Illumio Design Guide

Contents

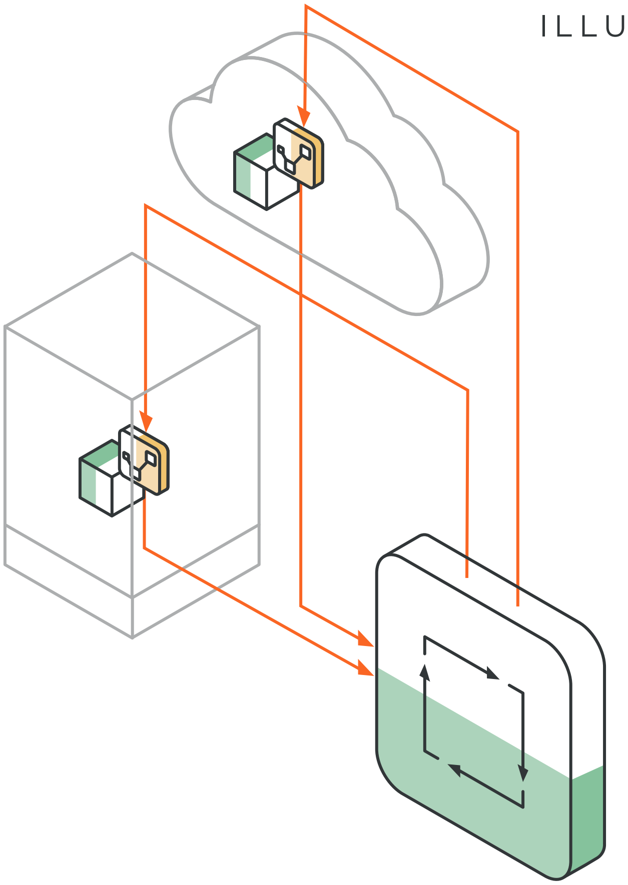
Introduction

Welcome, Illumio user! This design guide is aimed at users who are evaluating the Illumio Adaptive Security Platform® (ASP) and those who are planning or starting to execute their Illumio strategy. You'll be given a brief overview of several topics that are important to successfully deploying and using Illumio ASP. You'll also be given a lot of food for thought, including planning and design considerations and key decision points. This short guide is not a complete manual, but rather a primer to help you set off on the right foot.

In this guide we'll cover the basics of Illumio's architecture, deployment planning, an introduction to security policies, and some key operational topics. We'll also introduce Illumio's FIRST Principles of Security Segmentation, our unique approach to helping you quickly deliver value from your Illumio program.

More detailed guidance on each of these topics is available from our published documentation or from your Illumio representative. We hope that this guide gives you a solid foundation on which to build a visibility and segmentation platform that benefits your organization for years to come.

Architecture



Overview of Illumio Architecture

Illumio ASP is primarily made up of two components. Let's take a brief look at each of them:

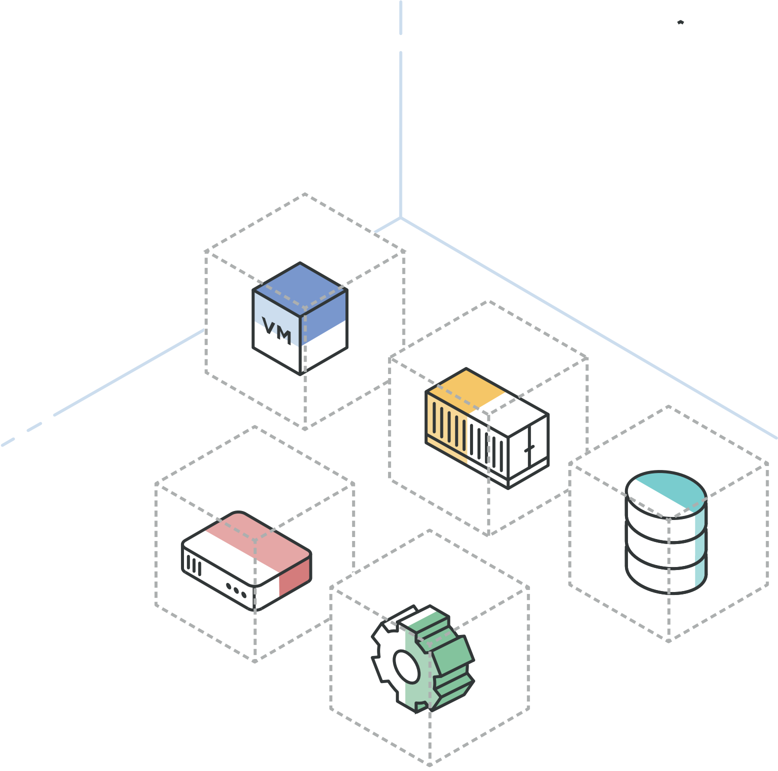
The Policy Compute Engine (PCE):

The PCE is the central brain of your Illumio ASP installation. This is our server component which stores data, calculates security policy, and provides the user interface. The PCE can be deployed as software running on physical or virtual servers in your data center or can be consumed from Illumio's cloud service. Either one can scale to support hundreds of thousands of workloads, and on-premise deployments can be distributed across data centers in different configurations to meet your high-availability and disaster recovery needs.

The Virtual Enforcement Node (VEN):

The VEN is the software component that runs on each managed workload. VEN software is available for Linux, Windows, Solaris, and AIX. Your VENs coordinate with your PCE to provide local monitoring and enforcement capabilities.

There may be additional components in your environment, but these are the primary components that we talk about most frequently.



What Is a Workload?

Before we get too far, it's important to understand how Illumio uses the term workload. When we talk about workloads, we're referring to endpoints on your network. A workload could be a physical or virtual server, a public cloud instance, a container, a storage appliance, a VIP on a load balancer or proxy device, or just about anything with an IP address. We'll be talking about two types of workloads:

Managed Workloads

A managed workload is a physical or virtual server, a public cloud instance, or a container that is running the Illumio VEN software component and is under management by Illumio ASP. When the VEN is installed and activated, it provides information about the workload to the PCE and implements security controls as instructed by the PCE.

Unmanaged Workloads

In any new Illumio ASP installation, it's typical to start with a small number of VENs and work your way up to full coverage. While you're just getting started, you might want your PCE to know about all of your endpoints, even the ones that do not yet have the VEN installed. And your environment is likely to contain some network appliances or other devices that aren't capable of running the VEN. These endpoints are represented by unmanaged workloads. You can tell the PCE about them and they form an important part of your model, even though they do not have the VEN software installed.

PCE Design Best Practices

Cloud Service vs. On-Premises

One of the first decisions you'll make in planning for your PCE is whether to use Illumio's cloud service or deploy the PCE software on-premises. The visibility and enforcement capabilities are the same, although some of the integration details might differ. An on-premise deployment requires Linux servers (physical or virtual) with specific operating system and storage requirements and may require load balancers or other local infrastructure. The PCE footprint can range from four servers to significantly more than four, depending on the scale and resiliency requirements.

Illumio's cloud service is recommended for:

- A quick start without buying any new hardware.
- Organizations that prefer a fully turnkey offering, where the PCE is managed by Illumio's operations team.
- Organizations that don't have substantial Linux expertise or that don't need the level of control that comes with operating the software in-house.
- Deployments from 0–10,000 workloads.

An on-premise deployment is recommended for:

- Deployments from 0–250,000 workloads.
- Organizations with specific data residency or custodianship requirements that preclude the use of a cloud security service
- Organizations with robust Linux expertise who require complete control over all aspects of PCE operations, including those who prefer to satisfy their high-availability and disaster recovery needs using in-house management, and those with strict maintenance and change control

Customers can choose to conduct a POC or evaluation using Illumio's cloud service, then transition to an on-premise deployment if warranted by their requirements. There are no differences in features or capabilities, although some integration points (e.g., connecting the PCE to your SIEM) might work differently.

Redundancy/Failover Considerations

Illumio's SOC 2-compliant cloud service is deployed in a highly resilient configuration distributed across multiple physical locations.

> On-premise PCEs are deployed in units called ***clusters***. A typical cluster consists of at least four physical or virtual servers supporting up to 10,000 managed workloads, with larger footprints able to support 25,000 managed workloads using a single cluster.

In a highly-available metro-area cluster, the nodes are split between two data centers such that the PCE will continue operating normally even in the event of a total failure of one data center. This hot-hot "split-cluster" deployment approach comes with specific latency requirements between the two data centers; the maximum permitted latency can vary based on the size of your deployment.

If your portfolio doesn't include two buildings that meet the metro-area latency requirements, or if you'd like additional redundancy, a cold-standby cluster can be used. This approach has no specific latency requirements and can also be combined with a split-cluster deployment to achieve full disaster recovery capabilities or to guard against a multi-site failure.

Larger On-Premise Installations

If you have more than 25,000 managed workloads, you may be a candidate for Supercluster. A Supercluster is a federated set of clusters connected with near-real-time replication to serve large organizations. Illumio's Supercluster capability provides single-pane-of-glass visibility and enforcement control over hundreds of thousands of managed workloads using independent clusters, offering high performance plus outage impact isolation. Each cluster is under central management but can continue operating normally in the event of an outage affecting one or more of its peers.

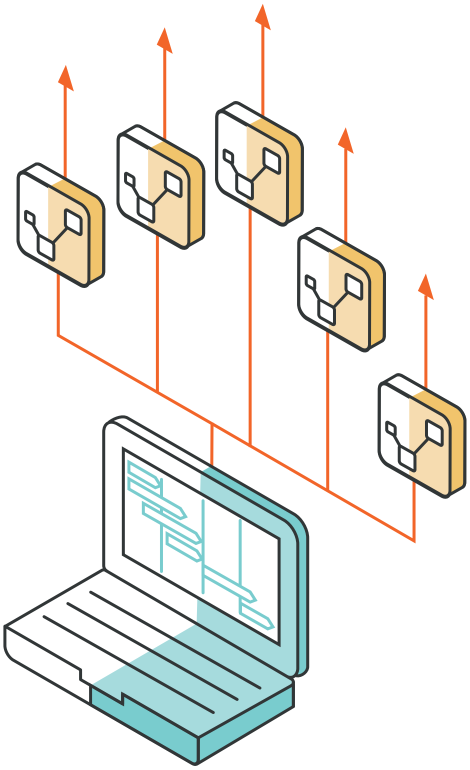
There is no specific latency requirement between the PCE and VEN, and many organizations deploy a single PCE cluster serving all global VENs. For example, we commonly see North America PCE deployments with VENs in Hong Kong and this configuration has no adverse effects. Larger organizations with global deployments often choose to deploy Supercluster with a local PCE in each region, however, to provide additional isolation and to follow their existing regional management models.

Securing Your PCE

This section applies only to on-premise customers. Security for Illumio's SOC 2-compliant cloud service is provided within the platform, but on-premise customers need to take specific steps to ensure that their PCE is protected.

The PCE is made up of several internal components with specific connectivity requirements. Illumio provides guidance on how best to harden the PCE servers, to protect against malicious connections and to safeguard the integrity of the data stored within.

Because the PCE software runs on your operating system, on a server inside your network, you may have different local security requirements than other customers. Illumio publishes a hardening guide and a set of tools that can be used to implement the recommended controls. Please discuss your specific situation with your Illumio representative to ensure that your PCE gets the correct level of protection.



VEN Deployment Strategy

Once your PCE deployment is nailed down, it's time to start planning for VENs. The VEN is the software component which performs all of Illumio's management functions on the protected workload. The VEN is offered as an installable package in a way that's suitable for each supported platform: RPM or DEB for Linux, MSI for Windows, PKG for Solaris, BFF for AIX. A list of all currently supported operating system versions is available on the Illumio website or from your Illumio representative.

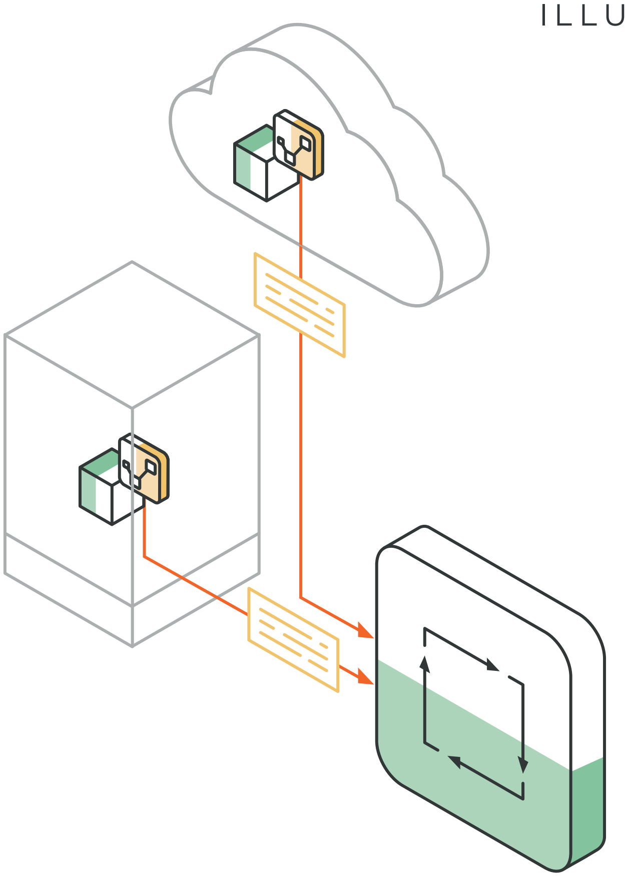
The VEN is designed to be lightweight, requiring minimal local configuration and consuming as little CPU and RAM as possible. The VEN manages its own logfiles and will not use too much disk space. Upgrades to new VEN releases can even be pushed centrally by the PCE, if desired.

The VEN is compatible with nearly all systems management and automation frameworks. Here are some approaches we commonly see employed for VEN distribution:

- Install the VEN using an automation tool like Chef or Puppet.
- Push the VEN to Windows workloads using Microsoft SCCM.
- Pre-install the VEN into a golden image, such that it activates when the image is deployed.
- Use custom scripting to SSH to each workload and pull the VEN package.

No matter how it gets onto the workload, the VEN immediately establishes communication with the PCE and can either begin providing visibility or start fully enforcing security policy, depending on how it's configured.

Data



Data You Supply

Getting the most value out of the Illumio ASP depends on pairing it with data about your environment. The “A” in ASP stands for ***Adaptive***, and Illumio ASP adapts by using data to drive your security policies. The data you supply is the foundation that your security model is built on.

All About Labels

Illumio’s policy model is based on pieces of metadata called labels. Each workload is identified by up to four labels:

R + A + E + L

Role: the role or function this workload performs, such as Web Server or Database Server

Application: the application this workload belongs to, such as Payroll or CRM

Environment: the environment this workload is part of, such as Development, QA, or Production

Location: where this workload is located, or the location it serves

The first step in thinking about security policies is to think about your label model. While Illumio defines the four label dimensions, there’s quite a bit of flexibility to tailor each dimension to your specific needs. For example: the location dimension might hold physical locations, like regions, cities, even down to racks in a data center. Or it could hold regulatory jurisdictions or other logical properties.

Security policies are written in terms of the four label dimensions. If two workloads have the exact same labels, the same security policies will apply to them. If two workloads need to be treated differently, their labels must reflect that.

It's important to remember that labels are not groups. Each label dimension is independent, and the labels can combine to form a unique set of security properties for each workload. A production HR webserver in London might inherit one set of security policies that applies to all London servers, another that applies to all HR servers, and a third for production webservers.

Illumio's four-dimensional model is a key enabler for building scalable and manageable security policies. Policies based on the intersections of these four dimensions are not only easy to implement, they are also easy to review and maintain. Models using flat groups or arbitrary levels of "tagging" often prove unwieldy and make it difficult for you to achieve your security goals.

One key principle is that each dimension should always be used to refer to the same logical concept. For example: if your Role label holds application tiers or components, it should always be used in this way, and not sometimes used to hold information sensitivity classifications. Using the labels consistently will help you write security policies that behave in predictable ways.

Your Source for Label Data

Most organizations have some source they can tap for at least one of the label dimensions. While a large enterprise usually has a CMDB or other full-featured catalog, a smaller organization might have a spreadsheet or even a hostname convention. It's probably not complete or 100% accurate, but that's ok.

Early in your Illumio adoption process, you'll want to identify what source or sources you have available that can supply some or all of the label data for your workloads. When it's time to give this data to the PCE, it can be entered by hand, or automatically synchronized using our API on a one-time or ongoing basis.

Label Data Quality

If you're worried that you don't have a complete catalog or a source for reliable label data, don't be. We find that customers are typically able to supply environment labels (is this a production workload or non-production?) with a 50-80% confidence rate, and it's often much lower for the other dimensions.

The Illumio ASP is built with the expectation that your existing catalog is incomplete or incorrect. The process of adopting Illumio's controls will help you to refine and improve your metadata. You aren't expected to have a high-quality catalog when you start, but you probably will by the time you finish.

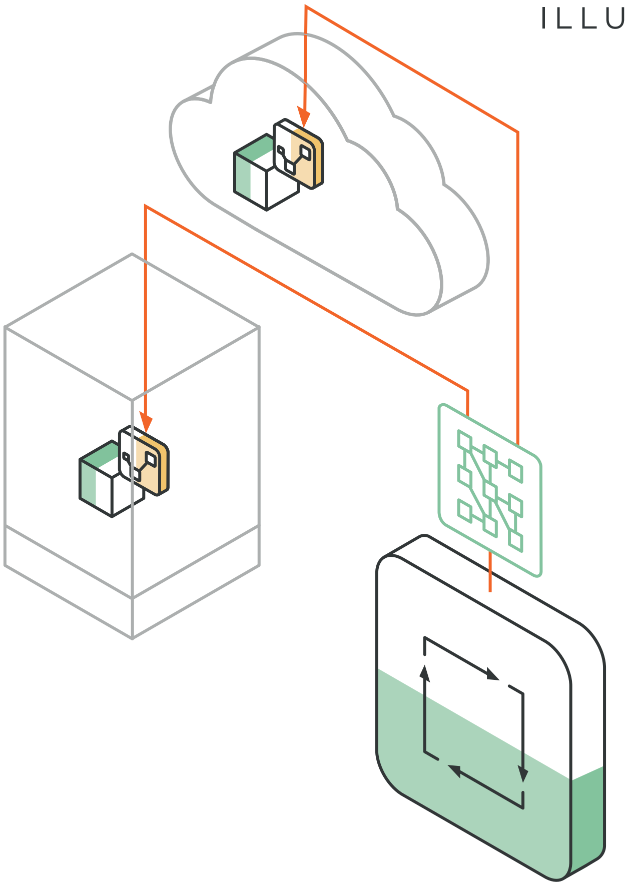
> Early in your Illumio adoption process, you'll want to identify what source or sources you have available that can supply some or all of the label data for your workloads.

Data Accuracy and Governance

For Illumio ASP to fulfill its adaptive mission, it needs to learn about changes in your environment that might affect security. This includes new workloads being provisioned, old ones being decommissioned, and also label changes, like the promotion of a non-production server to production.

If your label data is maintained by hand, you'll want to include a process to update the PCE whenever changes happen. For customers that load their label data into the PCE using our API, refreshes are typically done daily but can be done as often as needed, including in real time as changes occur.

Many organizations choose to implement a data governance process or appoint a data guardian who manages the quality of the data in the catalog. Our customers usually find that investments in metadata quality result in a wide range of security and operational benefits.



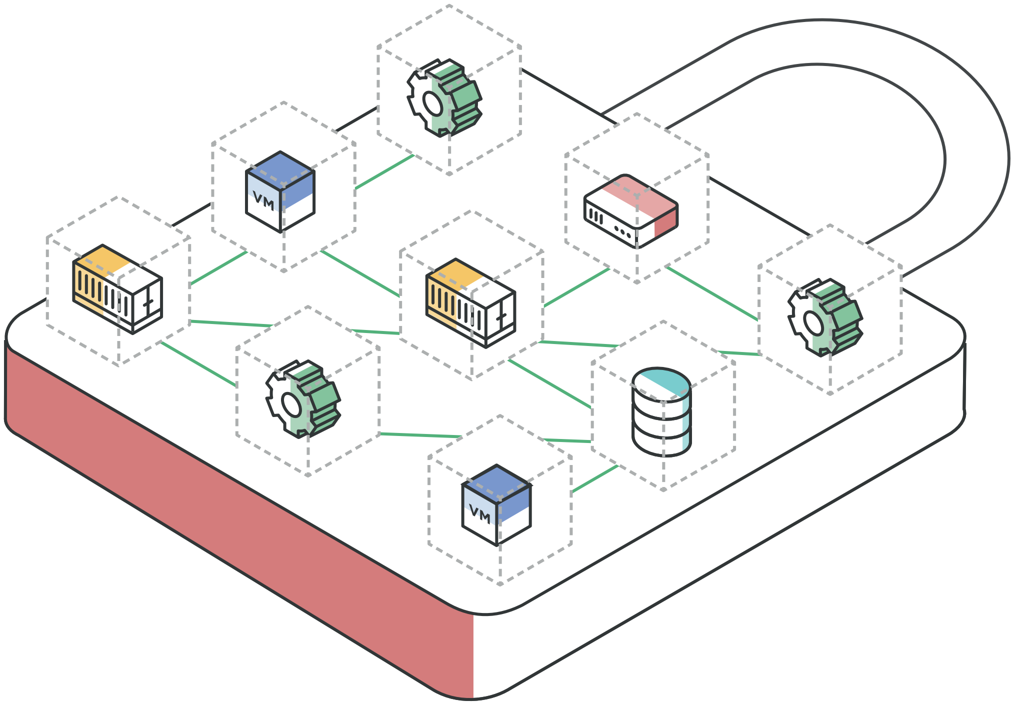
Data Provided by the PCE

Label data is the primary input you provide to the PCE, and in return you get the fruits of the VEN's observations: flow data. Illumio's flow data is the foundation for building the application dependency map. This map tells you which workloads are talking to which other workloads, not just in terms of their hostnames or IP addresses, but using their labels of course.

The PCE collects flow data for its own internal use and can also publish this data, as it's collected, to Splunk, ArcSight, or almost any other data warehouse or SIEM product. Each flow record gives you the source and destination enriched with labels where possible, so you can give context and meaning to your flows.

The application dependency map puts these flows in a context that your organization can understand. A connection between workloads can be expressed as a connection between applications or environments. This data can be inspected in real time for undesirable behavior or warehoused for future analysis or forensics.

Security Policies



What Is a Security Policy?

An Illumio security policy is a rule that allows a certain type of connection to occur. Security policies are written using labels, not hostnames or IP addresses. Typical security policies might say things like:

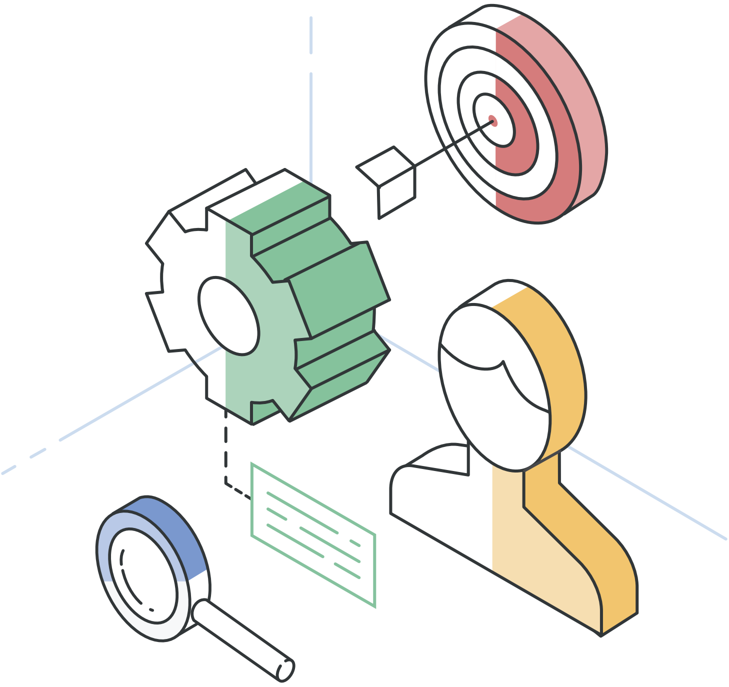
- Production workloads in North America can connect to the New York backup plant.
- Workloads running the payroll application can connect to HR database workloads.
- Certain UAT or prod-parallel workloads are permitted to query data from production.
- All of the workloads that make up one particular application can communicate amongst themselves but they are isolated from all others.

It's important to write these policies using labels so that they can be adaptive. Using an adaptive policy, whenever a new workload is added that satisfies the criteria, its policies and the policies of its peers are seamlessly recalculated.

This is one of the most significant differences between Illumio security policies and traditional firewalls. Firewall policies are typically written using IP addresses and are static for the life of the policy. Think about the changes that happen in your environment every day:

- A lifecycle event occurs, such as promoting a workload from test to production.
- A workload is decommissioned.
- A workload changes function or a new application is deployed.
- New workloads are added to an existing application for additional capacity.

In all of these cases, static firewall rules will remain and will continue to enforce a policy that is no longer appropriate for your environment. Illumio's adaptive policy model ensures that your security policy responds immediately to any such changes with no action required.



Illumio's FIRST Principles of Security Segmentation

Here's a quick primer on Illumio's recommended approach:

FIND
metadata sources.

IDENTIFY
a label design.

REACH OUT
to service owners early.

START
with core services.

TARGET
ringfencing for business applications

We've already covered the first two.

The third principle is all about communication: make sure your service owners are engaged early in the planning and deployment process. Service owners typically understand your applications and their interactions better than anyone, so they are perfectly positioned to review Illumio's dependency maps and match them up against the applications' expected behavior.

A successful Illumio deployment may require the participation of different stakeholders within your organization, and it's important to involve them from the start.

We'll cover the final two principles in the following sections.

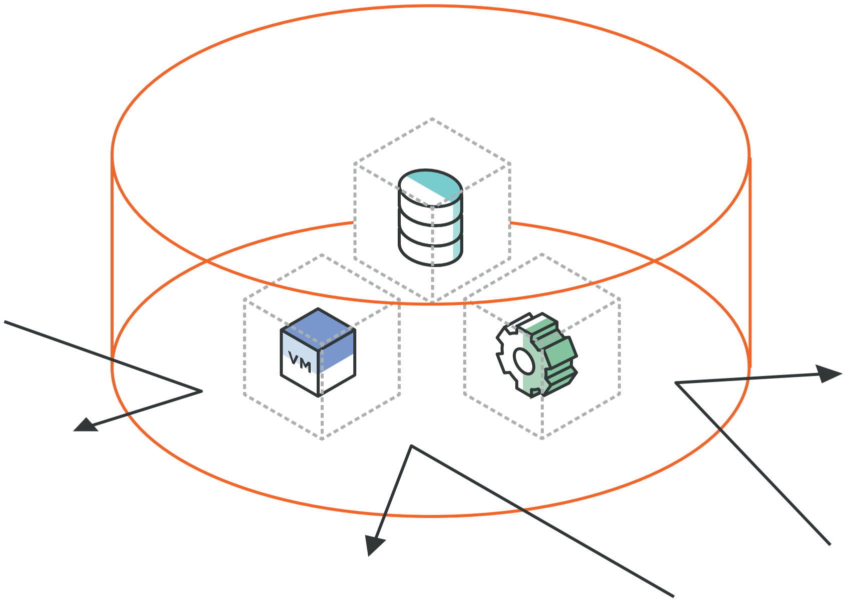
Core Services

Active Directory. NetBackup. Splunk. ServiceNow. What do these have in common? They are all frequently classified as ***core services***.

Core services are part of your infrastructure, usually managed centrally by your system or network administrators, to which most or all of your workloads will connect. Core services are typically those that provide platform or operating system-level functionality.

When writing security policies, it's helpful to start with core services, make sure they are properly labeled, and write rules that permit core service traffic.

Core service connections are typically a large percentage of your overall traffic, and they're consistent across all of your workloads, so it makes sense to tackle these first. It will help you get a feel for the rule writing process, and will also de-clutter your connectivity map, allowing application owners to focus on the traffic that is meaningful to them and specific to their applications.



Ringfencing

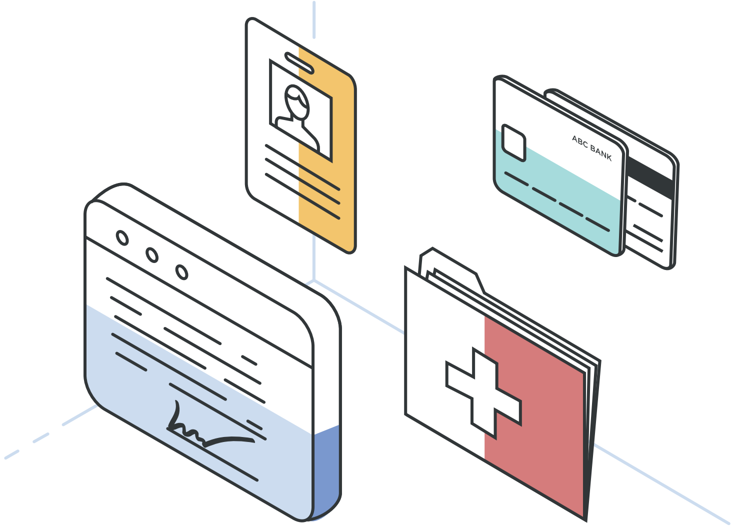
Now we're at the heart of Illumio's policy model: how do I write rules such that desirable connections are allowed, and all other connections are not?

Illumio supports many different levels of granularity in your security policies. Your policies can be as simple or as complex as they need to be to meet your control requirements. Here are some of the more common patterns:

- **App Group Level:** Also known as ***application ringfencing***. A virtual perimeter is drawn around the application's workloads and only permitted connections are allowed in from outside, but within the application level, all workloads can talk freely with each other.
- **Role Level – All Services:** In addition to the virtual perimeter, the application's workloads are divided up by role, and connections are only allowed between authorized roles. For example, your web servers may be able to talk to your application servers but not your database servers.
- **Role Level – Specified Services:** This is the most granular level of security policy. This takes the role separation down to the port, protocol, and process level, allowing connections between roles only on specific services. For example, your application servers may be able to talk to your database servers on TCP port 3306 but not any other ports.

Each of these patterns provides significantly more control than the multi-tier model offered by traditional firewalls. This list is not complete, and there are many ways to write security policies to achieve finer or looser levels of control. One key advantage of using Illumio is that you can mix and match. Different applications have different security and control requirements, and you don't have to use the same approach to cover all of your applications.

In our experience, application ringfencing offers the right balance between security and complexity. Your application's virtual perimeter prevents connections to sensitive internal components, while not restricting the application's internal connectivity. This is a simple model that can be implemented quickly, and it provides exceptional protection while requiring relatively little maintenance. Most Illumio customers settle on ringfencing as the base level of control for most of their applications.

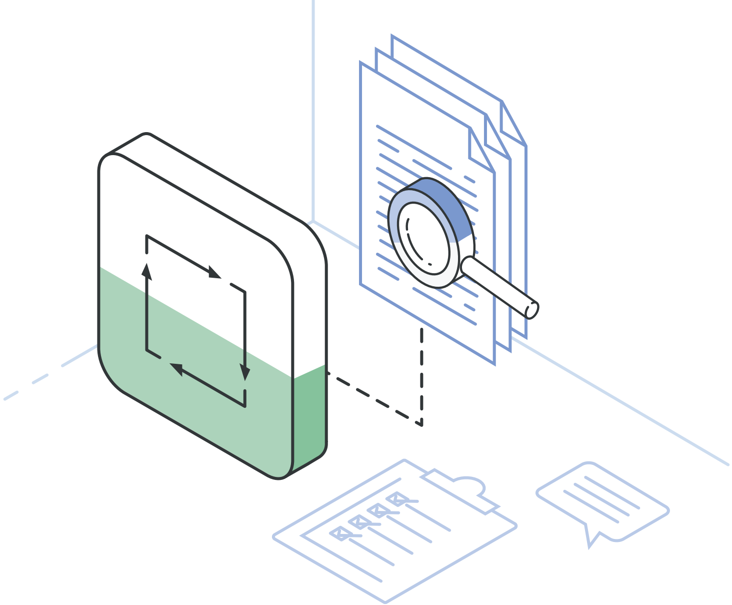


Digital Crown Jewels

Of course, Illumio provides the flexibility to implement different levels of control where needed. Have you identified your ***digital crown jewels***? These are the applications that store or process your most sensitive data, whether it's payment or financial data, personally identifiable information about your customers or employees, or anything else that you need to keep secret.

Once you've identified these applications, you might choose to protect them with a higher level of control, all the way down to the port, protocol, and process level. This approach can require more maintenance as an application changes over its lifecycle, because your security policies may need to be updated to reflect those changes. But it also provides the highest level of protection against lateral movement within your application's virtual perimeter.

Logging, Monitoring, and Alerting



PCE Logging Overview

Your PCE logs lots of useful information about what's happening in your environment. There are three different types of PCE log messages available:

1. **PCE internal messages, for support or troubleshooting purposes:** These are unstructured log records related to the inner workings of the PCE.
2. **Auditable events:** These are structured messages about significant events including agent activated, user password changed, security policy modified or provisioned, label created, etc. The PCE's auditable events comply with the Common Criteria Class FAU Security Audit standard.
3. **Traffic data records:** The PCE is able to produce periodic summaries of the connections observed on each managed workload, including the source and destination of each connection, enriched with labels when available.

Managing Logfiles

By default, the PCE includes its own logfile management system for internal messages. Logs are written to files in the PCE log directory and rotated automatically to conserve disk space. You should not need to look at these files on a regular basis, but they are available if needed for support or troubleshooting.

You also have the option of using your own syslog infrastructure in place of the PCE's default. The PCE can generate a high volume of log messages so please discuss this option with your Illumio representative if desired.

Monitoring, Alerting, and SIEM Integration

Auditable Events

Auditable events are structured records with well-defined data fields. There's a list of all events in the PCE documentation. These events are stored in the PCE's database and are available for query or review via the web interface.

The PCE can also be configured to publish these events via either syslog or Fluentd, a popular log aggregator. Events can be published in JSON, CEF, or LEEF formats, offering compatibility with most major SIEM vendors. For Splunk users, the Illumio App for Splunk, available in Splunkbase, offers turnkey integration. The PCE's events can also be used with HPE ArcSight, IBM QRadar, or any other SIEM that can accept messages in one of our supported formats.

As part of your deployment planning, Illumio recommends that you review the list of events in the PCE documentation and configure your SIEM to alert you as appropriate. Some customers choose to be alerted on individual events, such as “agent.tampering,” indicating that the VEN detected an unauthorized change to the local firewall rules. In our experience, it’s also very helpful to look at aggregates, such as the total number of messages per day by message type.

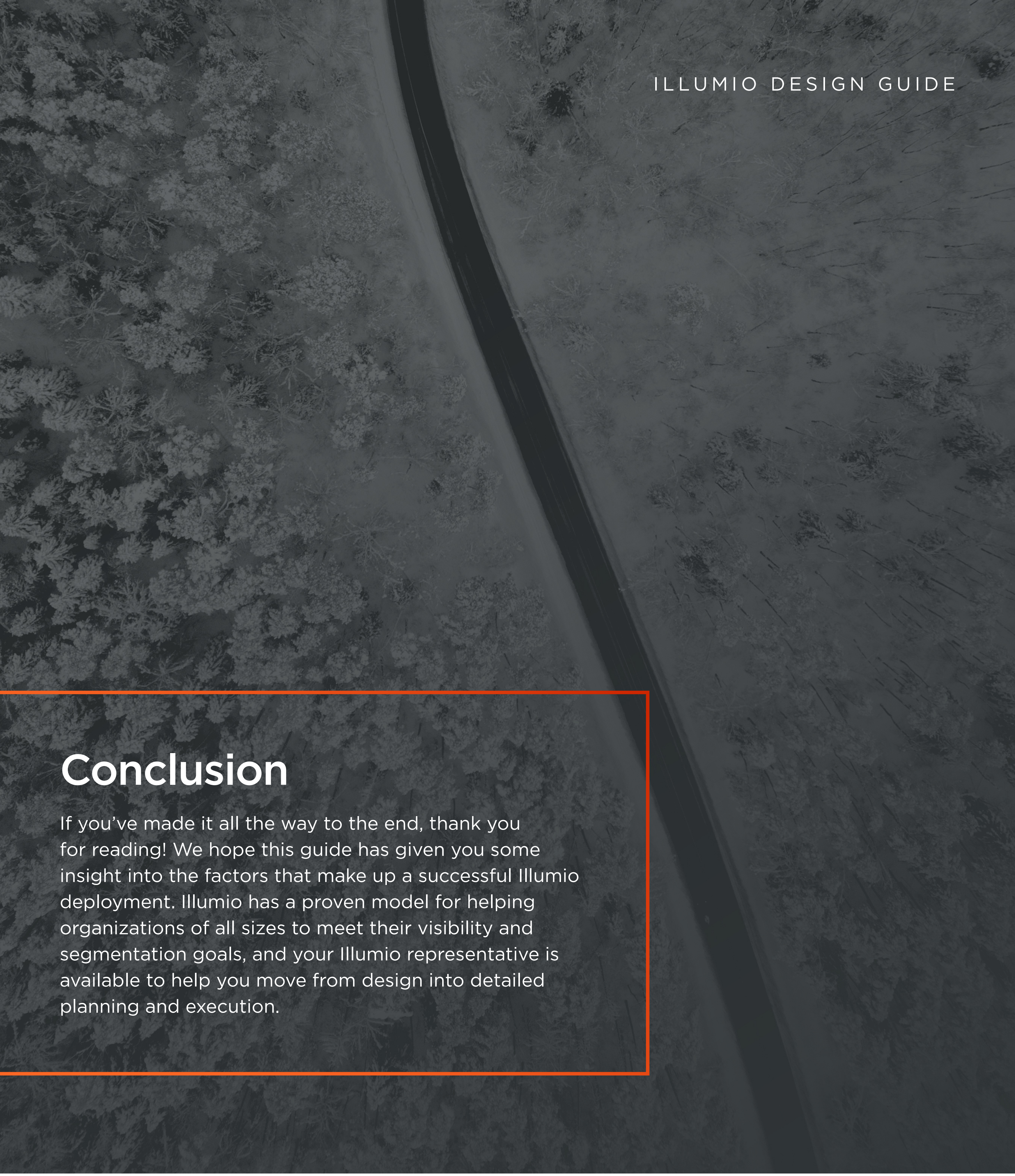
Traffic Data

In addition to auditable events, you can also publish traffic data records to your SIEM. The PCE can be configured to publish reports of accepted, blocked, and/or potentially blocked connections. (Potentially blocked connections are those that occur on workloads that are not yet enforcing; these are connections that aren’t allowed by policy, and would be blocked if the workload were moved into enforcement.)

Traffic data records contain important information about the connections taking place within your environment. In addition to IP addresses and hostnames, these records also contain labels for known workloads. This lets you easily identify connections that cross between development and production environments, for example.

Some customers have built ***targeted monitoring*** programs around this data. The VEN blocks unauthorized connections on workloads that are in enforcement mode, and targeted monitoring can alert you to suspicious lateral movement for non-enforcing workloads.
This is another way to get value from the data that Illumio provides.

Depending on how many managed workloads you have and how much traffic is observed, the amount of traffic data reported by the PCE can be quite significant. If you’d like to publish traffic data records, please consult with your Illumio representative to ensure your SIEM has enough capacity.



Conclusion

If you've made it all the way to the end, thank you for reading! We hope this guide has given you some insight into the factors that make up a successful Illumio deployment. Illumio has a proven model for helping organizations of all sizes to meet their visibility and segmentation goals, and your Illumio representative is available to help you move from design into detailed planning and execution.