

Deploy a Threat Defense Virtual Cluster on GCP

About Threat Defense Virtual Clustering on GCP 2

Licenses for Threat Defense Virtual Clustering 4

Requirements and Prerequisites for Threat Defense Virtual Clustering 4

Guidelines for Threat Defense Virtual Clustering 5

Deploy the Cluster in GCP 7

Threat Defense Virtual Clustering with Autoscale Solution in GCP 14

Add the Cluster to the Management Center (Manual Deployment) 26

Edit cluster health monitor settings 33

Manage Cluster Nodes 36

Monitoring the Cluster 39

Troubleshoot the cluster 44

Upgrading the Cluster 47

Reference for Clustering 47

History for Threat Defense Virtual Clustering on GCP 59

Revised: July 20, 2026

Clustering lets you group multiple Threat Defense Virtuals together as a single logical device. A cluster provides all the convenience of a single device (management, integration into a network) while achieving the increased throughput and redundancy of multiple devices.

Currently, only routed firewall mode is supported.

Note Some features are not supported when using clustering. See Unsupported Features and Clustering, on page 47.

About Threat Defense Virtual Clustering on GCP

This section describes the clustering architecture and how it works.

How the Cluster Fits into Your Network

The cluster consists of multiple firewalls acting as a single device. To act as a cluster, the firewalls need the following infrastructure:

- Isolated network for intra-cluster communication, known as the *cluster control link*, using VXLAN interfaces. VXLANs, which act as Layer 2 virtual networks over Layer 3 physical networks, let the Firewall Threat Defense Virtual send broadcast/multicast messages over the cluster control link.
- Load Balancer(s)—For external load balancing, you have the following options:
  - Native GCP load balancers, internal and external
  - Equal-Cost Multi-Path Routing (ECMP) using inside and outside routers such as Cisco Cloud Services Router

    ECMP routing can forward packets over multiple "best paths" that tie for top place in the routing metric. Like EtherChannel, a hash of source and destination IP addresses and/or source and destination ports can be used to send a packet to one of the next hops. If you use static routes for ECMP routing, then the Firewall Threat Defense failure can cause problems; the route continues to be used, and traffic to the failed Firewall Threat Defense will be lost. If you use static routes, be sure to use a static route monitoring feature such as Object Tracking. We recommend using dynamic routing protocols to add and remove routes, in which case, you must configure each Firewall Threat Defense to participate in dynamic routing.

Note Layer 2 Spanned EtherChannels are not supported for load balancing.

Individual Interfaces

You can configure cluster interfaces as *Individual interfaces*.

Individual interfaces are normal routed interfaces, each with their own local IP address. The IP address for the interface will be configured automatically via DHCP. Static IP configuration is not supported.

Control and data node roles

When you add a cluster in the Firewall Management Center , you choose one firewall to be the control node, and all additional firewalls to be data nodes. When you first create the cluster, the control node you specify will become the control node simply because it is the first node added to the cluster. Later, if multiple cluster nodes come online at the same time, the control node is determined by the priority setting; the priority is set between 1 and 100, where 1 is the highest priority.

When you create the cluster and deploy the cluster configuration, the Firewall Management Center deploys a special bootstrap configuration to each node that includes fundamental settings such as the cluster control link configuration. The bootstrap configuration lets each node join the cluster. Most bootstrap settings are defined in the cluster wizard; however, you define the cluster control link interface hardware configuration (for example, creating an EtherChannel or setting the Ethernet speed) on the control node before you create the cluster. The cluster control link settings are copied to the bootstrap configuration for each node.

All nodes in the cluster share the same policy configuration. The node that you initially specify as the control node will overwrite the configuration on the data nodes when they join the cluster, so you only need to perform initial configuration on the control node before you form the cluster. After you create the cluster, all configuration changes are made at the cluster level and are shared by all nodes.

Some features do not scale in a cluster, and the control node handles all traffic for those features.

Cluster Control Link

Each node must dedicate one interface as a VXLAN (VTEP) interface for the cluster control link.

VXLAN Tunnel Endpoint

VXLAN tunnel endpoint (VTEP) devices perform VXLAN encapsulation and decapsulation. Each VTEP has two interface types: one or more virtual interfaces called VXLAN Network Identifier (VNI) interfaces, and a regular interface called the VTEP source interface that tunnels the VNI interfaces between VTEPs. The VTEP source interface is attached to the transport IP network for VTEP-to-VTEP communication.

VTEP Source Interface

The VTEP source interface is a regular Firewall Threat Defense Virtual interface with which you plan to associate the VNI interface. You can configure one VTEP source interface to act as the cluster control link. The source interface is reserved for cluster control link use only. Each VTEP source interface has an IP address on the same subnet. This subnet should be isolated from all other traffic, and should include only the cluster control link interfaces.

VNI Interface

A VNI interface is similar to a VLAN interface: it is a virtual interface that keeps network traffic separated on a given physical interface by using tagging. You can only configure one VNI interface. Each VNI interface has an IP address on the same subnet.

Peer VTEPs

Unlike regular VXLAN for data interfaces, which allows a single VTEP peer, The Firewall Threat Defense Virtual clustering allows you to configure multiple peers.

Cluster Control Link Traffic Overview

Cluster control link traffic includes both control and data traffic.

Control traffic includes:

- Control node election.

- Configuration replication.
- Health monitoring.

Data traffic includes:

- State replication.
- Connection ownership queries and data packet forwarding.

Configuration Replication

All nodes in the cluster share a single configuration. You can only make configuration changes on the control node (with the exception of the bootstrap configuration), and changes are automatically synced to all other nodes in the cluster.

Management Network

You must manage each node using the Management interface; management from a data interface is not supported with clustering.

Licenses for Threat Defense Virtual Clustering

Performance tier license requirements

Each Firewall Threat Defense Virtual cluster node requires the same performance tier license. We recommend using the same number of CPUs and memory for all members, or else performance will be limited on all nodes to match the least capable member. The throughput level will be replicated from the control node to each data node so they match.

You assign feature licenses to the cluster as a whole, not to individual nodes. However, each node of the cluster consumes a separate license for each feature. The clustering feature itself does not require any licenses.

When you add the control node to the Firewall Management Center, you can specify the feature licenses you want to use for the cluster. You can modify licenses for the cluster in the **Devices** > **Device Management**, **Cluster** > **License** area.



Note If you add the cluster before the Firewall Management Center is licensed (and running in Evaluation mode), then when you license the Firewall Management Center, you can experience traffic disruption when you deploy policy changes to the cluster. Changing to licensed mode causes all data units to leave the cluster and then rejoin.

Requirements and Prerequisites for Threat Defense Virtual Clustering

Model Requirements

- FTDv5, FTDv10, FTDv20, FTDv30, FTDv50, FTDv100
- Maximum 16 nodes

See also the general requirements for the Firewall Threat Defense Virtual in the Secure Firewall Threat Defense Virtual getting started guides.

User roles

- Admin
- Access Admin
- Network Admin

Hardware and Software Requirements

All units in a cluster:

- Must be in the same performance tier. We recommend using the same number of CPUs and memory for all nodes, or else peformance will be limited on all nodes to match the least capable node.
- The Firewall Management Center access must be from the Management interface; data interface management is not supported.
- Must run the identical software except at the time of an image upgrade. Hitless upgrade is supported.
- Cluster control link interfaces of all units must be in the same subnet.

MTU

Make sure the ports connected to the cluster control link have the correct (higher) MTU configured. If there is an MTU mismatch, the cluster formation will fail. When a node joins the cluster, it checks MTU compatibility by sending a ping to the control node with a packet size matching the cluster control link MTU. If the initial ping fails, the node tries a ping using a smaller packet size (the MTU divided by 2, then by 4, then by 8) until a ping succeeds. A notification is generated so you can fix the MTU mismatch on connecting switches and try again.

The cluster control link MTU should be 154 bytes higher than the data interfaces. Because the cluster control link traffic includes data packet forwarding, the cluster control link needs to accommodate the entire size of a data packet plus cluster traffic overhead (100 bytes) plus VXLAN overhead (54 bytes).

Note We do not recommend setting the cluster control link MTU between 2561 and 8362; due to block pool handling, this MTU size is not optimal for system operation.

The following table shows the default values for the cluster control link MTU and the data interface MTU.

Table 1: Default MTU

Public Cloud	Cluster Control Link MTU	Data Interface MTU
GCP	1554	1400

Guidelines for Threat Defense Virtual Clustering

High Availability

High Availability is not supported with clustering.

IPv6

The cluster control link is only supported using IPv4.

Multi-zone clustering

Multi-zone clustering is supported for up to three zones.

Additional Guidelines

- When significant topology changes occur (such as adding or removing an EtherChannel interface, enabling or disabling an interface on the Firewall Threat Defense or the switch, adding an additional switch to form a VSS or vPC) you should disable the health check feature and also disable interface monitoring for the disabled interfaces. When the topology change is complete, and the configuration change is synced to all units, you can re-enable the interface health check feature.
- When adding a node to an existing cluster, or when reloading a node, there will be a temporary, limited packet/connection drop; this is expected behavior. In some cases, the dropped packets can hang your connection; for example, dropping a FIN/ACK packet for an FTP connection will make the FTP client hang. In this case, you need to reestablish the FTP connection.
- Do not power off a node without first disabling clustering on the node.
- For decrypted TLS/SSL connections, the decryption states are not synchronized, and if the connection owner fails, then decrypted connections will be reset. New connections will need to be established to a new node. Connections that are not decrypted (they match a do-not-decrypt rule) are not affected and are replicated correctly.
- Dynamic scaling is supported from Secure Firewall version 7.3.
- Perform a global deployment after the completion of each maintenance window.
- Ensure that you do not remove more than one device at a time from the instance group. We also recommend that you run the **cluster disable** command on the device before removing the device from the instance group.
- If you want to disable data nodes and the control node in a cluster, we recommend that you disable the data nodes before disabling the control node. If a control node is disabled while there are other data nodes in the cluster, one of the data nodes has to be promoted to be the control node. Note that the role change could disturb the cluster.
- If you experience CCL instability issues, such as intermittent ping failures, after deploying a Threat Defense Virtual cluster on a cloud platform, we recommend that you address the reasons that are causing CCL instability. Also, you can increase the hold time as a temporary workaround to mitigate CCL instability issues to a certain extent. For more information on how to change the hold time, see Edit Cluster Health Monitor Settings.
- When you are configuring your security firewall rule or security group for the Management Center virtual, you must include both Private and Public IP addresses of the Firewall Threat Defense Virtual in the Source IP address range. Also, ensure to specify the Private and Public IP addresses of the Firewall Management Center Virtual in the security firewall rule or security group of the Firewall Threat Defense Virtual. This is important to ensure proper registration of nodes during clustering deployment.

Defaults for Clustering

- The cLACP system ID is auto-generated, and the system priority is 1 by default.
- The cluster health check feature is enabled by default with the holdtime of 3 seconds. Interface health monitoring is enabled on all interfaces by default.
- The cluster auto-rejoin feature for a failed cluster control link is unlimited attempts every 5 minutes.
- The cluster auto-rejoin feature for a failed data interface is 3 attempts every 5 minutes, with the increasing interval set to 2.

• Connection replication delay of 5 seconds is enabled by default for HTTP traffic.

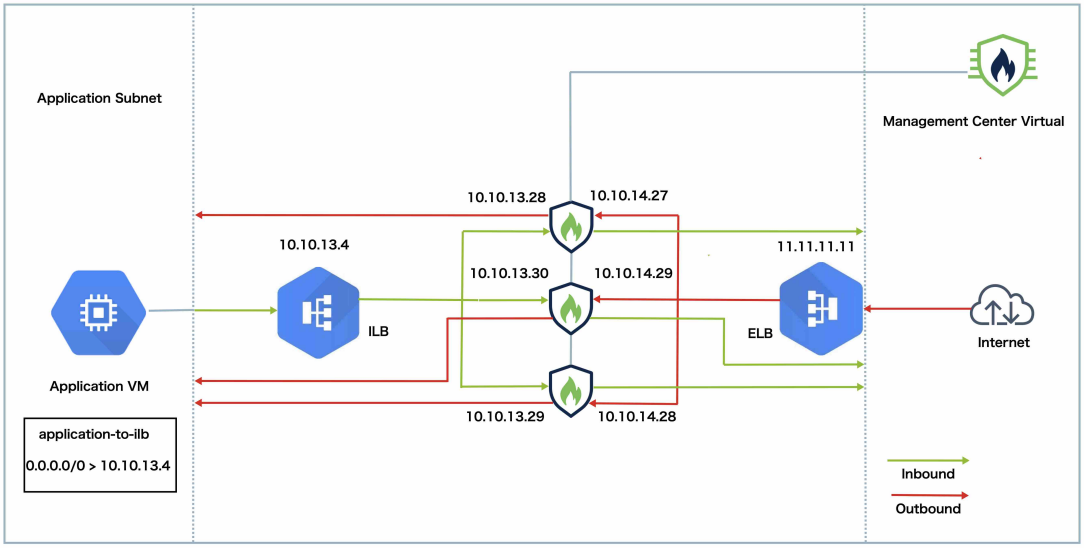
Deploy the Cluster in GCP

To deploy a cluster in GCP, you can either manually deploy or use an instance template to deploy an instance group. You can use the cluster with native GCP load-balancers, or non-native load balancers such as the Cisco Cloud Services Router.

Note Outbound traffic requires interface NAT and is limited to 64K connections.

Sample Topology of GCP Clustering Autoscale Solution

Figure 1: Sample Topology



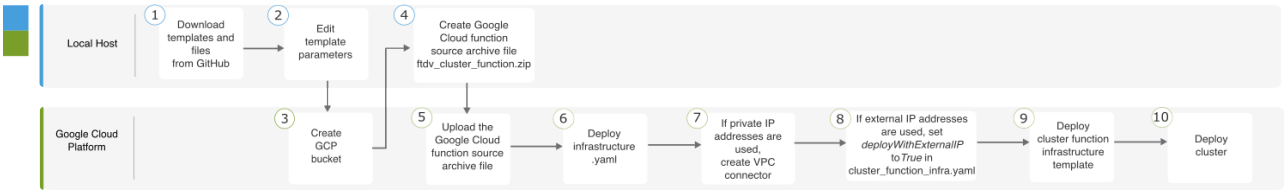
The topology shows both inbound and outbound traffic flow.

1. The Threat Defense Virtual cluster is placed between the internal and external load balancers. A Management Center Virtual instance is used to manage the cluster.
2. Inbound traffic from the internet goes to the external load balancer, which then transmits the traffic to the Threat Defense Virtual cluster.
3. The Threat Defense Virtual instance in the cluster inspects the traffic, and after inspection, forwards the traffic to the application VM.
4. Outbound traffic from the application VM goes to the internal load balancer. The load balancer forwards this traffic to the Threat Defense Virtual cluster, which sends it to the internet.

End-to-End Process for Deploying Threat Defense Virtual Cluster in GCP

Template-based Deployment

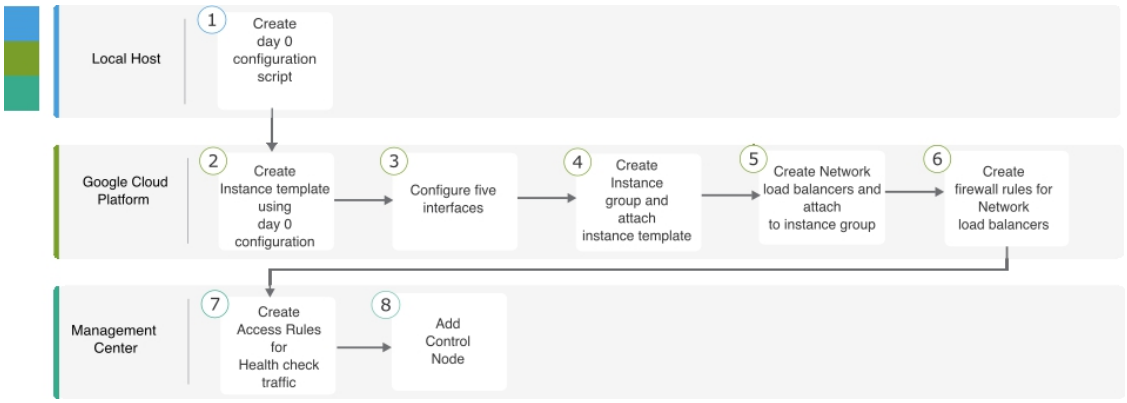
The following flowchart illustrates the workflow for template-based deployment of the Threat Defense Virtual cluster on GCP.



	Workspace	Steps
1	Local Host	Download templates and files from GitHub.
2	Local Host	Edit template parameters.
3	Google Cloud Platform	Create GCP bucket.
4	Local Host	Create Google Cloud function source archive file *ftdv_cluster_function.zip*.
5	Google Cloud Platform	Upload the Google function source archive file.
6	Google Cloud Platform	Deploy infrastructure.yaml.
7	Google Cloud Platform	If private IP addresses are used, create VPC connector.
8	Google Cloud Platform	If external IP addresses are used, set *deployWithExternalIP* to *True* in *cluster_function_infra.yaml*.
9	Google Cloud Platform	Deploy cluster function infrastructure template.
10	Google Cloud Platform	Deploy cluster.

Manual Deployment

The following flowchart illustrates the workflow for manual deployment of the Threat Defense Virtual cluster on GCP.



	Workspace	Steps
1	Local Host	Create day 0 configuration script.
2	Google Cloud Platform	Create instance template using day 0 configuration.
3	Google Cloud Platform	Configure the interfaces.
4	Google Cloud Platform	Create instance group and attach instance template.
5	Google Cloud Platform	Create NLB and attach to instance group.
6	Google Cloud Platform	Create firewall rules for NLB.
7	Management Center	Create access rules for health check traffic.
8	Management Center	Add control node.

Templates

The templates given below are available in GitHub. The parameter values are self-explanatory with the parameter names, and values, given in the template.

- Cluster deployment template for East-West traffic - *deploy_ngfw_cluster_yaml*
- Cluster deployment template for North-South traffic - deploy_ngfw_cluster.yaml

Deploy the Instance Group in GCP Using an Instance Template

Deploy the instance group in GCP using an instance template.

Before you begin

- Use Google Cloud Shell for deployment. Alternatively, you can use Google SDK on any macOS/Linux/Windows machine.

- To allow the cluster to auto-register with the Management Center, you need to create a user with administrative privileges on the Management Center that can use the REST API. See the *Cisco Secure Firewall Management Center Administration Guide*.
- Add an access policy in the Management Center that matches the name of the policy that you specified in *cluster_function_infra.yaml*.

Procedure

Step 1 Download the templates from GitHub to your local folder.

Step 2 Edit **infrastructure.yaml** , **cluster_function_infra.yaml** and **deploy_ngfw_cluster.yaml** with the required *resourceNamePrefix* parameter (for example, ngfwvcls) and other required user inputs.

From Secure Firewall version 7.4.1, you can deploy the cluster without the diagnostic interface. To deploy the cluster with only the Outside, Inside, Management, and CCL interfaces, set the *withDiagnostic* variable to **False** in both the **infrastructure.yaml** and the **deploy_ngfw_cluster.yaml** files.

Note that there is a **deploy_ngfw_cluster.yaml** file in both the **east-west** and **north-south** folders in GitHub. Download the appropriate template as per your traffic flow requirement.

Step 3 Create a bucket using Google Cloud Shell to upload the Google cloud function source archive file *ftdv_cluster_function.zip*.

gsutil mb --pap enforced gs://*resourceNamePrefix***-ftdv-cluster-bucket/**

Ensure that the *resourceNamePrefix* variable here matches the *resourceNamePrefix* variable that you specified in **cluster_function_infra.yaml**.

Step 4 Create an archive file for the cluster infrastructure.

Example:

```
zip -j ftdv_cluster_function.zip ./cluster-function/*
```

Step 5 Upload the Google source archive that you created earlier.

gsutil cp ftdv_cluster_function.zip gs://*resourceNamePrefix***-ftdv-cluster-bucket/**

Step 6 Deploy infrastructure for the cluster.

gcloud deployment-manager deployments create *cluster_name* **--config infrastructure.yaml**

Step 7 If you are using private IP addresses, perform the steps given below:

a) Launch and set up the Management Center Virtual with a Threat Defense Virtual management VPC.

b) Create a VPC connector to connect the Google Cloud functions with the Threat Defense Virtual management VPC.

gcloud compute networks vpc-access connectors create *vpc-connector-name* **--region us-central1 --subnet** *resourceNamePrefix***-ftdv-mgmt-subnet28**

Step 8 If the Management Center is remote from the Threat Defense Virtual, and the Threat Defense Virtual needs an external IP address, ensure that you set **deployWithExternalIP** to **True** in **cluster_function_infra.yaml**.

Step 9 Deploy the cluster function infrastructure.

gcloud deployment-manager deployments create *cluster_name* **--config cluster_function_infra.yaml**

Step 10 Deploy the cluster.

a. For North-South topology deployment:

   gcloud deployment-manager deployments create *cluster_name* **--config north-south/deploy_ngfw_cluster.yaml**

b. For East-West topology deployment:

   gcloud deployment-manager deployments create *cluster_name* **--config east-west/deploy_ngfw_cluster.yaml**

Deploy the Cluster in GCP Manually

To deploy the cluster manually, prepare the day0 configuration, deploy each node, and then add the control node to the Firewall Management Center.

Create the Day0 Configuration With a Fixed Configuration for GCP

The fixed configuration will auto-generate the cluster bootstrap configuration.

```
{
    "AdminPassword": "password",
    "Hostname": "hostname",
    "FirewallMode": "Routed",
    "ManageLocally": "No",
                "Diagnostic": "OFF",       //Optional user input from version 7.4.1 - use to deploy cluster
without Diagnostic interface
    "Cluster": {
        "CclSubnetRange": "ip_address_start ip_address_end",
        "ClusterGroupName": "cluster_name"
    }
}
```

For example:

```
{
    "AdminPassword": "DeanW1nche$ter",
    "Hostname": "ciscoftdv",
    "FirewallMode": "Routed",
    "ManageLocally": "No",
    "Cluster": {
        "CclSubnetRange": "10.10.55.2 10.10.55.253",     //mandatory user input
        "ClusterGroupName": "ftdv-cluster"               //mandatory user input
    }
}
```

Note If you are copying and pasting the configuration given above, ensure that you remove **//mandatory user input** from the configuration.

For the **CclSubnetRange** variable, note that you cannot use the first two IP addresses and the last two IP addresses in the subnet. See Reserved IP addresses in IPv4 subnets for more information. Ensure that you have at least 16 available IP addresses for clustering. Some examples of start and end IP addresses are given below.

Note All cluster infrastructure subnets must use /27 CIDR.

Table 2: Examples of Start and End IP addresses

CIDR	Start IP Address	End IP Address
10.1.1.0/27	10.1.1.2	10.1.1.29
10.1.1.32/27	10.1.1.34	10.1.1.61
10.1.1.64/27	10.1.1.66	10.1.1.93
10.1.1.96/27	10.1.1.98	10.1.1.125
10.1.1.128/27	10.1.1.130	10.1.1.157
10.1.1.160/27	10.1.1.162	10.1.1.189
10.1.1.192/27	10.1.1.194	10.1.1.221
10.1.1.224/27	10.1.1.226	10.1.1.253

Deploy Cluster Nodes Manually

Deploy the cluster nodes so they form a cluster. For clustering on GCP, you cannot use the 4 vCPU machine type. The 4 vCPU machine type only supports four interfaces, and five are needed. Use a machine type that supports five interfaces, such as c2-standard-8.

Procedure

Step 1 Create an instance template using the day 0 configuration (in the **Metadata** > **Startup Script** section) with 5 interfaces: outside, inside, management, diagnostic, and cluster control link.

See Secure Firewall Threat Defense Virtual getting started guides.

Step 2 Create an instance group, and attach the instance template.

Step 3 Create GCP network load balancers (internal and external), and attach the instance group.

Step 4 For GCP network load balancers, allow health checks in your security policy on the Management Center. See Allow Health Checks for GCP Network Load Balancers, on page 12.

Step 5 Add the control node to the Management Center. See Add the Cluster to the Management Center (Manual Deployment), on page 26.

Allow Health Checks for GCP Network Load Balancers

Google Cloud provides health checks to determine if backends respond to traffic.

See https://cloud.google.com/load-balancing/docs/health-checks to create firewall rules for network load balancers. Then in the Firewall Management Center, create access rules to allow the health check traffic. See https://cloud.google.com/load-balancing/docs/health-check-concepts for the required network ranges.

You also need to configure dynamic manual NAT rules to redirect the health check traffic to the Google metadata server at 169.254.169.254.

You can set up a route for GCP health checks across all interfaces that are used to configure their health probes. You can achieve this by creating a route with a higher metric on interfaces where a route for GCP health checks is not already available.

North-South NAT Rules Sample Configuration

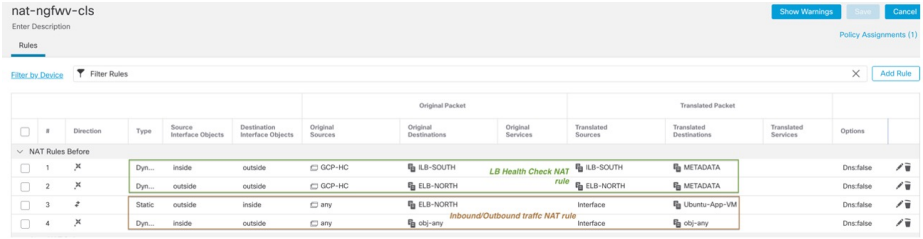
```
nat (inside,outside) source dynamic GCP-HC ILB-SOUTH destination static ILB-SOUTH METADATA
nat (outside,outside) source dynamic GCP-HC ELB-NORTH destination static ELB-NORTH METADATA

nat (outside,inside) source static any interface destination static ELB-NORTH Ubuntu-App-VM
nat (inside,outside) source dynamic any interface destination static obj-any obj-any

object network Metadata
 host 169.254.169.254

object network ILB-SOUTH
host <ILB_IP>
object network ELB-NORTH
host <ELB_IP>

object-group network GCP-HC
 network-object 35.191.0.0 255.255.0.0
 network-object 130.211.0.0 255.255.252.0
 network-object 209.85.204.0 255.255.252.0
 network-object 209.85.152.0 255.255.252.0
```



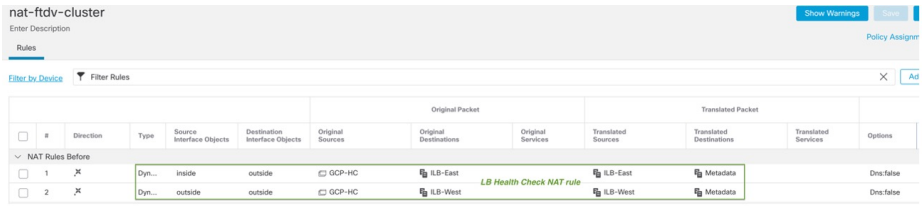
East-West NAT Rules Sample Configuration

```
nat (inside,outside) source dynamic GCP-HC ILB-East destination static ILB-East Metadata
nat (outside,outside) source dynamic GCP-HC ILB-West destination static ILB-West Metadata

object network Metadata
 host 169.254.169.254

object network ILB-East
host <ILB_East_IP>
object network ILB-West
host <ILB_West_IP>

object-group network GCP-HC
 network-object 35.191.0.0 255.255.0.0
 network-object 130.211.0.0 255.255.252.0
 network-object 209.85.204.0 255.255.252.0
 network-object 209.85.152.0 255.255.252.0
```



North-South and East-West Traffic Routing Configuration Sample

```
route outside 0.0.0.0 0.0.0.0 <Outside_Gateway> 1
route inside 35.191.0.0 255.255.0.0 <Inside_Gateway> 1
route inside 130.211.0.0 255.255.252.0 <Inside_Gateway> 1
route inside 209.85.152.0 255.255.252.0 <Inside_Gateway> 1
route inside 209.85.204.0 255.255.252.0 <Inside_Gateway> 1
```

If a default route is not available, then policy-based routing can be used to route the traffic for health checks.

Note Set the NAT > Translated destination port to 80.

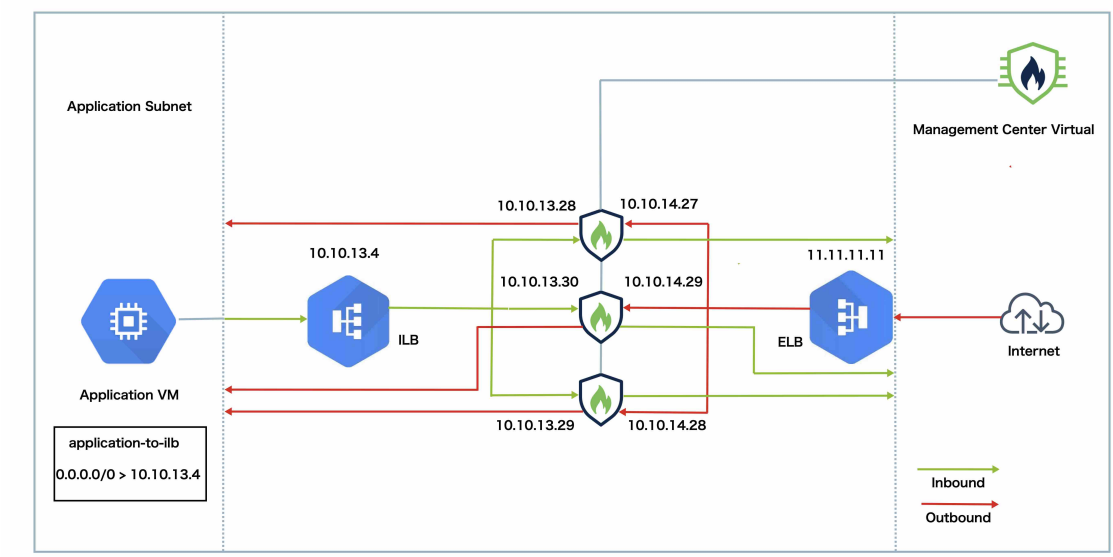
Threat Defense Virtual Clustering with Autoscale Solution in GCP

From Release 10.0.0, the Threat Defense Virtual clustering solution in Google Cloud Platform (GCP) is enhanced with dynamic autoscaling capabilities. The autoscale solution is based on CPU utilization metrics and helps to achieve optimal resource usage. It is deployed using Terraform-based templates.

You can choose the scaling option as either dynamic clustering or fixed-node clustering, depending on your requirements. You can also define the CPU threshold value, beyond which scaling will be initiated.

Sample Topology of GCP Clustering Autoscale Solution

Figure 2: Sample Topology



The topology shows both inbound and outbound traffic flow.

1. The Threat Defense Virtual cluster is placed between the internal and external load balancers. A Management Center Virtual instance is used to manage the cluster.
2. Inbound traffic from the internet goes to the external load balancer, which then transmits the traffic to the Threat Defense Virtual cluster.
3. The Threat Defense Virtual instance in the cluster inspects the traffic, and after inspection, forwards the traffic to the application VM.
4. Outbound traffic from the application VM goes to the internal load balancer. The load balancer forwards this traffic to the Threat Defense Virtual cluster, which sends it to the internet.

Requirements and Prerequisites

This section lists the requirements and supported configurations for Threat Defense Virtual clustering autoscale solution on GCP .

Model Requirements

- Supported Threat Defense Virtual models: FTDv20, FTDv30, FTDv50, and FTDv100.
- A valid Google Cloud Platform (GCP) account is required.
- Google Cloud SDK must be installed locally, or you should have access to GCP Cloud Shell.
- A Cisco Smart Licensing account is necessary for licensing the Management Center.

Management Center Requirements

- Make sure that an access control policy is configured.

- Make sure that the security zones for data interfaces are defined.
- A user with *Administrator* role is required for auto-registration.
- Make sure that the NAT rules are configured to respond to the health check of load balancers.

Licensing

- The license entitlement of the control node applies to all data nodes in a cluster, regardless of their initial performance tier configuration during boot-up.
- After cluster formation, changing the performance tier of any individual Threat Defense Virtual is not allowed.
- By default, each unit includes a 100 kbps evaluation license.
- License behavior after deregistration:
  - Until reboot: The unit retains the existing throughput.
  - After reboot: The throughput reverts to the default 100 kbps.
- Only BYOL (Bring Your Own License) license type is supported.
- PAYG (Pay-As-You-Go) license type is not supported.

Supported configurations

- Cluster size: 1-16 nodes.
- Traffic topologies: North-South and East-West.
- Deployment mode: Deployment using Terraform templates.

Templates

The following templates are available on GitHub.

- infrastructure/main.tf – Template for infrastructure deployment.
- cluster_deployment/main.tf – Template for cluster deployment.

Deploying Threat Defense Virtual Clustering Autoscale Solution on GCP

There are two steps in the deployment process:

1. Infrastructure deployment (optional) – to deploy the required network infrastructure.
2. Cluster deployment – to deploy the Threat Defense Virtual cluster.



Note If you choose to use your own infrastructure for deploying the Threat Defense Virtual, then you can skip the infrastructure deployment step. However, it is essential to make sure that all the required resources are provided for the proper functioning of the solution.

Infrastructure Deployment Parameters

The following parameters are used for the infrastructure deployment:

Table 3: List of template parameters for infrastructure deployment (infrastructure_params.tfvars)

Parameter	Description	Example
project_id	GCP project ID	test-project-12345
resource_name_prefix	String (lowercase alphabets only.)	demoftdv
region	GCP region to deploy	us-central1
mgmt_ip_cidr_range	Mgmt Subnet CIDR	10.112.0.0/27
vpc_connector_ip_cidr_range	Mgmt VPC Connector CIDR /28 subnet only	10.112.50.0/28
with_diagnostic	Whether to enable diagnostic interface.	true or false (false is supported only starting with version 7.4.1.)
diag_ip_cidr_range	Diagnostic Subnet CIDR	10.112.19.0/27
inside_ip_cidr_range	Inside Subnet CIDR	10.112.1.0/27
outside_ip_cidr_range	Outside Subnet CIDR	10.112.2.0/27
ccl_ip_cidr_range	CCL Subnet CIDR (/27 subnet recommended)	10.112.100.0/27

Deploy Infrastructure Using Terraform on GCP

The GCP Infrastructure Manager facilitates Terraform deployments as a stack. From this interface, you can manage resources and perform cleanups.

Procedure

Step 1 Navigate to the GCP console and log in using your credentials.

Step 2 Open Cloud Shell or local terminal.

If you are using local terminal, make sure that GCP CLI is configured on your system.

Step 3 Create a new directory.

Example:

Use the command **mkdir infra** to create a directory named infra in Cloud Shell or your local terminal.

Step 4 Navigate to the newly created directory.

Example:

Use the command **cd infra** to change to the new directory.

Step 5 Copy or download the infrastructure_params.tfvars file from Cisco GitHub repository and save it to the new infra folder.

Step 6 Open the infrastructure_params.tfvars file and update all the placeholder values as required. Refer to Infrastructure Deployment Parameters, on page 17.

Step 7 Use the following commands after updating placeholder values to start the deployment.

```
gcloud infra-manager deployments \
apply
"projects/<project_id>/locations/<region>/deployments
/<deployment_name>" \
--location="<region>" \
--git-source-repo="<repo name>" \
--git-source-directory="infrastructure” \
--git-source-ref="<branch name>" \
--serviceaccount="
projects/<project_id>/serviceAccounts/<servi
ce_account_name>" \
--artifacts-gcs-bucket="gs://<bucket_name>/artifacts"
\
--inputs-file="/path/to/your/infra_params.tfvars"
```

Table 4: Parameters description

Parameter	Description
deployment_name	Name of the Terraform deployment stack that will be shown in GCP Infrastructure Manager.
location	The location or region where the deployment will be applied.
git-source-repo	The name of the Git repository that contains the source code (Cisco GitHub Link).
git-source-directory	The directory in the Git repository where the infrastructure code is located.
git-source-ref	The branch name in the Git repository to be used for the deployment.
service-account	Name of the service account.
artifacts-gcs-bucket	The Google Cloud Storage bucket to store artifacts related to the deployment (optional).
inputs-file	The file path to the input parameters for the infrastructure configuration.

Note
If you do not use the template to deploy the infrastructure, then you must provide the following resources for proper deployment and functioning of the solution.

- Management VPC
  - Management Subnet
  - Management Subnet for VPC connector
  - Management Firewall rule name
- Inside VPC
  - Inside Subnet
  - Inside firewall rule name

- Outside VPC
  - Outside Subnet
  - Outside Firewall rule name
  - Nat GW
- CCL VPC
  - CCL Subnet
  - CCL firewall rule name.
- Diagnostic VPC (optional)
  - Diagnostic Subnet
  - Diagnostic Firewall rule name
- VPC Connector with management subnet

Cluster Deployment Parameters

The following input parameters are used to deploy the cluster.

Table 5: List of template parameters for cluster deployment

Parameter	Description	Example
General configuration		
type_of_deployment	Type of deployment (north_south or east_west)	north_south
Project information		
project_id	GCP project ID	test-project-12345
region	Deployment region	us-east1
zone1	Deployment zone	a
zone2	Deployment zone	b
zone3	Deployment zone	c
resource_name_prefix	Prefix for naming resources	democluster
service_account_mail_id	Service account email	service-account-mail@example-project.iam.gserviceaccount.com
VPC and subnet configuration		
mgmt_vpc_name	Management VPC name	<resource-name>-ftdv-mgmt-vpc

Parameter	Description	Example
mgmt_subnet_name	Management subnet name	<resource-name>-ftdv-mgmt-subnet
inside_vpc_name	Inside VPC name	<resource-name>-ftdv-inside-vpc
inside_subnet_name	Inside subnet name	<resource-name>-ftdv-inside-subnet
outside_vpc_name	Outside VPC name	<resource-name>-ftdv-outside-vpc
outside_subnet_name	Outside subnet name	<resource-name>-ftdv-outside-subnet
diag_vpc_name	Diagnostics VPC name	<resource-name>-ftdv-diag-vpc
diag_subnet_name	Diagnostics subnet name	<resource-name>-ftdv-diag-subnet
ccl_vpc_name	CCL VPC name	<resource-name>-ftdv-ccl-vpc
ccl_subnet_name	CCL subnet name	<resource-name>-ftdv-ccl-subnet
Firewall rules		
diag_firewall_rule_name	Firewall rule for diagnostics	<resource-name>-ftdv-diag-firewall-rule
ccl_firewall_rule_name	Firewall rule for CCL	<resource-name>-ftdv-ccl-firewall-rule
mgmt_firewall_rule_name	Firewall rule for management	<resource-name>-ftdv-mgmt-firewall-rule
inside_firewall_rule_name	Firewall rule for inside network	<resource-name>-ftdv-inside-firewall-rule
outside_firewall_rule_name	Firewall rule for outside network	<resource-name>-ftdv-outside-firewall-rule
inside_hc_firewall_rule_name	Firewall rule for inside network	<resource-name>-ftdv-inside-hc-firewall-rule
outside_hc_firewall_rule_name	Firewall rule for outside network	<resource-name>-ftdv-outside-hc-firewall-rule
Instance details		
machine_type	GCP supported machine type	n1-standard-8
source_image_url	Source image location for Threat Defense Virtual	projects/cisco-public/global/images/cisco-ftdv-10-0-0
public_key	Public Key for SSH access	ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAABAQC4v
Autoscale details		
auto_scaling	Enable autoscaling	true/false
cpu_utilization_target	Target CPU utilization for autoscaling [0.0-1.0]	0.60
min_ftd_count	Minimum number of Threat Defense Virtual instances	0
max_ftd_count	Maximum number of Threat Defense Virtual instances	2
Threat Defense Virtual specific configuration		

Parameter	Description	Example
ftd_password_secret_name	Name of the Threat Defense Virtual secret for new Admin password in Secret Manager, device will use this password after first time login.	ftd-password-secret
hostname	Hostname for Threat Defense Virtual	cisco-ngfwv
ccl_subnet_range	Subnet range for CCL, space separated	10.112.100.2 10.112.100.30
cluster_grp_name	Cluster group name for Threat Defense Virtual	ftdv-cluster
with_diagnostic	Whether to enable diagnostics	true/false
assign_public_ip_to_mgmt	Whether to assign public IP to management interface	true/false
ftd_reg_via_public_ip	Whether to register Threat Defense Virtual with public IP	true/false
Management Center information and Threat Defense Virtual configuration		
reg_id	Management Center registration ID	cisco
nat_id	NAT ID for Management Center	cisco
policy_id	Initial policy ID created in Management Center	ftdv-ini-pol
fmc_ip	Management Center IP address	10.112.0.2 / 34.113.15.29
fmc_password_secret_name	Name of the FMC secret for new Admin password in Secret Manager, device will use this password after first time login.	fmc-password-secret
fmc_username	Management Center login username	restapi
license_caps	License capabilities	BASE, MALWARE, URLFilter, THREAT
performance_tier	Performance tier for Threat Defense Virtual	FTDv20, FTDv30
vpc_connector_name	Name of VPC connector	<resource-name>-connector
Internal Load Balancer configuration		
ilb_frontend_protocol	Frontend protocol (TCP/UDP)	TCP
ilb_backend_protocol	Backend protocol	TCP

Parameter	Description	Example
ilb_health_check_port	ILB Health-check balancer port, NAT required in Management Center	8989
ilb_timeout_sec	Load balancer timeout (seconds)	5
ilb_draining_timeout_sec	Timeout for draining connections (seconds)	60
ilb_check_interval_sec	Interval between health checks for ILB (seconds)	10
ilb_unhealthy_threshold	Number of failed health checks before marking unhealthy	1
External Load Balancer specific configuration only in case of north_south deployment		
elb_frontend_protocol	Frontend protocol name for External Load Balancer (TCP/UDP)	TCP
elb_backend_protocol	Backend protocol name for External Load Balancer (TCP/UDP/UNSPECIFIED)	TCP
elb_front_end_ports	List of ELB frontend (listener) ports	all or [22, 80, 443]
elb_health_check_port	ELB health-check port, NAT required in Management Center	87878
elb_timeout_sec	Load balancer timeout (seconds)	5
elb_unhealthy_threshold	Number of failed health checks before marking unhealthy	2
elb_check_interval_sec	Interval between health checks for ELB (seconds)	10
elb_draining_timeout_sec	Timeout for draining connections (seconds)	60

Deploy Cluster Using Terraform on GCP

The GCP Infrastructure Manager facilitates Terraform deployments as a stack. From this interface, you can manage resources and perform cleanups.

Procedure

Step 1 Navigate to the GCP console and log in using your credentials.

Step 2 Open Cloud Shell or local terminal.

If you are using local terminal, make sure that GCP CLI is configured on your system.

Step 3 Create a new directory.

Example:

Use the command **mkdir cluster** to create a directory named cluster in Cloud Shell or your local terminal.

Step 4 Navigate to the newly created directory.

Example:

Use the command **cd cluster** to change to the new directory.

Step 5 Copy or download the cluster_params.tfvars file from Cisco GitHub repository and save it to the new cluster folder.

Step 6 Open the cluster_params.tfvars file and update all the placeholder values as required. Refer to Cluster Deployment Parameters, on page 19.

Step 7 Use the following commands after updating placeholder values to start the deployment.

```
gcloud infra-manager deployments \
apply "projects/<project_id>/locations/<region>/deployments/<deployment_name>" \
--location="<region>" \
--git-source-repo="<repo name>" \
--git-source-directory="cluster_deployment" \
--git-source-ref="<branch name>" \
--service-account="projects/<project_id>/serviceAccounts/<service_account_name>" \
--artifacts-gcs-bucket="gs://<bucket_name>/artifacts" \
--inputs-file="/path/to/cluster_params.tfvars"
```

Table 6: Parameters description

Parameter	Description
deployment_name	Name of the Terraform deployment stack that will be shown in GCP infrastructure manager.
location	The location or region where the deployment will be applied.
git-source-repo	The name of the Git repository that contains the source code (Cisco GitHub Link).
git-source-directory	The directory in the Git repository where the infrastructure code is located.
git-source-ref	The branch name in the Git repository to be used for the deployment.
service-account	Name of the service account.
artifacts-gcs-bucket	The Google Cloud Storage bucket to store artifacts related to the deployment, for example, Terraform state files.
inputs-file	The file path to the input parameters for the cluster deployment.

Deployment Configurations after Cluster Deployment

To create a route inside the VPC to forward the required traffic through an Internal Load balancer on Google Cloud Shell, follow these steps:

```
gcloud compute routes create <ilb-route-name>
--network=<inside-vpc-name> --priority=1000 --destination-range=0.0.0.0/0
--next-hop-ilb=<ilb-forwarding-rule-name> --next-hop-ilb-region=<region>
```



Note You can create route from the GCP UI as well. To create a route from the GCP UI, follow these steps:

1. In the GCP UI, navigate to `Routes > Route Management > Create route`.
2. Fill in the Create a route dialog box.
   a. **Name**: Specify a route name.
   b. **Network**: Select **<inside-vpc-name>**.
   c. **Route type**: Select **Static route**.
   d. **Destination IPv4 range**: Enter **0.0.0.0/0**.
   e. **Priority**: Set to **1000**.
   f. **Next Hop**: Select **Specify a forwarding rule of internal passthrough Network Load balancer**.
   g. **Forwarding rule name**: Select **<forwarding-rule-for-ilb>**.
3. Click **Create** to create the route.

Deploy the Cluster in GCP Manually

Create the Day0 Configuration for GCP:

To deploy the cluster manually, prepare the day0 configuration.

```
{
    "AdminPassword": "password",
    "Hostname": "hostname",
    "FirewallMode": "Routed",
    "ManageLocally": "No",
    "Diagnostic": "OFF",       //Optional user input from version 7.4.1
    "Cluster": {
      "CclSubnetRange": "ip_address_start ip_address_end",
      "ClusterGroupName": "cluster_name"
    }
}
```

For example:

```
{
"AdminPassword": "C15co@!23",
"Hostname": "cisco-ftd",
"FirewallMode": "routed",
"Diagnostic": "OFF",
"ManageLocally": "No",
"Cluster": {
  "CclSubnetRange": "10.10.65.2 10.10.65.29",
  "ClusterGroupName": "gcp-cls-ftd"
}
}
```

Allow Health Checks for GCP Network Load Balancers

Google Cloud provides health checks to determine if backends respond to traffic.

See https://cloud.google.com/load-balancing/docs/health-checks to create firewall rules for network load balancers. Then in the Firewall Management Center, create access rules to allow the health check traffic. See https://cloud.google.com/load-balancing/docs/health-check-concepts for the required network ranges.

You also need to configure dynamic manual NAT rules to redirect the health check traffic to the Google metadata server at 169.254.169.254.

You can set up a route for GCP health checks across all interfaces that are used to configure their health probes. You can achieve this by creating a route with a higher metric on interfaces where a route for GCP health checks is not already available.

North-South NAT Rules Sample Configuration

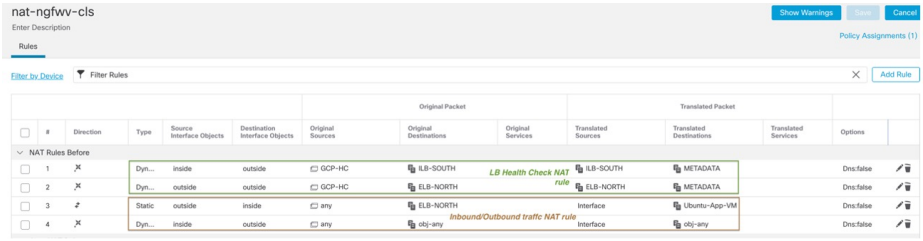
```
nat (inside,outside) source dynamic GCP-HC ILB-SOUTH destination static ILB-SOUTH METADATA
nat (outside,outside) source dynamic GCP-HC ELB-NORTH destination static ELB-NORTH METADATA

nat (outside,inside) source static any interface destination static ELB-NORTH Ubuntu-App-VM
nat (inside,outside) source dynamic any interface destination static obj-any obj-any

object network Metadata
 host 169.254.169.254

object network ILB-SOUTH
host <ILB_IP>
object network ELB-NORTH
host <ELB_IP>

object-group network GCP-HC
 network-object 35.191.0.0 255.255.0.0
 network-object 130.211.0.0 255.255.252.0
 network-object 209.85.204.0 255.255.252.0
 network-object 209.85.152.0 255.255.252.0
```



East-West NAT Rules Sample Configuration

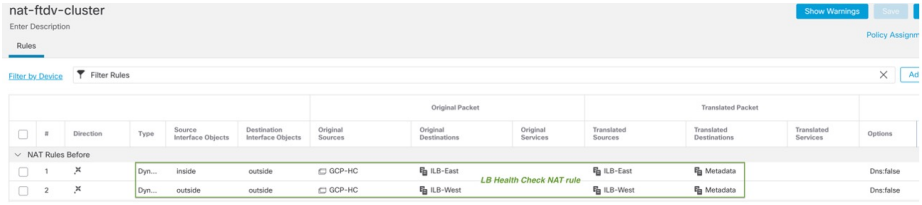
```
nat (inside,outside) source dynamic GCP-HC ILB-East destination static ILB-East Metadata
nat (outside,outside) source dynamic GCP-HC ILB-West destination static ILB-West Metadata

object network Metadata
 host 169.254.169.254

object network ILB-East
host <ILB_East_IP>
object network ILB-West
host <ILB_West_IP>

object-group network GCP-HC
```

```
network-object 35.191.0.0 255.255.0.0
network-object 130.211.0.0 255.255.252.0
network-object 209.85.204.0 255.255.252.0
network-object 209.85.152.0 255.255.252.0
```



North-South and East-West Traffic Routing Configuration Sample

```
route outside 0.0.0.0 0.0.0.0 <Outside_Gateway> 1
route inside 35.191.0.0 255.255.0.0 <Inside_Gateway> 1
route inside 130.211.0.0 255.255.252.0 <Inside_Gateway> 1
route inside 209.85.152.0 255.255.252.0 <Inside_Gateway> 1
route inside 209.85.204.0 255.255.252.0 <Inside_Gateway> 1
```

If a default route is not available, then policy-based routing can be used to route the traffic for health checks.

Note Set the NAT > Translated destination port to 80.

Troubleshooting GCP Clustering with Autoscale

Issue: The cluster is not formed.

Workaround:

- Check the IP address of NVE-only cluster interface. Make sure you can ping the NVE-only cluster interface of other nodes.
- Make sure that the IP addresses of the NVE-only cluster interfaces are part of the object group.
- Make sure that the NVE (Network Virtualization Edge) interface is configured with the corresponding object group.
- Make sure that the cluster interface in the cluster group has the correct VNI (Virtual Network Identifier) interface. This VNI interface should have the NVE with the corresponding object group.
- Each node has its own cluster interface IP address. Make sure the nodes can ping each other to verify connectivity.

Add the Cluster to the Management Center (Manual Deployment)

Use this procedure to add the cluster to the Firewall Management Center if you manually deployed the cluster. If you used a template, the cluster will auto-register on the Firewall Management Center.

Add one of the cluster units as a new device to the Firewall Management Center; the Firewall Management Center auto-detects all other cluster members.

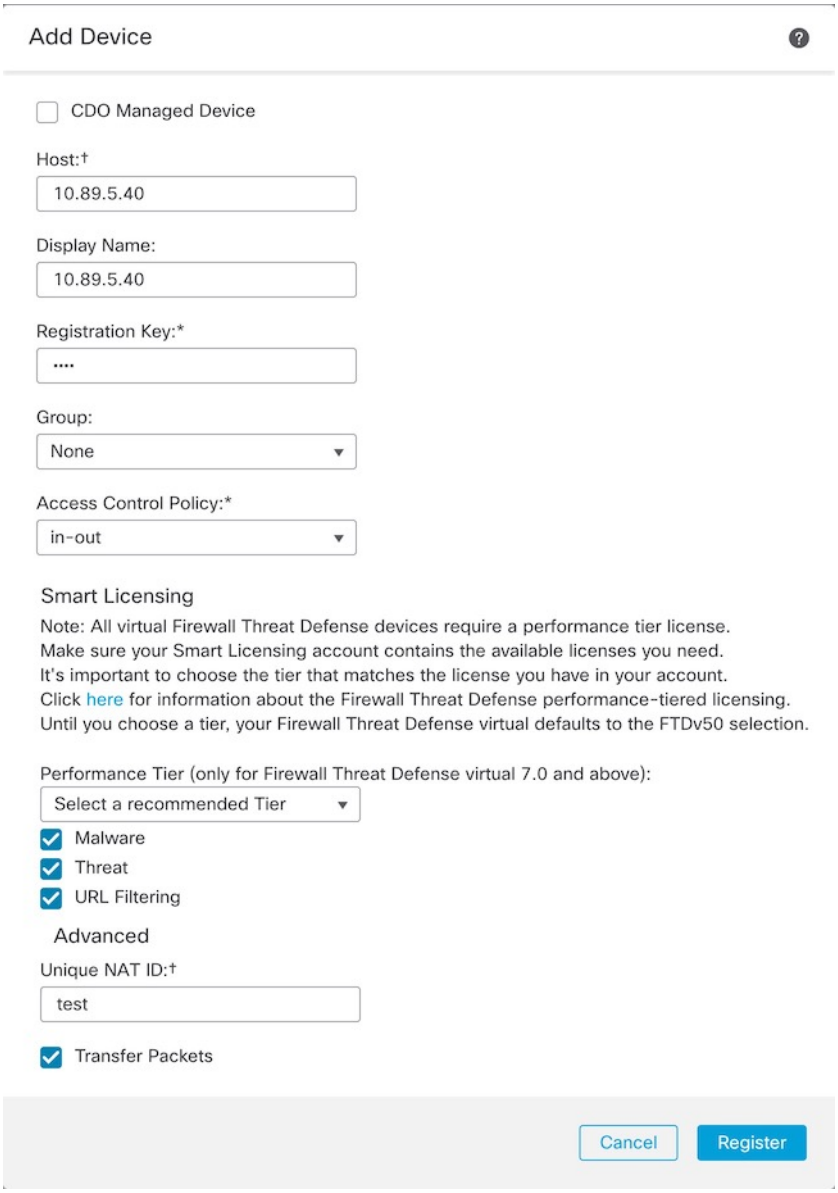
Before you begin

- All cluster units must be in a successfully-formed cluster prior to adding the cluster to the Firewall Management Center. You should also check which unit is the control unit. Use the Firewall Threat Defense **show cluster info** command.

Procedure

Step 1 In the Firewall Management Center, choose **Devices** > **Device Management**, and then choose **Add** > **Add Device** to add the control unit using the unit's management IP address.

Figure 3: Add Device



a) In the **Host** field, enter the IP address or hostname of the control unit.

We recommend adding the control unit for the best performance, but you can add any unit of the cluster.

If you used a NAT ID during device setup, you may not need to enter this field.

b) In the **Display Name** field, enter a name for the control unit as you want it to display in the Firewall Management Center.

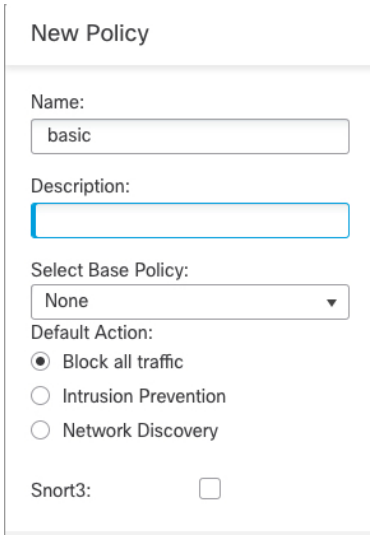
This display name is not for the cluster; it is only for the control unit you are adding. You can later change the name of other cluster members and the cluster display name.

c) In the **Registration Key** field, enter the same registration key that you used during device setup. The registration key is a one-time-use shared secret.

d) (Optional) Add the device to a device **Group**.

e) Choose an initial **Access Control Policy** to deploy to the device upon registration, or create a new policy.

If you create a new policy, you create a basic policy only. You can later customize the policy as needed.



f) Choose licenses to apply to the device.

g) If you used a NAT ID during device setup, expand the **Advanced** section and enter the same NAT ID in the **Unique NAT ID** field.

h) Check the **Transfer Packets** check box to allow the device to transfer packets to the Firewall Management Center.

This option is enabled by default. When events like IPS or Snort are triggered with this option enabled, the device sends event metadata information and packet data to the Firewall Management Center for inspection. If you disable it, only event information will be sent to the Firewall Management Center but packet data is not sent.

i) Click **Register**.

The Firewall Management Center identifies and registers the control unit, and then registers all data units. If the control unit does not successfully register, then the cluster is not added. A registration failure can occur if the cluster was not up, or because of other connectivity issues. In this case, we recommend that you try re-adding the cluster unit.

The cluster name shows on the **Devices** > **Device Management** page; expand the cluster to see the cluster units.

Figure 4: Cluster Management



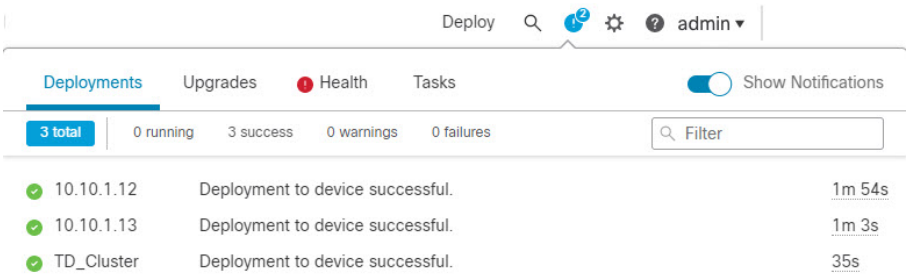
A unit that is currently registering shows the loading icon.

Figure 5: Node Registration



Note
GCP prioritizes nodes with public IP address during cluster node discovery. To ensure the Firewall Threat Defense Virtual cluster registers with the management center virtual using the private IP address, you must first disable the public IP address on the Firewall Threat Defense Virtual cluster node. This allows GCP node discovery to proceed using the private IP address for registration node with the management center virtual.

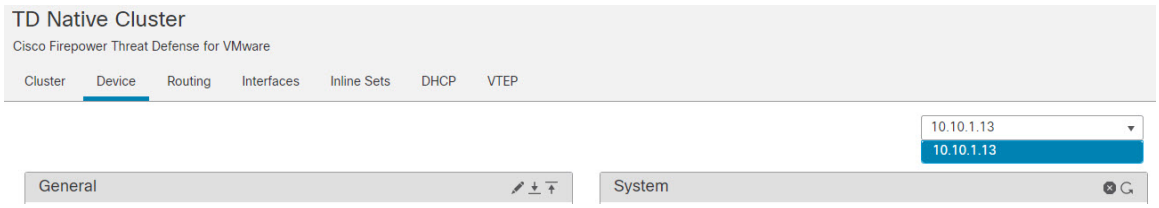
You can monitor cluster unit registration by clicking the **Notifications** icon and choosing **Tasks**. The Firewall Management Center updates the Cluster Registration task as each unit registers. If any units fail to register, see Reconcile Cluster Nodes, on page 37.



Step 2 Configure device-specific settings by clicking the **Edit** (✎) for the cluster.

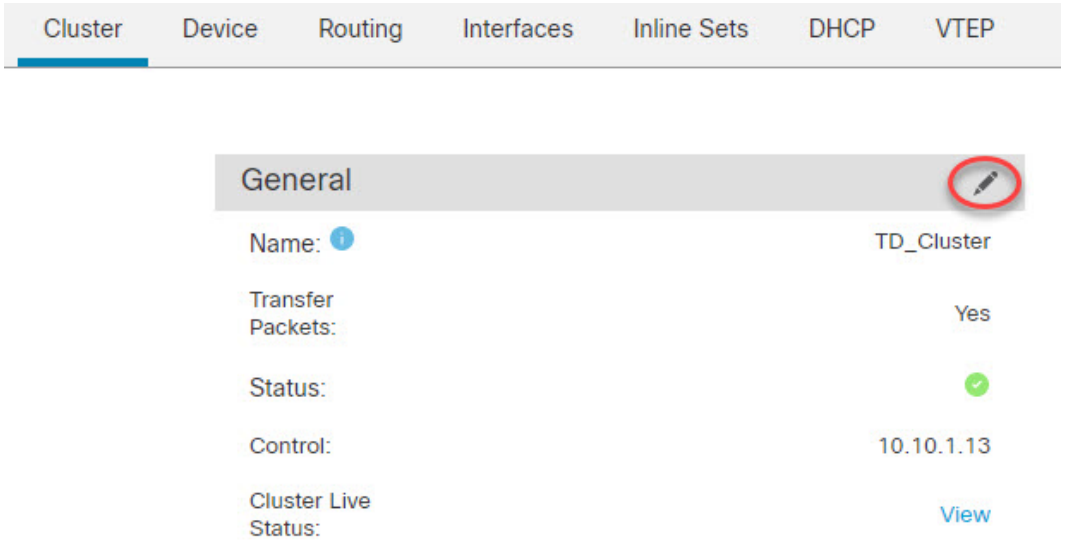
Most configuration can be applied to the cluster as a whole, and not nodes in the cluster. For example, you can change the display name per node, but you can only configure interfaces for the whole cluster.

Step 3 On the **Devices** > **Device Management** and then choose **Add**, **Cluster** screen, you see **General**, **License**, **System**, and **Health** settings.

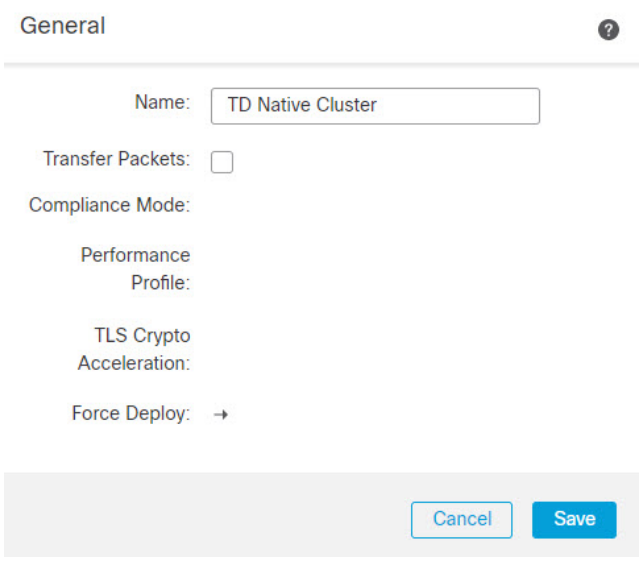


See the following cluster-specific items:

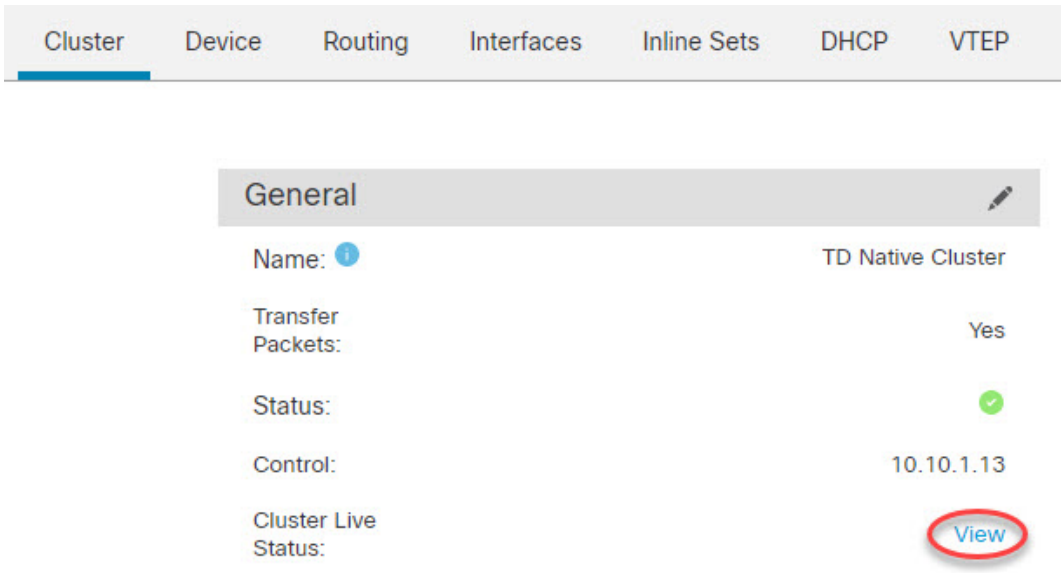
- **General > Name**—Change the cluster display name by clicking the **Edit** ( ).



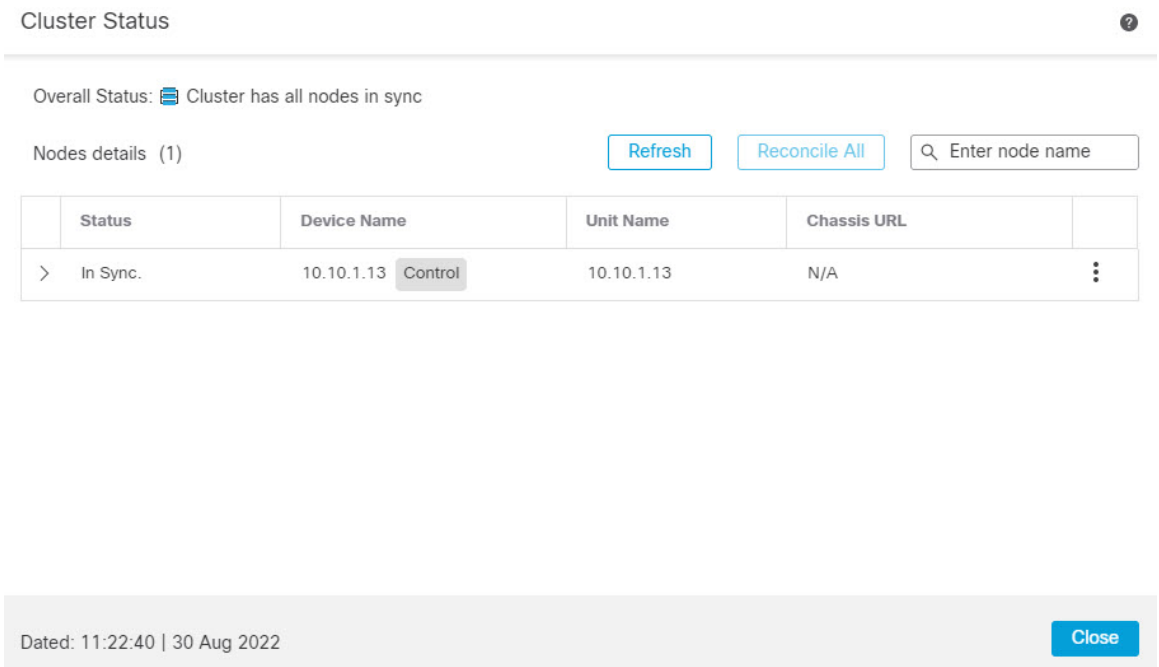
Then set the **Name** field.



- **General > Cluster Live Status**—Click the **View** link to open the **Cluster Status** dialog box.

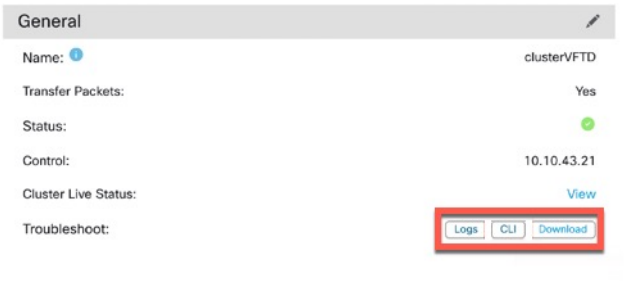


The **Cluster Status** dialog box also lets you retry data unit registration by clicking **Reconcile**.You can also ping the cluster control link from a node. See Perform a ping on the cluster control link, on page 45.



- **General > Troubleshoot**—You can generate and download troubleshooting logs, and you can view cluster CLIs. See Troubleshoot the cluster, on page 44.

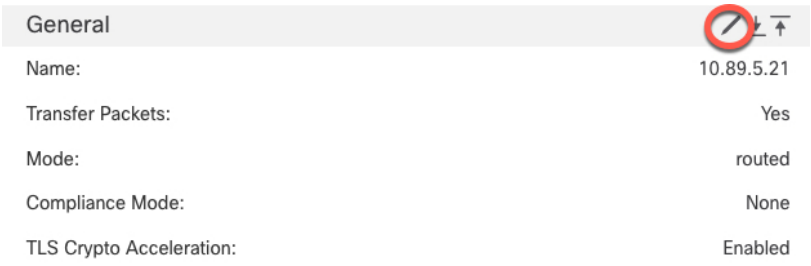
Figure 6: Troubleshoot



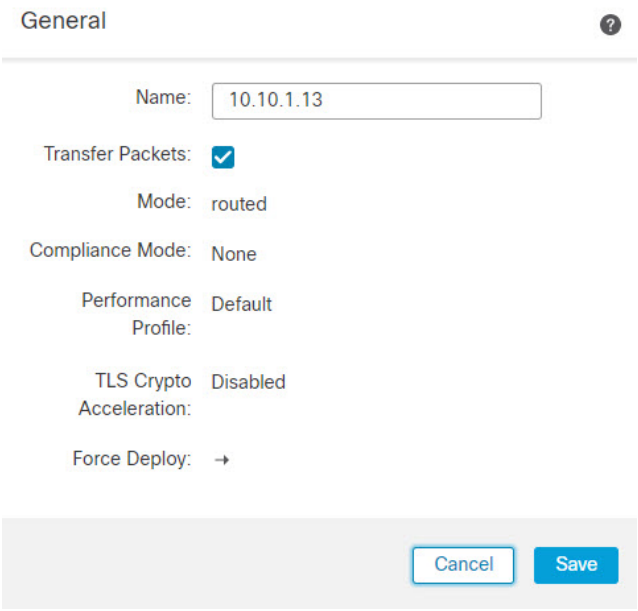
- **License**—Click **Edit** (✎) to set license entitlements.

Step 4 On the **Devices** > **Device Management** and then click **Add** > **Device**, you can choose each member in the cluster from the top right drop-down menu and configure the following settings.

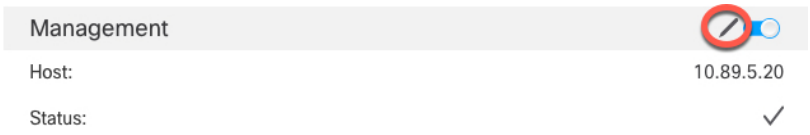
- **General > Name**—Change the cluster member display name by clicking the **Edit** (✎).



Then set the **Name** field.



- **Management > Host**—If you change the management IP address in the device configuration, you must match the new address in the Firewall Management Center so that it can reach the device on the network; edit the **Host** address in the **Management** area.



Edit cluster health monitor settings

This task allows you to modify cluster health monitor settings to control how cluster nodes monitor system health and automatically rejoin the cluster after failures.

The **Cluster Health Monitor Settings** section of the **Cluster** page displays the settings for monitoring cluster node health. You can monitor any port-channel ID, any single physical interface ID, and the Snort and disk-full processes. Health monitoring is not performed on VLAN subinterfaces or virtual interfaces such as VNIs or BVIs. You cannot configure monitoring for the cluster control link. It is always monitored.

When a topology change occurs (such as adding or removing a data interface, enabling or disabling an interface on the node or the switch, or adding an additional switch to form a VSS or vPC or VNet) you should disable the system health check feature. Also, disable interface monitoring for any interfaces that are disabled. When the topology change is complete, and the configuration change is synced to all nodes, you can re-enable the system health check feature and monitored interfaces.

Procedure

Step 1 Choose **Devices** > **Device Management**.

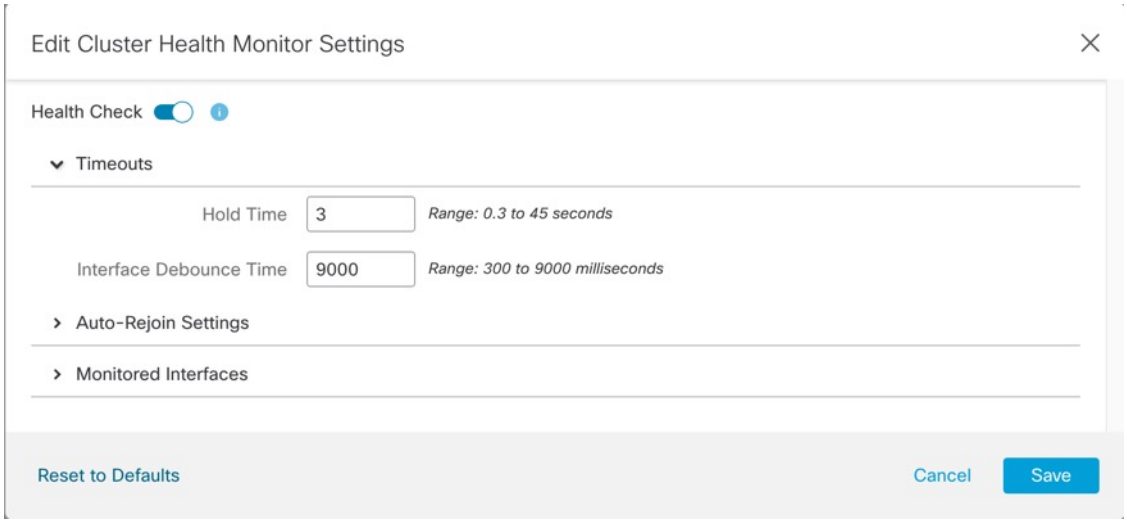
Step 2 Next to the cluster you want to modify, click **Edit** ( ).

Step 3 Click **Cluster**.

Step 4 In the **Cluster Health Monitor Settings** section, click **Edit** ( ).

Step 5 Disable the system health check by clicking the **Health Check** slider.

Figure 7: Disable the system health check



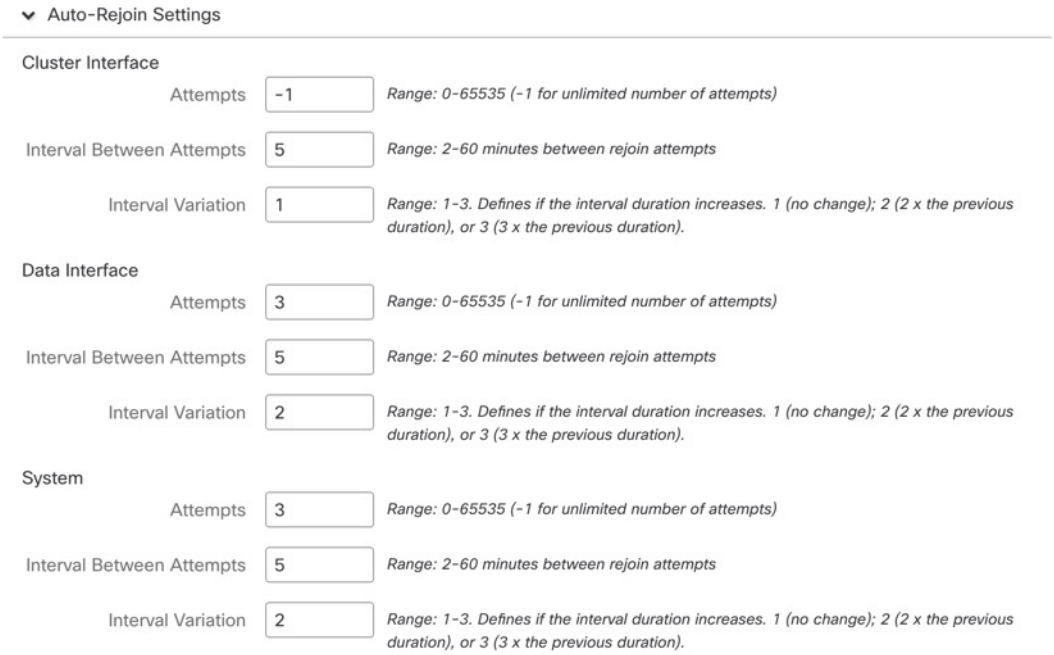
When any topology changes occur (such as adding or removing a data interface, enabling or disabling an interface on the node or the switch, or adding an additional switch to form a VSS or vPC or VNet) you should disable the system health check feature and also disable interface monitoring for the disabled interfaces. When the topology change is complete, and the configuration change is synced to all nodes, you can re-enable the system health check feature and monitored interfaces.

Step 6 Configure the hold time and interface debounce time.

- **Hold Time**—Set the hold time to determine the amount of time between node heartbeat status messages, between 0.3 and 45 seconds. The default is 3 seconds.
- **Interface Debounce Time**—Set the debounce time between 300 and 9000 ms. The default is 500 ms. Lower values allow for faster detection of interface failures. Note that configuring a lower debounce time increases the chances of false-positives. When an interface status update occurs, the node waits the specified number of milliseconds before marking the interface as failed and removing the node from the cluster. If an EtherChannel transitions from a down state to an up state, such as after the switch reloads or enables an EtherChannel, a longer debounce time can prevent the interface from appearing failed on a cluster node if another cluster node bundles the ports more quickly.

Step 7 Customize the auto-rejoin cluster settings after a health check failure.

Figure 8: Configure Auto-Rejoin settings

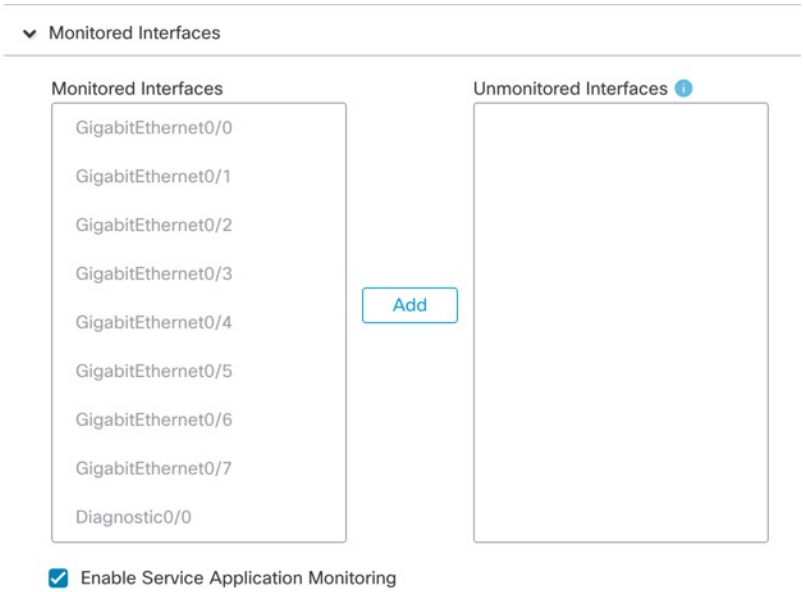


Set values for the **Cluster Interface**, **Data Interface**, and **System** (internal failures include: application sync timeout; inconsistent application statuses; and so on):

- **Attempts**—Sets the number of rejoin attempts, between -1 and 65535. **0** disables auto-rejoining. The default for the **Cluster Interface** is -1 (unlimited). The default for the **Data Interface** and **System** is 3.
- **Interval Between Attempts**—Defines the interval duration in minutes between rejoin attempts, between 2 and 60. The default value is 5 minutes. The maximum total time that the node attempts to rejoin the cluster is limited to 14400 minutes (10 days) from the time of last failure.
- **Interval Variation**—Defines if the interval duration increases. Set the value between 1 and 3: **1** (no change); **2** (2 x the previous duration), or **3** (3 x the previous duration). For example, if you set the interval duration to 5 minutes, and set the variation to 2, then the first attempt is after 5 minutes; the 2nd attempt is 10 minutes (2 x 5); the 3rd attempt 20 minutes (2 x 10), and so on. The default value is **1** for the **Cluster Interface** and **2** for the **Data Interface** and **System**.

Step 8 Configure monitored interfaces by moving interfaces in the **Monitored Interfaces** or **Unmonitored Interfaces** window. You can also check or uncheck **Enable Service Application Monitoring** to enable or disable monitoring of the Snort and disk-full processes.

Figure 9: Configure monitored interfaces



The interface health check monitors for link failures. If all physical ports for a given logical interface fail on a particular node, but there are active ports under the same logical interface on other nodes, then the node is removed from the cluster. The time required to remove a member from the cluster depends on the type of interface and whether the node is established or is joining the cluster. By default, health check is enabled for all interfaces and the Snort and disk-full processes.

You might want to disable health monitoring of non-essential interfaces.

When any topology changes occur (such as adding or removing a data interface, enabling or disabling an interface on the node or the switch, or adding an additional switch to form a VSS or vPC or VNet) you should disable the system health check feature and also disable interface monitoring for the disabled interfaces. When the topology change is complete, and the configuration change is synced to all nodes, you can re-enable the system health check feature and monitored interfaces.

Step 9 Click **Save**.

Deploy configuration changes.

Manage Cluster Nodes

Disable Clustering

You may want to deactivate a node in preparation for deleting the node, or temporarily for maintenance. This procedure is meant to temporarily deactivate a node; the node will still appear in the Firewall Management Center device list. When a node becomes inactive, all data interfaces are shut down.

Note Do not power off the node without first disabling clustering.

Procedure

Step 1 For the unit you want to disable, choose **Devices** > **Device Management**, click the **More** (⋮), and choose **Disable Node Clustering**.

Step 2 Confirm that you want to disable clustering on the node.

The node will show **(Disabled)** next to its name in the **Devices** > **Device Management** list.

Step 3 To reenable clustering, see Rejoin the Cluster, on page 37.

Rejoin the Cluster

If a node was removed from the cluster, for example for a failed interface or if you manually disabled clustering, you must manually rejoin the cluster. Make sure the failure is resolved before you try to rejoin the cluster. See Rejoining the Cluster, on page 54 for more information about why a node can be removed from a cluster.

Procedure

Step 1 For the unit you want to reactivate, choose **Devices** > **Device Management**, click the **More** (⋮), and choose **Enable Node Clustering**.

Step 2 Confirm that you want to enable clustering on the node.

Reconcile Cluster Nodes

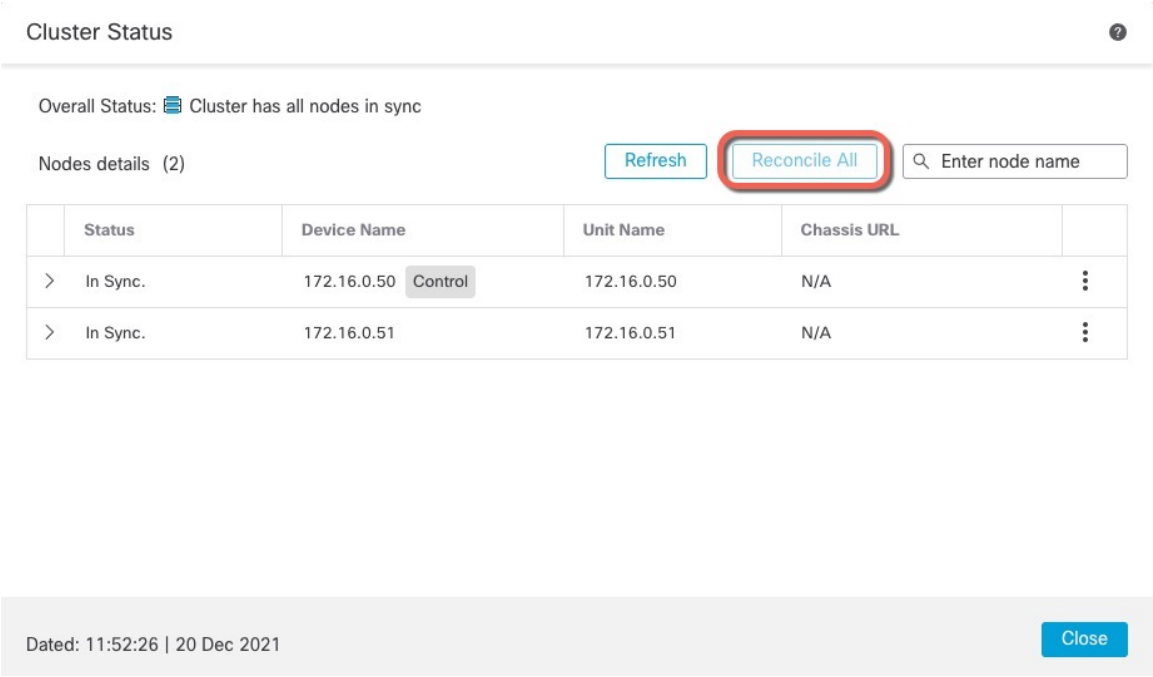
If a cluster node fails to register, you can reconcile the cluster membership from the device to the Firewall Management Center. For example, a data node might fail to register if the Firewall Management Center is occupied with certain processes, or if there is a network issue.

Procedure

Step 1 Choose **Devices** > **Device Management More** (⋮) for the cluster, and then choose **Cluster Live Status** to open the **Cluster Status** dialog box.

Step 2 Click **Reconcile All**.

Figure 10: Reconcile All



For more information about the cluster status, see Monitoring the Cluster, on page 39.

Unregister the Cluster or Nodes and Register to a New Firewall Management Center

You can unregister the cluster from the Firewall Management Center, which keeps the cluster intact. You might want to unregister the cluster if you want to add the cluster to a new Firewall Management Center.

You can also unregister a node from the Firewall Management Center without breaking the node from the cluster. Although the node is not visible in the Firewall Management Center, it is still part of the cluster, and it will continue to pass traffic and could even become the control node. You cannot unregister the current control node. You might want to unregister the node if it is no longer reachable from the Firewall Management Center, but you still want to keep it as part of the cluster while you troubleshoot management connectivity.

Unregistering a cluster:

- Severs all communication between the Firewall Management Center and the cluster.
- Removes the cluster from the **Device Management** page.
- Returns the cluster to local time management if the cluster's platform settings policy is configured to receive time from the Firewall Management Center using NTP.
- Leaves the configuration intact, so the cluster continues to process traffic.

  Policies, such as NAT and VPN, ACLs, and the interface configurations remain intact.

Registering the cluster again to the same or a different Firewall Management Center causes the configuration to be removed, so the cluster will stop processing traffic at that point; the cluster configuration remains intact so you can add the cluster as a whole. You

can choose an access control policy at registration, but you will have to re-apply other policies after registration and then deploy the configuration before it will process traffic again.

Before you begin

This procedure requires CLI access to one of the nodes.

Procedure

Step 1 Choose **Devices** > **Device Management**, click **More** (⋮) for the cluster or node, and choose **Unregister**.

Step 2 You are prompted to unregister the cluster or node; click **Yes**.

Step 3 You can register the cluster to a new (or the same) Firewall Management Center by adding one of the cluster members as a new device.

You only need to add one of the cluster nodes as a device, and the rest of the cluster nodes will be discovered.

a) Connect to one cluster node's CLI, and identify the new Firewall Management Center using the **configure manager add** command.

b) Choose **Devices** > **Device Management**, and then click **Add Device**.

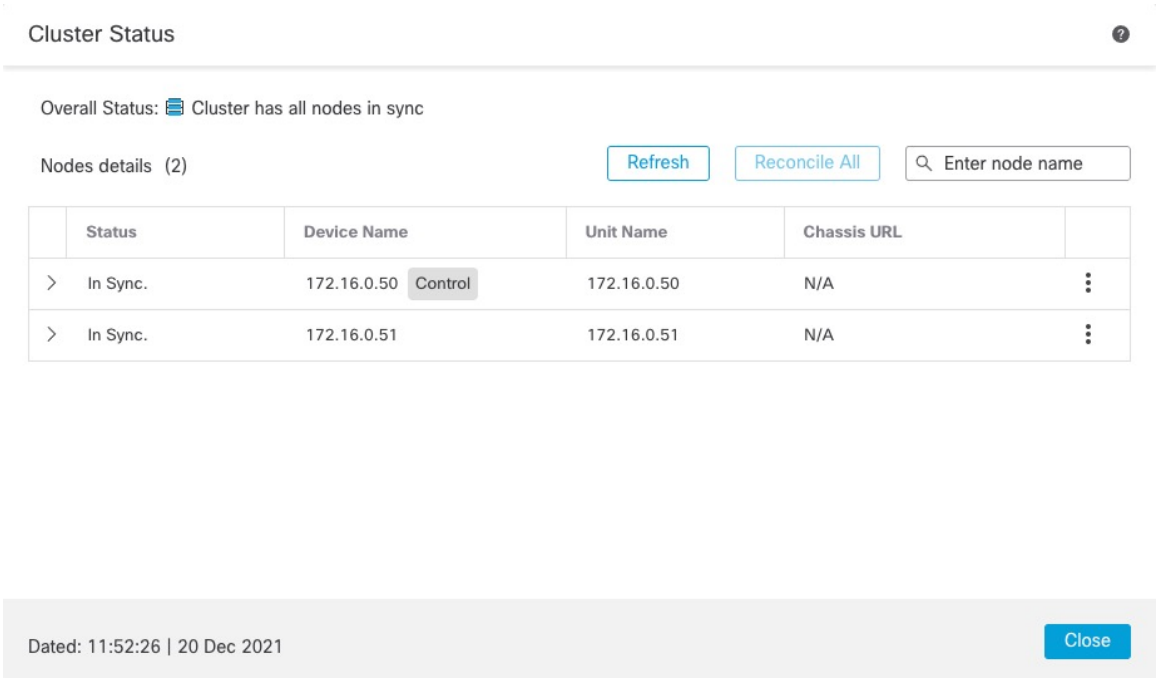
Step 4 To re-add a deleted node, see Reconcile Cluster Nodes, on page 37.

Monitoring the Cluster

You can monitor the cluster in the Firewall Management Center and at the Firewall Threat Defense CLI.

- **Cluster Status** dialog box, which is available from the **Devices** > **Device Management**, **More** (⋮) icon or from the **Devices** > **Device Management**, click **Add**, choose the **Cluster** page **General** area **Cluster Live Status** link.

Figure 11: Cluster Status



The Control node has a graphic indicator identifying its role.

Cluster member **Status** includes the following states:

- In Sync.—The node is registered with the Firewall Management Center.
- Pending Registration—The node is part of the cluster, but has not yet registered with the Firewall Management Center. If a node fails to register, you can retry registration by clicking **Reconcile All**.
- Clustering is disabled—The node is registered with the Firewall Management Center, but is an inactive member of the cluster. The clustering configuration remains intact if you intend to later re-enable it, or you can delete the node from the cluster.
- Joining cluster...—The node is joining the cluster on the chassis, but has not completed joining. After it joins, it will register with the Firewall Management Center.

For each node, you can view the **Summary** or the **History**.

Figure 12: Node Summary

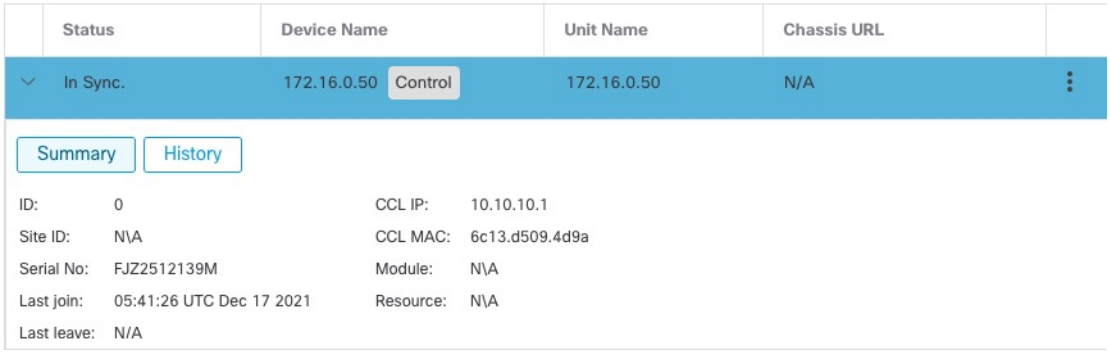
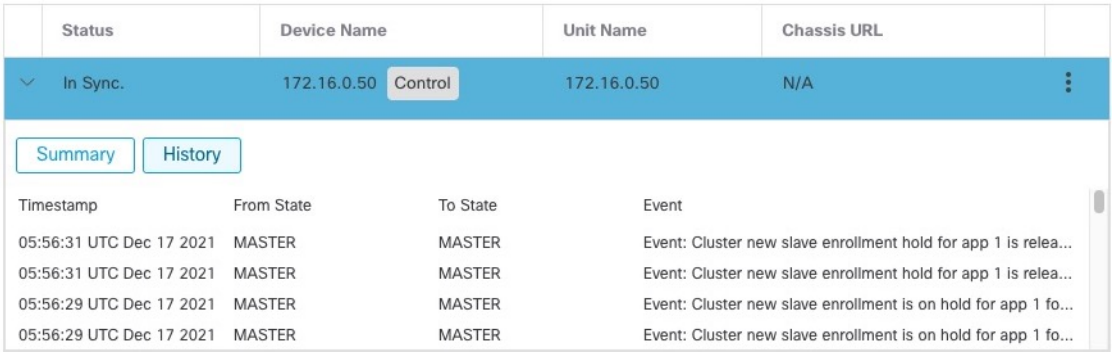


Figure 13: Node History



- **System** (⚙) > **Tasks** page.

  The **Tasks** page shows updates of the Cluster Registration task as each node registers.

- **Devices** > **Device Management** and then click **Add** > **Device** *cluster_name*.

  When you expand the cluster on the devices listing page, you can see all member nodes, including the control node shown with its role next to the IP address. For nodes that are still registering, you can see the loading icon.

- **show cluster** {**access-list** [*acl_name*] | **conn** [**count**] | **cpu** [**usage**] | **history** | **interface-mode** | **memory** | **resource usage** | **service-policy** | **traffic** | **xlate count**}

  To view aggregated data for the entire cluster or other information, use the **show cluster** command.

- **show cluster info** [**auto-join** | **clients** | **conn-distribution** | **flow-mobility counters** | **goid** [*options*] | **health** | **incompatible-config** | **loadbalance** | **old-members** | **packet-distribution** | **trace** [*options*] | **transport** { **asp** | **cp**}]

  To view cluster information, use the **show cluster info** command.

Cluster Health Monitor Dashboard

When a Firewall Threat Defense is the control node of a cluster, the Firewall Management Center collects various metrics periodically from the device metric data collector.

The cluster health monitor is comprised of these components:

- Overview dashboard—Displays information about the cluster topology, cluster statistics, and metric charts:

- The topology section displays a cluster's live status, the health of individual threat defense, threat defense node type (control node or data node), and the status of the device. The status of the device could be *Disabled* (when the device leaves the cluster), *Added out of box* (in a public cloud cluster, the additional nodes that do not belong to the Firewall Management Center), or *Normal* (ideal state of the node).
- The cluster statistics section displays current metrics of the cluster with respect to the CPU usage, memory usage, input rate, output rate, active connections, and NAT translations.



Note The CPU and memory metrics display the individual average of the data plane and snort usage.

- The metric charts, namely, CPU Usage, Memory Usage, Throughput, and Connections, diagrammatically display the statistics of the cluster over the specified time period.

- Load Distribution dashboard—Displays load distribution across the cluster nodes in two widgets:
  - The Distribution widget displays the average packet and connection distribution over the time range across the cluster nodes. This data depicts how the load is being distributed by the nodes. Using this widget, you can easily identify any abnormalities in the load distribution and rectify it.
  - The Node Statistics widget displays the node level metrics in table format. It displays metric data on CPU usage, memory usage, input rate, output rate, active connections, and NAT translations across the cluster nodes. This table view enables you to correlate data and easily identify any discrepancies.
- Member Performance dashboard—Displays current metrics of the cluster nodes. You can use the selector to filter the nodes and view the details of a specific node. The metric data include CPU usage, memory usage, input rate, output rate, active connections, and NAT translations.
- CCL dashboard—Displays, graphically, the cluster control link data namely, the input, and output rate.
- Troubleshooting and Links — Provides convenient links to frequently used troubleshooting topics and procedures.
- Time range—An adjustable time window to constrain the information that appears in the various cluster metrics dashboards and widgets.
- Custom Dashboard—Displays data on both cluster-wide metrics and node-level metrics. However, node selection only applies for the threat defense metrics and not for the entire cluster to which the node belongs.

Viewing Cluster Health

You must be an Admin, Maintenance, or Security Analyst user to perform this procedure.

The cluster health monitor provides a detailed view of the health status of a cluster and its nodes. This cluster health monitor provides health status and trends of the cluster in an array of dashboards.

Before you begin

Ensure you have created a cluster from one or more devices in the Firewall Management Center.

Follow these steps to view the cluster health monitor:

Procedure

Step 1 Choose **System** (⚙) > **Health** > **Monitor**.

Use the Monitoring navigation pane to access node-specific health monitors.

Step 2 In the device list, click **Expand**(>) and **Collapse** (⌄) to expand and collapse the list of managed cluster devices.

Step 3 Click on the cluster name to view the cluster health statistics.

The cluster monitor reports health and performance metrics in several predefined dashboards by default. The metrics dashboards include:

- Overview: Highlights key metrics from the other predefined dashboards, including its nodes, CPU, memory, input and output rates, connection statistics, and NAT translation information.
- Load Distribution: Traffic and packet distribution across the cluster nodes.
- Member Performance: Node-level statistics on CPU usage, memory usage, input throughput, output throughput, active connection, and NAT translation.
- CCL: Interface status and aggregate traffic statistics.

You can navigate through the various metrics dashboards by clicking on the labels. For a comprehensive list of the supported cluster metrics, refer to Cisco Secure Firewall Threat Defense Health Metrics.

Step 4 Configure the time range from the drop-down in the upper-right corner.

The time range can reflect a period as short as the last hour (the default) or as long as two weeks. Select **Custom** from the drop-down to configure a custom start and end date.

Click the refresh icon to set auto refresh to 5 minutes or to toggle off auto refresh.

Step 5 Click on deployment icon for a deployment overlay on the trend graph, with respect to the selected time range.

The deployment icon indicates the number of deployments during the selected time-range. A vertical band indicates the deployment start and end time. For multiple deployments, multiple bands/lines appear. Click on the icon on top of the dotted line to view the deployment details.

Step 6 View the **Health Alerts** for the node in the alert notification at the top of page, directly to the right of the device name.

Hover your pointer over the **Health Alerts** to view the health summary of the node. The popup window shows a truncated summary of the top five health alerts. Click on the popup to open a detailed view of the health alert summary.

Step 7 View device monitor health and performance metrics in several predefined dashboards.

The metrics dashboards include:

- Overview: Highlights key metrics from the other predefined dashboards, including CPU, memory, interfaces, connection statistics; plus disk usage and critical process information.
- CPU: CPU utilization, including the CPU usage by process and by physical cores.
- Memory: Device memory utilization, including data plane and Snort memory usage.
- Interfaces: Interface status and aggregate traffic statistics.
- Connections: Connection statistics (such as elephant flows, active connections, peak connections, and so on) and NAT translation counts.
- Snort: Statistics that are related to the Snort process.
- ASP drops: Statistics related to the dropped packets against various reasons.

You can navigate through the various metrics dashboards by clicking on the labels. Refer to Cisco Secure Firewall Threat Defense Health Metrics for a comprehensive list of the supported device metrics.

Step 8 Click the plus sign **Add New Dashboard**(+) in the upper right corner of the health monitor to create a custom dashboard by building your own variable set from the available metric groups.

For cluster-wide dashboard, choose Cluster metric group, and then choose the metric.

Cluster Metrics

The cluster health monitor tracks statistics that are related to a cluster and its nodes, and aggregate of load distribution, performance, and CCL traffic statistics.

Table 7: Cluster Metrics

Metric	Description	Format
CPU	Average of CPU metrics on the nodes of a cluster (individually for data plane and snort).	percentage
Memory	Average of memory metrics on the nodes of a cluster (individually for data plane and snort).	percentage
Data Throughput	Incoming and outgoing data traffic statistics for a cluster.	bytes
CCL Throughput	Incoming and outgoing CCL traffic statistics for a cluster.	bytes
Connections	Count of active connections in a cluster.	number
NAT Translations	Count of NAT translations for a cluster.	number
Distribution	Connection distribution count in the cluster for every second.	number
Packets	Packet distribution count in the cluster for every second.	number

Troubleshoot the cluster

Use these methods to troubleshoot the devices and clusters:

- Use the **CCL Ping** tool to make sure the cluster control link is operating correctly.
- Troubleshooting files—If a node fails to join the cluster, a troubleshooting file is automatically generated. You can also generate and download troubleshooting files from the **Devices** > **Device Management** and then choose **Add**, **ClusterGeneral** area.

  You can also generate files from the **Device Management** page. Click **More** (⋮) and choose **Troubleshoot Files**.
- CLI output—Choose **Devices** > **Device Management** and then choose **Add**, **ClusterGeneral** area, you can view a set of pre-defined CLI outputs that can help you troubleshoot the cluster. These commands are automatically run for the cluster:

Commands
show running-config cluster

Commands
show cluster info
show cluster info health
show cluster info transport cp
show version
show asp drop
show counters
show arp
show int ip brief
show blocks
show cpu detailed
show interface *ccl_interface*
ping *ccl_ip* **size** *ccl_mtu* **repeat 2**

You can also enter any **show** command in the **Command** field.

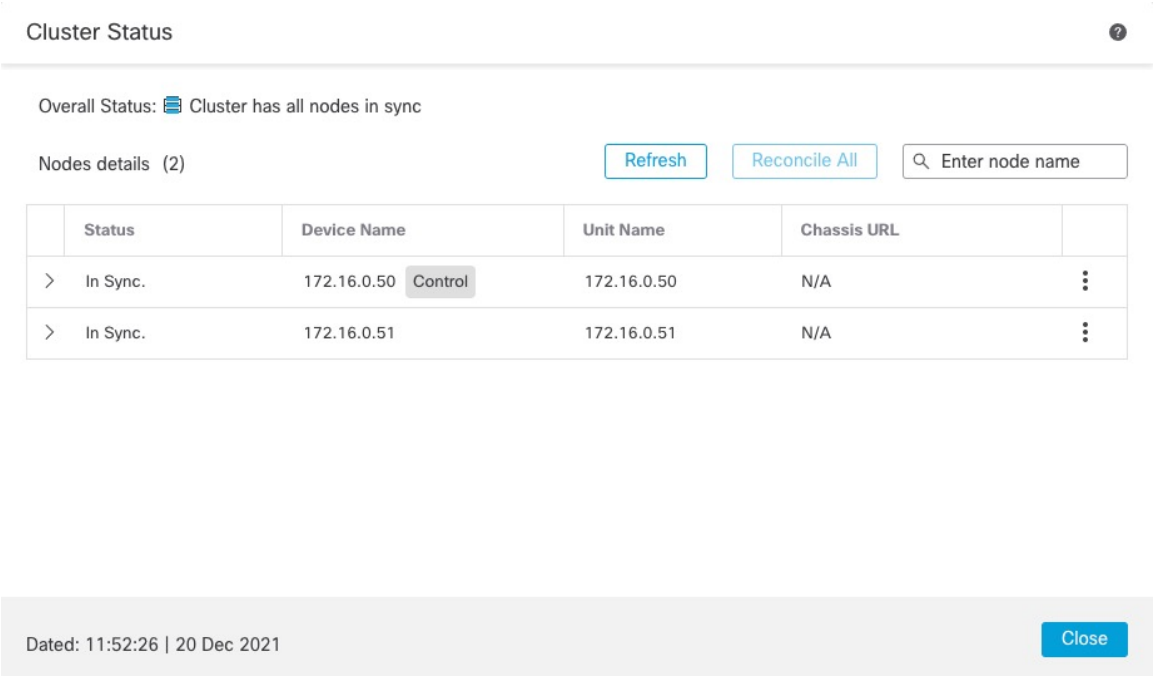
Perform a ping on the cluster control link

When a node joins the cluster, it checks MTU compatibility by sending a ping to the control node with a packet size matching the cluster control link MTU. If the ping fails, a notification is generated so you can fix the MTU mismatch on connecting switches and try again. This tool lets you manually ping all nodes that have already joined the cluster in case you are having cluster control link connectivity problems.

Procedure

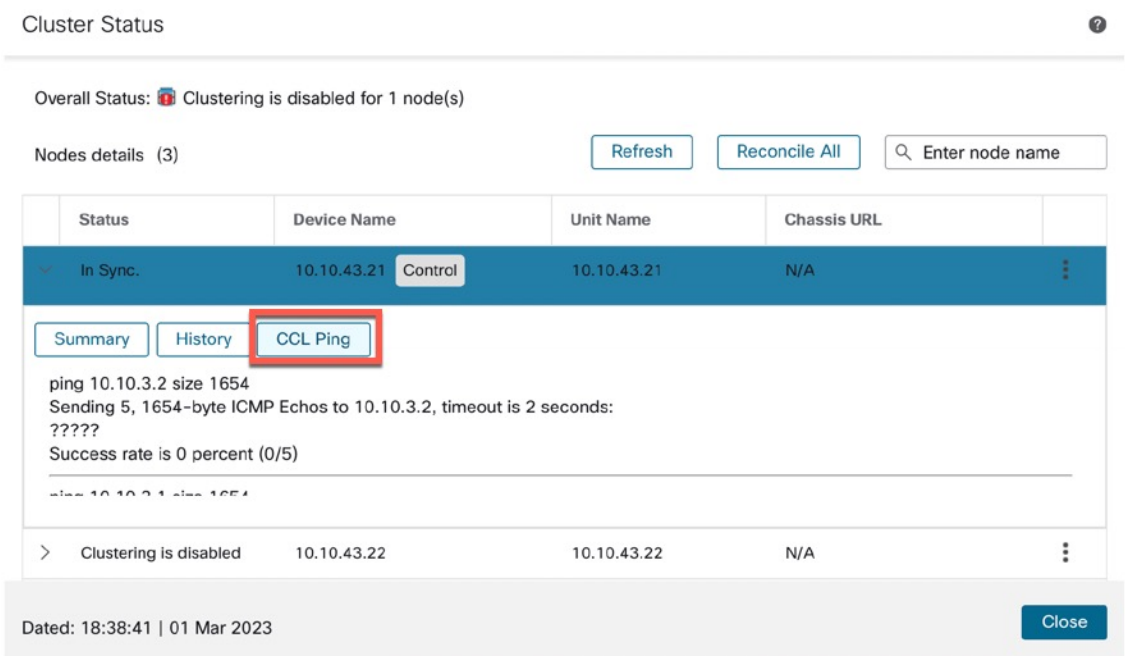
Step 1 Choose **Devices** > **Device Management**, click the **More** (⋮) icon next to the cluster, and choose **Cluster Live Status**.

Figure 14: Cluster status



Step 2 Expand one of the nodes, and click **CCL Ping**.

Figure 15: CCL ping



The node sends a ping on the cluster control link to every other node using a packet size that matches the maximum MTU.

Upgrading the Cluster

Perform the following steps to upgrade a Firewall Threat Defense Virtual cluster:

Before you begin

- Before you upgrade a cluster in the public cloud, copy the target version image to your cloud image repository and update the image ID in the cluster deployment template (we actually recommend replacing the existing template with a modified copy). This ensures that after the upgrade, new instances — for example, instances launched during cluster scaling — will use the correct version. If the marketplace does not have the image you need, such as when the cluster has been patched, create a custom image from a snapshot of a standalone Firewall Threat Defense Virtual instance running the correct version, with no instance-specific (day 0) configurations.

Procedure

Step 1 Upload the target image version to the cloud image storage.

Step 2 Update the cloud instance template of the cluster with the updated target image version.

a) Create a copy of the instance template with the target image version.

b) Attach the newly created template to cluster instance group.

Note
If the user wants to retain the old interface naming convention, use the "IfNamingConvention": "Old" key-value pair in the Day 0 configuration.

Step 3 Upload the target image version upgrade package to the Firewall Management Center.

Step 4 Perform readiness check on the cluster that you want to upgrade.

Step 5 After successful readiness check, initiate installation of upgrade package.

Step 6 The Firewall Management Center upgrades the cluster nodes one at a time.

Step 7 The Firewall Management Center displays a notification after successful upgrade of the cluster.

There is no change in the serial number and UUID of the instance after the upgrade.

Reference for Clustering

This section includes more information about how clustering operates.

Threat Defense Features and Clustering

Some Firewall Threat Defense features are not supported with clustering, and some are only supported on the control unit. Other features might have caveats for proper usage.

Unsupported Features and Clustering

These features cannot be configured with clustering enabled, and the commands will be rejected.

Note To view FlexConfig features that are also not supported with clustering, for example WCCP inspection, see the ASA general operations configuration guide. FlexConfig lets you configure many ASA features that are not present in the Firewall Management Center GUI.

- Remote access VPN (SSL VPN and IPsec VPN)
- Site-to-site VPN (Policy-based and route-based) is not supported in public clouds.
- DHCP client, server, and proxy. DHCP relay is supported.
- Virtual Tunnel Interfaces (VTIs)
- High Availability
- Integrated Routing and Bridging
- Firewall Management Center UCAPL/CC mode

Centralized features for clustering

These features are only supported on the control node, and are not scaled for the cluster.

Note Traffic for centralized features is forwarded from member nodes to the control node over the cluster control link.

If you use the rebalancing feature, traffic for centralized features may be rebalanced to non-control nodes before the traffic is classified as a centralized feature; if this occurs, the traffic is then sent back to the control node.

For centralized features, if the control node fails, all connections are dropped, and you have to re-establish the connections on the new control node.

Note To view FlexConfig features that are also centralized with clustering, for example RADIUS inspection, see the ASA general operations configuration guide. FlexConfig lets you configure many ASA features that are not present in the Firewall Management Center GUI.

- These application inspections are supported:
  - DCERPC and ESMTP
  - NetBIOS and PPTP
  - RSH
  - SQLNET and SUNRPC
  - TFTP
  - XDMCP
- Static route monitoring

- Site-to-site VPN
- IGMP multicast control plane protocol processing (data plane forwarding is distributed across the cluster)
- PIM multicast control plane protocol processing (data plane forwarding is distributed across the cluster)
- Dynamic routing (Spanned EtherChannel mode only)

Cisco Trustsec and Clustering

Only the control node learns security group tag (SGT) information. The control node then populates the SGT to data nodes, and data nodes can make a match decision for SGT based on the security policy.

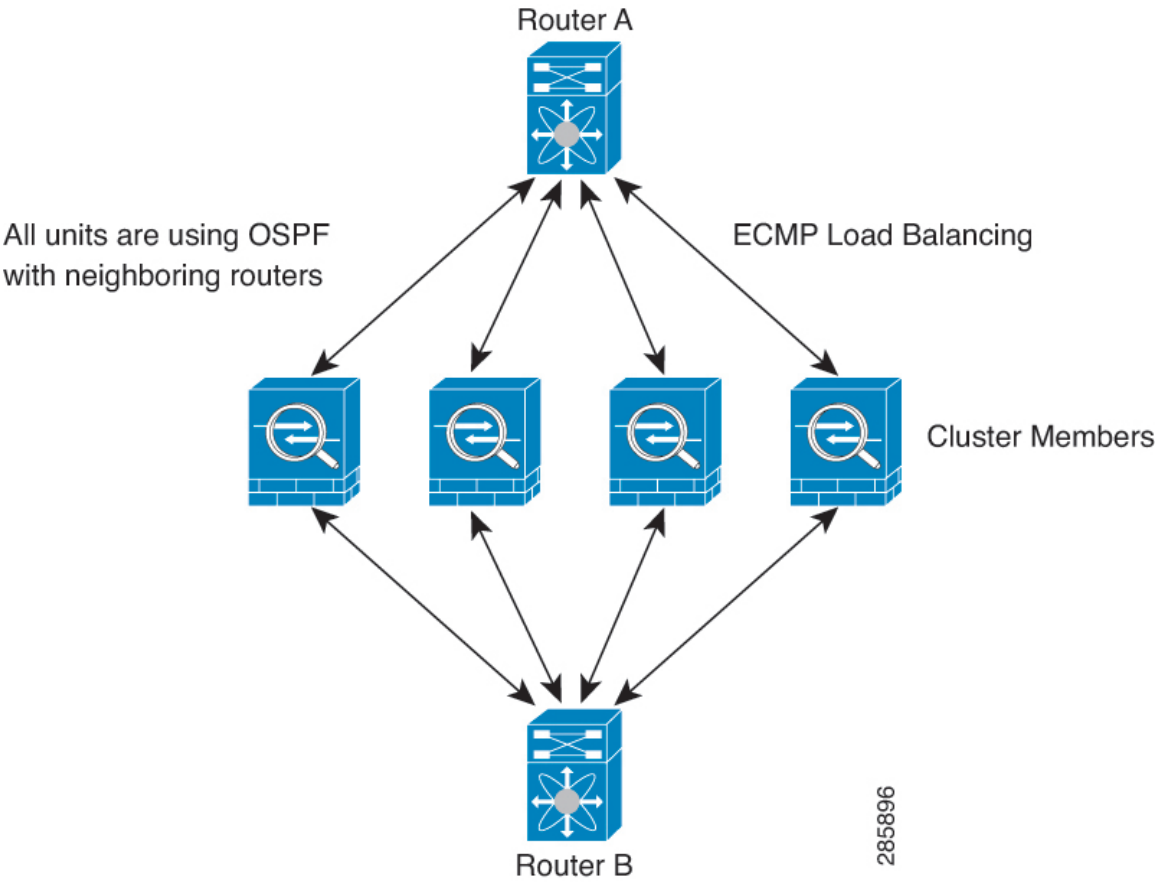
Connection Settings and Clustering

Connection limits are enforced cluster-wide. Each node has an estimate of the cluster-wide counter values based on broadcast messages. Due to efficiency considerations, the configured connection limit across the cluster might not be enforced exactly at the limit number. Each node may overestimate or underestimate the cluster-wide counter value at any given time. However, the information will get updated over time in a load-balanced cluster.

Dynamic Routing and Clustering

In Individual interface mode, each node runs the routing protocol as a standalone router, and routes are learned by each node independently.

Figure 16: Dynamic Routing in Individual Interface Mode



In the above diagram, Router A learns that there are 4 equal-cost paths to Router B, each through a node. ECMP is used to load balance traffic between the 4 paths. Each node picks a different router ID when talking to external routers.

You must configure a cluster pool for the router ID so that each node has a separate router ID.

FTP and Clustering

- If FTP data channel and control channel flows are owned by different cluster members, then the data channel owner will periodically send idle timeout updates to the control channel owner and update the idle timeout value. However, if the control flow owner is reloaded, and the control flow is re-hosted, the parent/child flow relationship will not longer be maintained; the control flow idle timeout will not be updated.

NAT and Clustering

For GCP, outbound traffic requires interface NAT. Outbound traffic with interface NAT is limited to 64k connections. For other NAT uses, see the following limitations.

NAT can affect the overall throughput of the cluster. Inbound and outbound NAT packets can be sent to different Firewall Threat Defenses in the cluster, because the load balancing algorithm relies on IP addresses and ports, and NAT causes inbound and outbound packets to have different IP addresses and/or ports. When a packet arrives at the Firewall Threat Defense that is not the NAT owner, it is forwarded over the cluster control link to the owner, causing large amounts of traffic on the cluster control link. Note that the receiving node does not create a forwarding flow to the owner, because the NAT owner may not end up creating a connection for the

packet depending on the results of security and policy checks. For GCP, outbound traffic requires interface NAT. Outbound traffic with interface NAT is limited to 64k connections.

If you still want to use NAT in clustering, then consider the following guidelines:

- No Proxy ARP—For Individual interfaces, a proxy ARP reply is never sent for mapped addresses. This prevents the adjacent router from maintaining a peer relationship with an ASA that may no longer be in the cluster. The upstream router needs a static route or PBR with Object Tracking for the mapped addresses that points to the Main cluster IP address.
- PAT with Port Block Allocation—See the following guidelines for this feature:
  - Maximum-per-host limit is not a cluster-wide limit, and is enforced on each node individually. Thus, in a 3-node cluster with the maximum-per-host limit configured as 1, if the traffic from a host is load-balanced across all 3 nodes, then it can get allocated 3 blocks with 1 in each node.
  - Port blocks created on the backup node from the backup pools are not accounted for when enforcing the maximum-per-host limit.
  - On-the-fly PAT rule modifications, where the PAT pool is modified with a completely new range of IP addresses, will result in xlate backup creation failures for the xlate backup requests that were still in transit while the new pool became effective. This behavior is not specific to the port block allocation feature, and is a transient PAT pool issue seen only in cluster deployments where the pool is distributed and traffic is load-balanced across the cluster nodes.
  - When operating in a cluster, you cannot simply change the block allocation size. The new size is effective only after you reload each device in the cluster. To avoid having to reload each device, we recommend that you delete all block allocation rules and clear all xlates related to those rules. You can then change the block size and recreate the block allocation rules.
- NAT pool address distribution for dynamic PAT—When you configure a PAT pool, the cluster divides each IP address in the pool into port blocks. By default, each block is 512 ports, but if you configure port block allocation rules, your block setting is used instead. These blocks are distributed evenly among the nodes in the cluster, so that each node has one or more blocks for each IP address in the PAT pool. Thus, you could have as few as one IP address in a PAT pool for a cluster, if that is sufficient for the number of PAT'ed connections you expect. Port blocks cover the 1024-65535 port range, unless you configure the option to include the reserved ports, 1-1023, on the PAT pool NAT rule.
- Same translation object in static and dynamic NAT rules—Do not use the same network object or object group as the translated address in both static NAT rules and dynamic NAT rules in a cluster. Similarly, avoid using network objects or object groups with overlapping address ranges across static and dynamic NAT rules. Use separate network objects or object groups for static translations and dynamic translations. Mixing these configurations can cause translation ownership inconsistencies in the cluster, which can result in traffic drops or intermittent connection failures.
- Reusing a PAT pool in multiple rules—To use the same PAT pool in multiple rules, you must be careful about the interface selection in the rules. You must either use specific interfaces in all rules, or "any" in all rules. You cannot mix specific interfaces and "any" across the rules, or the system might not be able to match return traffic to the right node in the cluster. Using unique PAT pools per rule is the most reliable option.
- No round-robin—Round-robin for a PAT pool is not supported with clustering.
- No extended PAT—Extended PAT is not supported with clustering.
- Dynamic NAT xlates managed by the control node—The control node maintains and replicates the xlate table to data nodes. When a data node receives a connection that requires dynamic NAT, and the xlate is not in the table, it requests the xlate from the control node. The data node owns the connection.
- Stale xlates—The xlate idle time on the connection owner does not get updated. Thus, the idle time might exceed the idle timeout. An idle timer value higher than the configured timeout with a refcnt of 0 is an indication of a stale xlate.

- No static PAT for the following inspections—
  - FTP
  - RSH
  - SQLNET
  - TFTP
  - XDMCP
  - SIP
- If you have an extremely large number of NAT rules, over ten thousand, you should enable the transactional commit model using the **asp rule-engine transactional-commit nat** command in the device CLI. Otherwise, the node might not be able to join the cluster.

SIP Inspection and Clustering

A control flow can be created on any node (due to load balancing); its child data flows must reside on the same node.

SNMP and Clustering

You should always use the Local address, and not the Main cluster IP address for SNMP polling. If the SNMP agent polls the Main cluster IP address, if a new control node is elected, the poll to the new control node will fail.

Syslog and Clustering

- Each node in the cluster generates its own syslog messages. You can configure logging so that each node uses either the same or a different device ID in the syslog message header field. For example, the hostname configuration is replicated and shared by all nodes in the cluster. If you configure logging to use the hostname as the device ID, syslog messages generated by all nodes look as if they come from a single node. If you configure logging to use the local-node name that is assigned in the cluster bootstrap configuration as the device ID, syslog messages look as if they come from different nodes.

Performance scaling factor

When you combine multiple units into a cluster, you can expect the total cluster performance to be approximately 80% of the maximum combined throughput.

Throughput calculation

For example, if your model can handle approximately 10 Gbps of traffic when running alone, then for a cluster of 8 units, the maximum combined throughput will be approximately 80% of 80 Gbps (8 units x 10 Gbps): 64 Gbps.

Control node election

Summary

The key components involved in the control node election process are:

- Cluster nodes: Communicate over the cluster control link and participate in the election process

- Election requests: Broadcast every 3 seconds by nodes when clustering is enabled
- Priority values: Set between 1 and 100 where 1 is the highest priority
- Control node: The elected node that coordinates cluster operations

Workflow

These stages describe how cluster nodes elect a control node:

1. When clustering is enabled for a node (or when it first starts up with clustering already enabled), the node broadcasts an election request every 3 seconds.
2. Any other nodes with a higher priority respond to the election request. The priority is set between 1 and 100, where 1 is the highest priority.
3. If after 45 seconds, a node does not receive a response from another node with a higher priority, then it becomes the control node.



Note If multiple nodes tie for the highest priority, the cluster node name and then the serial number is used to determine the control node.

4. If a node later joins the cluster with a higher priority, it does not automatically become the control node. The existing control node always remains as the control node unless it stops responding, at which point a new control node is elected.
5. In a "split brain" scenario when there are temporarily multiple control nodes, the node with highest priority retains the role while the other nodes return to data node roles.



Note You can manually force a node to become the control node. For centralized features, if you force a control node change, then all connections are dropped, and you have to re-establish the connections on the new control node.

High Availability within the Cluster

Clustering provides high availability by monitoring node and interface health and by replicating connection states between nodes.

Node health monitoring

Each node periodically sends a broadcast heartbeat packet over the cluster control link. If the control node does not receive any heartbeat packets or other packets from a data node within the configurable timeout period, then the control node removes the data node from the cluster. If the data nodes do not receive packets from the control node, then a new control node is elected from the remaining nodes.

Split brain scenario

If nodes cannot reach each other over the cluster control link because of a network failure and not because a node has actually failed, then the cluster may go into a "split brain" scenario where isolated data nodes will elect their own control nodes. For example, if a router fails between two cluster locations, then the original control node at location 1 will remove the location 2 data nodes from the cluster. Meanwhile, the nodes at location 2 will elect their own control node and form their own cluster. Note that asymmetric traffic may fail in this scenario. After the cluster control link is restored, then the control node that has the higher priority will keep the control node's role.

Interface Monitoring

Each node monitors the link status of all named hardware interfaces in use, and reports status changes to the control node.

All physical interfaces are monitored; only named interfaces can be monitored. You can optionally disable monitoring per interface.

A node is removed from the cluster if its monitored interfaces fail. The node is removed after 500 ms.

Status After Failure

If the control node fails, then another member of the cluster with the highest priority (lowest number) becomes the control node.

The Firewall Threat Defense automatically tries to rejoin the cluster, depending on the failure event.



Note When the Firewall Threat Defense becomes inactive and fails to automatically rejoin the cluster, all data interfaces are shut down; only the Management interface can send and receive traffic.

Rejoining the Cluster

After a cluster member is removed from the cluster, how it can rejoin the cluster depends on why it was removed:

- Failed cluster control link when initially joining—After you resolve the problem with the cluster control link, you must manually rejoin the cluster by re-enabling clustering.
- Failed cluster control link after joining the cluster—The Firewall Threat Defense automatically tries to rejoin every 5 minutes, indefinitely.
- Failed data interface—The Firewall Threat Defense automatically tries to rejoin at 5 minutes, then at 10 minutes, and finally at 20 minutes. If the join is not successful after 20 minutes, then the Firewall Threat Defense application disables clustering. After you resolve the problem with the data interface, you have to manually enable clustering.
- Failed node—If the node was removed from the cluster because of a node health check failure, then rejoining the cluster depends on the source of the failure. For example, a temporary power failure means the node will rejoin the cluster when it starts up again as long as the cluster control link is up. The Firewall Threat Defense application attempts to rejoin the cluster every 5 seconds.
- Internal error—Internal failures include: application sync timeout; inconsistent application statuses; and so on. After you resolve the problem, you must manually rejoin the cluster by re-enabling clustering.
- Failed configuration deployment—If you deploy a new configuration from Firewall Management Center, and the deployment fails on some cluster members but succeeds on others, then the nodes that failed are removed from the cluster. You must manually rejoin the cluster by re-enabling clustering. If the deployment fails on the control node, then the deployment is rolled back, and no members are removed. If the deployment fails on all data nodes, then the deployment is rolled back, and no members are removed.

Data Path Connection State Replication

Every connection has one owner and at least one backup owner in the cluster. The backup owner does not take over the connection in the event of a failure; instead, it stores TCP/UDP state information, so that the connection can be seamlessly transferred to a new owner in case of a failure. The backup owner is usually also the director.

Some traffic requires state information above the TCP or UDP layer. See the following table for clustering support or lack of support for this kind of traffic.

Table 8: Features Replicated Across the Cluster

Traffic	State Support	Notes
Up time	Yes	Keeps track of the system up time.
ARP Table	Yes	Transparent mode only.—
MAC address table	Yes	Transparent mode only.—
User Identity	Yes	—
IPv6 Neighbor database	Yes	—
Dynamic routing	Yes	—
SNMP Engine ID	**No**	—

How the Cluster Manages Connections

Connections can be load-balanced to multiple nodes of the cluster. Connection roles determine how connections are handled in both normal operation and in a high availability situation.

Connection Roles

See the following roles defined for each connection:

- Owner—Usually, the node that initially receives the connection. The owner maintains the TCP state and processes packets. A connection has only one owner. If the original owner fails, then when new nodes receive packets from the connection, the director chooses a new owner from those nodes.
- Backup owner—The node that stores TCP/UDP state information received from the owner, so that the connection can be seamlessly transferred to a new owner in case of a failure. The backup owner does not take over the connection in the event of a failure. If the owner becomes unavailable, then the first node to receive packets from the connection (based on load balancing) contacts the backup owner for the relevant state information so it can become the new owner.

  As long as the director (see below) is not the same node as the owner, then the director is also the backup owner. If the owner chooses itself as the director, then a separate backup owner is chosen.

  For clustering on the Firepower 9300, which can include up to 3 cluster nodes in one chassis, if the backup owner is on the same chassis as the owner, then an additional backup owner will be chosen from another chassis to protect flows from a chassis failure.
- Director—The node that handles owner lookup requests from forwarders. When the owner receives a new connection, it chooses a director based on a hash of the source/destination IP address and ports (see below for ICMP hash details), and sends a message to the director to register the new connection. If packets arrive at any node other than the owner, the node queries the director about which node is the owner so it can forward the packets. A connection has only one director. If a director fails, the owner chooses a new director.

  As long as the director is not the same node as the owner, then the director is also the backup owner (see above). If the owner chooses itself as the director, then a separate backup owner is chosen.

  ICMP/ICMPv6 hash details:

  - For Echo packets, the source port is the ICMP identifier, and the destination port is 0.
  - For Reply packets, the source port is 0, and the destination port is the ICMP identifier.

- For other packets, both source and destination ports are 0.

- Forwarder—A node that forwards packets to the owner. If a forwarder receives a packet for a connection it does not own, it queries the director for the owner, and then establishes a flow to the owner for any other packets it receives for this connection. The director can also be a forwarder. Note that if a forwarder receives the SYN-ACK packet, it can derive the owner directly from a SYN cookie in the packet, so it does not need to query the director. (If you disable TCP sequence randomization, the SYN cookie is not used; a query to the director is required.) For short-lived flows such as DNS and ICMP, instead of querying, the forwarder immediately sends the packet to the director, which then sends them to the owner. A connection can have multiple forwarders; the most efficient throughput is achieved by a good load-balancing method where there are no forwarders and all packets of a connection are received by the owner.

  Note We do not recommend disabling TCP sequence randomization when using clustering. There is a small chance that some TCP sessions won't be established, because the SYN/ACK packet might be dropped.

- Fragment Owner—For fragmented packets, cluster nodes that receive a fragment determine a fragment owner using a hash of the fragment source IP address, destination IP address, and the packet ID. All fragments are then forwarded to the fragment owner over the cluster control link. Fragments may be load-balanced to different cluster nodes, because only the first fragment includes the 5-tuple used in the switch load balance hash. Other fragments do not contain the source and destination ports and may be load-balanced to other cluster nodes. The fragment owner temporarily reassembles the packet so it can determine the director based on a hash of the source/destination IP address and ports. If it is a new connection, the fragment owner will register to be the connection owner. If it is an existing connection, the fragment owner forwards all fragments to the provided connection owner over the cluster control link. The connection owner will then reassemble all fragments.

Port Address Translation Connections

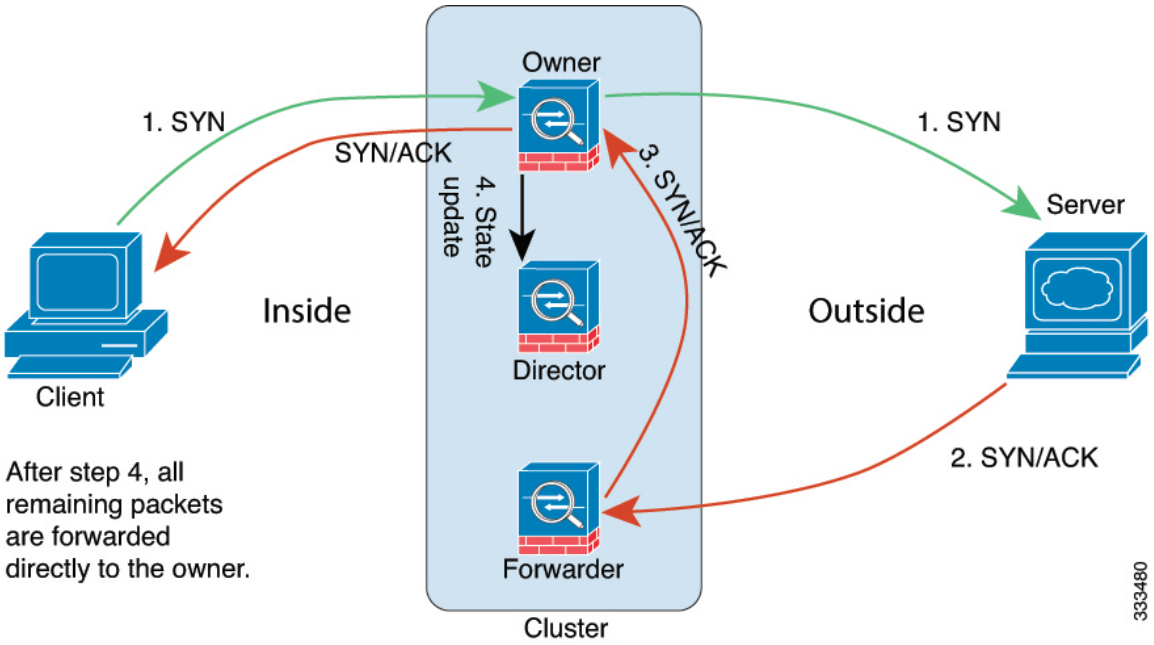
New connection ownership

When a new connection is directed to a node of the cluster via load balancing, that node owns both directions of the connection. If any connection packets arrive at a different node, they are forwarded to the owner node over the cluster control link. If a reverse flow arrives at a different node, it is redirected back to the original node.

Traffic redirection is not supported in this release. When a new connection is directed to a node of the cluster via load balancing, that node owns both directions of the connection. All the subsequent packets for the same connection should arrive the same node. If any connection packets arrive at a different node, they will be dropped. If a reverse flow arrives at a different node, it will be dropped as well. For centralized features, if the connections do not arrive on the control node, they will be dropped.

Sample Data Flow for TCP

The following example shows the establishment of a new connection.

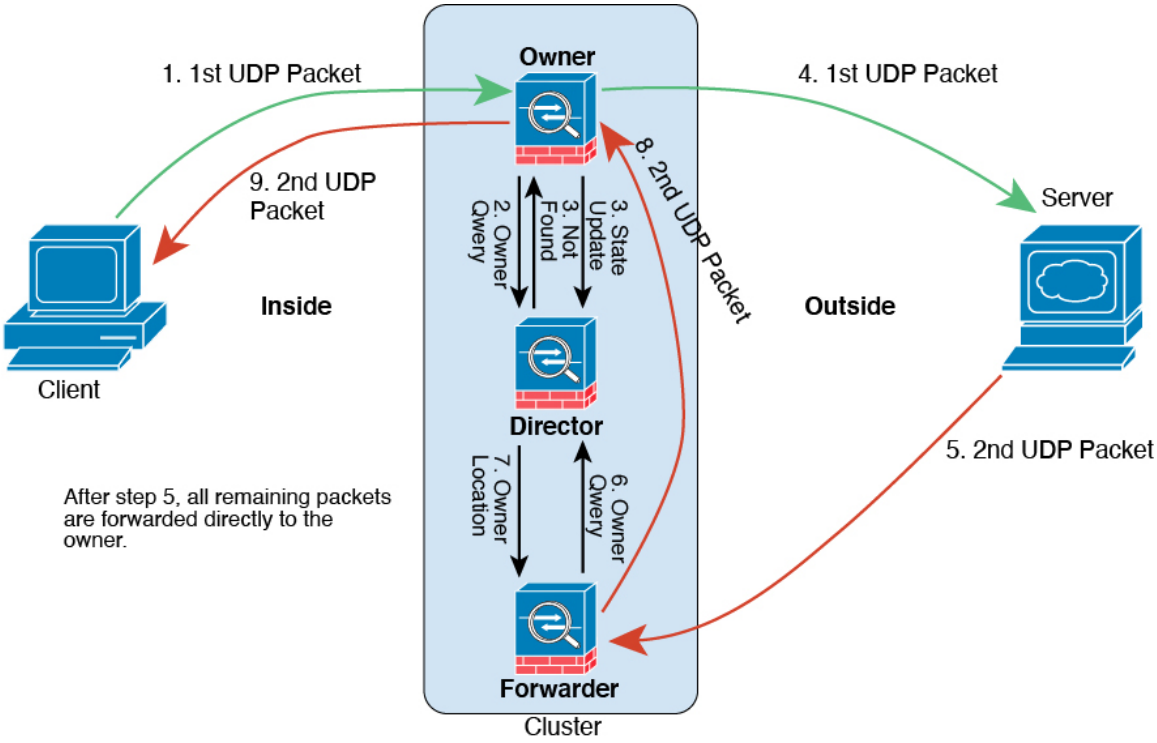


1. The SYN packet originates from the client and is delivered to one Firewall Threat Defense (based on the load balancing method), which becomes the owner. The owner creates a flow, encodes owner information into a SYN cookie, and forwards the packet to the server.
2. The SYN-ACK packet originates from the server and is delivered to a different Firewall Threat Defense (based on the load balancing method). This Firewall Threat Defense is the forwarder.
3. Because the forwarder does not own the connection, it decodes owner information from the SYN cookie, creates a forwarding flow to the owner, and forwards the SYN-ACK to the owner.
4. The owner sends a state update to the director, and forwards the SYN-ACK to the client.
5. The director receives the state update from the owner, creates a flow to the owner, and records the TCP state information as well as the owner. The director acts as the backup owner for the connection.
6. Any subsequent packets delivered to the forwarder will be forwarded to the owner.
7. If packets are delivered to any additional nodes, it will query the director for the owner and establish a flow.
8. Any state change for the flow results in a state update from the owner to the director.

Sample Data Flow for ICMP and UDP

The following example shows the establishment of a new connection.

1. ***Figure 17: ICMP and UDP Data Flow***



The first UDP packet originates from the client and is delivered to one Firewall Threat Defense (based on the load balancing method).

2. The node that received the first packet queries the director node that is chosen based on a hash of the source/destination IP address and ports.
3. The director finds no existing flow, creates a director flow and forwards the packet back to the previous node. In other words, the director has elected an owner for this flow.
4. The owner creates the flow, sends a state update to the director, and forwards the packet to the server.
5. The second UDP packet originates from the server and is delivered to the forwarder.
6. The forwarder queries the director for ownership information. For short-lived flows such as DNS, instead of querying, the forwarder immediately sends the packet to the director, which then sends it to the owner.
7. The director replies to the forwarder with ownership information.
8. The forwarder creates a forwarding flow to record owner information and forwards the packet to the owner.
9. The owner forwards the packet to the client.

History for Threat Defense Virtual Clustering on GCP

Table 9:

Feature	Min. Firewall Management Center	Min. Firewall Threat Defense	Details
MTU ping test from data node on node join	7.6.0	7.6.0	When a node joins the cluster, it checks MTU compatibility by sending a ping to the control node with a packet size twice the cluster control link MTU. Previously, only the control node sent a ping. If the ping fails, a notification is generated so you can fix the MTU mismatch on connecting switches and try again. Added/modified commands: **show cluster history**
Cluster control link ping tool.	7.4.1	Any	You can check to make sure all the cluster nodes can reach each other over the cluster control link by performing a ping. One major cause for the failure of a node to join the cluster is an incorrect cluster control link configuration; for example, the cluster control link MTU may be set higher than the connecting switch MTUs. New/modified screens: **Devices** > **Device Management** > **More** > **Cluster Live Status.**
Troubleshooting file generation and download available from Device and Cluster pages.	7.4.1	7.4.1	You can generate and download troubleshooting files for each device on the Device page and also for all cluster nodes on the Cluster page. For a cluster, you can download all files as a single compressed file. You can also include cluster logs for the cluster for cluster nodes. You can alternatively trigger file generation from the **Devices** > **Device Management** > **More** > **Troubleshoot Files** menu. New/modified screens: • **Devices** > **Device Management** > **Device** > **General** • **Devices** > **Device Management** > **Cluster** > **General**
View CLI output for a device or device cluster.	7.4.1	Any	You can view a set of pre-defined CLI outputs that can help you troubleshoot the device or cluster. You can also enter any **show** command and see the output. New/modified screens: **Devices** > **Device Management** > **Cluster** > **General**
Cluster health monitor settings	7.3.0	Any	You can now edit cluster health monitor settings. New/Modified screens: **Devices** > **Device Management** > **Cluster** > **Cluster Health Monitor Settings** **Note** If you previously configured these settings using FlexConfig, be sure to remove the FlexConfig configuration before you deploy. Otherwise the FlexConfig configuration will overwrite the management center configuration.

Feature	Min. Firewall Management Center	Min. Firewall Threat Defense	Details
Cluster health monitor dashboard	7.3.0	Any	You can now view cluster health on the cluster health monitor dashboard. New/Modified screens: **System** > **Health** > **Monitor**
Clustering for the Firewall Threat Defense Virtual on the Google Cloud Platform (GCP)	7.2.0	7.2.0	The Firewall Threat Defense Virtual supports Individual interface clustering for up to 16 nodes on GCP. New/Modified screens: • **Devices** > **Device Management** > **Add Device** • **Devices** > **Device Management** > **More** menu • **Devices** > **Device Management** > **Cluster**

© 2026 Cisco Systems, Inc. All rights reserved.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA 95134-1706
USA

Asia Pacific Headquarters
CiscoSystems(USA)Pte.Ltd.
Singapore

Europe Headquarters
CiscoSystemsInternationalBV
Amsterdam,TheNetherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.