Illumio Adaptive Security Platform 18.2.1 PCE Supercluster Deployment and Usage Guide

01/24/2019

60000-100-18.2.1

Table of Contents

Product Version

Illumio® Adaptive Security Platform®

Current PCE Version: 18.2.1

Current VEN Version: 18.2.1

Note: 18.2.1 has not been designated as a Long Term Support (LTS) release. In the future an 18.2.x LTS release will be designated.

About Illumio

Illumio products and services are built on Illumio's patented technologies. For more information, see Illumio Patents.

Illumio Professional Services for Deployment

To ensure optimal deployment of the Illumio Adaptive Security Platform, contact your Illumio Professional Services representative.

Preview Features Only for Evaluation Before General Availability

Any preview features in this release of Illumio Adaptive Security Platform are for your evaluation only.

> **Do not deploy preview features in a production environment**
> Be sure to install these preview features only on non-production systems. To avoid inadvertently impacting your current operations, do *not* install the preview features on production systems.
> The purpose of preview features is to make them more useful for your needs before general availability.

Illumio welcomes your comments and suggestions for improving preview features and documentation. For more information and to send feedback, contact Illumio Customer Support.

Illumio Adaptive Security Platform Training

Illumio offers a wide yet focused training curriculum for Illumio Adaptive Security Platform, from beginning to advanced topics.

To see available courses, log into your Illumio support account and select the **Training** tab.

Search Knowledge Base and Documentation

For useful short articles about Illumio Adaptive Security Platform, log into your Illumio support account and select the **Knowledge Base** or **Documentation** tab.

Illumio Adaptive Security Platform Support

If you cannot find what you are looking for in this document or in support Knowledge Base and Documentation, contact us at:

- support@illumio.com
- +1-888-631-6354
- +1-408-831-6354

Recommended Skills

Illumio recommends that you be familiar with the following:

- Your organization's security goals.
- Solid understanding of Illumio Adaptive Security Platform®ASP.
- General computer system administration of Linux and Windows operating systems, including startup/shutdown, common processes or services.
- Linux shell (bash), Windows PowerShell, or both.
- Understanding TCP/IP networks, including protocols, well-known ports, and the Domain Name System (DNS).
- Familiarity with 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:
  . . .
  some command or command output
  . . .
- References to section titles in this guide are in double quotation marks. Example: See "Basic Theory of Operation".
- Reference to other guides in the Illumio library are *italicized*. Example: See the *PCE Web Console User 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, backup and restore, and workload pairing considerations.
- Basic Theory of PCE Supercluster Operations

Use this guide in conjunction with the *PCE Deployment Guide* and *PCE Operations Guide*.

Overview to PCE Supercluster

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*. The other PCEs are Supercluster *Members*. There is only one Leader in a Supercluster. 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, include an aggregated Illumination map to visualize network traffic and policy coverage for all workloads. Members in the Supercluster have a mostly read-only web console and API for viewing local data.

Workload Management

All PCEs in the Supercluster can manage workloads. You can deploy a "workload-less" Leader with no managed workloads to reduce the load on the Leader to 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 Member, you can generate pairing keys and pairing scripts tied to the Member itself for activation, not the Leader.

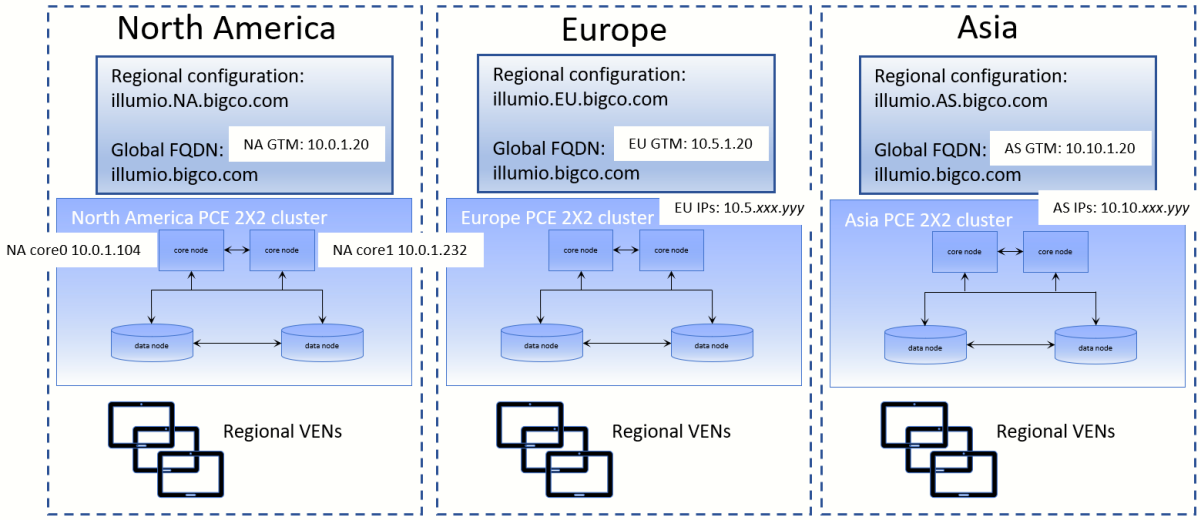
Security Policy Replication

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 ensures the Supercluster can dynamically adapt the policy to changes in the environment, even if 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 if the Leader is unavailable.

Possible Supercluster Logical Network Architectures

The diagram below shows shows geographically distributed data centers for a fictitious company called BigCo.com. This is only one of many possible Supercluster configurations.



PCE Supercluster Deployment Planning

These are preparations you should consider for deploying a PCE Supercluster

Plan Your Supercluster FQDNs With Care

Be sure of the fully qualified domain names you want to use with your Supercluster PCEs. It is important that you have defined these names exactly how you want them before you deploy. Changing FQDNs after creating a Supercluster is possible but time-consuming.

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.

Number of Supercluster PCEs

A PCE Supercluster consists of a minimum of two and a maximum of six PCEs. One of the PCEs is always the Supercluster Leader. 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.

Exact requirements vary on a large number of factors, including, but not limited to:

- Number of managed workloads.
- Number of unmanaged workloads and other labeled objects, such as Bound Services.
- Policy complexity, which includes the following:
    - 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 workload 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.
- Total number of unique flows reported to the PCE from all VENs.

Recommended vs minimum sizes

The capacity planning table below shows minimal and recommended sizes. Illumio encourages you to plan for the recommended sizes. In addition, based on your actual usage and the various factors listed above, your capacity needs might be even greater than the recommended sizes.

There are two configurations for data nodes:

1. A single storage device shared between the data nodes.
2. A dedicated storage device for each data node. This configuration is to accommodate growth in traffic data, which is used by the Explorer. See also "PCE Storage Device Partitions".

MNC Type + Workloads/ VENs[1]	Cores/Clock Speed[2]	RAM per Node[3]	Storage Device Size[4] and IOPS[5]	
			Core Nodes	Data Nodes

2X2
• 2,500 VENs
• 12,500 workloads
	• Four cores per node.
	• Intel® Xeon(R) CPU E5-2695 v4 at 2.10GHz or equivalent.
		32 GB	• 1 x 100 GB
			• 100 IOPS
				• Recommended:
				  • 2 x 250 GB
				  • 600 IOPS per device
				• Minimum:
				  • 1 x 250 GB
				  • 600 IOPS

2X2
• 10,000 VENs
• 50,000 workloads
	• 16 cores per node
	• Intel® Xeon(R) CPU E5-2695 v4 at 2.10GHz or equivalent.
		• Recommended: 128 GB
		• Minimum: 64 GB
			• 1 x 200 GB
			• 100 IOPS
				• Recommended:
				  • 2 x 1 TB
				  • 1,800 IOPS per device
				• Minimum:
				  • 1 x 1 TB
				  • 1,800 IOPS

4X2 • 25,000 VENs • 125,000 workloads	• 16 cores per node • Intel® Xeon(R) CPU E5-2695 v4 at 2.10GHz or equivalent.	128 GB	• 1 x 200 GB • 100 IOPS	• 2 x 1 TB • 5,000 IOPS per device

Footnotes

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

[2] CPUs:

- The recommended number of cores is based only on physical cores from allocated CPUs, irrespective of hyper-threading or virtual cores. For example, in AWS one vCPU is only a single hyperthread running on a physical core. that is. half a core. So 16 physical cores equates to 32 vCPUs in AWS.
- Full reservations for vCPU. No overcommit.

[3] Full reservations for vRAM. No overcommit.

[4] Storage requirements for network traffic data can increase rapidly as the amount of network traffic increases. Allocating a separate, large storage device for traffic data can accommodate these rapid changes without potentially interrupting service.

[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.

For more than 100 IOPS, either mixed-use Solid-State Disk (SSD) or Storage Area Network (SAN) is required. Locally attached, spinning hard disk drives (HDD) are not sufficient.

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".

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

Device	Partition mount point	Size to Allocate	Node Types	Notes
Device 1	`/`	16 GB	Core, Data	
Device 1	`/var/log`	32 GB	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	Partition mount point	Size to Allocate	Node Types	Notes
Device 1	`/var/lib/illumio_pce`	Balance of Device 1	Core, Data	
Device 2 in two-storage-device configuration	`/var/lib/illumio_pce/data/traffic_datastore`	All of Device 2 (250GB)	Data	For network traffic data, in a two-device configuration for the data nodes, this should be a separate device that is mounted on this directory. `/var/lib/illumio_pce/data/traffic_datastore` is the default value of network traffic datastore's `traffic_datastore:data_dir` runtime environment setting. When you change the defined path in this parameter, make sure that the new value matches the path you actually mount.

PCE 2x2 Multi-Node Cluster for 10,000 VENs or PCE 4x2 Multi-Node Cluster for 25,000 VENs

Storage Device	Partition mount points	Size to Allocate	Node Types	Notes
Device 1	`/`	16 GB	Core, Data	
Device 1	`/var/log`	32 GB	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.

Storage Device	Partition mount points	Size to Allocate	Node Types	Notes
Device 1	`/var/lib/illumio_pce`	Balance of Storage Device 1	Core, Data	
Device 2 in two-storage-device configuration	`/var/lib/illumio_pce/data/traffic_datastore`	All of Storage Device 2 (1 TB)	Data	For network traffic data, in a two-device configuration for the data nodes, this should be a separate device that is mounted on this directory. `/var/lib/illumio_pce/data/traffic_datastore` is the default value of network traffic datastore's `traffic_datastore:data_dir` runtime environment setting. When you change the defined path in this parameter, make sure that the new value matches the path you actually mount.

Runtime parameters for Two-Storage-Device Configuration

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

If you are deploying the two-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 following are recommended values for these parameters based on cluster type and estimated number of workloads.

Setting	PCE 2x2 Multi-Node Cluster for 2,500 VENs	PCE 2x2 Multi-Node Cluster for 10,000 VENs	PCE 4x2 Multi-Node Cluster for 25,000 VENs	Note
`traffic_datastore:max_disk_usage_gb`	100 GB	400 GB	400 GB	This size reflects only part of the required total size, as detailed in "PCE Capacity Planning".
`traffic_datastore:partition_fraction`	0.5	0.5	0.5	
`traffic_datastore: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` **Same port for all PCEs in Supercluster** 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	PCE FQDN of all other PCEs

Ports	Sources	Destinations
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 PCEs

Load Balancers

As with 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.

Global Server Load Balancing (GSLB) or a manual DNS update can be used to fail over VENs to a different PCE. See "GSLB Requirements" and "Supercluster HA".

Traffic Load Balancer Configuration

If 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 the following additional inbound connections originating from the other PCEs to an available, healthy core node.

- TCP 5432
- TCP 8302

GSLB Requirements

Workloads can be paired to a specific PCE, or you can optionally use a Global Server Load Balancer (GSLB) to route workloads to the desired PCE in your Supercluster.

If 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/ geolocation.
- GSLB persistence (also known as ‘stickiness’) must be enabled so workloads are always routed to the same PCE 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 for each PCE to have a configured "buddy" PCE that the GSLB uses for failover.
- If a PCE is unavailable for an extended period of time (such as 24 hours), its workloads can be routed to one of the following:
- Workload failover time depends on the DNS time-to-live (TTL) configured in the GSLB.
- Illumio strongly recommends that workload failover using GSLB not be fully automated but instead be initiated by a human.

Optionally configure SAML IdP for User Login

After installation, you can configure the PCE to rely on an external, third-party SAML identity provider system. See the section "Single Sign-On" in the *PCE Web Console Guide*. The guide has step-by-step details 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 Web Console User 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 if you move VENs from one PCE to another via the REST API.

An exception is made if 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 a temporary exceeding of the VENs per PCE limit if there is an extended outage to a PCE in the Supercluster.

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

RBAC Permissions: Leader or Member

In general, if you are using the Illumio PCE web console or the Illumio Adaptive Security Platform REST API, the types of operations you can perform depend on your PCE Role-Based Access Controls (RBAC) permissions and whether you have logged in to 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

User Role	Operations	Leader	Members
Ruleset Manager	Create, update, and delete Rulesets within defined Scopes.	Yes	No
Ruleset Provisioner	Provision Rulesets within defined Scopes.	Yes	No

Optionally Configure PCE Internal syslog on Leader

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

> 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 login to the Supercluster leader and configure internal syslog for each member individually.
> If the events prior to joining a PCE as a member are important to preserve, backup the PCE before you join it to the Supercluster.

See details on the PCE internal syslog in the *PCE Deployment Guide*.

PCE Control Interface illumio-pce-ctl and other 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.

The PCE also includes two other command-line utilities used to setup and operate your PCE:

- `illumio-pce-env`. Used for verifying and collecting information about the PCE runtime environment.
- `illumio-pce-db-management`. Used for PCE database management.
- `supercluster-sub-command`. Used for Supercluster specific operations.

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

In this document, all command-line examples assume an RPM installation. If you installed the PCE tarball, you will need to modify the commands based on your PCE user account and the directory where you installed the software.

Control command access via /usr/bin. By default, for easier command execution, the installation of the PCE creates softlinks in `/usr/bin` for the Illumio PCE control commands. The `/usr/bin` directory is usually included by default in the PATH environment variable in most Linux systems. If for some reason 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

In this document, all command-line examples assume a RPM installation. If you installed the PCE tarball, you will need to modify the commands based on your PCE user account and the directory where you installed the software.

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

```
$ 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 Supercluster

You can deploy the Illumio Adaptive Security Platform Supercluster in several ways:

- **New**
  You have never deployed a PCE and want to deploy a brand new Supercluster. See "Deploy a new Supercluster".
- **Expand**
  You have already deployed a standalone PCE and want to expand it to a Supercluster. See "Expand an existing PCE to a 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_env.yml 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.

Additionally, 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:

If the nodes of each PCE use multiple IPs addresses, or if they will 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, then configure this optional parameter:

- `supercluster.node_public_ip`. The public IP address of this node to advertise 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. If your PCE is deployed in a public cloud, such as AWS, this must be a public IP address.

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

- `supercluster.fqdn`. If set, 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 instructions on how to install a PCE, see the *PCE Deployment Guide*.

Install Members

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

For instructions on how to install a PCE, see the *PCE Deployment Guide*.

Initialize Supercluster Leader

After the Leader has been installed, configured, and verified, the next step is to initialize the Leader.

Note: You must initialize the Leader *before* you start joining any Members.

1. Run the following command on any node to bring all nodes to runlevel 2:

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

2. Setting runlevel might take some time to complete. 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
```

When all of the nodes have reached runlevel 2, the output displays the following:

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

3. After the cluster is at runlevel 2, run the following command on any node to initialize the Leader:

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

4. Bring this new Leader PCE to runlevel 5 by running the following command:

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

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.

In the next step, you will join the new Member to the Supercluster.

1. Run the following command on any node in the cluster to bring all nodes to runlevel 2:

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

2. Then, run the following command on any node as you wait for all nodes to reach runlevel 2 before proceeding:

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

3. On any Core node or the Data0 node of the Member cluster, run the following command to join the Member to the Supercluster (identified by the Leader's FQDN).

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

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

4. Bring this new Member PCE to runlevel 5 by running the following command:

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

5. Repeat for each Member you want to join to the Supercluster.
6. Finally, restart all PCEs in the Supercluster by running the following command on each PCE:

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

Verify Supercluster is ready to use

Before you begin using your Supercluster, you should verify that the Leader and Members are all joined together and all PCEs in the Supercluster have a good health status.

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

To verify that your Supercluster is ready to use:

1. Log in to the Leader.
2. Run the following command on any Core node to show Supercluster membership:

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

   The output should show all PCEs in your Supercluster.
3. Log in to the PCE web console of the Leader.
4. Click the Health status button at the top of the web console. You should see all PCEs in your Supercluster with **Normal** health.

Expand Standalone PCE to Supercluster

If you want 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 is as follows:

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 Before Expansion, Change Parameter pce_fqdn and Verify Health of the Standalone PCE.
2. Verify network connectivity to the database nodes. See "Before Expansion Ensure Network Connectivity from All Standalone PCEs to Database Nodes".
3. Initialize your existing PCE as the Supercluster Leader. See "Initialize Supercluster Leader".
4. Install and configure the new PCEs that will become Members of the new Supercluster. See the PCE Deployment guide for instructions.
5. Join Members to the Supercluster. See "Join each Member to Supercluster".

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 Parameter pce_fqdn and Verify Health of the Standalone PCE

Read this KB article about how to change the PCE's FQDN.

Verify standalone PCE health
After changing the `pqce_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.

Before Expansion, Ensure Network Connectivity from All Standalone PCEs to Database Nodes

Before expansion to 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, you can 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
```

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, suppose you have three regions. With the following `nc` commands on Data 0 and Data1 in each

region, test the connection to the other regions by connecting to port 5432 for the other regions' Supercluster FQDN.

From Region 1: Set the PCE from which you are testing to runlevel 2.

- `nc -zv region2fqdn 5432`
- `nc -zv region3fqdn 5432`

From Region 2: Set the PCE from which you are testing to runlevel 2.

- `nc -zv region1fqdn 5432`
- `nc -zv region3fqdn 5432`

From Region 3: Set the PCE from which you are testing to runlevel 2.

- `nc -zv region1fqdn 5432`
- `nc -zv region2fqdn 5432`

Migrate Existing Supercluster to New Supercluster

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

1. In `runtime_env.yml` on all nodes, pre-configure the IP addresses of the new Supercluster. See "Before Migration, Pre-configure New IP addresses for DNS-based load balancing".
2. Backup the current Supercluster. See "Backup Supercluster".
3. Restore the old Supercluster configuration and data to new systems. See "Restore PCE or Entire Supercluster".

Before Migration, Pre-configure New IP addresses for DNS-based load balancing

If you rely on DNS-based load balancing (such as round-robin DNS) and you are using new IP addresses for the restored PCE, before the migration, 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.

> **Traffic load balancers**
> If you rely on 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, and under the `supercluster.node_public_ip` parameter, add the new IP addresses of all new core nodes, as shown in this snippet:

```
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>
```

To send the configuration update to all members, on the leader, run the following command to restart the Supercluster:

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

Upgrade Supercluster

Before upgrade -- backup

Before the upgrade, backup of the leader and all Member databases and each PCE's `runtime_env.yml` file. See "Backup Supercluster".

Before upgrade – All PCEs in 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 Heath page to make sure the PCE health status is Normal.

Types of Upgrade

The following are the types of upgrade:

- Supercluster simple upgrade: The Supercluster simple upgrade procedure requires you shut down the entire Supercluster for the duration of the upgrade. During a simple upgrade, the Supercluster is not fully operational.
- Supercluster rolling upgrade: Rolling upgrade keeps the Supercluster operational while individual PCEs are successively upgraded.

Upgrading from Supercluster Version 18.1

> **Version-dependent upgrade paths**
> If you plan on upgrading from Supercluster version 18.1 to version 18.2.1, use the simple upgrade. Do *not* use the rolling upgrade.
> Rolling upgrade is supported from version 18.2.0 forward

Supercluster simple upgrade

A Supercluster simple upgrade follows these general steps.

> **On all PCEs in the Supercluster**
> Do these steps on all PCEs in the Supercluster.

- Bring all PCEs in the Supercluster to runlevel 2.
- On all PCEs, quiesce the data replication.
- Upgrade the software on all nodes of all clusters.
- Promote replication on the leader of the Supercluster.
- Bring all clusters back to runlevel 5.

Steps

1. On any node in the cluster, set runlevel 2.

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

2. Setting runlevel might take some time to complete. 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
   ```

3. To determine the database master, run the following command on any data node.

   ```
   $ sudo -u ilo-pce illumio-pce-db-management show-master
   ```

4. On the database master node, run the following command. This command waits for data replication to finish, which can take some time. To stop waiting, you can timeout the command with the final argument in seconds shown below. Default timeout is 600 seconds. If the command times out, you must rerun it.
   $ **sudo -u ilo-pce illumio-pce-db-management supercluster-quiesce *timeout_in_seconds***
   When the Supercluster is quiesced, the command prints the following message:

   ```
   Replication is complete.
   ```

5. Stop the PCE cluster by running the following command on any node:
   $ **sudo -u ilo-pce illumio-pce-ctl cluster-stop**
6. On every node of every PCE, install the new version of PCE.

7. Start the cluster with runlevel 1 by running the following command on any node:
   $ **sudo -u ilo-pce illumio-pce-ctl start --runlevel 1**
8. On any node of every upgraded PCE, migrate the PCE database:

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

9. On every upgraded PCE, set runlevel 2:
   `$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2`
10. Setting runlevel might take some time to complete. 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`

11. Run the following command on one of the leader's Core nodes to apply remaining replication-related changes on the leader. This command cannot be run on a Member PCE.

    `$ sudo -u ilo-pce illumio-pce-ctl supercluster-promote-replication`

12. On each PCE, set runlevel 5 by running the following command any node:

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

13. Verify that you can log in to the PCE web console on each PCE in the Supercluster.

In rare cases you might receive an error when attempting to log in to the PCE web console. If this happens, run the following command on **all** nodes, and then try logging in again:
`$ sudo -u ilo-pce illumio-pce-ctl restart`

Supercluster rolling upgrade

With a rolling upgrade, the Supercluster continues to operate, but the upgraded PCEs are isolated from the other PCEs that are not yet upgraded. Additionally, all PCEs are switched into PCE Listen Only mode and the Leader switches to Read Only, which means that during the upgrade, no changes can be made on the Leader PCE and no policy changes will be provisioned to any of your VENs. Your current policy enforcement remains intact while you upgrade the Supercluster. After you have finished, you can begin making changes and provision policy as normal.

Rolling upgrade requirements, constraints, and conditions

Before starting the Supercluster rolling upgrade, be aware of the following requirements and conditions:

- You *must* upgrade the Leader first, followed by each Member. The rolling upgrade fails if you start by upgrading a Member first.
- Make sure all PCEs in the Supercluster are online and running.
- Any failed Supercluster members must be restored and running.
- You must manually turn off Listen Only mode on all PCEs in the Supercluster. See "Disable Listen Only Mode on All PCEs".
- Each individual PCE must be completely upgraded before you upgrade the next one.

- VEN failover between PCEs is not supported during a rolling upgrade.
- While a rolling upgrade is in progress, assigning a new Leader in the Supercluster is not supported.
- There are no explicit time constraints for completing the Supercluster rolling upgrade.

During rolling upgrade

During the rolling upgrade process, each PCE in the Supercluster is placed into Listen Only mode, and the Leader is placed into Read Only mode for the duration of the upgrade. Consequently, if you attempt to use the Illumio ASP REST API, any PUT, POST, or DELETE API calls will return a 406 HTTP response. GET API calls function as normal.

While a rolling upgrade is in progress, when you log in to the Leader, two notifications appear in the upper right corner. One states that the Supercluster rolling upgrade is in progress, and the other indicates that the Leader is in Listen Only mode. until the upgradee has finished, when you log in to the Leader, these dialogues continue to be displayed.

Additionally, 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 will show Complete when the upgrade is complete. When the upgrade is finished, the Upgrade Status column will no longer appear.

- Upgrade the Leader first before any Members.
- The steps for upgrading a member are identical to the steps for upgrading the leader.
- Be sure to run these commands in the order shown. Running them in a different order can cause problems with the upgrade.

Steps for rolling upgrade of the Supercluster leader

1. Bring the PCE software to runlevel 2 by running the following command on any node:
   `$ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2`
2. Setting runlevel might take some time to complete. Check the progress with the following command to see when the status is **Running**.
   `$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w`
3. Run the following command *on any node except Data1* to prepare the database migration:
   `$ sudo -u ilo-pce illumio-pce-ctl supercluster-upgrade-prepare`
4. Run the following command on any node to stop the PCE:
   `$ sudo -u ilo-pce illumio-pce-ctl cluster-stop`
5. On all nodes of the PCE, upgrade to the new PCE RPM version:
   `$ rpm -Uvh path_to_illumio_pce_rpm`
6. Bring the PCE to runlevel 1 by running the following command on all nodes.
   `$ sudo -u ilo-pce illumio-pce-ctl start --runlevel 1`
7. Setting runlevel might take some time to complete. Check the progress with the following command to see when the status is **Running**.
   `$ sudo -u ilo-pce illumio-pce-ctl cluster-status -w`

8. On any node, start database migration with the following command:
   $ **sudo -u ilo-pce illumio-pce-db-management migrate**
9. When the migration has completed, set the PCE to runlevel 2 by running the following command on any node.
   $ **sudo -u ilo-pce illumio-pce-ctl set-runlevel 2**
10. Setting runlevel might take some time to complete. Check the progress with the following command to see when the status is **Running**.
    $ **sudo -u ilo-pce illumio-pce-ctl cluster-status -w**
11. On any core node, run the following command to rejoin the PCE to the Supercluster:
    $ **sudo -u ilo-pce illumio-pce-ctl supercluster-upgrade-rejoin**
12. Bring the Leader to runlevel 5 by running the following command on any node:
    $ **sudo -u ilo-pce illumio-pce-ctl set-runlevel 5**
13. Repeat these same steps on all members of the Supercluster.

Disable Listen Only Mode on All PCEs

Next, you need to disable PCE Listen Only mode on all PCEs in your Supercluster.

> ⚠ Normally you wait until the entire Supercluster is upgraded and then disable Listen Only mode on all PCEs. If you need to push policy changes from a PCE while an upgrade is in progress, you can disable Listen Only mode on that PCE during the upgrade. However, be aware that the new policy might not be applied consistently across your environment if the changes have not replicated to the other PCEs or the other PCEs have Listen Only mode enabled.

1. On a Core node, run the following command to disable Listen Only node:

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

2. To determine if your PCE is in Listen Only mode, run the following command:

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

Verify Supercluster is Working

The final step performing a Supercluster rolling upgrade is to verify that the health of each PCE is normal.

To validate that the upgraded Supercluster is fully operational, log in to the Leader and check the PCE health for the Supercluster. Each PCE's health should be shown as Normal.

If Rolling Upgrade Fails

- In the event of a problem, all commands in this guide for rolling upgrade can be repeated.
- The Supercluster can remain in the upgraded state while you troubleshoot to pinpoint and resolve the underlying issue.
- Until the upgrade is successfully completed, performance and functions of some aspects of the Supercluster might be degraded.
- If the issue cannot be resolved, the entire Supercluster must be reinstalled.

Supercluster Listen Only Mode – stop sending policy

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 will not modify the firewall rules on any workloads or send any updates from the PCE to the VENs. Also, the PCE will not mark workloads as Offline and will 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.

PCE Listen Only mode and rolling upgrade

During a Supercluster rolling upgrade, the entire Supercluster is placed into Listen Only mode for the entire duration of the upgrade. The leader is also set to Read Only for the duration of the upgrade. After performing a Supercluster rolling upgrade, each PCE must manually be taken out of Listen Only mode. For more information, see "Supercluster Rolling Upgrade".

During a rolling upgrade, if you log in to one of the PCEs, you will see two banners: one that states the Supercluster is in the process of a rolling upgrade, and another that states the PCE is in Listen Only mode. Once the upgrade has finished and you have manually disabled Listen Only mode on each PCE in the Supercluster, the banners will not display.

To enable PCE Listen Only mode:

1. On each of the nodes in the cluster, run the following command to stop the PCE software:

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

2. Set each node in the PCE cluster at run level 1.

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

3. From one of the Data nodes, run the following command to enable Listen Only node:

```
$ 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
```

To 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 each of the nodes in the cluster, run the following command to stop the PCE software:

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

2. Set each node in the PCE cluster at run level 1.

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

3. From one of the Data nodes, run the following command to enable Listen Only node:

```
$ 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
```

Assign New Leader

A Supercluster can only have one Leader at a time. The following section shows you how to permanently choose a new Leader or temporarily assign a new Leader if the old Leader has failed and you need to make changes to the Supercluster before it can be recovered.

Assign New Leader When Old Leader Is Connected

Use the following command if you want to choose a new Supercluster Leader and the old Leader is still running and connected to the rest of the Supercluster. When you choose the new Leader, the former Leader will become a Member in the Supercluster.

Before you begin, choose a PCE that will be the new Leader, and then run the following commands:

1. On both the current Leader PCE and the new PCE you want to make the Supercluster Leader, run the following command on any node to bring both PCEs to runlevel 2:

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

2. Next run the check cluster status command to ensure that the software is running on all nodes. Make sure you wait until the software is running before you proceeed.

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

3. On the PCE you want to become the new Leader, run the following command on any node to make it the new Leader.

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

4. On both the new Leader and the old Leader, run the following command on any node to bring both PCEs to runlevel 5:

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

5. Next run the check cluster status command to ensure that the software is running on all nodes.

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

Assign New Leader When Leader Has Failed

In the event that your Supercluster Leader has failed, you must first drop the failed Leader from the Supercluster before you can assign a new Leader.

⚠ When the new Leader is promoted, you must isolate the former Leader from the network and not allow it to be brought back online. If the former Leader is not isolated it will incorrectly re-join the Supercluster as 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 the section titled "Restore an Individual PCE", 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. Log into a core node of **each** surviving PCE in the Supercluster and set their run level to 2:

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

2. Run the following command on the PCE you assigned as the new Leader to drop the failed Leader from the Supercluster:

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

3. On the PCE that you have designated as the new Leader, run the following command:

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

4. Next, run the following command and wait until the Cluster status returns: RUNNING.

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

5. Next, run the following command on the newly designated Leader PCE to assign is as the new Supercluster Leader:

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

6. Next, run the following command on the new Leader PCE to set it back to runlevel 5:

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

7. Next run the check cluster status command to ensure that the software is running on all nodes. Make sure you wait until the software is running before you proceeed.

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

Supercluster VEN Management

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 US.

In some cases, however, 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. Additionally, in cases of a PCE failure, you may want to be able to temporarily failover the workloads to another healthy PCE. In both cases, the result is that a set of workloads are managed by a "new" PCE.

Unmanaged Workloads

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

Dealing with VENs paired to a disconnected PCE

A PCE that has lost connectivity to the VEN maintains the "online" status of the VENs and retains the workloads in policy. This condition can be corrected with the following general steps:

1. Determine the cause of the PCE failure and correct it.
    a. You can restore the failed PCE.
    b. 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. Optionally, with either the Web Console or the REST API, you can delete records of the incorrectly marked "online" VENs.

The final step is optional, because after VEN heartbeating 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 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 Membe
- 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.

Example 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
```

Example 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 with Leader or Member

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 details about pairing, see the *VEN Deployment Guide*.

Pair workloads with a GSLB-determined PCE

If you rely on a Global Services Load Balancer (GSLB) to control which specific PCE a workload communicates with, 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.

> 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 is a snippet from the generated pairing script after the `supercluster_fqdn` parameter has been set.

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

Reassign Workloads with REST API

See "Reassign VENs to a Different PCE using the REST API".

Workload Uptime/Last Heartbeat in a Supercluster

Each workload managed by your Supercluster provides the latest 'Uptime' of the workload, which is 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 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.

Blocked Traffic in a Supercluster

For each PCE in a Supercluster, Leader or Members, the **Blocked Traffic** page shows blocked traffic only from workloads that have been paired with that PCE.

Delete Blocked Traffic on Members and Leader

One operation you can perform on both Members and the Leader in a Supercluster is deleting blocked traffic events, using either the PCE web console or the Illumio Adaptive Security Platform REST API.

Your user account must have the Global Administrator user role in order to delete blocked traffic.

- In the PCE web console, from the left navigation menu, select Troubleshooting → Blocked Traffic. From this page you can filter the blocked traffic list and select blocked traffic events you want to remove. **Note**: The Blocked Traffic page in a Supercluster only shows blocked traffic events from the PCE you are logged in to.
- With the Blocked Traffic API, you can GET and DELETE one or multiple blocked traffic events,

The Blocked Traffic API is Public Experimental, which means the API is fully functional but is subject to change in later releases.

If you are using the Illumio Adaptive Security Platform REST API to delete blocked traffic, the URIs re as follows (allowed on Members and Leader):

Get a collection of blocked traffic events:

```
GET [api_version][org_href]/blocked_traffic
```

Get an individual blocked traffic event:

```
GET [api_version][blocked_traffic_href]
```

Delete a collection of blocked traffic events:

```
PUT [api_version][org_href]/blocked_traffic
```

Delete an individual blocked traffic event:

```
DELETE [api_version][blocked_traffic_href]
```

Workload Support Reports in a Supercluster

If you are logged into the Leader of a Supercluster, you can generate and download workload support reports for any workload in the Supercluster. This includes workloads that have been paired with and are being managed by other Members.

From a Member PCE you can generate a support report for all workloads connected to that PCE. However, you cannot generate a support report from a Member PCE for any workloads connected to different PCE

As soon as support reports are finished, you can download them from the Leader PCE web console.

For information on running workload support reports from the command line on the host, see the *PCE Operations Guide*.

Updating workloads on leader during member Failure

If one of your Member PCEs goes down, any changes you make to workloads managed by the affected Member (while logged into the Leader) will be immediately reflected in the Leader UI, even though the change has not been replicated to the Member and applied on the workload.

For example, if one Member of your Supercluster fails, and 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

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

To failover 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.

If you have configured the `supercluster.fqdn` parameter in your `runtime_env.yml` file, then the target PCE of all workloads will be 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 will be lost when VENs failover 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 failover to. If a VEN fails over and cannot validate the certificate of the new PCE, it will not be able to authenticate and enter 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. This scenario is grave, however, because the PCE that was managing the VEN is still running and will mark the workload as offline in 1 hour, which in turn isolates it from all other workloads.

VEN Failover When PCE Fails Immediately After Pairing

In rare cases, if 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 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 HTTP error 401 Unauthorized and eventually is moved into the Lost Agent state.

To recover from this situation, you have two options:

1. Uninstall the VEN completely from the workload, then repair with a functioning PCE.
   OR
2. Recover the affected PCE and once it is fully functional and online, the VEN will automatically come out of the Lost Agent state when it successfully heartbeats to a PCE.

   This recovery will only work if the affected PCE had information about that VEN before the failure. If you recover the PCE from a backup that was taken before the VEN was paired, then the VEN will have to be uninstalled and the workload repaired.

Supercluster Health Monitoring

There are two general methods for monitoring the health of your PCE Supercluster:

- REST API calls that determine the Supercluster leader and a PCE's health.
- The PCE web console, which displays the health of the entire Supercluster from the leader, or individual Member health if logged in to a Member.

This section discusses health monitoring specifically for a PCE Supercluster. In addition to the information here, you should also follow the PCE health monitoring guidelines in the *PCE Operations Guide*.

REST API for Supercluster health

REST API mechanisms for working with Supercluster health are detailed in this section.

REST API /health

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

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

REST API /supercluster/leader – determine leader

You can use this Public Stable REST API call to determine if PCE in a Supercluster is a leader or Member. This call can be made by your GSLB in order to monitor health of the leader.

```
GET [api_version]/supercluster/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 knows which is the Supercluster leader, make the following REST API call to monitor the leader's availability.

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

HTTP response code from /node_available

There can be up to a 30 second delay for the health check API 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 on a Supercluster provides health information about your on premise PCE, whether you deployed a 2X2, 4X2, Supercluster, or PCE virtual appliance.

- **General PCE Health**. Shows general health information for each PCE in your Supercluster, such as health status, node status and uptime, as well as system health information for each node (CPU usage, memory, disk usage, and more. If you deployed a PCE Supercluster, then this page will list 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, with the ability to click to view the health each individual 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 also 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 also 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. Specifically:

- **Supercluster Replication** (**Lag**). Indicates how long it took for one PCE to receive replicated data from another PCE in the Supercluster. For example, if someone on the leader created a new IP List and saved it, and the change took 4 seconds to replicate to a Member1, then the Member health page will show that the replication lag for Member1 is 4 seconds behind the leader. Replication lag is shown 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 on Members, who periodically send traffic data to the leader for a full picture of Illumination traffic for your entire

Supercluster. You can initiate a sync of Illumination data on demand by clicking the small link at the lower right of the Illumination map.

Supercluster PCE health icon badge

If the PCE health button has a small badge with a number it means that one or more of the PCEs in your Supercluster have a health status of ***not*** "Normal". The color of the badge indicates the type of warning.

For example, a yellow warning badge on the button with the number 1: means that one of the PCEs in the Supercluster has a health status of Warning.

If the badge is red and showed the number 1, it means that one of the Supercluster PCEs has failed and/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 **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.

Status	Color	Definition
Warning	Yellow	A PCE is considered to be in **warning** state when: • One or more nodes are unreachable. • One or more optional services are missing, or one or more required services have been 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 >30 seconds.
Critical	Red	A PCE is considered to be in **Error** state when one or more required services are missing. **Note**: In this scenario it may not be possible to authenticate to the PCE or get an API response depending on which services are missing from the PCE.

PCE health on workload details

If your workloads have been paired to a Supercluster leader or Member, then you can also view PCE health on the workload details page, Summary tab. On this page, there is a section named PCE which lists the hostname and health of the PCE that this workload is paired with.

PCE health on Illumination workload command panel

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

Command-line show all Supercluster members

Run the following command on any Core node or the Data0 node in a cluster to display all members of the Supercluster, including the Leader and all Member PCEs.

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

Backup Supercluster

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

Different data is backed up depending on if you run the backup from the Supercluster Leader or a Member:

- **Leader backup**: Contains all of Supercluster replicated data, including workloads, Labels, Rulesets/Rules, Services, Organization Events, workload traffic data, and Supercluster configuration data.
- **Members 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`, but for the exact location on your systems, check the value of the `log_dir` parameter

When to Backup

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

Illumio also recommends taking backups in the following scenarios:

- 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, applying Labels)
- After provisioning major policy changes
- After making major changes in your environment that affects workload information (such as an IP address change)
- 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: migrate, upgrade, and so forth.

Determine Data node of each PCE for Backup

For each PCE, you must run the backup command on the node that runs the `agent_traffic_redis_server` service.

Note: Check for `agent_traffic_redis_server` on a Data node before every backup, because this service can be running on either Data node. To find out which node runs the service, use the `cluster-status` command. The output indicates which Data node to backup.

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

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          <=== backup command should run on this
 node
agent_traffic_service NOT RUNNING
...
```

Backup each PCE's data

For the Leader and every Member PCE in your Supercluster, perform these steps.

1. Login to 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 execute 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.

Backup Leader and Member PCE runtime_env.yml file

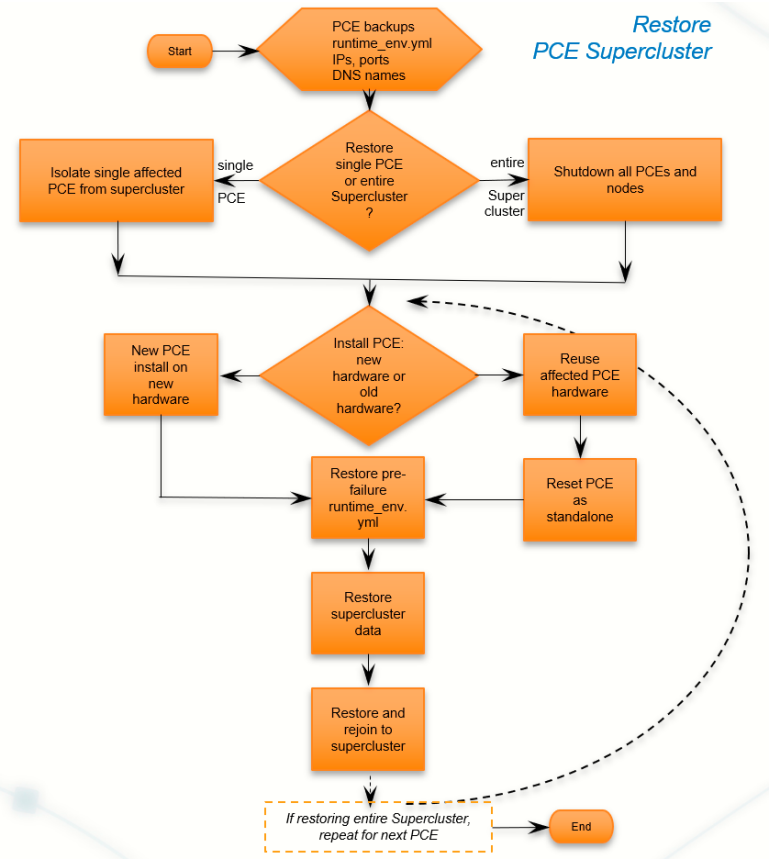
Store a copy of each node's `runtime_env.yml` file on a system that is not part of the Supercluster. The default location of the PCE Runtime Environment File is `/etc/illumio-pce/ runtime_env.yml`, but for the exact location on your systems, check the value of the `log_dir` parameter

Restore Single PCE or Entire Supercluster

This section describes how to restore a single failed PCE, either Leader or Member, and rejoin it to a Supercluster or how to restore the entire Supercluster.

- **Restore an Individual PCE**. Only one PCE in the Supercluster has failed and needs to be restored, which could be the Leader or one of the Members. You isolate that PCE from the Supercluster, restore it, and rejoin it to the Supercluster. If more than one PCE has failed you must restore the entire Supercluster.

- **Restore Entire Supercluster**. More than one PCE or the entire Supercluster has failed and needs to be restored.
  To restore an entire failed Supercluster, for each affected PCE of the Supercluster, follow the same general process while *cycling in turn through each affected PCE*.
  a. Stop all PCEs in the Supercluster.
  b. Start with one of the affected PCEs and restore it. This PCE is now restored and running.
  c. Move to the next PCE and restore it. This second PCE is now restored and running.
  d. Repeat the above for all remaining affected PCEs. At the end, all affected PCEs are restored and running.



The general process for restoring is as follows:

1. Preparation:
   a. Have your backups and your copy of the affected PCE's `runtime_env.yml` configuration file ready to use.
   b. Be sure you know the IP address, ports and DNS name of the affected PCE or the same information for all PCEs in the Supercluster. You must use the same values for rejoining to the Supercluster.
2. Isolate the single affected PCE from the Supercluster or shutdown the entire Supercluster
3. Decide: do a new PCE installation on new hardware or reuse the installation on the affected PCE.
4. Restore the failed PCE's `runtime_env.yml` file from backup.

5. Restore the Supercluster data from backup.
6. Join the repaired PCE to the Supercluster.

Prepare for Restore

Have the following ready:

- The backup of the failed PCE from section "Backup Supercluster".
- The backup copy of the failed PCE's `runtime_env.yml` file
- The IP address, ports, and fully qualified domain name of the failed PCE to reconfigure for the repaired PCE or a list of the new IP addresses for all Supercluster members

Isolate a single affected PCE or Shutdown Entire Supercluster

Single PCE: To restore a single affected PCE, isolate that single PCE from the Supercluster:

1. Shut down the affected PCE:
   ```
   $ sudo shutdown -h now
   ```
2. Log into a core node of **each** surviving PCE in the Supercluster and set their runlevel to 2:
   ```
   $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
   ```
3. Log into any core node of a surviving PCE and drop the failed PCE from that surviving PCE :
   ```
   $ sudo -u ilo-pce illumio-pce-ctl supercluster-drop fqdn_of_failed_pce
   ```

Entire Supercluster: Shutdown all PCEs in the Supercluster.

- Shut down all the PCEs in the Supercluster:
  ```
  $ sudo shutdown -h now
  ```

Decide - New PCE on New Hardware or Reuse Affected PCE

Decide whether you want to do a completely new installation of the PCE on new hardware or to reuse the PCE installation already on the affected system.

- Do a new installation, see "Deploy Supercluster".

or

- Reuse the affected PCE installation. In this case, you must run an additional command to delete some pre-failure directories. See below.

In either case, you must reestablish fully qualified domain name of the affected PCE, so that VENs can continue to communicate with the Supercluster. The IP addresses can be different. If you rely on DNS-based load balancing, the new IP addresses must be recorded in the `runtime_env.yml` file on all member PCE core nodes. See "Before Migration, Pre-configure New IP addresses for DNS-based load balancing".

Option: On Reused PCE hardware, refresh PCE as standalone

If you decide to reuse the PCE's pre-failure installation, you need to refresh the installation as a standalone PCE.

1. On every node of the affected PCE, run the following command.

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

   Note: *Reset all nodes before going to the next step.*

2. After you have reset all nodes of the affected PCE in the preceding step, bring every node to runlevel 1:

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

3. Verify runlevel 1 on any node.
   ```
   $ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
   ```

4. On any node, run the following:

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

Restore the Affected PCE's runtime_env.yml file

Put your backed up copy of the failed PCE's `runtime_env.yml` file to its location on the newly repaired PCE. See "Backup PCE runtime_env.yml file on Leader and Members".

The default location of the PCE Runtime Environment File is `/etc/illumio-pce/ runtime_env.yml`, but for the exact location on your systems, check the value of the `log_dir` parameter

Restore the Affected PCE's Supercluster Data

1. Bring the PCE to runlevel 1.
   ```
   $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 1
   ```
2. Verify runlevel 1 on any node.
   ```
   $ sudo -u ilo-pce illumio-pce-ctl cluster-status -w
   ```

3. Restore the Supercluster data. The data restore can take up to one hour to complete.

On Data 0 node:

- For single PCE restore , on any core node:

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

- For entire Supercluster restore: Use the `--restore-type entire_supercluster` option.

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

On Data 1 node. The data has already been restored by the previous command. Below, we use the `--skip-db-restore true` option to recreate only the Supercluster metadata.

- For single PCE restore:

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

- For entire Supercluster restore: Use the `--restore-type entire_supercluster` option.

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

Restore and Rejoin the PCE to the Supercluster

1. On all Supercluster PCEs, set runlevel to 2.
   ```
   $ sudo -u ilo-pce illumio-pce-ctl set-runlevel 2
   ```
2. Setting runlevel might take some time to complete. 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
   ```
3. Restore and rejoin the PCE to the Supercluster. This command can take up to one hour depending on the number of PCEs in the Supercluster and size of the PCE databases.

Restoring and rejoining the Leader PCE:

- For single PCE, on any core node:

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

- For entire Supercluster restore, on any core node, use the `--restore-type entire_supercluster` option.

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

Restoring and rejoining a Member PCE. On one of the Data nodes:

- For single PCE single, on any core node:

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

- For entire Supercluster restore, on any core node, use the `--restore-type entire_supercluster` option.

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

4. After restoring all failed PCE(s) in the Supercluster, set runlevel to 5:

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

5. Verify runlevel:

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

The restore is complete.

Verify that the restored PCEs have rejoined the Supercluster and are fully operational.

a. Login to the PCE.
b. Check the PCE Heath page to make sure the PCE health status is Normal.

Supercluster PCE Web Console

Many uses of the Supercluster PCE web console are detailed in this guide. For general information about the PCE web console, including the traffic explorer, see the *PCE Web Console User Guide*.

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, will look different depending on if 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
- Rule Builder
- App Group configuration

The following data are not replicated in a Supercluster. They are available only on the Leader itself and the individual Members themselves:

- VEN heartbeat and uptime

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 may take several minutes to complete.

Depending on your network speeds and possible latency, the Illumination map's traffic data may 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. If you are viewing the Illumination map on a Member, you a message indicating that you are viewing a local set of traffic data.

Web Console Filtering Problem in Supercluster Member, with Workaround

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. The filter combination is as follows:

Policy Sync: Active and PCE:Member PCE FQDN

REST API and Supercluster

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

Regardless or your user's permissions, you can only perform "read" operations on a Member, which in terms of REST means you can perform GET operations on Members, but not any POST, PUT, or DELETE operations.

On the Leader, you can perform full CRUD (GET, POST, PUT, DELETE) operations, given your user has the permissions to do so. Other API calls 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
Generate a workload support report	Yes	Yes

REST Operation	Leader	Members
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 cannot make any PUT, POST, or DELETE API requests, which will return a 406 HTTP response if. You can, however, perform any GET calls. Note that during a simple upgrade, you cannot use the REST API at all until the upgrade has finished. For more information, see "Upgrade Supercluster".

REST API Login Response

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

For example, if you make this call to log in to 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 will have a value of either member or leader. For example, you will see this response from a Leader when you log in:

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

Reassign VENs to a Different PCE using the REST API

When deploying a Supercluster, you might want to "move" workloads that have been paired to one PCE so that they are managed a different PCE in the Supercluster. For example, if you currently have a single standalone PCE and then expand that PCE into a Supercluster, you may want to reassign some of your existing VENs to be managed by the nearest PCE. In these cases you can reconfigure the VEN on a paired workload so that it uses a different FQDN to communicate with the proper PCE.

Using the Illumio Agent API ("agent" is the API name used for the Illumio VEN), you change the target PCE of the workload to be the PCE you want to reassign the workload to. The PCE that is currently managing the workload will send the workload the FQDN of the new target PCE, after which the workload will begin heartbeating to and receiving its policy updates from this PCE. At this point, the active PCE of the workload is the same as the target PCE.

Be aware that manually moving a VEN to a different PCE via the REST API is subject to the object limit `active_agents_per_pce`. For more discussion, see "Object Limits and Supercluster".

Terms: Active and Target PCE

To manage moving VENs from one PCE to another, you need to be familiar with these two terms: active PCE and target PCE, which correspond to two properties that are added to a workload's VEN upon pairing.

- "`active_pce_fqdn`", the PCE that is currently managing this workload. I.e. 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 (if you have configured the `supercluster.fqdn` property in your `runtime_env.yml` file)

Before you Begin

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

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

Workload Reassignment Workflow

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

1. **GET workloads**. In order to find the HREF of the agent on a workload, you need to get a collection of workloads from the PCE. Or, if you already know the HREF of a workload, you can also get an individual instance of the workload which will return the HREF of the agent that was used to pair that workload.
2. **Identify agent HREF**. Included in the response of getting a workload or multiple workloads is a property named 'agent', which represents the VEN that has been installed on the workload as part of the pairing process. The agent is identified by its HREF.
3. I**dentify 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 used 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 the VEN can be managed by a different PCE in your Supercluster.

Get workloads

In order to get the HREF of an agent (VEN) on a workload, you need to get a collection of workloads. You can GET up to a maximum of 500 workloads at a time, or if you know the HREF of an individual workload you can get just the single workload.

To get a collection of workloads, you can 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, notice the section that begins with the '`agent`' property, which shows the HREF of the VEN (`href":"/orgs/3/agents/40916`). Notice also in the response that the Active PCE (`active_pce_fqdn`) and the Target PCE (`target _pce_fqdn`) are the same. This will 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

Now that you have the agent HREF, you can update the the target pce with the PCE FQDN for the VEN to use. In your JSON request body, pass the following data:

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

The URI for this operation is as follows:

```
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:'xxxxxxx5048a6a85ce846a706e134ef1d4bf2ac1f253b84c1bf8df6b83c70d95'
-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, check that the active PCE matches the target PCE. You can perform a GET on the agent again, and both target and active PCE FQDN should be the same. If the operation is successful, the response will 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
```


Basic Theory of PCE Supercluster Operations

To illustrate how a PCE Supercluster works, we will use the example of a three-tier application (web, processing, database) that is deployed across three data centers in the US, Europe, and Asia. Each data center 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 data center while allowing the database tier to replicate across data centers.

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 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 our example, when we pair a new instance of the database in the US, the following occurs:

1. The US PCE sends firewall rules to the US database workload.
2. The US PCE also sends send new firewall rules to the US web and processing workloads since 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 since the policy allows these database 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 (i.e., 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 then propagated to all members. On the 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 Adaptive Security Platform 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 data centers 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 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. Upon receiving the new policy, each of these PCEs re-computes policy and sends new firewall rules to their database workloads.

Adapting to Changes in the Environment

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.

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. If 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.

High Availability and Disaster Recovery

A PCE Supercluster provides multiple levels of redundancy and failover for high availably (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). The PCE can survive a building failure but manual action (issuing a PCE administrative command) may be necessary, depending on which building is lost.

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

Cross-PCE Failover and Recovery (Optional)

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.

> **Manual Failover**
> 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, not automated.

Design Considerations

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

- **How many total workloads will 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 there is an extended PCE outage?** By default, VENs will continue to enforce the current policy when their PCE is unavailable. If 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.

Supercluster command-line reference

The Illumio PCE control interface for Supercluster commands often have restrictions on the type of node they can be executed on. For example, setting a cluster's runlevel can be exeuted from any node, Core or Data. 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 executed on.

Supercluster commands to node reference

Main command	Option	Function	Node to run on
`illumio-pce-ctl`	`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-join`	Joins a PCE cluster to a Supercluster. **Note**: Executing this command can take up to 30 minutes depending on the number of PCEs in your Supercluster and size of the PCE database.	On any Core node
	`supercluster-members`	Displays all current active Supercluster PCEs, Members and Leader.	On any Core node or the Data0 node.
	`supercluster-promote-replication`	Establishes the database replication for the entire Supercluster.	On the Data node that is the database master
	`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. **Note**: 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-upgrade-prepare`	Prepares the cluster for database migration after the a new version of the PCE software is installed during an upgrade.	Any node

Main command	Option	Function	Node to run on
	`supercluster-upgrade-rejoin`	Rejoins a PCE to a Supercluster after a database migration during a software upgrade.	Any core node
`illumio-pce-db-management`	`supercluster-data-restore`	Restores a failed Leader from a database backup.	On one of the Data nodes only
	`supercluster-quiesce`	Pauses all the pending database replication, e.g., during a software upgrade.	On the Data node that is the database master

Rerunnable arguments on illumio-pce-ctl

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

Argument on illumio-pce-ctl	Description	Rerunnable?
`supercluster-init-leader`	Configures this PCE as the Supercluster leader.	Yes
`supercluster-join [supercluster_leader_fqdn]`	Joins this PCE into the Supercluster specified by FQDN of the Supercluster leader.	Yes
`supercluster-assign-leader`	Assigns a new Supercluster leader.	Yes
`supercluster-restore [failed_pce_fqdn] [supercluster_leader_fqdn] [--restore-type single_pce|entire_supercluster]`	Restores a failed PCE and re-joins to the Supercluster .	Yes
`supercluster-drop [failed_pce_fqdn]`	Temporarily drops the failed PCE from the Supercluster so it is no longer replicated.	Yes

supercluster-promote-replication	**Applies changes to the PCE replication configuration following an upgrade.**	**Yes**
supercluster-members	Shows the members in the Supercluster.	Yes
supercluster-config	Shows the Supercluster configuration.	Yes
supercluster-upgrade-prepare	Unjoins the PCE from the Supercluster and prepares data for upgrade.	Yes
supercluster-upgrade-rejoin	Rejoins the PCE to the Supercluster.	Yes

Rerunnable arguments on illumio-pce-db-management

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

Argument on illumio-pce-db-managment	Description	Rerunnable?
supercluster-data-dump	Writes the database other persistent state in Supercluster to a file.	Yes
supercluster-data-restore [restore-type single_pce|entire_supercluster] [skip-db-restore true|false]	Restores data and persistent state of a Supercluster database.	Yes
show-supercluster-replication-info	Display Supercluster node replication information.	Yes
supercluster-quiesce [wait_timeout]	Quiesces all pending replication.	Yes
supercluster-replication-debug [--detailed]	Show replication-related information for debugging.	Yes
supercluster-replication-check	Determines the state of replication.	Yes

Revision History

Illumio Adaptive Security Platform PCE Supercluster Deployment and Usage

Document ID: 60000-100-18.2.1

Date	Description
2019-01-23	Updated for Illumio Adaptive Security Platform version 18.2.1. • New recommended storage capacity for deployment: two-storage-device configuration, with a separate device for large amounts of network traffic data. • New storage device layout/partitioning recommendations. • Change in procedure for Supercluster simple upgrade. • Configure PCE internal syslog on Supercluster Leader.
2018-09-13	Added details about optionally configuring a SAML IdP.
2018-09-12	Minimum number of PCEs to make a Supercluster is two.
2018-09-11	Reorganized for clarity.
2018-09-06	First publication to all customers with Illumio Adaptive Security Platform version 18.2.
2018-07-23	Verified required open ports. See "Networking requirements between PCEs".
2018-07-02	Added single PCE cluster sizing requirements in "PCE Cluster Capacity Requirements". These are the same for clusters in a Supercluster deployment as for the standard PCE.
2018-06-29	Corrected syntax of backup command for "Backup each PCE's Data",
2018-06-26	• Corrected details for running `illumio-pce-db-management setup` and `illumio-pce-db-management migrate`. These commands can be run on any node of the PCE cluster. • Corrected syntax for `illumio-pce-ctl supercluster-restore` in "Join the PCE to the Supercluster".
2018-06-19	• Propagation to members of changes to object limits on the leader occurs with 30 minutes. • Minor changes to "Supercluster simple upgrade".

Date	Description
2018-05-11	• Updated for Illumio Adaptive Security Platform version 18.1 • Start of revision history