Illumio Core®

Version 21.2

PCE Supercluster Deployment Guide

November 2022

60000-100-21.2

Legal Notices

The content in this documentation is provided for informational purposes only and is provided "as is," without warranty of any kind, expressed or implied of Illumio. The content in this documentation is subject to change without notice.

Product Version

PCE Version: 21.2 (LTS release)

For the complete list of Illumio Core components compatible with Core PCE, see the Illumio Support portal (log in required).

For information on Illumio software support for Standard and LTS releases, see Versions and Releases on the Illumio Support portal.

Resources

Legal information, see https://www.illumio.com/legal-information

Trademarks statements, see https://www.illumio.com/trademarks

Patent statements, see https://www.illumio.com/patents

License statements, see https://www.illumio.com/eula

Open source software utilized by the Illumio Core and their licenses, see Open Source Licensing Disclosures

Contact Information

To contact Illumio, go to https://www.illumio.com/contact-us

To contact the Illumio legal team, email us at legal@illumio.com

To contact the Illumio documentation team, email us at doc-feedback@illumio.com

Contents

Chapter 1

Overview of Supercluster Deployment

This chapter contains the following topics:

This section introduces concepts that you need to understand in order to achieve a successful PCE Supercluster deployment.

About This Supercluster Guide

The following sections provide useful information to help you get the most out of this guide.

How to Use This Guide

This guide includes several major sections:

- Overview to PCE Supercluster possible architectures and components
- General tasks required to deploy, operate, and use a PCE Supercluster: health monitoring, PCE web console and API access, back up and restore, and workload pairing considerations
- Basic theory of PCE Supercluster operations

Use this guide in conjunction with the *PCE Installation and Upgrade Guide* and *PCE Administration Guide*.

Before Reading This Guide

Before attempting the procedures in this guide, you should be familiar with the following technology:

- Your organization's security goals
- Illumio Core
- General computer system administration of Linux and Windows operating systems, including startup/shutdown, common processes or services
- Linux shell (bash) and Windows PowerShell
- TCP/IP networks, including protocols and well-known ports
- PKI certificates

Notational Conventions

- Newly introduced terminology is italicized. Example: *activation code* (also known as pairing key)
- Command-line examples are monospace. Example: `illumio-ven-ctl --activate`
- Arguments on command lines are monospace italics. Example: `illumio-ven-ctl --activate` *`activation_code`*
- In some examples, the output might be shown across several lines but is actually on one single line.
- Command input or output lines not essential to an example are sometimes omitted, as indicated by three periods in a row. Example:

```
...
some command or command output
...
```

PCE Supercluster Concepts

A Policy Compute Engine (PCE) Supercluster consists of a single administrative domain that spans two or more replicating PCEs. One PCE in the Supercluster is the Supercluster leader and the other PCEs are Supercluster members. A Supercluster deployment has only one leader. Any member can be manually promoted to be the leader.

The leader has a central PCE web console and REST API endpoint for configuring and provisioning security policy. The web interface on the leader also provides other

centralized management functions, including an aggregated Illumination map to visualize network traffic and policy coverage for all workloads. Members in the Supercluster mostly have a read-only PCE web console and REST API for viewing local data.

To illustrate how a PCE Supercluster works, consider this example three-tier application (web, processing, database) that is deployed across three datacenters in the US, Europe, and Asia. Each datacenter has its own PCE, and the US PCE is the leader. The policy for this application is designed to micro-segment the application in each datacenter while allowing the database tier to replicate across datacenters.

Workload Management

All PCEs in the Supercluster can manage workloads. You can deploy a leader without managed workloads to reduce the load on the leader and maintain performance for policy computation and other tasks.

Pairing profiles must always be created on the leader, from which they are replicated to all members. On the members, you can generate pairing keys and pairing scripts tied to the members themselves for activation and not the leader.

Pairing Workloads

Before workloads can be paired, a pairing profile must be created on the leader, which is then replicated to all other PCEs in the Supercluster. Workloads can be paired to a specific PCE FQDN or to the Supercluster FQDN. In the latter case, you must use a Global Server Load Balancing (GSLB) or DNS server that supports persistent routing of workloads to the nearest PCE based on geolocation.

When a workload is paired with a PCE, a managed workload object is created on the PCE and its labels are assigned based on the settings in the pairing profile. The PCE calculates policy and distributes firewall rules to the newly paired workload and other managed workloads so that these workloads can communicate with the newly paired workload. The PCE also replicates the information about the new workload to the other PCEs, which in turn re-compute and re-distribute firewall rules to their managed workloads that are allowed to communicate with the newly paired workload.

In this example, when you pair a new instance of the database in the US, the following events occurs:

1. The US PCE sends firewall rules to the US database workload.
2. The US PCE sends send new firewall rules to the US web and processing workloads because the policy allows these workloads to communicate.

3. The US PCE replicates information about the new US database workload to the PCEs in Europe and Asia.
4. The PCEs in Europe and Asia re-calculate policy and send new firewall rules to their database workloads because the policy allows these databases to communicate with the US database.

There might be a short time period when one of the database workloads has received rules allowing outbound traffic, but the other database workloads have not yet received their corresponding inbound rules to allow the connection. This condition can occur with a single PCE (for example, a non-Supercluster deployment) but can take slightly longer with a PCE Supercluster due to replication delays between PCEs.

Pairing with Specific Members

A pairing profile must always be created on the Supercluster leader. This pairing profile is propagated to all members. On a member, you can generate new pairing keys from the propagated profile. The pairing script generated from a pairing profile pairs the workload to the specific member.

Making Policy Modifications

Changes to your policy are made and provisioned on the leader using the PCE web console or the Illumio Core REST API, which in turn is replicated to all other PCEs in the Supercluster. Whenever a PCE receives updated policy, it re-computes policy for its own managed workloads and sends firewall rules to any other affected managed workloads.

Example: The original policy was written to allow the database workloads to communicate across datacenters using all ports. The organization has decided to tighten this policy and restrict it to just the port needed for database replication.

When the new policy is provisioned on the leader, the following actions occurs:

1. The US PCE recalculates policy and sends new firewall rules to its database workload.
2. The US PCE replicates the policy to the PCEs in Europe and Asia.
3. On receiving the new policy, each of these PCEs re-computes policy and sends new firewall rules to their database workloads.

Adapting to Environmental Changes

Changes to a workload's assigned labels, IP address changes, or when a workload goes offline, are handled similarly to pairing a new workload. The PCE managing the

workload detects the changes and re-calculates and re-distributes new firewall rules for its managed workloads. It also replicates information about the change to the other PCEs, and these PCEs re-calculate policy and send new firewall rules to any of their managed workloads that are affected by the change.

Security Policy Replication

Security policy provisioned on the leader is replicated to all other PCEs in the Supercluster. All Supercluster leader and members replicate copies of each workload's context, such as IP addresses, to all other PCEs in the Supercluster. This behavior ensures the Supercluster can dynamically adapt the policy to changes in the environment, even when the leader is down. Policy and workload replication is performed using standard database replication technology of the PCE databases. The replication is trigger-based and only the deltas are transmitted to minimize delays and make efficient use of bandwidth.

Each member PCE in the Supercluster computes and distributes the firewall rules to its managed workloads based on the replicated policy and workload information. This design leverages the full computing power of the Supercluster to minimize policy convergence times for organization-wide policy changes affecting large numbers of workloads. Distributed policy computation also allows each member PCE to continuously enforce the latest policy, even when the leader is unavailable.

Flow Data and Illumination

Each PCE processes the summarized flow data reported by its managed workloads and stores a computed view of the traffic in memory, just as if each were a standalone PCE. The leader periodically queries this data from each PCE to generate an aggregated Illumination map for the entire Supercluster. The raw summarized flow data is not sent to the leader, only the computed view of the flow data. When the raw flow data is needed, it can be streamed from each individual PCE in the Supercluster to one or more log collectors using either syslog or Fluentd.

Supercluster in PCE Web Console

This section describes how to use the PCE web console with a PCE Supercluster.

Overview of Supercluster in Web Console

Each PCE in the Supercluster processes the summarized traffic data reported by its managed workloads and stores a computed view of the traffic in memory, just as on a standalone PCE. The display of this data in the Illumination map, however, looks

different depending on whether you are logged into the leader or one of the members:

- The Illumination map on the leader shows an aggregated view of traffic data for the entire Supercluster. The leader periodically queries traffic data from each PCE to generate this map.
- The Illumination map on Supercluster members only shows data from workloads that have been paired with that member PCE.

The following Illumination features are not available in a Supercluster (leader or member):

- Clear traffic for one traffic link
- Increase the VEN reporting rate

These features are only available on a leader (and not available on a member):

- Add a rule from Illumination
- Policy Generator
- App Group configuration

VEN heartbeat and uptime data is not replicated in a Supercluster. It is available only on the leader itself and the individual members themselves:

Leader: Aggregated Illumination Data

The leader of the Supercluster shows a complete picture of all aggregated traffic from all PCEs in your Supercluster. Traffic data from members is refreshed periodically and then cached on the leader.

The refresh interval increases with the number of workloads that you pair with the Supercluster, with a minimum sync interval of 10 minutes and up to 24 hours, depending on how many workloads are paired with your Supercluster. You can force a sync of traffic data from members to the leader at any time, but the sync can take several minutes to complete.

Depending on your network speeds and possible latency, the Illumination map's traffic data can be delayed temporarily while the data is syncing.

Supercluster Illumination Sync with Members

In the lower right of the Illumination map on the leader, a small timer indicates when the Illumination map data was last refreshed.

Click the timer to launch a dialog from which you can refresh the Illumination map data so all traffic from all PCEs in the Supercluster is displayed.

Member: Local Illumination Data

The Illumination map on a member displays traffic information only from those workloads that have been paired with the member PCE. When viewing the Illumination map on a member, you can see a message indicating that you are viewing a local set of traffic data.

Web Console Filtering Problem

In the PCE web console on a Supercluster member, filtering the workload view with **Policy Sync: Active** displays the workloads for the entire Supercluster, instead of workloads for the member on which the report is run. This filter includes workloads marked as "Unavailable."

Workaround: In addition to **Policy Sync: Active**, use the PCE member FQDN filter to exclude all workloads not paired with the desired member. This filter combination is available:

Policy Sync: Active and PCE:Member PCE FQDN

REST API and Supercluster

The types of operations you can perform with the Illumio Core REST API are determined by the permissions granted to your user account by a PCE administrator.

Available Operations on Leader vs. Members

Regardless or your user's permissions, you can only perform read operations on a member, which means you can perform GET operations on members, but not any POST, PUT, or DELETE operations using the REST API.

On the leader, you can perform full CRUD (GET, POST, PUT, DELETE) operations when your user account has the permissions to do so. Other REST API requests that assist in PCE operations, such as checking a node's availability, or determining the Supercluster leader, are available on the leader and members.

REST Operation	Leader	Members
POST, PUT, DELETE	Yes	No
GET	Yes	Yes
DELETE blocked traffic	Yes	Yes

REST Operation	Leader	Members
Generate a workload support report	Yes	Yes
Asynchronous GET collections	Yes	Yes
GET product version	Yes	Yes
Check node availability	Yes	Yes
Determine Supercluster leader	Yes	Yes

During a Supercluster rolling upgrade, you can use the REST API on all PCEs except the one that is currently being upgraded. During a Supercluster simple upgrade, you cannot use the REST API until the upgrade has finished on all PCEs. For more information, see Upgrade Supercluster.

REST API Login Response

When you have deployed a PCE Supercluster and use the REST API to connect to a PCE in the Supercluster, the response indicates when the PCE is a member of the Supercluster.

For example, when you log into a PCE in a Supercluster:

```
GET https://my.pce.supercluster:443/api/v1/login
```

The response contains a JSON property named `pce_cluster_type` and has a value of either member or leader. For example, you see this response from a leader when you log in:

```
"pce_cluster_type": "leader"
```

Design Supercluster Deployment

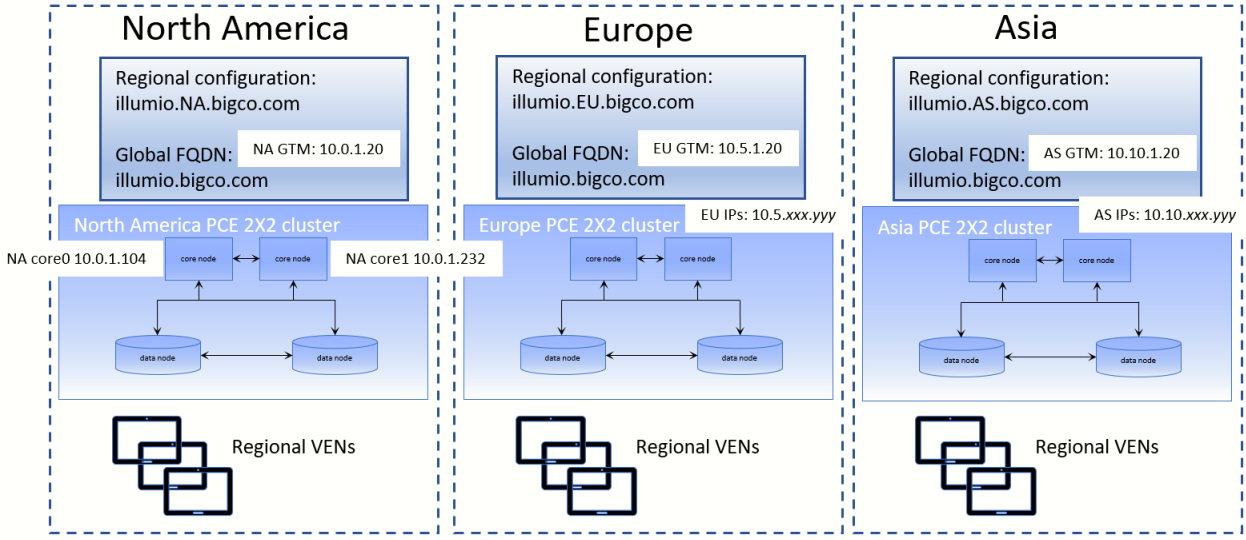
A PCE Supercluster consists of a single administrative domain that spans two or more replicating PCEs. One PCE in the Supercluster is the Supercluster *leader* and the other PCEs are Supercluster *members*. A Supercluster deployment has only one leader. Any member can be manually promoted to be the leader.

The leader has a central PCE web console and REST API endpoint for configuring and provisioning security policy. The PCE web console on the leader provides other centralized management functions, including an aggregated Illumination map to visualize network traffic and policy coverage for all workloads. Members in the

Supercluster mostly have a read-only PCE web console and REST API for viewing local data.

Supercluster Logical Architectures

The diagram below shows geographically distributed datacenters for a fictitious company called BigCo.com. This example presents only one of many possible Supercluster configurations.



Supercluster Design Considerations

When planning a PCE Supercluster deployment, consider these important factors:

- **How many total workloads does the PCE Supercluster need to support?** Scale constraints apply to both the number of managed workloads connected to each PCE and the total number of replicated workloads and other policy objects in the PCE's database.
- **How many managed workloads will be connected to each PCE?** Deployments should be sized such that each PCE is able to support the required number of locally-connected workloads and influx of workloads from a different PCE cross-PCE failover is configured.
- **What level of isolation is needed to support PCE outages (failures and maintenance)?** Each PCE in the Supercluster is independent and even a complete failure will not affect other PCEs. Deploying more PCEs in a Supercluster increases the number of failure domains.
- **What should happen to VENs when an extended PCE outage occurs?** By default, VENs continue to enforce the current policy when their PCE is unavailable. When

you need to provision policy changes during an extended PCE outage, you can use a GSLB to route orphaned VENs to another PCE in the Supercluster.

- **Which PCE in the PCE Supercluster will be the Leader?** The leader should be in a central location that can be readily accessed by PCE users and REST API clients. The leader should have reliable connectivity to all other PCEs in the Supercluster. Some organizations choose to deploy a leader with no managed workloads to reduce load on this PCE and optimize for REST API data loading.

High Availability and Disaster Recovery

A PCE Supercluster provides multiple levels of redundancy and failover for high availability (HA) and disaster recovery (DR).

Local Recovery

Each PCE in the Supercluster is a multi-node cluster (MNC) that can automatically survive a hardware or software failure affecting any one node. Each half of the PCE can be split across multiple LAN-connected buildings or availability zones, with 10 milliseconds latency between availability zones. Proper operation of Illumination and Explorer is assured when latency is 10ms or less. The PCE can survive a building failure, but manual action (issuing a PCE administrative command) might be necessary, depending on which building is lost.

When a complete failure of a PCE in the Supercluster occurs, its VENs continue to enforce the last known good policy until the PCE is restored or rebuilt from backup. When the leader becomes unavailable, each PCE operates autonomously and continues to distribute the latest provisioned policy to existing and newly paired workloads.

(Optional) Cross-PCE Failover and Recovery

During an extended outage of a PCE, workloads can optionally be failed over to any other PCE to continue to receive policy. Cross-PCE failover requires a GSLB or manual DNS. During failover, a workload's reported traffic flows are streamed via syslog and Fluentd but are not recorded by the PCE.

IMPORTANT:
Failover must be carefully managed to ensure the PCE does not exceed its capacity and become overloaded. For this reason, Illumio strongly recommends that failover be done manually and not automated.

Chapter 2

PCE Supercluster Deployment

This chapter contains the following topics:

This section describes how to set up a PCE Supercluster, a single administrative domain that encompasses two or more replicating PCEs. Before you start, be sure to understand the concepts explained in Overview of Supercluster Deployment.

PCE Supercluster Deployment Planning

This section describes requirements that you need to follow before deploying a PCE Supercluster.

Plan Supercluster FQDNs Carefully

Be sure to plan the fully qualified domain names (FQDNs) you want to use with your Supercluster PCEs. Be careful to define these names exactly how you want them before you deploy the Supercluster. Changing FQDNs after deploying a Supercluster is possible but time-consuming. The PCE FQDNs are set in the `pce_fqdn` parameter in `runtime_env.yml`.

For example, you might want to have identifying strings in the FQDNs that indicate the geographic location of the various members of the Supercluster, such as the following examples:

- `illumio-eu.bigco.com`: `eu` in the hostname indicates Europe.
- `illumio.na.bigco.com`: North America as a separate domain.

You can also configure a global FQDN for the Supercluster. The global FQDN is used by the VENs rather than individual PCE FQDNs. The global Supercluster FQDN is set in the `supercluster_fqdn` parameter in `runtime_env.yml`. When set, the PCE provides this FQDN instead of its own FQDN to VENs during pairing. This parameter must be set on *all* nodes in *each* PCE of the Supercluster. When you configure this option, each PCE server certificate must include the global FQDN in the SAN field. For example:

- `illumio-supercluster.bigco.com`

Number of Supercluster PCEs

A PCE Supercluster consists of a minimum of two and a maximum of eight PCEs for version 21.2.1 and greater (in version 21.2.0, the maximum is six PCEs). One of the PCEs is always the Supercluster leader, while the others are Supercluster members.

Capacity Planning for Supercluster PCEs

Use these guidelines and requirements to estimate host system capacity based on typical usage patterns.

The exact requirements vary based on a large number of factors, including, but not limited to:

- Whether you are using physical or virtual hardware
- Number of managed workloads
- Number of unmanaged workloads and other labeled objects, such as virtual services
- Policy complexity, which includes the following factors:
  - Number of rules in your rulesets
  - Number of labels, IP lists, and other objects in your rules
  - Number of IP ranges in your IP lists
  - Number of workloads affected by your rules
- Frequency at which your policies change
- Frequency at which workloads are added or deleted, or workload context changes, such as, change of IP address
- Volume of traffic flows per second reported to the PCE from all VENs

See the “Maximum Flow Capacity” table for information about maximum flow capacity of the PCE.

- Total number of unique flows reported to the PCE from all VENs

Recommended CPU, Memory, and Storage

The capacity planning tables in this section list the minimum recommended sizes for CPU, memory, and storage. This section provides two tables, one for physical hardware and one for virtual machines. Use these tables to plan your deployment.

NOTE:
Based on your actual usage and other factors, your capacity needs might be greater than the recommended sizes. For example, if you have installed additional software along with the PCE, such as application performance management (APM) software or an endpoint protection agent, this consumes additional system resources.

Data nodes are configured with a dedicated storage device for each database on the data nodes. This configuration accommodates growth in traffic data, which is used by Explorer. See Runtime Parameters for Traffic Datastore on Data Nodes.

For more than 150 IOPS, locally attached, spinning hard disk drives (HDD) are not sufficient. You will require either mixed-use Solid-State Disk (SSD) or Storage Area Network (SAN).

The PCE does not require that you set up swap memory, but it is permissible to enable swap memory. As long as the PCE nodes are provisioned with the recommended memory (RAM) as shown in the tables below, the use of swap memory should not cause any issues.

Physical Hardware

Use this table if you are installing the PCE on physical hardware. If you are using virtual machines, see the table Virtual Hardware.

MNC Type + Workloads/VENs	Cores/Clock Speed	RAM per Node	Storage Device Size and IOPS	
			Core Nodes	Data Nodes
SNC • 250 VENs[1] • 2500 work-	• 3 cores[2] • Intel® Xeon(R) CPU E5-2695	16GB	A single node including both	N/A

MNC Type + Work-loads/VENs	Cores/Clock Speed	RAM per Node	Storage Device Size and IOPS	
			Core Nodes	Data Nodes
loads	v4 at 2.10GHz or equivalent		core and data: • 1 x 50GB[4] • 100 IOPS per device[5]	
2x2 • 2,500 VENs[1] • 12,500 work-loads	• 4 cores per node[2] • Intel® Xeon(R) CPU E5-2695 v4 at 2.10GHz or equivalent	32GB	Minimum: • Disk: 50GB[3, 4] • 150 IOPS per device[5]	Minimum: • Disk 1: 250GB[4] • Disk 2: 250GB[4] • 600 IOPS per device[5]
2x2 • 10,000 VENs[1] • 50,000 work-loads	• 16 cores per node[2, 6] • Intel® Xeon(R) CPU E5-2695 v4 at 2.10GHz or equivalent	• Recommended: 128GB[6] • Minimum: 64GB	Minimum: • Disk: 50GB[3, 4] • 150 IOPS per device[5]	Minimum: • Disk 1: 1TB[4] • Disk 2: 1TB[4] • 1,800 IOPS per device[5]
4x2 • 25,000 VENs[1] • 125,000 work-loads	• 16 cores per node[2, 6] • Intel® Xeon(R) CPU E5-2695 v4 at 2.10GHz or equivalent.	128GB[6]	Minimum: • Disk: 50GB[3, 4] • 150 IOPS per device[5]	• Disk 1: 1TB[4] • Disk 2: 1TB[4] • 5,000 IOPS

MNC Type + Workloads/VENs	Cores/Clock Speed	RAM per Node	Storage Device Size and IOPS	
			Core Nodes	Data Nodes
				per device[5]

Footnotes:

[1] Number of VENs/workloads is the sum of both the number of managed VENs and the number of unmanaged workloads.

[2] CPUs:

- The recommended number of cores is based only on physical cores from allocated CPUs, irrespective of hyper-threading.

[3] This is the absolute minimum needed. In the future, other applications, support reports, or new features may require additional disk.

[4] Additional disk notes:

- Storage requirements for network traffic data can increase rapidly as the amount of network traffic increases.
- Network File Systems (NFS) is not supported for Illumio directories specified in runtime; for example, data_dir, persistent_data_dir, ephemeral_data_dir.

[5] Input/output operations per second (IOPS) are based on 8K random write operations. IOPS specified for an average of 300 flow summaries (80% unique `src_ip`, `dest_ip`, `dest_port`, `proto`) per workload every 10 minutes. Different traffic profiles might require higher IOPS.

[6] In the case of fresh installs or upgrades of a 2x2 for 10,000 VENs or a 4x2 for 25,000 VENs, if you deploy a system without sufficient cores, memory, or both, then the PCE will automatically reduce the object limits to 2,500 workloads. Object limit is the number of VENs (agents) per PCE. Adding more than 2,500 workloads will fail and an event is logged indicating that object limits have been exceeded. The workaround is to increase the number of cores, memory, or both to the recommended specifications and then increase the object limits manually. See PCE Default Object Limits in the *PCE Administration Guide*.

Virtual Hardware

Use this table if you are installing the PCE on virtual machines. If you are using physical hardware, see the table Physical Hardware.

MNC Type + Workloads/VENs	Virtual Cores/Clock Speed	RAM per Node	Storage Device Size and IOPS	
			Core Nodes	Data Nodes
SNC • 250 VENs[1] • 2500 workloads	• 6 virtual cores (vCPU)[2] • Intel® Xeon (R) CPU E5-2695 v4 at 2.10GHz or higher	16GB[7]	Minimum: • Disk: 50GB[3, 4] • 150 IOPS per device[5]	N/A
2x2 • 2,500 VENs[1] • 12,500 workloads	• 8 virtual cores (vCPU) per node[2] • Intel® Xeon (R) CPU E5-2695 v4 at 2.10GHz or higher	32GB[7]	Minimum: • Disk: 50GB[3, 4] • 150 IOPS per device[5]	Minimum: • Disk 1: 250GB • Disk 2: 250GB • 600 IOPS per device
2x2 • 10,000 VENs[1] • 50,000 workloads	• 32 virtual cores (vCPU) per node[2, 6] • Intel® Xeon (R) CPU E5-2695 v4 at 2.10GHz or higher	• Recommended: 128GB[6, 7] • Minimum: 64GB	Minimum: • Disk: 50GB[3, 4] • 150 IOPS per device[5]	Minimum: • Disk 1: 1TB[4] • Disk 2: 1TB[4] • 1,800 IOPS per device[5]
4x2 • 25,000 VENs[1]	• 32 virtual cores (vCPU) per	128GB[6, 7]	Minimum: • Disk:	• Disk 1: 1TB[4] • Disk 2:

MNC Type + Workloads/VENs	Virtual Cores/Clock Speed	RAM per Node	Storage Device Size and IOPS	
			Core Nodes	Data Nodes
• 125,000 workloads	node[2, 6] • Intel® Xeon (R) CPU E5-2695 v4 at 2.10GHz or higher		50GB[3, 4] • 150 IOPS per device[5]	1TB[4] • 5,000 IOPS per device[5]

Footnotes:

[1] Number of VENs/workloads is the sum of both the number of managed VENs and the number of unmanaged workloads.

[2] Full reservations for vCPU. No overcommit.

[3] This is the absolute minimum needed. In the future, other applications, support reports, or new features may require additional disk.

[4] Additional disk notes:

- Storage requirements for network traffic data can increase rapidly as the amount of network traffic increases.
- Network File Systems (NFS) is not supported for Illumio directories specified in runtime; for example, data_dir, persistent_data_dir, ephemeral_data_dir.

[5] Input/output operations per second (IOPS) are based on 8K random write operations. IOPS specified for an average of 300 flow summaries (80% unique `src_ip`, `dest_ip`, `dest_port`, `proto`) per workload every 10 minutes. Different traffic profiles might require higher IOPS.

[6] In the case of fresh installs or upgrades of a 2x2 for 10,000 VENs or a 4x2 for 25,000 VENs, if you deploy a system without sufficient cores, memory, or both, then the PCE will automatically reduce the object limits to 2,500 workloads. Object limit is the number of VENs (agents) per PCE. Adding more than 2,500 workloads will fail and an event is logged indicating that object limits have been exceeded. The workaround is to increase the number of cores, memory, or both to the recommended specifications and then increase the object limits manually. See PCE Default Object Limits in the *PCE Administration Guide*.

[7] Full reservations for vRAM. No overcommit.

Maximum Flow Capacity

The following table shows the maximum capacity of the PCE to accept flow data from all VENs.

MNC Type + Workloads/VENs	Flow Rate (flow-summaries/second)	Equivalent Flow Rate (flows/second)[2]
SNC • 250 VENs • 2500 workloads	100	1,030
2x2 • 2,500 VENs • 12,500 workloads	1,000	10,300
2x2 • 10,000 VENs • 50,000 workloads	4,100	422,000
4x2 • 25,000 VENs • 125,000 workloads	10,400[1]	1,070,000

Footnotes:

[1] The PCE might need to be tuned to achieve this rate. If you need to tune the PCE, please contact Illumio Support for assistance.

[2] Real-world observation shows that 102 flows result in one flow summary on average.

Storage Device Layout

You should create separate storage device partitions to reserve the amount of space specified below. These recommendations are based on PCE Capacity Planning.

The values given in these recommendation tables are guidelines based on testing in Illumio's labs. If you wish to deviate from these recommendations based on your own platform standards, please first contact your Illumio support representative for advice and approval.

PCE Single-Node Cluster for 250 VENs

Storage Device	Partition mount point	Size to Allocate	Node Types	Notes
Device 1, Partition A	/	8GB	Core, Data	The size of this partition assumes the system temporary files are stored in /tmp and core dump file size is set to zero. The PCE installation occupies approximately 500MB of this space.
Device 1, Partition B	`/var/log`	16GB	Core, Data	The size of this partition assumes that PCE application logs and system logs are both stored in `/var/log`. PCE application logs are stored in the `/var/log/illumio-pce` directory. The recommended size assumes average use by the OS with common packages installed and logging levels set to system defaults. Log size limits are configurable, so your system may require more or less log space. To find the potential maximum disk space required for your logs, use this command: `$ sudo -u ilo-pce illumio-pce-env logs --diag`
Device 1, Partition C	`/var/lib/illumio-pce`	Balance of Device 1	Core, Data	The size of this partition assumes that Core nodes use local storage for application code in /var/lib/illumio-pce, and also assumes that PCE support report files, and other temporary (ephemeral) files, etc., are stored in /var/lib/illumio-pce/tmp.

PCE 2x2 Multi-Node Cluster for 2,500 VENs

Storage Device	Partition mount point	Size to Allocate	Node Types	Notes
Device 1, Partition A	/	16GB	Core, Data	The size of this partition assumes the system temporary files are stored in /tmp

Storage Device	Partition mount point	Size to Allocate	Node Types	Notes
				and core dump file size is set to zero.
Device 1, Partition B	`/var/log`	32GB	Core, Data	The size of this partition assumes that PCE application logs and system logs are both stored in `/var/log`. PCE application logs are stored in the `/var/log/illumio-pce` directory.
Device 1, Partition C	`/var/lib/illumio-pce`	Balance of Device 1	Core, Data	The size of this partition assumes that Core nodes use local storage for application code in /var/lib/illumio-pce, and also assumes that PCE support report files, and other temporary (ephemeral) files, etc. are stored in /var/lib/illumio-pce/tmp.
Device 2, Single partition. Applicable in a two-storage-device configuration	`/var/lib/illumio-pce/data/Explorer`	All of Device 2 (250GB)	Data	For network traffic data in a two-storage-device configuration for the data nodes, it should be a separate device that is mounted on this directory. Set the `runtime_emv.yml` to `data_dir: /var/lib/illumio-pce/data/Explorer`, which will automatically create a sub-directory called `/var/lib/illumio-pce/data/Explorer/traffic_datastore` The partition mount point and the runtime setting must match. If you customize the mount point, make sure that

Storage Device	Partition mount point	Size to Allocate	Node Types	Notes
				you also change the runtime setting accordingly.

PCE 2x2 Multi-Node Cluster for 10,000 VENs and

PCE 4x2 Multi-Node Cluster for 25,000 VENs

Storage Device	Partition mount point	Size to Allocate	Node Types	Notes
Device 1, Partition A	/	16GB	Core, Data	The size of this partition assumes the system temporary files are stored in /tmp and core dump file size is set to zero.
Device 1, Partition B	`/var/log`	32GB	Core, Data	The size of this partition assumes that PCE application logs and system logs are both stored in `/var/log`. PCE application logs are stored in the `/var/log/illumio-pce` directory.
Device 1, Partition C	`/var/lib/illumio-pce`	Balance of Device 1	Core, Data	The size of this partition assumes that Core nodes use local storage for application code in /var/lib/illumio-pce, and also assumes that PCE support report files, and other temporary (ephemeral) files, etc. are stored in /var/lib/illumio-pce/tmp.
Device 2, Single Partition Applicable in a two-storage-device configuration	`/var/lib/illumio-pce/data/traffic`	All of Device 2 (1TB)	Data	For network traffic data in a two-storage-device configuration for the data nodes, it should be a separate device that is mounted on this directory. In `runtime_env.yml`, set the

Storage Device	Partition mount point	Size to Allocate	Node Types	Notes
				`traffic_datastore : data_dir` parameter to match the value of the partition mount point (see previous column) as follows: `traffic_datastore: data_dir: /var/lib/illumio-pce/data/traffic`. The partition mount point and the runtime setting must match. If you customize the mount point, make sure that you also change the runtime setting accordingly.

Runtime Parameters for Two-Storage-Device Configuration

In the two-storage-device configuration, to accommodate growth in the traffic data store, set the following parameters in `runtime_env.yml`:

NOTE:
When you are deploying the two-storage-device configuration, you must set these parameters.

```
traffic_datastore:
    data_dir: path_to_second_disk
    max_disk_usage_gb: Set this parameter according to the table below.
    partition_fraction: Set this parameter according to the table below.
    time_bucket_type: Set this parameter according to the table below.
```

The recommended values for the above parameters based on PCE node cluster type (2x2 or 4x2) and the estimated number of workloads (VENs) are as follows:

Setting	2x2 | 2,500 VENs	2x2 | 10,000 VENs	4x2 | 25,000 VENs	Note
`traffic_data-store:max_disk_usage_`	100 GB	400 GB	400 GB	This size reflects only part of the required total size, as detailed in

Setting	2x2 | 2,500 VENs	2x2 | 10,000 VENs	4x2 | 25,000 VENs	Note
`gb`				PCE Capacity Planning in the *PCE Installation and Upgrade Guide*.
`traffic_data-store:partition_frac-tion`	0.5	0.5	0.5	
`traffic_data-store:time_bucket_type`	Day	Day	Day	

Network Traffic Between PCEs

PCEs in the Supercluster communicate via the following ports. Any network firewalls between the PCEs must be configured to allow this traffic.

Ports	Sources	Destinations
The default TCP 8443 or the management port configured for the PCE Web Console and REST API in `runtime_env.yml`. This port must be the same on all PCEs in the Supercluster.	Core nodes of leader PCE	PCE FQDN of all member PCEs
TCP 5432	All nodes of all PCEs	IP addresses of all other PCE data nodes
TCP 5532	Core nodes of leader PCE	IP addresses of all other PCE data nodes
TCP 8302	All nodes of all PCEs	PCE FQDN of all other PCEs and IP address of all nodes of all other PCEs
UDP 8302	All nodes of all PCEs	IP address of all nodes of all other PCEs
TCP 8300	All nodes of all PCEs	IP address of all nodes of all other

Ports	Sources	Destinations
		PCEs

Load Balancers

Similar to a single PCE, all PCEs in the Supercluster must be front-ended with a load balancer (DNS or L4) to distribute requests across the PCEs' core nodes.

GSLB or a manual DNS update can be used to fail over VENs to a different PCE. See GSLB Requirements and High Availability and Disaster Recovery.

Traffic Load Balancer Configuration

When you use L4 load balancers in front of the PCEs, the load balancers should already be configured to forward inbound connections on the default TCP 8443 or the management port configured for the PCE web console and REST API in `runtime_env.yml` and 8444 to an available, healthy core node.

In a Supercluster, the L4 load balancer must also be configured to forward additional inbound TCP 8302 connections originating from the other PCEs to an available, healthy core node.

GSLB Requirements

Workloads can be paired to a specific PCE, or you can optionally use a GSLB to route workloads to the required PCE in your Supercluster.

When you are using a GSLB to route workloads, consider the following general guidelines.

For normal operations:

- When all PCEs are available, workloads should be routed to the nearest PCE based on proximity and geolocation.
- GSLB persistence (also known as "stickiness") must be enabled so workloads are always routed to the same PCE that they are paired with (non-failure case). Balancing workloads across multiple PCEs is not supported.

For failover:

- Recommended: A dedicated failover PCE joined to the Supercluster that has no other VENs.
- Failover to any other PCE in the Supercluster. In this case, take care to prevent overloading the PCE beyond its rated capacity and to avoid cascading failures.

One strategy is to configure a “buddy” PCE for each PCE that the GSLB uses for failover.

- Workload failover time depends on the DNS time-to-live (TTL) configured in the GSLB.
- Illumio strongly recommends that you do not automate workload failover using GSLB and instead initiate it manually.

Configure SAML IdP for User Login

After installation, you can configure the PCE to rely on an external, third-party SAML identity provider system IIdP). See "Single Sign-On Configuration" in the *PCE Administration Guide*. The guide provides set up instructions for a wide variety of IdPs.

For the PCE Supercluster, you configure the details in the leader PCE web console exactly as you do for the standalone PCE, with one exception: you are presented an intermediate page that lists all the PCEs in the Supercluster, including the leader and all members. Follow the same processes detailed in the *PCE Administration Guide* to configure all the Supercluster PCEs, both leader and members.

Certificate Requirements

PCE-to-PCE communication is done over TLS v1.2. The root CA certificate that signed each PCEs certificate must be in the root CA bundle on all other PCEs in the Supercluster.

Object Limits and Supercluster

The PCE enforces certain soft and hard limits to restrict the total number of system objects you can create. These limits are based on tested performance and capacity limits of the PCE. Most PCE object limits apply to the entire Supercluster. The limits are enforced by the leader when objects are created.

The object limit for number of VENs per PCE (`active_agents_per_pce`) is not cluster-wide and applies to each individual PCE. When the VENs per PCE limit is reached, no more VENs can be paired to that PCE. This limit is enforced when you move VENs from one PCE to another via the REST API.

An exception is made when VENs are failed over by the system itself from one PCE to a different PCE in the cluster. The VENs that failover do not count towards the limit, allowing you to temporarily exceed the limit of VENs per PCE when an extended outage to a PCE in the Supercluster occurs.

Changes to the object limit for number of VENs per PCE (`active_agents_per_pce`) made on the Supercluster leader are propagated to the members within 30 minutes.

For more information on object limits and how to view your current object limit usage, see the *PCE Administration Guide* and the `illumio-pce-ctl obj-limits list` command.

RBAC Permissions: Leader or Member

In general, when you are using the Illumio PCE web console or the Illumio REST API, the types of operations you can perform depend on your PCE role-based access control (RBAC) permissions and whether you have logged into the leader or a member, as shown in the table below.

User Role	Operations	Leader	Members
Any Role	View objects	Yes	Yes
Global Administrator & User Manager (Organization Owner)	Add, delete users Add, modify, delete, and provision system objects and rulesets (includes creating a pairing script).	Yes	No
Global Administrator	Add, modify, delete, and provision system objects and rulesets (includes creating a pairing script)	Yes	No
Global read only	View all objects	Yes	Yes
Global Policy Object Provisioner	Provision system objects	Yes	No
Ruleset Manager	Create, update, and delete rulesets within defined scopes.	Yes	No
Ruleset Provisioner	Provision rulesets within defined scopes.	Yes	No

Configure PCE Internal Syslog on Leader

You can configure the PCE's internal syslog service in the PCE web console on the Supercluster leader, for both the leader and the member PCEs. The internal syslog cannot be configured on a member PCE.

NOTE:
When a standalone PCE is installed, a local destination for the PCE internal syslog is created for recording events. When the PCE is joined as member of the Supercluster, this local destination is removed.

After joining a member, you have to log into the Supercluster leader and configure the internal syslog for each member individually.

When the events occurring prior to joining a PCE as a member are important to preserve, back up the PCE before you join it to the Supercluster.

See the *PCE Installation and Upgrade Guide* for information about the PCE internal syslog.

PCE Control Interface and Commands

The Illumio PCE control interface `illumio-pce-ctl` is a command-line tool for performing key tasks for operating your PCE cluster, such as starting and stopping nodes, setting cluster runlevels, and checking the cluster status.

IMPORTANT:
In this guide, all command-line examples based on an RPM installation. When you install the PCE using the tarball, you must modify the commands based on your PCE user account and the directory where you installed the software.

The PCE includes other command-line utilities used to set up and operate your PCE:

- `illumio-pce-env`: Verify and collect information about the PCE runtime environment.
- `illumio-pce-db-management`: Manage the PCE database.
- `supercluster-sub-command`: Manage specific Supercluster operations.

The PCE control interface can only be executed by the PCE runtime user (`ilo-pce`), which is created during the PCE RPM installation.

Control Command Access with `/usr/bin`

For easier command execution, PCE installation creates softlinks in `/usr/bin` by default for the Illumio PCE control commands. The `/usr/bin` directory is usually included by default in the `PATH` environment variable in most Linux systems. When your `PATH` does

not include `/usr/bin`, add it to your `PATH` with the following command. You might want to add this command to your login files (`$HOME/.bashrc` or `$HOME/.cshrc`).

```
export PATH=$PATH:/usr/bin
```

Syntax of `illumio-pce-ctl`

To make it simpler to run the PCE command-line tools, you can run the following Linux softlink commands or add them to your `PATH` environment variable.

```
$ cd /usr/bin
$ sudo ln -s /opt/illumio-pce/illumio-pce-ctl ./illumio-pce-ctl
$ sudo ln -s /opt/illumio-pce/illumio-pce-db-management ./illumio-pce-db-
management
$ sudo ln -s /opt/illumio-pce/illumio-pce-env ./illumio-pce-env
```

After these commands are executed, you can run the PCE command-line tools using the following syntax:

```
$ sudo -u ilo-pce illumio-pce-ctl sub-command --option
```

Where:

`sub-command` is an argument displayed by `illumio-pce-ctl --help`.

Deploy a PCE Supercluster

You can deploy the Illumio Supercluster in several ways:

- **New:** You have never deployed a PCE and want to deploy a new Supercluster. See Deploy New Supercluster.
- **Expand:** You have already deployed a standalone PCE and want to expand it to a Supercluster. See Expand Standalone PCE to Supercluster.
- **Join:** You already have more than one standalone PCE and you want to join them together into a Supercluster. Contact your Illumio Customer Support for assistance.

Deploy New Supercluster

Deploying a new PCE Supercluster follows this general workflow:

1. Install the leader PCE as a standalone PCE.
2. Install and configure each member PCE as a standalone PCE.
3. Initialize the Supercluster leader.
4. Join members to the Supercluster.
5. Bring the leader and members to a fully operational state.
6. Verify that the Supercluster is ready for use.

NOTE:
The sequence of events for deploying a Supercluster is not bound by any time requirements; for example, there is no time limit between initializing a Supercluster leader and joining individual members.

Before You Begin: Runtime Configuration

Before you deploy your PCE Supercluster, be aware of the following `runtime_env.yml` configurations:

- The value of the parameter `service_discovery_encryption_key` in the `runtime_env.yml` file must be exactly the same on all nodes on all PCEs in your Supercluster.
- You do not need to configure the public IP addresses of other PCEs under the `cluster_public_ips` parameter. Supercluster PCEs automatically exchange their configured public IP addresses with each other, which get programmed by the VEN to allow workloads to migrate between PCEs.

Optional

Depending on your deployment environment, you might need to make the following changes to the `runtime_env.yml` file on each PCE in the Supercluster.

When the nodes of each PCE use multiple IP addresses or they use IP addresses other than the one advertised on the node for communication with other PCEs, such as having a NAT between the PCEs in your Supercluster, configure this optional parameter:

- `supercluster.node_public_ip`: The public IP address of this node is advertised to other PCEs in your Supercluster deployment. This IP address must be reachable from all other Supercluster PCEs that you want to join. This parameter must be set on *all* nodes in *each* PCE. When your PCE is deployed in a public cloud, such as AWS, this must be a public IP address.

When you configure your GSLB for routing VENs to the appropriate PCE, configure this optional parameter on each node in a PCE:

- `supercluster.fqdn`: The PCE responds to this FQDN, instead of its own canonical FQDN to VENs during pairing. This parameter must be set on *all* nodes in *each* PCE of the Supercluster.

For example:

```
supercluster:
   node_public_ip: 192.168.33.10
   fqdn: global-pce.mycompany.com
```

Install Leader

The first step to deploy a new Supercluster is to install and configure the leader PCE, just as you would install a standalone PCE.

For detailed information about installing a PCE, see the *PCE Installation and Upgrade Guide*.

Install Members

Install each member of your Supercluster by following the exact same procedures you use installing a standalone PCE, except *do not* create a domain during deployment.

For information about installing a PCE, see the *PCE Installation and Upgrade Guide*.

Initialize Supercluster Leader

After the leader has been installed, configured, and verified, you initialize the leader.

NOTE:
You must initialize the leader *before* you start joining any members.

1. On *any node*, bring all nodes to runlevel 2:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
   ```

   Setting the run level might take some time to complete.

2. Check the progress with `illumio-pce-ctl cluster-status -w` to see when the status is Running:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
   ```

The nodes must be at runlevel 2 before you run the next command. When all of the nodes have reached runlevel 2, you see the following output:

```
Illumio Runtime System                              RUNNING [2] 34.28s
```

3. On *any node*, initialize the leader:

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-init-leader
```

Join Each Member to Supercluster

IMPORTANT:
You must join only one member one at a time, and complete all steps before joining the next member. Ensure that each member is at runlevel 2 before joining.

In this procedure, you join the new member to the Supercluster.

All nodes must start at runlevel 2. The nodes should already be at runlevel 2 from the previous procedure.

1. If necessary, on *any node*, bring all nodes to runlevel 2:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
```

2. On *any node*, run the following command while you wait for all nodes to reach runlevel 2:

```
$ sudo -u ilo-pce illumio-pce-ctl status --wait
```

3. On *any core node* or the *data0 node of the member cluster*, join the member to the Supercluster (identified by the leader's FQDN):

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-join leader_pce_fqdn
```

While this command runs, the PCE temporarily sets the runlevel to 1. When the command is interrupted, you might unexpectedly see runlevel 1 .

IMPORTANT:
Running this command can take an hour or more depending on the number of PCEs in your Supercluster and size of the PCE database. When this command fails due to network latency, do not proceed until you can run the command again and it executes successfully.

4. Repeat step 3 for all members you want to join to the Supercluster.
5. On *all PCEs*, restart the PCEs in the Supercluster:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-restart
```

6. On *all PCEs*, bring the PCEs to runlevel 5:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
```

Verify Supercluster Readiness

Before you begin using your Supercluster, verify that the leader and members are all joined and all PCEs in the Supercluster have a good health status.

NOTE:
It can take up to 10 minutes for all PCEs in your Supercluster to achieve full healthy status.

To verify that your Supercluster is ready to use:

1. Log into the leader.
2. On *any core node*, show Supercluster membership:

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-members
```

   The output should show all PCEs in your Supercluster.

3. Log into the PCE web console of the leader.
4. Click the Health status icon at the top of the PCE web console. You should see all PCEs in your Supercluster with **Normal** health status.

Expand Standalone PCE to Supercluster

To expand your existing standalone PCE to a Supercluster, the steps are similar to the steps for installing a new Supercluster, with additional checks of the standalone PCE before the expansion.

The general workflow for expanding an existing PCE deployment into a Supercluster follows these steps:

1. Change the `pce_fqdn` on your standalone PCE; then log into the standalone PCE's web console to verify that the standalone PCE is healthy and is working correctly. See Change FQDN and Verify PCE Health for information.
2. Verify network connectivity to the database nodes. See Network Connectivity from PCEs to Data Nodes for information.
3. Initialize your existing PCE as the Supercluster leader. See Initialize Supercluster Leader for information.
4. Install and configure the new PCEs that will become members of the new Supercluster. See the *PCE Installation and Upgrade Guide* for information.
5. Join members to the Supercluster. See Join Each Member to Supercluster for information.

Illumio recommends that you perform each these operations during different change windows.

After your Supercluster is operational, you can reassign workloads connected to the leader to a different PCE in the Supercluster.

Change FQDN and Verify PCE Health

See How to Rename the PCE in the Illumio Knowledge Base for information. (Login required)

WARNING:
Verify standalone PCE health

After changing the `pce_fqdn` parameter and before preceding with the expansion, you must log into the standalone PCE's web console to verify that the standalone PCE is healthy and is working correctly.

Network Connectivity from PCEs to Data Nodes

Before expansion of a Supercluster, ensure that every data node in the standalone cluster can connect to the database nodes via the Supercluster FQDN.

To verify the connections, use `telnet` or the `nc` (netcat) utility, which is part of the NMAP set of tools. If not already installed, install NMAP with the following command:

```
# yum install nmap
```

IMPORTANT:
Required runlevels

Be sure that the PCEs are set to the following runlevels before checking connectivity:

- On the PCE from which you run the check: Runlevel 2
- On the PCEs in other regions that you are checking: Runlevel 2 or higher

For example, you have three regions. With the following `nc` commands on data0 and data1 in each region, test the connection to the other regions by connecting to port 5432 for the other regions' data0 and data1 nodes.

- From region 1: Set the PCE from which you are testing to runlevel 2:

```
nc -zv region2_data0_ip 5432
nc -zv region3_data0_ip 5432
```

- From region 2: Set the PCE from which you are testing to runlevel 2:

```
nc -zv region1_data0_ip 5432
nc -zv region3_data0_ip 5432
```

- From region 3: Set the PCE from which you are testing to runlevel 2:

```
nc -zv region1_data0_ip 5432
nc -zv region2_data0_ip 5432
```

Migrate to New Supercluster

When you need to migrate your existing Supercluster to a new set of machines, follow these general steps:

1. On *all nodes*, pre-configure the IP addresses of the new Supercluster in the `runtime_env.yml` file. See Pre-configure New IP Addresses for information.
2. Back up the current Supercluster. See Back Up Superclusterfor information.
3. Restore the old Supercluster configuration and data to the new systems. See Restore an Entire Supercluster for information.

Pre-configure New IP Addresses

Before the migration: When you use DNS-based load balancing (such as round-robin DNS) and are using new IP addresses for the restored PCE, be sure to record those new IP addresses in the `runtime_env.yml` file on all Supercluster core nodes. This allows VENs to continue to communicate with the PCEs after migration.

NOTE:
When you use traffic-based load balancing, such as with the F5, you do *not* need to add the new IP addresses to `runtime_env.yml`. The VENs communicate exclusively with the traffic load balancers' virtual IP addresses, and not directly with the PCEs.

To update `runtime_env.yml` with additional IP addresses:

1. On *all existing core nodes* in your cluster, edit the `runtime_env.yml` file; under the `cluster_public_ips.cluster_fqdn` parameter, add the new IP addresses of all new core nodes:

```
cluster_public_ips:
  cluster_fqdn:
    - <old IP address>
    - <old IP address>
    - <new IP address>
    - <new IP address>
  cluster_event_service_fqdn:
    - <old IP address>
    - <old IP address>
    - <new IP address>
    - <new IP address>
```

2. On *the leader*, restart the Supercluster to send the configuration update to all

members:

```
sudo -u ilo-pce illumio-pce-ctl restart
```

Supercluster Command-line Reference

The Illumio PCE control interface for Supercluster commands often have restrictions on the type of node they can be run on. For example, setting a cluster's runlevel can be run from any core or data node. Other database specific commands must only be run on specific data nodes. The following tables list the different command line operations you can perform and the specific node (or nodes) the commands must be run on.

Supercluster Commands to Node Reference

This section lists commands you can use to control behavior of PCEs and PCE databases in a Supercluster.

Supercluster PCE Control Commands

The commands have the following general syntax. The `--retry-count` is optional and defaults to 5.

```
# sudo -u ilo-pce illumio-pce-ctl sub-command --option [--retry-count]
```

The following table shows commands you can use to control PCE behavior:

Command	Description	Run on Node
`reset`	Revert a PCE to standalone state.	The affected node being repaired to join the Supercluster
`supercluster-assign-leader`	Designate an existing member PCE cluster to be a Supercluster leader.	Any node
`supercluster-drop`	Removes a PCE cluster from a Supercluster.	Any node
`supercluster-init-leader`	Assign a PCE cluster as the leader of your Supercluster.	Any node
`supercluster-`	Joins a PCE cluster to a Supercluster.	On any core node

Command	Description	Run on Node
`join`	Running this command can take up to 30 minutes depending on the number of PCEs in your Supercluster and size of the PCE database.	
`supercluster-members`	Displays all current active Supercluster PCEs, members and leader.	On any core node or the data0 node
`supercluster-restore`	Restores a formerly failed leader or member to be restored to the Supercluster. This command can be run for a leader or a member. Executing this command can take up to 1 hour depending on the number of PCEs in the Supercluster and size of the PCE database.	On any core node

Supercluster PCE Database Commands

The commands have the following general syntax. The `--retry-count` is optional and defaults to 5.

```
# sudo -u ilo-pce illumio-pce-db-management sub-command --option [--retry-count]
```

The following table shows commands you can use to control PCE database behavior:

Command	Description	Run on Node
`supercluster-data-restore`	Restores a failed PCE's data, using a backup taken from that PCE before the failure.	On one of the data nodes only
`supercluster-quiesce`	Pauses all the pending database replication; for example, during a software upgrade.	Any node

Re-runnable illumio-pce-ctl Arguments

All arguments to `illumio-pce-ctl` are re-runnable in case of a command failure.

Argument on illumio-pce-ctl	Description
`supercluster-init-leader`	Configures this PCE as the Supercluster leader.
`supercluster-join [supercluster_leader_fqdn]`	Joins this PCE into the Supercluster specified by the FQDN of the Supercluster leader. While this command is running, it temporarily sets the runlevel to 1. When the command is interrupted, you

Argument on illumio-pce-ctl	Description
	might see runlevel 1 unexpectedly.
`supercluster-assign-leader`	Assigns a new Supercluster leader.
`supercluster-restore` *`failed_pce_fqdn`* `[`*`supercluster_leader_fqdn`*`] [--restore-type single_pce|entire_supercluster]`	Restores a failed PCE and rejoins it to the Supercluster. Replace *`failed_pce_fqdn`* with the FQDN of the PCE you are restoring. The *`supercluster_leader_fqdn`* is required when you are restoring a member PCE, not the leader PCE. While this command is running, it temporarily sets the runlevel to 1. When the command is interrupted, you might see runlevel 1 unexpectedly.
`supercluster-drop [`*`failed_pce_fqdn`*`]`	Temporarily drops the failed PCE from the Supercluster, so it is no longer replicated.
`supercluster-members`	Shows the members in the Supercluster.
`supercluster-config`	Shows the Supercluster configuration.
`supercluster-upgrade-prepare`	Unjoins the PCE from the Supercluster and prepares data for upgrade.
`supercluster-upgrade-rejoin`	Rejoins the PCE to the Supercluster. While this command is running, it temporarily sets the runlevel to 1. When the command is interrupted, you might see runlevel 1 unexpectedly.
`supercluster-replication-check [--detailed] [--show-data-detail]`	Displays the state of replication. The `--detailed` option displays more verbose output. The `--show-data-detail` option displays primary keys for the replication data check.

Re-runnable illumio-pce-db-management Arguments

All arguments to `illumio-pce-db-management` are re-runnable in case of a command failure.

Argument on illumio-pce-db-managment	Description
`supercluster-data-dump --file` *`desired_location_of_backup_file`*	Writes the database persistent state in Supercluster to a file.
`supercluster-data-restore --restore-type single_pce|entire_supercluster [skip-db-restore]`	Restores data and persistent state of a Supercluster database.
`show-supercluster-replication-info`	Display Supercluster node rep-

Argument on illumio-pce-db-managment	Description
	lication information.
`supercluster-quiesce [wait_timeout]`	Quiesces all pending replication.
`supercluster-replication-debug [--detailed]`	Show replication-related information for debugging.

Upgrade Supercluster

This section describes how to install a newer software version on PCEs in a Supercluster.

Before Upgrading

Before you upgrade the Supercluster, perform the following:

1. Back up the PCE.

   Before the upgrade, back up the leader and all member databases and each PCE's `runtime_env.yml` file. See Back Up Superclusterfor information.

2. Ensure all PCEs are in a healthy state.

   Before upgrading, make sure all PCEs in the entire Supercluster are in a healthy state. In the PCE web console, check the PCE Health page to make sure the PCE health status is **Normal**.

Types of Supercluster Upgrade

You can choose to perform a simple upgrade or a rolling upgrade.

- Supercluster simple upgrade: The Supercluster simple upgrade procedure requires you to set all the PCEs in the Supercluster to runlevel 1 for the duration of the upgrade. During a simple upgrade, the Supercluster is not fully operational. See Supercluster Simple Upgrade.
- Supercluster rolling upgrade: Rolling upgrade keeps the Supercluster operational while individual PCEs are upgraded one at a time. With a rolling upgrade, the Supercluster continues to operate. See Supercluster Rolling Upgrade.

NOTE:
Supercluster rolling upgrade is supported only for a hotfix or a maintenance release. The major and minor release numbers in the installed and upgrade versions must match. For example, you can do a rolling upgrade from 21.2.0 to 21.2.1.

Supercluster Simple Upgrade

A Supercluster simple upgrade follows these general steps:

1. On all PCEs, quiesce the data replication.
2. Upgrade the software on all nodes of all PCEs.
3. Migrate the database on all PCEs.
4. Bring all PCEs back to runlevel 5.

Steps for Upgrade

1. **Quiesce data replication.**
   a. On *any node* in the PCE cluster, bring all PCEs to runlevel 2:

      ```
      $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
      ```

   b. In *the PCE clusters*, repeat step (a) for all leaders and all members.

      The cluster status should be RUNNING.

   c. On *any node in all PCE clusters*, verify that the `set-runlevel` command finished and the cluster status is `RUNNING`:

      ```
      $ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
      ```

      Do not proceed to the next step until the `set-runlevel` command finishes.

   d. Quiesce database replication.

      On *any node*, run the following command. Repeat this command on every PCE.

      ```
      $ sudo -u ilo-pce illumio-pce-db-management supercluster-
      quiesce timeout_in_seconds
      ```

This command waits for data replication to finish, which can take some time. To set a time limit, use `timeout_in_seconds` (default: 600). If the command doesn't complete within this time, it will stop. You must then run the command again.

Expected output when database replication is successfully quiesced:

```
Replication is complete.
```

2. **Upgrade the software.**

   Because this is a simple upgrade, you upgrade the software on all nodes of all PCEs in parallel.

   a. On *any node*, stop the PCE cluster:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl cluster-stop
   ```

   b. The packages to install depend on the type of PCE node:
      - Core nodes: Two packages, the PCE RPM and UI RPM.
      - Data nodes: One package, the PCE RPM.

   On *each core node* in the cluster, log in as root and install the PCE RPM and UI RPM. Be sure to specify both of the RPM file names on the command line:

   ```
   $ rpm -Uvh illumio_pce_rpm illumio_ui_rpm
   ```

   For `illumio_pce_rpm` and `illumio_ui_rpm`, substitute the paths and filenames of the two RPM files you downloaded from the Illumio Support portal.

   c. On *each data node* in the cluster, log in as root and install the PCE RPM:

   ```
   $ rpm -Uvh <illumio_pce_rpm>
   ```

   For `illumio_pce_rpm`, substitute the path and filename of the software you downloaded from the Illumio Support portal.

   d. On *any node*, start each cluster at runlevel 1:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl start --runlevel 1
   ```

3. **Update the runtime environment file** (`runtime_env.yml`). See What's New and Changed in This Release to determine whether any changes to `runtime_env.yml` are required to upgrade. If changes are required:
   a. On *all nodes* in the cluster, update the `runtime_env.yml` file.
   b. On *all nodes* in the cluster, check the validity of the `runtime_env.yml` file:

      ```
      $ sudo -u ilo-pce illumio-pce-ctl check-env
      ```

      If any issues are reported by this command, correct them before moving on to the next step.
4. **Migrate the PCE database.**
   a. On *any node* of *every upgraded PCE*, run the following command:

      ```
      $ sudo -u ilo-pce illumio-pce-db-management migrate --upgrade-type
      simple
      ```

   b. The migration might take some time to complete. Check the progress with the following command:

      ```
      $ sudo -u ilo-pce illumio-pce-db-management supercluster-upgrade-status
      ```

   c. On *any node* in the *first PCE cluster*, bring all PCEs to runlevel 2. Repeat this step on all the other PCEs.

      ```
      $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
      ```

   d. For all leader and member PCE clusters, repeat step b. Verify that all PCEs in the Supercluster are at runlevel 2.
   e. Wait until `agent_slony_service` and `login_slony_service` are up and running. These service names appear in bright blue or may have a pound character (#) appended, depending on which color option was chosen when starting the PCE, `--color` or `--no-color`. **Do not restart the PCE.** This step could take some time, depending on how recently you upgraded the PCE software. Run the following command to monitor the progress:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

Issue the command again, when needed, until the services are ready

5. **Bring PCEs back to operational status.**
   a. On *any node for each PCE*, set the runlevel to 5 :

   ```
   $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
   ```

   Setting the runlevel can take time to complete.

   b. On *any node in all PCE clusters*, verify that the `set-runlevel` command finished and the cluster status is `RUNNING`:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
   ```

   NOTE:
   Due to the time it takes to replicate new database tables across all the PCEs, the upgrade might take longer than usual. The delay occurs when you bring the PCE to runlevel 2 or 5 from runlevel 1 after upgrading the software. The wait time depends on the number of new tables that are part of the upgrade. The wait might be up to 20 minutes.

   c. Verify that you can log into the PCE web console on each PCE in the Supercluster.

Upgrade is complete.

NOTE:
In rare cases, you might receive an error when attempting to log into the PCE web console. When this happens, run the following command on all nodes, and try logging in again:

```
$ sudo -u ilo-pce illumio-pce-ctl restart
```

Supercluster Rolling Upgrade

In a rolling upgrade, the PCEs are upgraded one by one. The PCE that is being upgraded is at runlevel 1, while all the other PCEs are fully operational (runlevel 5).

NOTE:
Supercluster rolling upgrade is supported only for a hotfix or a maintenance release. The major and minor release numbers in the installed and upgrade versions must match. For example, you can do a rolling upgrade from 21.2.0 to 21.2.1. Also, due to a software change in 21.2.2, you can only do a rolling upgrade when the installed and upgrade versions are both either before 21.2.2 or after it. For example, you can do a rolling upgrade from 21.2.3 to 21.2.7, but not from 21.2.1 to 21.2.7.

A Supercluster rolling upgrade follows these general steps:

1. Upgrade the software on all nodes of the leader PCE.
2. Migrate the database on the leader PCE.
3. Bring the leader PCE back to runlevel 5.
4. Repeat these steps for each member PCE.

Steps for Upgrade

1. **Upgrade the software on the leader PCE.**
   a. On *any node* of the leader PCE, stop the PCE cluster:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl cluster-stop
   ```

   b. The packages to install depend on the type of PCE node:
      - Core nodes: Two packages, the PCE RPM and UI RPM.
      - Data nodes: One package, the PCE RPM.

   On *each core node* in the cluster, log in as root and install the PCE RPM and UI RPM. Be sure to specify both of the RPM file names on the command line:

   ```
   $ rpm -Uvh illumio_pce_rpm illumio_ui_rpm
   ```

For `illumio_pce_rpm` and `illumio_ui_rpm`, substitute the paths and filenames of the two RPM files you downloaded from the Illumio Support portal.

c. On *each data node* in the cluster, log in as root and install the PCE RPM:

```
$ rpm -Uvh <illumio_pce_rpm>
```

For `illumio_pce_rpm`, substitute the path and filename of the software you downloaded from the Illumio Support portal.

d. On *any node*, start the cluster at runlevel 1:

```
$ sudo -u ilo-pce illumio-pce-ctl start --runlevel 1
```

2. **Update the runtime environment file** (`runtime_env.yml`). See What's New and Changed in This Release to determine whether any changes to `runtime_env.yml` are required to upgrade. If changes are required:
   a. On *all nodes* in the cluster, update the `runtime_env.yml` file.
   b. On *all nodes* in the cluster, check the validity of the `runtime_env.yml` file:

```
$ sudo -u ilo-pce illumio-pce-ctl check-env
```

   If any issues are reported by this command, correct them before moving on to the next step.

3. **Migrate the PCE database on the leader PCE.**
   a. On *any node* of the *leader PCE*, run the following command:

```
$ sudo -u ilo-pce illumio-pce-db-management migrate --upgrade-type
rolling
```

   b. The migration might take some time to complete. Check the progress with the following command:

```
$ sudo -u ilo-pce illumio-pce-db-management supercluster-upgrade-status
```

4. **Bring the leader PCE back to operational status.**
   a. On *any node of the leader PCE*, set the runlevel to 5 :

      ```
      $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
      ```

      Setting the runlevel can take time to complete.

   b. On *any node of the leader PCE*, verify that the `set-runlevel` command finished and the cluster status is `RUNNING`:

      ```
      $ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
      ```

5. **Upgrade the software on a member PCE.**
   a. On *any node* of the member PCE, stop the PCE cluster:

      ```
      $ sudo -u ilo-pce illumio-pce-ctl cluster-stop
      ```

   b. On *all nodes* of the member PCE, install the new version of the PCE. For information, see the *PCE Installation and Upgrade Guide*.
   c. On *any node*, start the cluster at runlevel 1:

      ```
      $ sudo -u ilo-pce illumio-pce-ctl start --runlevel 1
      ```

6. **Migrate the PCE database on the member PCE.**
   a. On *any node* of the *member PCE*, run the following command:

      ```
      $ sudo -u ilo-pce illumio-pce-db-management migrate
      ```

   b. The migration might take some time to complete. Check the progress with the following command:

      ```
      $ sudo -u ilo-pce illumio-pce-db-management supercluster-upgrade-status
      ```

7. **Bring the member PCE back to operational status.**
   a. On *any node of the member PCE*, set the runlevel to 5 :

      ```
      $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
      ```

      Setting the runlevel can take time to complete.

   b. On *any node of the member PCE*, verify that the `set-runlevel` command finished and the cluster status is `RUNNING`:

      ```
      $ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
      ```

8. **Repeat steps 4 through 6 for each additional member PCE.**
9. **Verify that you can log in to the PCE web console on each PCE in the Supercluster.**

Upgrade is complete.

NOTE:
In rare cases, you might receive an error when attempting to log into the PCE web console. When this happens, run the following command on all nodes, and try logging in again:

```
$ sudo -u ilo-pce illumio-pce-ctl restart
```

During Supercluster Upgrade

During a rolling upgrade, if you log in to one of the PCEs, you will see a banner that states the Supercluster is in the process of a rolling upgrade.

The PCE Health page on the Leader displays the upgrade status for each PCE. The Upgrade Status column shows Pending if the PCE is in the process of being upgraded, and it shows Complete when the upgrade is complete. When the upgrade is finished, the Upgrade Status column no longer appears.

Supercluster Listen Only Mode

The PCE “Listen Only” mode allows you stop the PCE from sending policy changes to your VENs. Enabling Listen Only mode for the PCE is typically used in these situations:

- During PCE maintenance windows, and when starting the PCE back up.
- After restoring the PCE from a backup.
- During maintenance windows for other parts of your network environment.

In Listen Only mode, VENs still report updated workload information to the PCE, but the PCE does not modify the firewall rules on any workloads or send any updates from the PCE to the VENs. Also, the PCE does not mark workloads as Offline and does not remove them from policy when Listen Only mode is enabled.

When this mode is enabled, you can still write policy, pair new workloads, provision policy changes, assign or change workload Labels, but changes will not be sent to the VENs until you disable Listen Only mode. You can disable Listen Only mode when you are ready to resume normal policy operations.

Enable PCE Listen Only Mode

1. On *all nodes* in the PCE cluster, stop the PCE software:

```
$ sudo -u ilo-pce illumio-pce-ctl stop
```

2. On *all nodes* in the PCE cluster, set the node at runlevel 1:

```
$ sudo -u ilo-pce illumio-pce-ctl start --runlevel 1
```

3. On *any data node*, enable Listen Only mode:

```
$ sudo -u ilo-pce illumio-pce-ctl listen-only-mode enable
```

4. Set the PCE runlevel to 5:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
```

Disable PCE Listen Only Mode

NOTE:
The command to disable PCE Listen Only mode can be executed at either runlevel 1 or 5.

1. On *all nodes* in the PCE cluster, stop the PCE software:

```
$ sudo -u ilo-pce illumio-pce-ctl stop
```

2. On *all nodes* in the PCE cluster, set the node to runlevel 1:

```
$ sudo -u ilo-pce illumio-pce-ctl start --runlevel 1
```

3. On *any data node*, disable Listen Only mode:

```
$ sudo -u ilo-pce illumio-pce-ctl listen-only-mode disable
```

4. Set the PCE runlevel to 5:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
```

Chapter 3

PCE Supercluster VEN Management

This chapter contains the following topics:

This section describes how Virtual Enforcement Nodes (VENs) are managed by a PCE Supercluster, and what tasks you need to perform to successfully pair VENs and manage them in a Supercluster deployment.

Pair VENs in a Supercluster

A Supercluster allows you to control which PCE you want your workloads to pair with and be managed by, depending on your needs. You can pair one set of workloads with a PCE in Europe, for example, and you can pair another set of workloads with a PCE in the United States.

In some cases, you might need to reassign some workloads (and their VENs) to be managed by a different PCE than the one they were initially paired with. In certain cases of a PCE failure, you might want workloads to temporarily fail over to another healthy PCE. In both cases, a set of workloads are managed by a another PCE.

VENs Paired to Disconnected PCE

A PCE that loses connectivity to its VENs maintains its “online” status from the VENs perspective and retains the workloads in policy. This condition can be corrected taking the following general steps:

1. Determine the cause of the PCE failure and correct it.
    - Restore the failed PCE.
    - In the case of a failed leader, promote a member to leader.
2. For a failed member, uninstall or unpair the VEN on the affected workloads.
3. (Optional) Using the PCE web console or the REST API, delete records of the incorrectly marked “online” VENs. This step is option because after VEN heart-beating resumes, the proper state of the VEN will be reestablished.

Pair Workloads with Leader or Member

This section discusses how to pair your workloads with a Supercluster leader or member.

Pairing workloads with the leader or member follows nearly the same process as a standalone PCE cluster:

You create a pairing profile in the Supercluster leader's PCE web console.

- Member PCEs can be offline when this profile is created.
- Pairing profiles must always be created on the Supercluster leader.
- This pairing profile is propagated to all members.

With this pairing profile, you generate a pairing script.

- The pairing script can be configured to pair either with the Supercluster leader or with a member PCE:
    - A pairing script generated on the leader includes the FQDN of the leader.
    - A pairing script generated on a member includes the FQDN of that member.
- The pairing script includes the option `--management-server` with the domain name and port of the leader or the member.
- The paring script includes a pairing key (`--activation-code` option) that can be used to pair with any member.
- Members can create new pairing keys from pairing profiles replicated from the leader.
- Members can be isolated from the Supercluster but still continue to pair with workloads.
- You run the pairing script on the workload to pair.

Pairing Script Examples for Supercluster

Pairing Script to Pair with Leader

The leader's FQDN is `supercluster-pce-LEADER.BigCo.com:8443`.

```
rm -fr /opt/illumio/scripts && umask 026 && mkdir -p /opt/illumio/scripts &&
curl https://repo.illum.io/sPl1tOExo0FIEphoewIujIucrLaTOAS3/pair.sh -o
/opt/illumio/scripts/pair.sh &&
chmod +x /opt/illumio/scripts/pair.sh && /opt/illumio/scripts/pair.sh
--management-server supercluster-pce-LEADER.BigCo.com:8443
--activation-code xxyyzzyywwxx654321
```

Pairing Script to Pair with Member

The member's FQDN is `supercluster-pce-MEMBER.BigCo.com:8443`.

```
rm -fr /opt/illumio/scripts && umask 026 && mkdir -p /opt/illumio/scripts &&
curl https://repo.illum.io/sPl1tOExo0FIEphoewIujIucrLaTOAS3/pair.sh -
o /opt/illumio/scripts/pair.sh &&
chmod +x /opt/illumio/scripts/pair.sh && /opt/illumio/scripts/pair.sh
--management-server supercluster-pce-MEMBER.BigCo.com:8443
--activation-code xxyyzzyywwxx654321
```

Run Pairing Script on Workloads

As with the standalone PCE configuration, you run the Supercluster-generated pairing script directly on the workload itself.

Linux environment variables and Windows command-line variables allow you to specify the management server to pair with.

For more information about pairing, see the *VEN Installation and Upgrade Guide*.

Pair Workloads with GSLB PCE

When you rely on a Global Services Load Balancer (GSLB) to control which specific PCE a workload communicates with or to pair workloads to a generic name for the Supercluster, set the FQDN value of the `supercluster_fqdn` parameter in each PCE's `runtime_env.yml` file.

This value is used as the argument to the pairing script's `--management-server` option, which is the name of FQDN you define.

NOTE:
Do not put the port number at the end of the `supercluster_fqdn` value. The system itself adds the port number to the pairing script.

Example

This snippet from the generated pairing script shows how the `supercluster_fqdn` parameter is set.

```
...
--management-server MyBigSuperclusterFQDN-from-supercluster-fqdn-
parameter.BigCo.com:8444
...
```

VEN Failover After PCE Failure

In rare cases, when you pair workloads with a Supercluster PCEs and that PCE fails immediately after you run the workload pairing script, the information about that workload's pairing does not get replicated to the other PCEs in the Supercluster. When that workload's VEN tries to retrieve policy from the PCE or sends a heartbeat, the VEN receives an HTTP 401 Unauthorized error and is eventually moved into the Lost Agent state.

To recover from this situation, you can perform one of these actions:

- Uninstall the VEN completely from the workload and repair it with a functioning PCE.
- Recover the affected PCE so that it is fully functional and online. After the VEN successfully heartbeats to the recovered PCE, it automatically comes out of the Lost Agent state.

  This second option to recover the PCE only works when the affected PCE had information about that VEN before the failure. When you recover the PCE from a backup that was taken before the VEN was paired, the VEN will have to be uninstalled and the workload repaired.

Pairing Container Clusters

You can pair workloads as part of a container cluster on supercluster member regions. Container clusters can be managed in a member region as well as all the resources

attached to this container cluster: container workloads (pods), virtual services (services) and workloads (nodes).

Manage VENs in a Supercluster

This section describes how to manage VENs in a PCE Supercluster. Some of the management tasks are affected by Supercluster considerations, such as whether the task is performed on a leader or member PCE.

Unmanaged Workloads

When you need to create unmanaged workloads for assets that do not have a VEN installed, they must be created on the leader.

VEN Uptime and Heartbeat in Supercluster

Each workload managed by your Supercluster provides the latest "Uptime" of the workload. Uptime is defined as the amount of time that has passed in seconds since the workload reported its first heartbeat to the PCE, either after being paired or after a workload system restart.

Depending on which PCE you are logged into while viewing this information, the Uptime field might display the following:

```
Unavailable. Viewable on nameOfPCE
```

This message means that the PCE that you are currently logged into does not manage this workload. Instead, the Uptime and Last Heartbeat properties on the Workload details page indicate the name of the PCE that this workload was paired with.

Workload Support Reports in Supercluster

When you are logged into the leader of a Supercluster, you can generate and download Workload Support Reports for any workload in the Supercluster. This report includes workloads that have been paired with and are being managed by other members.

From a member PCE you can generate a Workload Support Report for all workloads connected to that PCE. However, you cannot generate a Workload Support Report from a member PCE for any workloads connected to a different PCE.

When the Workload Support Report is finished, you can download it from the leader PCE web console.

For information on running Workload Support Reports from the command line on the host, see the *PCE Administration Guide*.

Workloads on Leader When Member Fails

When one of your member PCEs goes down, any changes you make to workloads managed by the affected member (while logged into the leader) are immediately reflected in the leader PCE web console, even though the change has not been replicated to the member and applied on the workload.

For example, one member of your Supercluster fails. While you are logged into the leader, you make a change to a workload that was paired with that affected member, such as changing the workload's policy state. The Workload's details page on the leader will show the policy state change. However, the actual workload policy state will not be changed until the member is recovered.

VEN Failover

When a PCE in your Supercluster fails, its workloads continue to enforce the latest policy and buffer traffic data until the PCE is recovered. When you need to modify policy on the workload before the affected PCE can be recovered, you can fail over its workloads to a different PCE in the Supercluster. Workload failover is managed outside the Supercluster and requires either a GSLB or an update to your DNS infrastructure.

To fail over a workload to a different PCE, configure your GSLB or DNS to resolve the FQDN of the workload's target PCE to the public IP addresses of another PCE in your Supercluster.

When you configure the `supercluster.fqdn` parameter in your `runtime_env.yml` file, the target PCE of all workloads is the Supercluster FQDN. The next time the workload resolves this FQDN, it will receive the updated IP addresses and begin heartbeating to and receiving policy from the new PCE.

To validate that the VEN reassignment was successful, check that the active PCE now corresponds to the FQDN the workload should have failed-over to.

VEN Failover Impact on Traffic Data

Be aware that some traffic data can be lost when VENs fail over to a different PCE:

- Traffic data used for Illumination and blocked traffic is lost and will be missing from Illumination.

- Traffic data that is exported to syslog or Fluentd is not lost, as long as the PCE has the capacity to handle all incoming flow summaries from all VENs.

VEN Failover and Certificates

A VEN must be able to validate the certificate of the PCE that is managing it and any other PCEs it will fail over to. When a VEN fails over and cannot validate the certificate of the new PCE, it cannot authenticate and enters the Lost Agent state. In this state, just as in a failure scenario, the VEN is disconnected from the PCE and it cannot receive policy updates. In this scenario, because the PCE that was managing the VEN is still running, it will mark the workload as offline in 1 hour, which in turn isolates it from all other workloads.

Reassign VENs in Supercluster Using REST API

When deploying a Supercluster, you might want to "move" workloads that have been paired to one PCE so that they are managed by a different PCE in the Supercluster. For example, you expand your single standalone PCE into a Supercluster and you want to reassign some of your existing VENs to be managed by the nearest PCE. In this case, you can reconfigure the VENs on paired workloads so that they use a different FQDN to communicate with the proper PCE.

Using the Illumio Agent API (the REST API refers to VENs as "agents"), you change the target PCE of the workload to the PCE you want to reassign the workload to. The PCE that is currently managing the workload sends the workload the FQDN of the new target PCE; the workload begins heartbeating to and receiving its policy updates from that PCE. The active PCE of the workload is now the same as the target PCE.

NOTE:
Manually moving a VEN to a different PCE using the REST API is subject to the object limit `active_agents_per_pce`. For more information, see Object Limits and Supercluster.

Active and Target PCE

Before reassigning VENs to another PCE, you need to understand these terms: active PCE and target PCE. These terms correspond to two properties that are added to a workload's VEN on pairing.

- `active_pce_fqdn`: The PCE that is currently managing a workload; namely, the PCE the workload has last heartbeat to.

- `target_pce_fqdn`: The PCE that is configured to manage this workload or the FQDN of the Supercluster (when you configured the `supercluster.fqdn` property in your `runtime_env.yml` file).

Workload Reassignment Workflow

This section assumes you are familiar with the basic concepts and usage of the Illumio Core REST API.

IMPORTANT:
Before reassigning workloads to a new PCE, make sure that the active and target PCE are fully operational and at runlevel 5.

The workflow to reassign workloads to a different PCE consists of these general tasks:

1. **GET workloads:** To find the HREF of the agent on a workload, get a collection of workloads from the PCE. When you already know the HREF of a workload, you can get an individual instance of that workload, which returns the HREF of the agent that was used to pair that workload.
2. **Identify agent HREF:** The workloads' GET response include the `agent` property, which represents the VENs that are installed on the workloads as part of the pairing process. An agent is identified by its HREF.
3. **Identify active PCE FQDN of agent:** The workloads GET schema returns two properties that indicate the FQDN of the PCE that is actively managing the agent (`active_pce_fqdn`) and a second property that allows you to use a different "target" PCE FQDN (`target_pce_fqdn`) to manage the agent.
4. **Change target PCE FQDN of agent:** Update (PUT) the `target_pce_fqdn` property so that the VEN can be managed by a different PCE in your Supercluster.

Get Workloads

To get the HREF of an agent (VEN) on a workload, get a collection of workloads. You can GET up to 500 workloads at a time. When you know the HREF of an individual workload, you can get just the single workload.

To get a collection of workloads, you use this URI:

```
GET [api_version][org_href]/workloads
```

For example, using curl:

```
curl -u api_
xxxxxxx64fcee809:'xxxxxx5048a6a85ce846a706e134ef1d4bf2ac1f253b84c1bf8df6b83c70d95'
-H "Accept: application/json" -X GET
https://my.pce.supercluster:443/api/v1/orgs/7/workloads
```

Identify Agent HREF in Response

The JSON response from getting workloads provides information about the VEN ("agent") that was installed when the workload was paired with the PCE. In this response, you identify the workload's VEN (`agent`) by its HREF.

For example, the section that begins with the `agent` property shows the HREF of the VEN (`href":"/orgs/3/agents/40916`). In the response, the active PCE (`active_pce_fqdn`) and the target PCE (`target _pce_fqdn`) are the same. This does not change until you perform the reassignment.

```
"agent": {
            "config": {
                "log_traffic": false,
                "visibility_level": "flow_summary",
                "mode": "illuminated",
                "security_policy_update_mode": "adaptive"
            },
            "href": "/orgs/3/agents/40916",
            "status": {
                "uid": "e6c21a34-ebc2-4cf4-834e-3ec5df31d6ed",
                "last_heartbeat_on": "2016-02-11T12:22:32.91936Z",
                "instance_id": "perf_instance_1289213668111202403-
1821@1455178338188",
                "managed_since": "2016-02-11T08:13:19.482909Z",
                "fw_config_current": false,
                "firewall_rule_count": null,
                "security_policy_refresh_at": null,
                "security_policy_applied_at": null,
                "security_policy_received_at": null,
                "uptime_seconds": 95819257,
                "status": "active",
                "agent_version": "2.10.0-20150715010305",
                "agent_health_errors": {
                    "errors": [],
```

```
                "warnings": []
            },
            "agent_health": [],
            "security_policy_sync_state": "syncing"
        },
        "active_pce_fqdn": current-pce-fqdn.example.com,
        "target_pce_fqdn": current-pce-fqdn.example.com,
```

Change Target PCE

When you have the agent HREF, you can update the the target PCE with the PCE FQDN the VEN will use. In your JSON request body, pass the following data:

```
{
"target_pce_fqdn": "new-pce-fqdn.example.com"
}
```

The URI for this operation:

```
PUT [api_version][agent_href]/update
```

This curl example show how you can pass the `target_pce_fqdn` property containing the FQDN of the new PCE:

```
curl -u  api_
xxxxxxx64fcee809:'xxxxxxx5048a6a85ce846a706e134ef1d4bf2ac1f253b84c1bf8df6b83c70d9
5'
-H "Accept: application/json" -H "Content-Type:application/json" -X PUT
-d '{"target_pce_fqdn":"target-pce.example.com"}'
https://my.pce.supercluster:443/api/v1/orgs/3/agents/40916/update
```

Validate VEN Reassignment

To validate that the VEN reassignment was successful, verify the active PCE matches the target PCE. Perform a GET request on the agent again. The target and active PCE FQDN should be the same. When the operation is successful, the response return an HTTP 204 code indicating success.

NOTE:
Reassigning a VEN to a different PCE can take up to 10 minutes to complete.

For example:

```
"agent": {
            "config": {
                "log_traffic": false,
                "visibility_level": "flow_summary",
                "mode": "illuminated",
                "security_policy_update_mode": "adaptive"
            },
            "href": "/orgs/3/agents/40916",
            "status": {
                "uid": "e6c21a34-ebc2-4cf4-834e-3ec5df31d6ed",
                "last_heartbeat_on": "2016-02-11T12:22:32.91936Z",
                "instance_id": "perf_instance_1289213668111202403-
1821@1455178338188",
                "managed_since": "2016-02-11T08:13:19.482909Z",
                "fw_config_current": false,
                "firewall_rule_count": null,
                "security_policy_refresh_at": null,
                "security_policy_applied_at": null,
                "security_policy_received_at": null,
                "uptime_seconds": 95819257,
                "status": "active",
                "agent_version": "2.10.0-20150715010305",
                "agent_health_errors": {
                    "errors": [],
                    "warnings": []
                },
                "agent_health": [],
                "security_policy_sync_state": "syncing"
            },
            "active_pce_fqdn": new-pce-fqdn.example.com,
            "target_pce_fqdn": new-pce-fqdn.example.com
```

Chapter 4

PCE Supercluster Administration

This chapter contains the following topics:

This section explains how to perform common administration tasks for a PCE Supercluster.

Monitor Supercluster Health

You can use these two general methods for monitoring the health of your PCE Supercluster:

- REST API calls to determine the Supercluster leader and a PCE member's health
- The PCE web console to view the health of the entire Supercluster from the leader or for the member you are logged into.

This section discusses health monitoring specifically for a PCE Supercluster. Additionally, follow the PCE health monitoring guidelines in the *PCE Administration Guide*.

REST API for Supercluster Health

You can monitor Supercluster health using the following REST API mechanisms.

REST API `/health`

Using the PCE Health API, you can get current health information about all PCEs in your Supercluster, including the leader and members.

```
GET [api_version]/health
```

REST API to Determine Supercluster Leader

Use this Public Stable REST API request to determine whether the PCE in a Supercluster is a leader or member.

```
GET [api_version]/supercluster/leader
```

Your GSLB can issue this request to monitor health of the leader.

HTTP Response Code from `/supercluster/leader`

Response	Meaning
202	The PCE is the leader.
404	The PCE is a member.

REST API `/node_available`

After your GSLB determines the Supercluster leader, issue the following REST API request to monitor the leader's availability:

```
GET [api_version]/node_available
```

HTTP response code from `/node_available`

The Health REST API can take up to 30 seconds to reflect the actual status of the node.

Response	Meaning
202	The node is healthy and is connected to the rest of the cluster.
404 or no response	The node is unhealthy and cannot accept requests. Such a node should be removed from the load balancing pool.

PCE Web Console for Supercluster Health

The Health page in the PCE web console in a Supercluster provides health information about your on-premises PCE, whether you deployed an SNC, 2x2, 4x2, or

Supercluster.

- **General PCE Health:** Shows general health information for each PCE in your Supercluster, such as health status, node status and uptime, and system health information for each node (CPU usage, memory, and disk usage). When you deployed a PCE Supercluster, the Health page lists all PCEs in the Supercluster with individual health information for each PCE.
- **Supercluster Leader Health:** Displays the health status of the leader PCE in the Supercluster. You can view the health of each PCE in the Supercluster.
- **Supercluster Member Health:** Shows health information about the member you are logged into, including a timer that indicates the amount of time since Illumination data was synced across the Supercluster. The Health page shows the database replication lag for each PCE relative to all other PCEs in the Supercluster, indicating how long it took for data to be replicated from one PCE to another.

The PCE health page indicates the current state of database replication across the Supercluster and how recently each member PCE's Illumination data has been synced with the leader.

- **Supercluster Replication (Lag):** Indicates how long it took for one PCE to receive replicated data from another PCE in the Supercluster. For example, a user created a new IP list in the leader and saved it. The change took 4 seconds to replicate to Member1 and Member1's Health page showed that its replication lag is 4 seconds behind the leader. The PCE web console shows replication lag for each PCE in the Supercluster.
- **Supercluster Illumination Sync (Members only):** Shows the last time since a member PCE replicated its Illumination traffic data with the Supercluster leader. This information only appears for members that periodically send traffic data to the leader. This information provides a full picture of Illumination traffic for your entire Supercluster. You can initiate a sync of Illumination data on demand by clicking the link in the lower right of the Illumination map.

Supercluster PCE Health Icon

When the PCE Health button has a badge with a number, one or more of the PCEs in your Supercluster have a health status that is *not* "Normal." The badge color indicates the type of warning.

For example, a yellow warning badge with the number 1 indicates that one of the PCEs in the Supercluster has a health warning status.

When the badge is red and shows the number 1, one of the Supercluster PCEs has failed or is down.

Supercluster Web Console Health Page

The Supercluster Health page on the leader displays a high-level view of each PCE's health. You can click a PCE to view individual health information. The information on this page is refreshed every 60 seconds.

Individual PCE Health Status

The following table lists the possible health statuses for a PCE: Normal, Warning, or Critical.

Status	Color	Definition
Normal (healthy)	Green	A PCE is considered to be in a **normal** state when: • All required services are running. • All nodes are running. • CPU usage of all nodes is less than 95%. • Memory usage of all nodes is less than 95%. • Disk usage of all nodes is less than 95%. • Database replication lag is less than or equal to 30 seconds.
Warning	Yellow	A PCE is considered to be in a **warning** state when: • One or more nodes are unreachable. • One or more optional services are missing, or one or more required services are degraded. • The CPU usage of any node is greater than or equal to 95%. • Memory usage of any node is greater than or equal to 95%. • Disk usage of any node is greater than or equal to 95%. • Database replication lag is greater than 30 seconds.
Critical	Red	A PCE is considered to be in a **critical** state when one or more required services are missing. In this scenario, it might not be possible to authenticate to the PCE or get a REST API response depending on which services are missing from the PCE.

PCE Health on Workload Details

When your workloads have been paired with a Supercluster leader or member, you can view PCE health on the Summary tab of the Workload details page. This page includes the PCE section, which lists the hostname and health of the PCE that this workload is paired with.

PCE Health on Illumination Command Panel

When you select a workload in the Illumination map in a Supercluster, the command panel that displays workload details includes the health of the PCE that the workload is paired with. For example, you can see the health status of the PCE the workload is paired with in the PCE Health field.

Command to Show All Supercluster Members

On *any core node* or the *data0 node* in a cluster, run the following command to display the leader and all member PCEs of the Supercluster.

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-members
```

Back Up Supercluster

You need to perform regular backups on all PCEs in the Supercluster.

Different data is backed up depending on whether you run the backup from the Supercluster leader or a member:

- **Leader backup:** Contains all Supercluster replicated data, including workloads, labels, rulesets, rules, services, organization events, workload traffic data, and Supercluster configuration data.
- **Member backup:** Contains the member's local data, including login information, workload traffic data, and Supercluster configuration data.
- **All PCE nodes' runtime environment file:** The `runtime_env.yml` is not included in the backup and must be backed up separately for each node. The default location of the PCE Runtime Environment File is `/etc/illumio-pce/runtime_env.yml`. When the location is different on your system, you can find it by checking the value of the `ILLUMIO_RUNTIME_ENV` environment variable.

When to Back Up

Follow your own organization's policies and procedures for backup, including frequency (such as, hourly, daily, or weekly) and retention of backups offsite or on a system other than any of the Supercluster nodes.

Illumio recommends taking backups in the following situations:

- Before and after a PCE version upgrade
- After pairing a large number of VENs
- After updating a large number of workloads (such as, changing workload policy state or applying labels)
- After provisioning major policy changes
- After making major changes in your environment that affect workload information (such as, an IP address changes)
- Before and after adding new PCEs to your Supercluster
- After you assign a new leader
- On-demand backups before the procedures documented in this guide, such as migration and upgrade

Determine Data Node of All PCEs

The data node is the node that runs the `agent_traffic_redis_server` service. To determine the data node, run the following command:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status
```

Expected output:

```
SERVICES (runlevel: 5) NODES (Reachable: 1 of 1)
====================== ==========================
agent_background_worker_service 192.168.33.90
agent_service NOT RUNNING
agent_slony_service 192.168.33.90
agent_traffic_redis_cache 192.168.33.90
agent_traffic_redis_server 192.168.33.90          <=== Run backup command on this
node
agent_traffic_service NOT RUNNING
...
```

NOTE:
Check for `agent_traffic_redis_server` on a data node before every backup, because this service can be running on either data node.

Back Up Each PCE's Data

For the leader and every member PCE in your Supercluster, perform these steps:

1. Log into the node running the `agent_traffic_redis_server` service.
2. Create a directory for the backup file that is not one of the PCE software's installation directories.
3. Grant both the `ilo-pce` user and the user who will run the backup command Read and Write permissions to this directory.
4. Run the following command:

```
$ sudo -u ilo-pce install_root/illumio-pce-db-management supercluster-data-dump --file desired_location_of_backup_file
```

5. Repeat these steps for every PCE in the Supercluster.

Copy Leader Backup to Members

Copy the backup file that you just made on the leader PCE to the data0 node of each member PCE. The leader's data is readily available to every member so that you can more quickly restore the entire Supercluster. You can copy the file to any file system location of the member data0 node, except for the PCE software's installation directories.

Back Up Leader and Member Runtime Environment Files

Store a copy of each node's `runtime_env.yml` file on a system that is not part of the Supercluster. By default, the PCE Runtime Environment File is stored in `/etc/illumio-pce/runtime_env.yml`. When the location is different on your system, locate the file by checking the `ILLUMIO_RUNTIME_ENV` environment variable.

Assign New Leader

A Supercluster can only have one leader at a time. The following section describes how to choose a new leader or temporarily assign a new leader when the leader has failed and you need to change the Supercluster before it can be recovered.

Assign Leader When a Leader Is Connected

You can choose a new Supercluster leader when the existing leader is still running and connected to the rest of the Supercluster. When you choose a new leader, the former leader becomes a member in the Supercluster.

1. Decide which PCE member you want to be the new leader.
2. On *any node on current leader* and on *any node on the new leader*, bring the nodes to runlevel 2:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
   ```

   Make sure you wait until the software is running before you proceed.
3. Check the progress to see when the status is `RUNNING` on all nodes:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
   ```

4. On *any node on the new leader*, promote it to be the leader:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl supercluster-assign-leader
   ```

5. On *any node* on the new leader and the former leader, bring both PCEs to run-level 5:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
   ```

6. Check the progress to see when the status is `RUNNING` on all nodes:

   ```
   $ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
   ```

Assign New Leader When Leader Has Failed

When your Supercluster leader has failed, you must drop the failed leader from the Supercluster before you can assign a new leader.

WARNING:
When the new leader is promoted, you must isolate the former leader from the network and not allow it to be brought back online. When the former leader is not isolated, it will incorrectly re-join the Supercluster as the leader. Having two leaders in a Supercluster is not supported and can lead to data corruption.

When you are ready to restore the failed PCE and rejoin it to the Supercluster, follow the procedures in Restore a PCE or Entire Supercluster, which will bring the PCE back as a member. After it has been brought back as a member, you can assign it to be the leader again.

To drop the failed leader and assign a new leader:

1. On a *core node* of *each surviving PCE* in the Supercluster, set the runlevel to 2:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
```

2. On the PCE you *assigned as the new leader*, drop the failed leader from the Supercluster:

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-drop failed_PCE_fqdn
```

3. Check the progress to see when the status is `RUNNING`:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

4. On the *newly designated leader PCE*, assign is as the new Supercluster leader:

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-assign-leader
```

5. On the *new leader PCE* and *all member PCEs*, set them to runlevel 5:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
```

Make sure you wait until the software is running before you proceed.

6. Check the progress to see when the status is `RUNNING` on all nodes:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

Restore a PCE or Entire Supercluster

This section describes how to restore a single failed PCE, either leader or member, and rejoin it to a Supercluster. It also describes how to restore the entire Supercluster.

- Restore a Single PCE in a Supercluster. Follow this procedure when a single leader or member PCE has failed and needs to be restored.
- Restore an Entire Supercluster. Follow this procedure when more than one PCE has failed.

Restore a Single PCE in a Supercluster

This section explains how to restore a leader or member PCE in a Supercluster. Isolate that PCE from the Supercluster, restore it, and rejoin it to the Supercluster.

Summary

The following steps are an overview of how to restore a single PCE in a Supercluster. For detailed instructions, read the rest of this section.

1. Preparation:
   a. Have the backups and the copy of the affected PCE's `runtime_env.yml` configuration file ready to use.
   b. Know the IP address, ports, and DNS name of the affected PCE. You must use the same values when you rejoin the PCE to the Supercluster.
2. Isolate the affected PCE from the Supercluster.
3. Install the PCE on new hardware or reuse the installation and `runtime_env.yml` file of the affected PCE.
4. Restore the Supercluster data from backup.
5. Join the repaired PCE to the Supercluster.

Prepare to Restore a Single PCE

Have the following items available:

- Back up the failed PCE; see Back Up Supercluster for information.
- Back up the failed PCE's `runtime_env.yml` file.

- Make a list of the IP address, ports, and FQDN of the failed PCE. You will use these values to reconfigure the repaired PCE.

Isolate the Affected PCE

Before restoring a single PCE, isolate that PCE from the Supercluster.

1. On *all nodes*, shut down the affected PCE:

```
$ sudo -u ilo-pce illumio-pce-ctl stop
```

2. On *a core node of each surviving PCE* in the Supercluster, set the PCE to runlevel to 2:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
```

> NOTE:
> You *must* set all PCEs to runlevel 2 before proceeding to the next step.

3. On *any core node of a surviving PCE*, drop the failed PCE from the Supercluster:

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-drop fqdn_of_failed_pce
```

Install New PCE or Reuse Affected PCE

Decide whether to completely reinstall the PCE on new hardware or reuse the PCE installation that is already on the affected system.

> NOTE:
> In both cases, you must reestablish the FQDN of the affected PCE so that VENs can continue to communicate with the Supercluster. When you have any VENs in enforcement, or you rely on DNS-based load balancing, the new IP addresses of the PCE nodes can be different, as long as the new IP addresses were already in the appropriate settings in the `runtime_env.yml` file on all PCE core nodes. See Pre-configure New IP addresses for information.

- To reinstall the PCE on new hardware, see Deploy a PCE Supercluster.
- To reuse the affected PCE installation, complete the following steps.

When you decide to reuse the PCE's pre-failure installation, refresh the installation as a standalone PCE:

1. Power on the PCE nodes.
2. On *all nodes of the affected PCE*, run the following command to delete pre-failure directories:

```
$ sudo -u ilo-pce illumio-pce-ctl reset
```

NOTE:
You must run this command on *all nodes* before proceeding to the next step.

3. Copy your backed-up copy of the failed PCE's `runtime_env.yml` file to its location on the newly repaired PCE. See Back Up Leader and Member Runtime Environment Files for information.

   The default location of the PCE Runtime Environment File is `/etc/illumio-pce/runtime_env.yml`. When the location is different on your system, locate the file by checking the value of the `ILLUMIO_RUNTIME_ENV` environment variable.

4. On *all nodes*, bring the nodes to runlevel 1:

```
$ sudo -u ilo-pce illumio-pce-ctl start --runlevel 1
```

5. On *any node*, verify the nodes are at runlevel 1:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

6. On *any node*, run the following command:

```
$ sudo -u ilo-pce illumio-pce-db-management setup
```

Restore Affected PCE's Supercluster Data

1. On *any node of the affected PCE*, verify the runlevel is still 1:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

2. On the *data0 node*, run the following command:

```
$ sudo -u ilo-pce illumio-pce-db-management supercluster-data-restore --
local-pce-file path_to_backup_file --restore-type single_pce
```

The restore operation can take some time to complete. Wait until it finishes before proceeding to the next step.

3. On the *data1 node*, run the following command:

```
$ sudo -u ilo-pce illumio-pce-db-management supercluster-data-restore --skip-
db-restore --local-pce-file path_to_backup_file --restore-type single_pce
```

The `--skip-db-restore` option prevents the command from unnecessarily repeating work that has already been done by previous commands.

Rejoin PCE to Supercluster

1. On *all Supercluster PCEs*, set the runlevel to 2:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
```

Setting the runlevel might take some time to complete.

2. Check the progress to see when the status is `RUNNING` on all nodes:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

3. Rejoin the PCE to the Supercluster. This command can take some time, depending on the number of PCEs in the Supercluster and the size of the PCE databases.

Choose one of the following options, depending on whether you are working on a leader or member.

Rejoining the Leader PCE

On *any core node*, run the following command:

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-restore fqdn_of_failed_cluster
--restore-type single_pce
```

While this command is running, the PCE temporarily sets the runlevel to 1. When the command is interrupted, you might see runlevel 1 unexpectedly.

Rejoining a Member PCE

On *any core node*, run the following command:

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-restore fqdn_of_failed_cluster fqdn_of_supercluster_leader --restore-type single_pce
```

While this command is running, the PCE temporarily sets the runlevel to 1. If the command is interrupted, you might see runlevel 1 unexpectedly.

4. On *every PCE*, set the runlevel to 5:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
```

5. Verify the run level:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

6. Verify that the restored PCE has rejoined the Supercluster and is fully operational:
   a. Log in to the leader PCE web console.
   b. Go to the PCE Health page and verify that the PCE health status is Normal.
7. Check the status of the paired VENs on each PCE. From the PCE web console, choose Workloads and VENs > VENs. After all VENs change status from Active (Syncing) to Active, run the following command *on one PCE at a time*:

```
$ sudo -u ilo-pce illumio-pce-ctl listen-only-mode disable
```

Restore an Entire Supercluster

Restoring an entire Supercluster follows this high-level process:

1. Preparation:
   a. Have the backups of all PCEs ready to use. For each member PCE, have that member's backup and, on the data0 node, a copy of the backup file from the leader. The leader only needs its own backup.
   b. Have copies of every PCE's `runtime_env.yml` configuration file ready to use.

c. Know the IP address, ports, and DNS name of all PCEs in the Supercluster. You must use the same values when you rejoin the PCEs to the Supercluster.

2. Shut down the entire Supercluster.
3. Restore the PCEs. Repeat the following steps for all PCEs in the Supercluster, either one at a time or in parallel:
   a. Reinstall the PCE on new hardware or reuse the installations and `runtime_env.yml` files.
   b. Restore the Supercluster data from backup.
4. Join the repaired PCEs to the Supercluster one at a time.

Prepare to Restore Entire Supercluster

Have the following items ready:

- Backup of each PCE, and the leader's backup copied to each member. See Back Up Supercluster for information.
- Copy of each PCE's `runtime_env.yml` file.
- List of the new IP address, ports, and DNS name for all Supercluster members.

Shut Down Entire Supercluster

On *all nodes of every PCE in the Supercluster*, run the following command:

```
$ sudo -u ilo-pce illumio-pce-ctl stop
```

Install New PCEs or Reuse PCEs

Decide whether you want to completely reinstall the PCEs on new hardware or to reuse the PCE installations.

NOTE:
In both cases, you must reestablish the FQDN of the affected PCE so that VENs can continue to communicate with the Supercluster. When you have any VENs in enforcement, or you rely on DNS-based load balancing, the new IP addresses of the PCE nodes can be different, as long as the new IP addresses were already in the appropriate settings in the `runtime_env.yml` file on all PCE core nodes. See Pre-configure New IP addresses for information.

- To reinstall the PCEs on new hardware, see Deploy a PCE Supercluster for information.
- To reuse the PCE installations, complete the following steps.

When you decide to reuse the PCE's pre-failure installation, refresh the installation as a standalone PCE:

1. On *all nodes of the PCE*, reset the nodes:

```
$ sudo -u ilo-pce illumio-pce-ctl reset
```

NOTE:
You must run this command on *all nodes* before proceeding to the next step.

2. Copy your backed-up copy of the failed PCE's `runtime_env.yml` file to its location on the newly repaired PCE. See Back Up Leader and Member Runtime Environment Files. The default location of the PCE Runtime Environment File is `/etc/illumio-pce/runtime_env.yml`. When the location is different on your system, locate the file by checking the value of the `ILLUMIO_RUNTIME_ENV` environment variable.
3. Bring *all nodes* to runlevel 1:

```
$ sudo -u ilo-pce illumio-pce-ctl start --runlevel 1
```

4. On *any node*, verify runlevel 1:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

5. On *any node*, run the following command:

```
$ sudo -u ilo-pce illumio-pce-db-management setup
```

6. Repeat these steps for all PCEs in the Supercluster.

Restore Supercluster Data

Perform the following steps for *all PCEs in the Supercluster* one at a time or all in parallel.

1. On *any node of the PCE*, verify the runlevel is still 1:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

2. On the *data0 node*, run the following depending on whether you are restoring a member or leader PCE.

   Member PCE

   In `--local-pce-file`, enter the path to the member PCE's backup file. In `--restoring-pce-file`, enter the path to the leader PCE's backup file, which should already be present on the PCE from when you followed the steps in Copy Leader Backup to Members.

   NOTE:
   If necessary, copy the leader PCE's backup file to the data0 node of this PCE before running this command.

```
$ sudo -u ilo-pce illumio-pce-db-management supercluster-data-restore --
local-pce-file path_to_backup_file --restoring-pce-file path_to_leader_pce_
backup_file  --restore-type entire_supercluster
```

   Leader PCE

   In `--local-pce-file`, enter the path to this leader PCE's backup file:

```
$ sudo -u ilo-pce illumio-pce-db-management supercluster-data-restore --
local-pce-file path_to_backup_file --restore-type entire_supercluster
```

   NOTE:
   The restore operation can take some time to complete.

3. On *the data1 node*, run the following command after the restore operation finishes:

```
$ sudo -u ilo-pce illumio-pce-db-management supercluster-data-restore --skip-
db-restore --local-pce-file path_to_backup_file --restore-type entire_
supercluster
```

The `--skip-db-restore` option prevents the command from unnecessarily repeating work that has already been done by previous commands.

4. Set the runlevel to 2:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
```

Setting the run level might take some time to complete.

5. Check the progress to see when the status is `RUNNING` on all nodes:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

For all PCEs in the Supercluster, you must complete all steps in Install New PCEs or Reuse PCEs and Restore Supercluster Data. When finished, proceed to the next task.

Rejoin the PCEs to the Supercluster

Rejoin the leader PCE, then rejoin the member PCEs one at a time in any order.

1. Rejoin the leader PCE to the Supercluster. This command can take some time depending on the number of PCEs in the Supercluster and size of the PCE databases.

   On *any core node*, run the following command:

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-restore fqdn_of_failed_cluster
--restore-type entire_supercluster
```

   While this command is running, it temporarily sets the runlevel to 1. When the command is interrupted, you might see runlevel 1 unexpectedly.

2. Rejoin each member PCE to the Supercluster. This command can take some time depending on the number of PCEs in the Supercluster and size of the PCE databases.

   On *any core node*, run the following command:

```
$ sudo -u ilo-pce illumio-pce-ctl supercluster-restore fqdn_of_failed_cluster
fqdn_of_supercluster_leader --restore-type entire_supercluster
```

   While this command is running, it temporarily sets the runlevel to 1. When the

command is interrupted, you might see runlevel 1 unexpectedly.

3. Repeat step 2 until all PCEs are rejoined to the Supercluster.

Finish and Verify Full Supercluster Restore

After rejoining all PCEs in the Supercluster:

1. On *all PCEs*, set the runlevel to 5:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
```

2. On *all PCEs*, verify the runlevel:

```
$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
```

3. Verify that the restored PCEs have rejoined the Supercluster and are fully operational.
   a. Log into the leader PCE web console.
   b. Go to the PCE Health page and verify that the PCE health status is Normal.
4. Check the status of the paired VENs on each PCE. From the PCE web console, choose Workloads and VENs > VENs. After all VENs change status from Active (Syncing) to Active, run the following command *on one PCE at a time*:

```
$ sudo -u ilo-pce illumio-pce-ctl listen-only-mode disable
```

Import Database to Another Supercluster

This topic explains how to import data from one Supercluster to another Supercluster. For example, you might want to synchronize a test Supercluster with production Supercluster data.

The procedure makes use of two scripts, `remap_supercluster_backup.rb` and `update_supercluster_login.rb`, which are found in `$INSTALL_ROOT/illumio/scripts`.

Back Up Source Supercluster

Back up each PCE in the source Supercluster. See Back Up Supercluster for information.

Restore Backup to Target Supercluster

To import the database, you use a procedure that is similar to restoring a backup with a few extra steps.

Prepare Target Supercluster

1. On *each node in the target Supercluster*, install the PCE software. Use the same software version on the target Supercluster that was installed on the source Supercluster.
2. Copy the backup files to the data0 node of each corresponding PCE in the target Supercluster. When the target Supercluster has fewer PCEs, decide which backups you want to restore.
3. Collect the public IP addresses for each target PCE. You can find them in the `cluster_public_ips` section of each PCE's Runtime Environment File.
4. When you have configured the following settings, verify that they are identical on the source and target Superclusters:
   - `front_end_https_port`
   - `front_end_event_service_port`
   - `front_end_management_https_port`

Remap Supercluster Backup Files

Perform the following steps on each PCE in the target Supercluster.

1. On the *data0 node of the PCE in the target Supercluster*, run the script `remap_supercluster_backup.rb` with the name of the backup file to be remapped and the name of the file in which to write the remapped backup data. You can optionally include the flag `--pce-fqdns-to-skip` with a comma-separated list of fictitious FQDNs that you do not want to include in the remapped database.

   ```
   $ sudo -u ilo-pce remap_supercluster_backup.rb --pce-fqdns-to-skip
   FQDN1,FQDN2... source_backup_file remapped_backup_file
   ```

   For each source PCE FQDN in the backup, the script prompts for the following values:

   - The corresponding target FQDN
   - The corresponding target `cluster_public_ips` from that PCE's runtime environment file

When you have more source PCE FQDNs than target FQDNs, use fictitious names and IP addresses for the extras.

2. When prompted, enter the FQDN of the target PCE where this backup will be restored.
3. After the remapped target backup file is written, copy the file to the data1 node.
4. Repeat these steps on each PCE in the target Supercluster.

Restore Remapped Backup Files

Follow the steps for restoring an entire Supercluster in Restore a PCE or Entire Supercluster, but do not bring the PCEs to runlevel 5. Leave the PCEs at runlevel 2.

Update Login Service on Leader

1. On the *leader PCE*, update login service properties by running the script `update_supercluster_login.rb` and specify the full path to the remapped backup file generated by `remap_supercluster_backup.rb`:

```
$ sudo -u ilo-pce update_supercluster_login.rb remapped_backup_file
```

2. Bring the PCEs to runlevel 5:

```
$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 5
```