. If **Share after password reset** or **Share template credentials** is chosen, the end user needs to specify a username and password for the chosen templates. |
| **VM Network Interfaces** field | Expand **VM Network Interfaces**, and click **Add (+)** to add the VM network interface. If you are adding another network interface, click **Add (+)**.<br>To add a new VM network interface, complete the following fields:<br>• **VM Network Interface Name** field—The name of the VM network interface.<br>• **Select the Network** drop-down list—Choose a network.<br>• **Adapter Type** drop-down list —Choose the type of the adapter.<br>• **IP Address** field —The IP address of the network. |
| **Maximum Quantity** field | States the maximum number of instances that can be added in this container after it is created. |
| **Initial Quantity** field | States the number of VM instances to provision when the container is created. |

**Step 11** Click **Submit**.

**Step 12** In the **Application Container Template - F5 Application Service** screen, complete the following fields:

| Name | Description |
|---|---|
| **Application Service Name** field | The name of the application service. |
| **Template** field | Choose a template. |
| **IP Address** field | The IP address of the network. |
| **Virtual Server IP** field | IP address of the virtual server. |
| **Virtual Server Port** field | Port used on the virtual server. |
| **FQDN names of Virtual Server** field | Name of the FQDN virtual server.<br>**Note** Separate each FQDN name with a comma. |

| Name | Description |
|---|---|
| **Nodes List** drop-down list | Select a node from the Nodes list and click **Submit**. If the Nodes list fails to present a node that you want to associate with the virtual server:<br>• Click **Add (+)** icon to add the nodes. The **Add Entry to Nodes list** dialog box appears.<br>• Provide the Node IP address, the port, and the connection limit; then click **Submit**. |

**Step 13** Click **Next**.

**Step 14** The **Application Container Template - Deployment Policies** screen appears.

You must select the compute, storage, network, system policy, and cost model required for VM provisioning. A policy is a group of rules that determine where and how a new VM is to be provisioned within an application container (based on the availability of system resources).

- The network policy is used only to deploy the outside interface of the virtual firewall (container gateway).
- The selected *Portgroup in Network Policy* should be on the host on which the Gateway VM is provisioned.
- The network policy can use either a *Static IP Pool* or *DHCP*. However, for container type VSG the network policy should use a Static IP Pool only. The VSG VM requires IP addresses as input. There is no current provision to specify DHCP for deploying a VSG VM.
- The network adapter settings for a provisioned VM (container gateway) should be similar to the settings in the template. You may or may not have to check the *Copy Adapter from Template* check box in the network policy used for this application container.

Complete the following fields:

| Name | Description |
|---|---|
| **Compute Policy** drop-down list | Choose a computer policy. |
| **Storage Policy** drop-down list | Choose a storage policy. |
| **Network Policy** drop-down list | Choose a network policy. |
| **Systems Policy** drop-down list | Choose a systems policy. |
| **Cost Model** drop-down list | Choose a cost model. |

**Step 15** Click **Next**. The **Application Container Template - Options** screen appears.

You can select options to enable or disable certain privileges for the self-service end user.

Complete the following fields:

| Name | Description |
|---|---|
| **Enable Self-Service Power Management of VMs** check box | If checked, enables self-service power management of VMs. |

| Name | Description |
|---|---|
| **Enable Self-Service Resizing of VMs** check box | If checked, enables self-service resizing of VMs. |
| **Enable Self-Service VM Snapshot Management** check box | If checked, enables self-service VM snapshot management. |
| **Enable VNC Based Console Access** check box | If checked, enables self-service VNC based console access. |
| **Enable Self-Service Deletion of Containers** check box | If checked, enables self-deletion of containers. |
| **Technical Support Email Addresses** field | The technical support email address. A detailed technical email is sent to one or more email addresses entered into this field after a container is deployed. |

**Step 16** Click **Next**. The **Application Container Template - Setup Workflows** screen appears. Complete the following fields:

| Name | Description |
|---|---|
| **Container Setup Workflow** drop-down list | Click **Select** to view a list of available container setup workflow. Check the check box of the workflow to establish the application container and click **Select**. |

**Step 17** Click **Next** to complete the creation of the application container template and review the **Summary** screen.

**Note** Notice the inclusion of a Load Balancing Criteria Summary entry.

**Step 18** Click **Submit**.

# Creating an Application Container Using a Template

After creating an application container template, you can use the template administrator to create other application containers. If you want to create a template for use in a VSG environment, see Creating an Application Template for a VSG.

**Note** An application container must use a unique VLAN for its own network. There can be no other port group on (VMware) vCenter using VLAN.

**Step 1** Choose **Policies** > **Application Containers**.

**Step 2** On the **Application Containers** page, click **Application Container Templates**.

**Step 3** Choose a template.

**Step 4** Click **Create Container**.

**Step 5** On the **Create container from template** screen, complete the following fields:

**Note** The combined length of the private network name, service name, and container name must not exceed 32 characters.

| Name | Description |
|---|---|
| **Container Name** field | The name of the container. This name must be unique. |
| **Container Label** field | The label for the container. |
| **Tenant** field | Click **Select** to view the list of available tenants. Check the check box of the tenant that you want to use and then click **Select**. |
| **Private Network Name** field | This field appears only when the tenant with private network is selected. Click **Select** and choose the name of a private network associated with the selected tenant. |
| **Network Throughput** drop-down list | Choose the throughput of the network.<br>**Note** This field is not supported for a tenant with private network. |
| The following fields appear only when the tenant with private network is selected in the **Tenant** field. That is, these fields appear only for the tenant which has the capability of supporting varied VM instances per internal tier. | |
| **VM Labels Prefix Customization** field | The customized prefix name of the VM for each tier.<br>**Note** The prefix name is obtained from the VM label prefix of the application profile where the administrator defines the prefix for each tier. During application container deployment, you can update the prefix from what is defined in the application profile. |
| The maximum quantity of the internal tiers such as WEB/APP/DB tiers are illustrated only as examples in the following fields. There can be more than three tiers as per the definition in the application profile. The number of fields appear based on the number of internal tiers that are defined as part of the application profile. | |
| **Maximum Quantity for <WEB> tier** field | The maximum number of VM instances in the <*web*> tier. |
| **Maximum Quantity for <APP> tier** field | The maximum number of VM instances in the <*application*> tier. |
| **Maximum Quantity for <DB> tier** field | The maximum number of VM instances in the <*database*> tier. |
| **Note** The internal tier names given in the application profile are dynamically fetched and displayed in the fields during create container action. If the maximum quantity for any internal tier is changed during create container action, this value is considered as maximum number of host size for allocating subnet per tier. | |
| The following fields will not appear if the tenant with private network is selected in the **Tenant** field. | |
| **Enable Disaster Recovery** check box | Check this check box to enable the disaster recovery service for the container. |
| **Enable Resource Limits** check box | Check this check box to specify the number of vCPUs, memory, maximum storage, and maximum number of servers for the container. |
| **Enable Network Management** check box | Check this check box to enable the network management service for the container.<br>**Note** This field is applicable only for Layer 4 and Layer 7 devices that are managed by APIC device package. |

| Name | Description |
|---|---|
| **Tier Label Customization** area | This field appears only for APIC containers. The customized name of the tier label. |

**Step 6** Click **Submit**. The **Submit Result** dialog box appears.

**Note** Remember to make a note of the service request presented in the **Submit Result** dialog box.

**Step 7** Click **OK**.

**Note** You can view the progress of the container's creation by viewing the details of the service request.

**Step 8** Click the **Application Containers** tab.
The new container appears in the **Application Containers** pane.

# Initiating a Service Request

**Note** F5 Load Balancing is only supported on fenced virtual containers.

**Step 1** Choose **Organizations** > **Service Requests**.
**Step 2** Click the **Advanced Filter** button (far right side of interface).
**Step 3** Choose **Request Type** from the **Search in Column** drop-down list.
**Step 4** Enter `Advanced` in the **Text** field.
**Step 5** Click **Filter**.
**Step 6** Double-click the **Fenced Container Setup** workflow.