

Connection to Private Subnets

Date published: 2019-08-22

Date modified:



Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Cluster Connectivity Manager.....	4
Upgrading from CCMv1 to CCMv2.....	9
Outbound network access destinations for CCMv2.....	10
Configuring a private VPC in AWS.....	11
Configuring a VNet with private IPs in Azure.....	11
Configuring a VPC with private IPs in GCP.....	12
Enabling CCM in the Management Console.....	12
Troubleshooting CCMv1.....	12
Troubleshooting CCMv2.....	15
Data Warehouse and private networking.....	16
 Public Endpoint Access Gateway.....	 16
Enable Public Endpoint Access Gateway for AWS.....	17
Enable Public Endpoint Access Gateway for Azure.....	19
Enable Public Endpoint Access Gateway for GCP.....	20
 Azure Load Balancers in Data Lakes and Data Hubs.....	 21

Cluster Connectivity Manager

CDP can communicate with Data Lake, Data Hubs, CDP data services workload clusters, and on-prem classic clusters that are on private subnets. This communication occurs via the Cluster Connectivity Manager (CCM). CCM is available for CDP deployments on AWS, Azure, and GCP.

**Note:**

When you register a new AWS, Azure, or GCP environment in CDP via web interface, Cluster Connectivity Manager (CCM) is enabled by default and the option to disable it has been removed from the web interface. CCM is also enabled by default when registering an environment via CDP CLI.

CCM enables the CDP Control Plane to communicate with workload clusters that do not expose public IPs. Communication takes place over private IPs without any inbound network access rules required, but CDP requires that clusters allow outbound connections to CDP Control Plane.

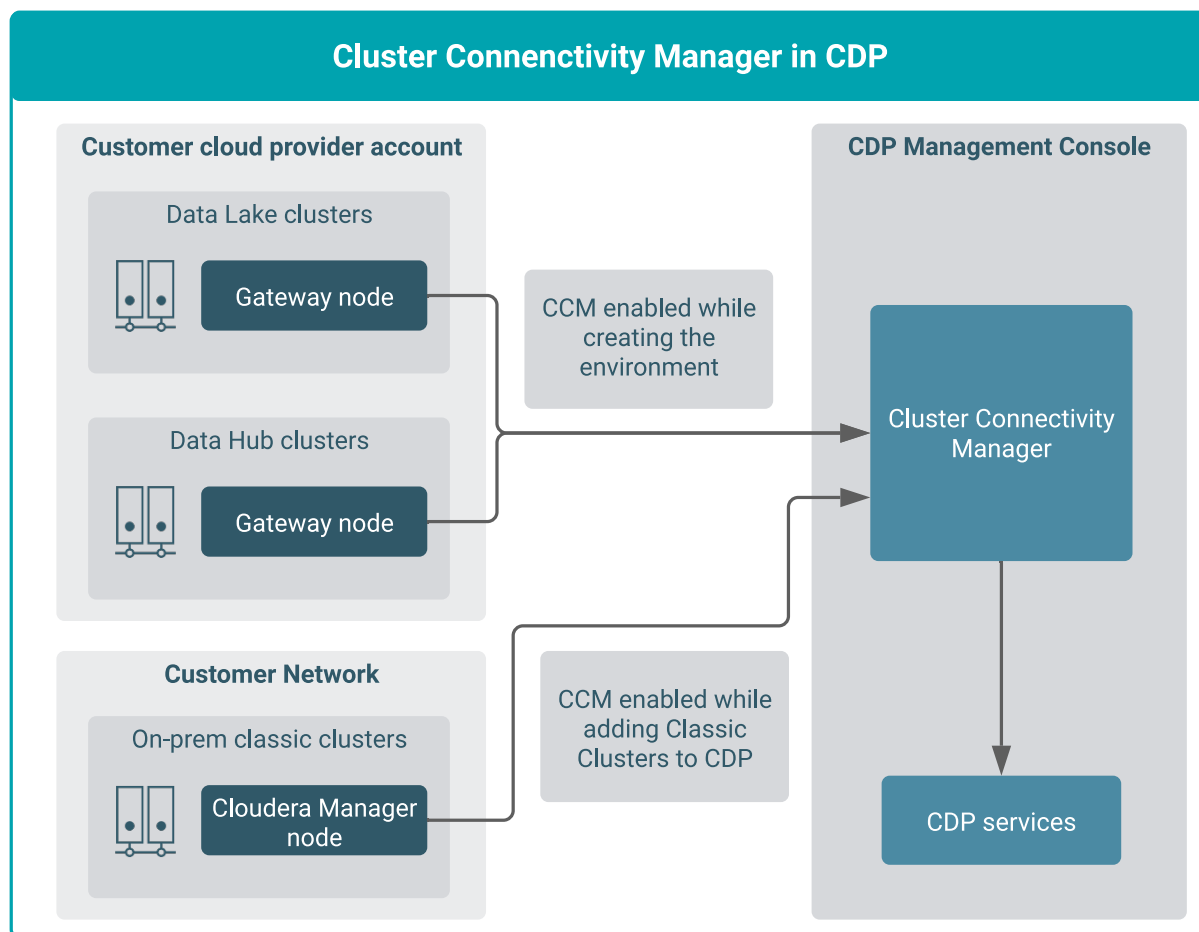
CCM provides enhanced security for communication between customer workload clusters and the CDP Control Plane. A CDP environment supports public, private, and semi-private networks for running workloads. In a public network, CDP Control Plane initiates a connection to the nodes in a workload cluster; However, when using a private or semi-private environment, this option is not available due to the private nature of the subnet and some of the hosts. In such cases, CCM is required to simplify the network configuration in the customer's subnet.

CCM implements an inverted proxy that initiates communication from the secure, private workload subnet to CDP Control Plane. With CCM enabled, the traffic direction is reversed so that the private workload subnet does not require inbound access from Cloudera's network. In this setup, configuring security groups is not as critical as in the public network setup. All communication via CCM is encrypted via TLS v1.2.

From a data security perspective, no data or metadata leaves the workload subnet. The CCM connection is used to send control signals, logs and heartbeats, and communicate the health status of various components with the CDP Control Plane.

When deploying environments without public IPs, a mechanism for end users to connect to the CDP endpoints should already be established via a Direct Connection, VPN or some other network setup. In the background, the CDP Control Plane must also be able to communicate with the entities deployed in your private network.

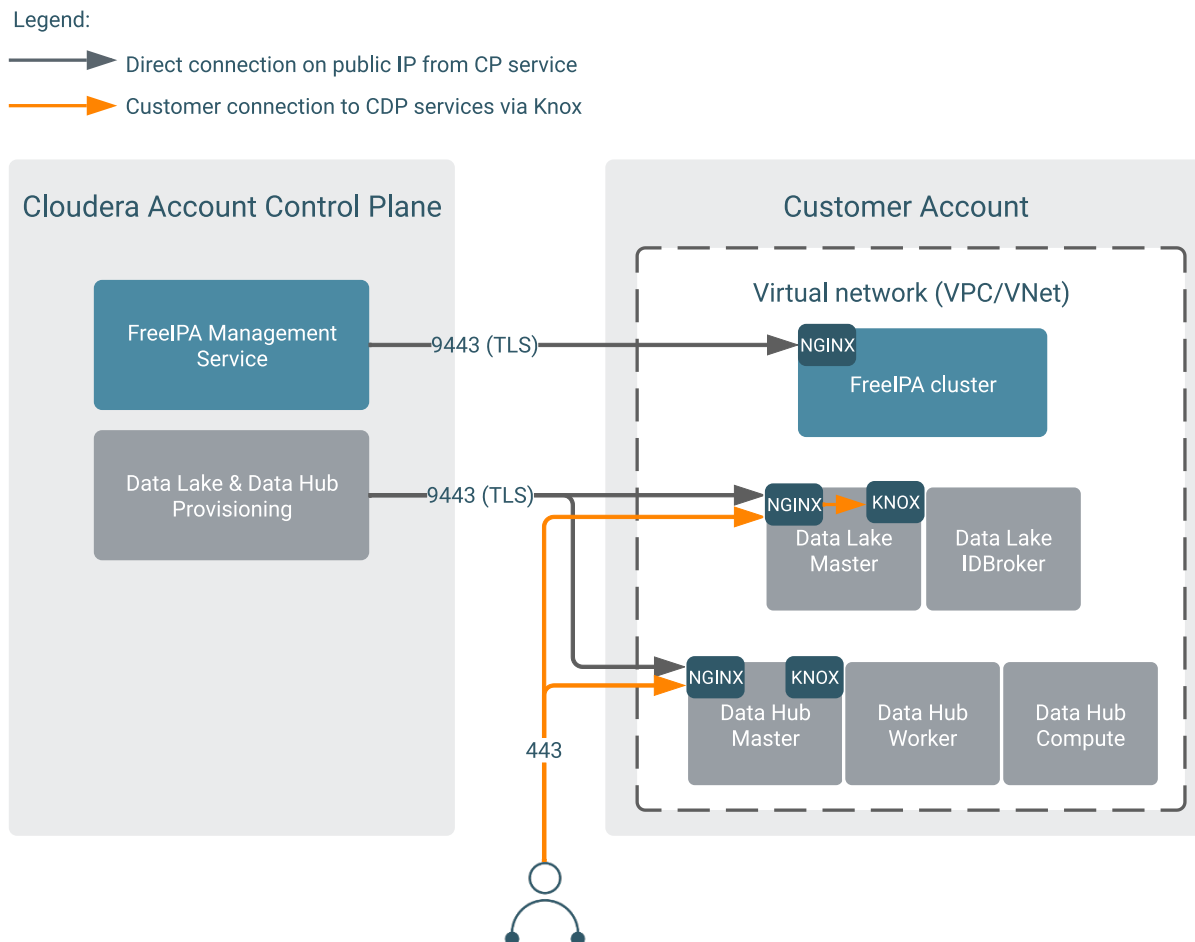
The following diagram illustrates the CDP components that are responsible for communication via CCM:



CCM was initially released as CCMv1 and later CCMv2 was released to replace it. While CCMv1 establishes and uses a tunnel based on the SSH protocol, with CCMv2 the connection is via HTTPS. All new environments created with Runtime 7.2.6 or newer use CCMv2. Existing environments and new environments created with Runtime older than 7.2.6 continue to use CCMv1. All newly registered classic clusters use CCMv2, but previously registered classic clusters continue to use CCMv1.

The following diagram illustrates connectivity to a customer account without using CCM:

Figure 1: Connectivity to customer account with CCM disabled



CCMv2

CCMv2 agents deployed on FreeIPA nodes initiate an HTTPS connection to the CDP Control Plane. This connection is then used for all communication thereafter. Data Lake and Data Hub instances receive connections from the CDP Control Plane via the agents deployed onto FreeIPA nodes. This is illustrated in the diagram below.

CCMv2 also supports classic clusters. You can use Replication Manager with your on-premise CDH, HDP, and CDP Private Cloud Base clusters accessible via a private IPs to assist with data migration and synchronization to cloud storage by first registering your cluster using classic cluster registration.

When CCMv2 is enabled, the traffic direction is reversed so the environment does not require inbound access from Cloudera's network. Since in this setup, inbound traffic is only allowed on the private subnets, configuring security groups is not as critical as in the public IP mode outlined in the previous diagram; However, in case of bridged networks it may be useful to restrict access to a certain range of private IPs.

The following diagram illustrates connectivity to a customer account using CCMv2:

Figure 2: Connectivity to customer account with CCMv2 enabled

CCMv2 with Token Authentication

In the current scheme of things, the communication between agent and CDP Control Plane services uses a two-way SSL or client certificate based authentication mechanism.

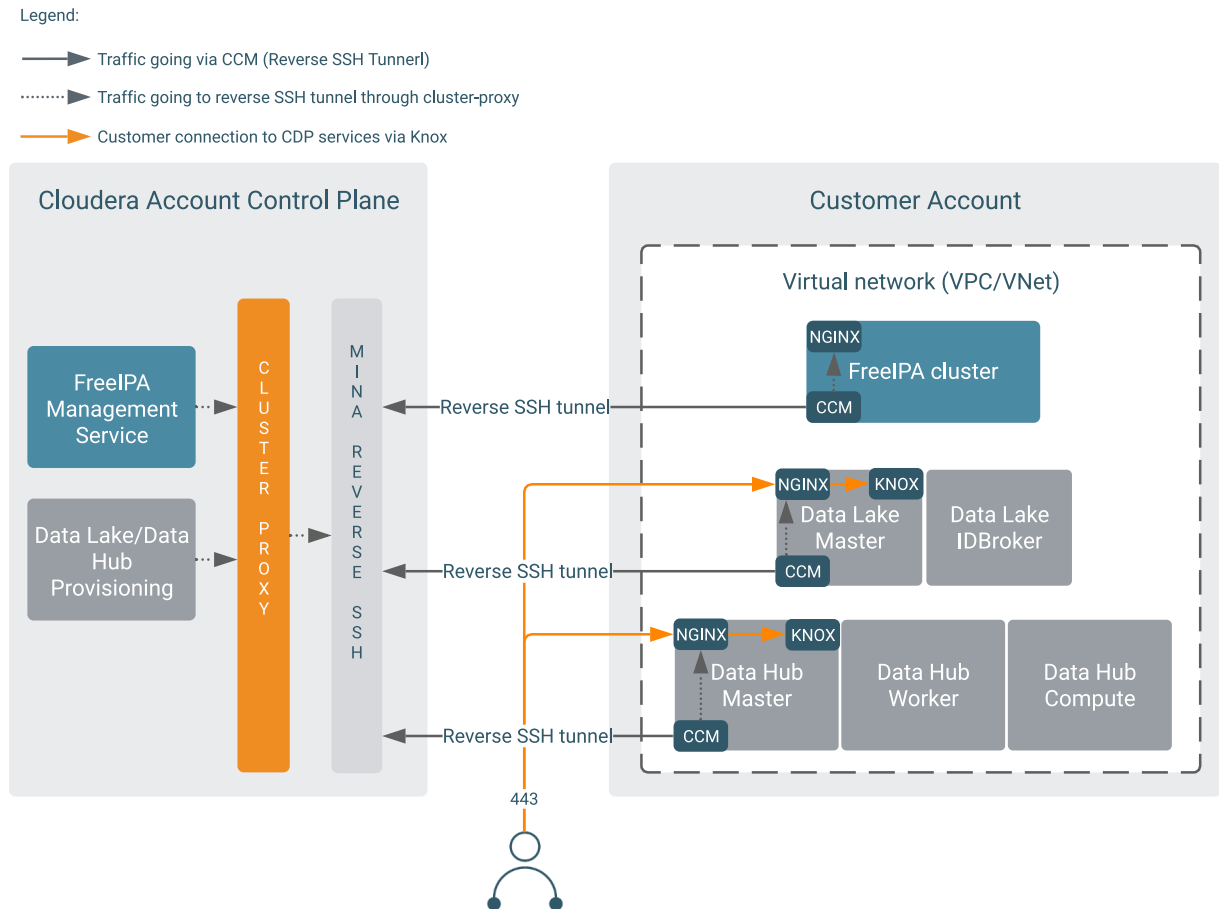
In order to enable traffic inspection which could further pave the way for traffic monitoring and anomaly detection in traffic, the communication between agent and CDP Control Plane can optionally be configured to use a combination of TLS (to validate the server) and bespoke validation (to validate the client).

This approach does away with the client certificate based agent authentication on the Control Plane side and instead uses request signing and authorization to validate incoming requests from the CCM agent.

CCMv1

The below diagram illustrates the CDP connectivity to a customer account with CCMv1 enabled. CCMv1 agents are deployed not only on the FreeIPA cluster (like in CCMv2), but also on the Data Lake and Data Hub. While CCMv2 establishes a connection via HTTPS, CCMv1 uses a tunnel based on the SSH protocol. Workload clusters initiate an SSH tunnel to the CDP control plane, which is then used for all communication thereafter.

Figure 3: Connectivity to customer account with CCMv1 enabled



Supported services

The following CDP services are supported by CCM:

CCMv2

Supports environments with Runtime 7.2.6+

CDP service	AWS	Azure	GCP
Data Lake	GA	GA	GA
FreeIPA	GA	GA	GA
Data Engineering	Preview	Preview	
Data Hub	GA	GA	GA
Data Warehouse		GA	
DataFlow	GA	GA	
Machine Learning	GA	GA	
Operational Database	GA	GA	GA

CCMv1

Supports environments with Runtime <7.2.6 and environments created prior to CCMv2 GA.

CDP service	AWS	Azure	GCP
Data Lake	GA	GA	GA
FreeIPA	GA	GA	GA
Data Engineering			
Data Hub	GA	GA	GA
Data Warehouse			
DataFlow			
Machine Learning			
Operational Database			

To learn more about CCM, refer to the following documentation:

Upgrading from CCMv1 to CCMv2

Upgrading from CCMv1 to CCMv2 is available for customers who created their environments with CCMv1. The upgrade requires no downtime. If your environment needs to be updated, you will see a notification printed in your environment details.

Both CCMv1 and CCMv2 provide the same functionality but version 2 is an improvement over version 1. CCMv2 is more efficient in communicating with the CDP Control Plane and has a smaller footprint on the customer network.

The differences between CCMv2 and CCMv1 are related to improving reliability and efficiency. These differences are as follows:

- CCMv2 uses fewer agents (its agents are on FreeIPA nodes only), whereas CCMv1 has a larger footprint, with its agents running across not only on FreeIPA nodes but also on Data Lake and Data Hub nodes.
- CCMv2 uses the HTTPS protocol while CCMv1 is based on the SSH protocol. The use of HTTPS makes CCMv2 much more customer firewall friendly; it uses standard outbound ports which are allowed in almost all customer networks.
- Since CCMv2 uses HTTPS, it is more proxy-friendly. Customers can set up a proxy server such as Squid and inspect the underlying traffic to be more confident about the data that is being sent to the CDP Control Plane.
- CCMv2 uses token authentication.
- Newer Cloudera services, such as Machine Learning, Data Engineering, DataFlow, Data Warehouse, and others, only support CCMv2.

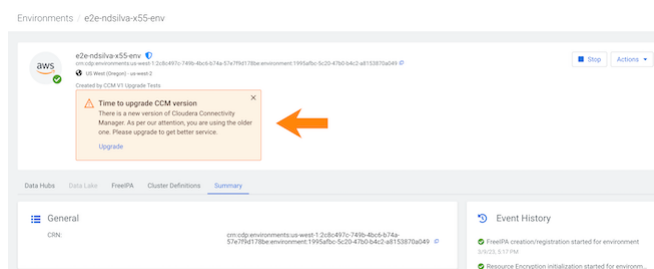
Before you begin

Prior to upgrading CCM, you should ensure that:

1. You are using the latest FreeIPA image. To upgrade your FreeIPA, follow the steps described in [Upgrade FreeIPA](#).
2. You have the network configured so that CCMv2 endpoints are available. See "Cloudera CCMv2" in the [AWS Azure GCP](#) outbound network access destinations.

Steps

1. If your environment should be updated from CCMv1 to CCMv2, you will see a "Time to upgrade CCM version" notification similar to the following in your environment details:



2. Upgrade CCM either via CDP web interface or CLI:
 - a. CDP web interface: Navigate to the Management Console > Environments and click on the tile corresponding to a specific environment. Trigger the upgrade by clicking on the Upgrade button present on the UI element.
 - b. CDP CLI: Use the following command:

```
cdp environments upgrade-ccm --environment <ENV_NAME_OR_CRN>
```

3. The status of your environment changes while the upgrade is pending. All environment components get upgraded (FreeIPA, Data Lake, and then Data Hubs). Check the event history to confirm that the upgrade has been completed. Once the upgrade has been completed, the status of the environment changes back to "Running". If an error occurs during the process, CDP rolls back the upgrade and reverts the environment to CCMv1.

Outbound network access destinations for CCMv2

When using CCM in an environment with limited internet access or proxy, make sure to add the following outbound network access destinations to the allow list.

Outbound network access destinations for CCMv2

Description/Usage	CDP Service	Destination	Protocol & Authentication	IP Protocol / Port	Comments
Cloudera CCMv2 Persistent Control Plane connection	All services	US-based Control Plane: *.v2.us-west-1.ccm.cdp.cloudera.com 35.80.24.128/27 EU-based Control Plane: *.v2.ccm.eu-1.cdp.cloudera.com 3.65.246.128/27 AP-based Control Plane: *.v2.ccm.ap-1.cdp.cloudera.com 3.26.127.64/27	HTTPS with mutual authentication	TCP/443	Multiple long-lived/persistent connections
Cloudera CCMv1 Persistent Control Plane connection	All services	*.ccm.cdp.cloudera.com 44.234.52.96/27	SSH public/private key authentication	TCP/6000-6049	One connection per cluster configured; persistent.



Note: If you have existing environments using CCMv1, you shouldn't remove the previously added CCMv1 specific outbound rules (ports 6000-6049).

Configuring a private VPC in AWS

When you create your CDP environment, you have two options: Have CDP set up your private network and security groups or set up the VPC with your private IPs and security groups. Either way, you must enable CCM.

Have CDP set up a private VPC network and security groups

If you choose to have CDP create your private network and security groups when you are testing or sandboxing CCM, when you enable CCM, CDP performs the following:

- Creates 3 public and 3 private subnets
- Creates the Data Lake and Data Hub clusters in the private subnet.

The public subnets have an internet gateway attached. The private subnets have a NAT gateway attached. When CDP creates the security group, it opens two ports to the CIDR range that you specify, port 22 and port 443. Use these ports only to access these clusters. For the list of security group rules that CDP creates, see [Default security group settings on AWS](#).

Set up your own private VPC network and security groups

If you choose to configure your own VPCs with private IPs, you will need the following:

- At least two private subnets in at least two different availability zones (AZs).
- Outbound traffic via the HTTP secure connection initiated by CCM to the Cloudera hosted NLBs on workload nodes.

In the AWS console, configure the following:

1. Create one public subnet and place the NAT gateway there to allow outbound connectivity in the private subnet.
2. Assign an internet gateway to the public subnet.
3. All inbound traffic must be on private subnets.
4. Create three private subnets. The private subnets must be in different availability zones (AZs).
5. Route the private subnet to the NAT gateway.
6. You must configure outbound traffic for CDP resources.
7. The workload clusters containing CCM (Knox, master, or CM for Classic Cluster) must be able to reach the Network Load Balancers (NLBs). You can use port 443 to connect to the NLBs.

Create your security groups as described in [Security groups](#).

Configuring a VNet with private IPs in Azure

When you create your CDP environment without public IPs, you have two options: Have CDP set up your private network and security groups, or set up the VNet with your private IPs and security groups. Either way, you must enable CCM.

Have CDP set up a VNet with private IPs and security groups

If you choose to have CDP create your private network and security groups when you are testing or sandboxing CCM, when you enable CCM, CDP performs the following:

- Creates more than 30 subnets. Azure does not distinguish between public and private subnets. By default they are all private.
- When CDP creates the security group, it opens two ports to the CIDR range that you specify, port 22 and port 443. Use these ports only to access these clusters. For a list of security group rules that CDP creates, see [Default security group settings on Azure](#).

Set up your own VNet with private IPs and security groups

If you choose to configure your own VNets without public IPs, you will need to configure the following in your Azure Portal:

1. Create at least one virtual network subnet. See [VNet and subnet planning](#) to determine the exact number of subnets needed.
2. You must configure outbound traffic for CDP resources.
3. The workload clusters containing CCM (Knox, master, or CM for Classic Cluster) must be able to reach the Network Load Balancers (NLBs).
4. You can use port 443 to connect to the NLBs.
5. Create your security groups as described in [Network security groups](#).

Configuring a VPC with private IPs in GCP

Prior to registering your GCP environment in CDP, you should set up a VPC network with private IPs, and create firewall rules.

You need the following:

- At least one subnet for hosts that will use CCM.
- Outbound traffic via the HTTP connection initiated by CCM should be allowed to the Cloudera hosted Network Load Balancers (NLBs) on workload nodes.

In the Google Cloud console, configure the following:

1. Create a VPC with custom (preferred) subnet configuration in the desired GCP region.
2. Create a GCP cloud router in the desired region.
3. Create a GCP NAT gateway specifying the previously created GCP cloud router.
4. You must configure outbound traffic for CDP resources.
5. The workload clusters containing CCM (Knox, master, or CM for Classic Clusters) must be able to reach the Network Load Balancers (NLBs).
6. You can use port 443 to connect to the NLBs.
7. Create your firewall rules as described in [Firewall rules](#).

Enabling CCM in the Management Console

When you register a new AWS, Azure, or GCP environment in CDP via web interface, Cluster Connectivity Manager (CCM) is enabled by default and the option to disable it has been removed from the web interface. CCM is also enabled by default when registering an environment via CDP CLI.

Related Information

[Register an AWS environment from CDP UI](#)

[Register an AWS environment from CDP CLI](#)

[Register an Azure environment from CDP UI](#)

[Register an Azure environment from CDP CLI](#)

[Register a GCP environment from CDP UI](#)

[Register a GCP environment from CDP CLI](#)

Troubleshooting CCMv1

You can troubleshoot cluster connection issues that might occur when you use CCMv1 by referring to the syslog on your Cloudera Manager node for classic clusters and your gateway and Knox nodes for data lake and data hub clusters.

Search for auto-ssh messages to diagnose any connection problems.

For example, CDP can communicate with clusters that are on private subnets with only private IPs without any additional network configuration or setup. However, CDP requires that these clusters have outbound connections to AWS NLBs hosted in the Cloudera's AWS account.

Common cases when connection via CCMv1 fails

The following list includes the most common cases when connection via CCM fails and steps for how to fix the problems.

In all the cases listed below, it is essential to check if the CCM domain name “.ccm.cdp.cloudera.com” has been whitelisted and the port range 6000-6049 has been whitelisted for SSL connections. This is because by default SSL connections are allowed only on the port 443 but connections to the Control Plane via CCM go through the ports in the range 6000-6049.

Reason for connection failure	How to fix the problem
You have not whitelisted the NLB domains or the port range 6000-6049 on your VPC.	Both the incoming and the outgoing traffic must be allowed from CCM NLB and port ranges.
You are using a transparent proxy and you did not explicitly whitelist NLB domains and the ports on the transparent proxies.	If you are using a transparent proxy, you need to explicitly whitelist the NLB domains and the ports on the transparent proxies.
You are using a non-transparent proxy but you did not register it in CDP.	If you are using a non-transparent proxy, you need to perform the steps in Setting up a non-transparent proxy in CDP documentation.
Your firewall doesn't allow traffic to/from NLBs.	The firewall configured on your VPC should allow the traffic to/from the NLBs and the port range 6000-6049.
The IPs of the NLB do not fall under the elastic IP CIDR block assigned to CCM NLBs.	CCM NLBs have IPs assigned to them which are taken from the CIDR block 44.234.52.96/27. If the NLB assigned to the environment/cluster returns IPs that do not fall under this CIDR block when an nslookup call is made, then the IPs returned need to be explicitly whitelisted so that the traffic from these IPs is allowed in your VPCs.

If you are still unable to determine the root cause and fix the issue, you may consider collecting more information.

Collecting information from the cluster where connection fails

Obtaining relevant information such as the NLB domain name, tunnel type (for example, CM), autossh client logs, and the syslogs from a cluster helps narrow down the root cause.

The following steps explain how to obtain this information.

1. Check if the autossh process is running on the cluster:

- a. SSH into the node on which the issue is being seen. Execute `ps -aux | grep autossh` to see if the autossh client responsible for creating the tunnel from the node to CDP is running.
- b. If the process is running, it would show the command similar to:

```
autossh -M 0 -o ServerAliveInterval 30 -o ServerAliveCountMax 3 -o UserKnownHostsFile=/etc/ccm/ccm.pub -N -T -R 0.0.0.0:0:localhost:<LOCAL_PORT> -i /tmp/pk.key -p <NLB_PORT> <INITIATOR_ID>@<NLB_HOST_NAME> -vvv
```

- c. Collect the values of NLB_HOST_NAME, NLB_PORT, INITIATOR_ID from the above step.
- d. Share the collected information with Cloudera support.

2. Check if the NLB is reachable from the node:

- a.** On the node, execute the command `nslookup <NLB_HOST_NAME>` to check if the NLB is reachable from the node.
- b.** Use tools such as `telnet` or `netcat` to check if it is possible to reach the NLB on the specific `NLB_PORT`. For example:

```
telnet <NLB_HOST_NAME> <NLB_PORT>
```

- c.** If `telnet` is not installed on the node, you can use `curl telnet` as an alternative. For example:

```
curl -vv telnet://<NLB_HOST_NAME>:<NLB_PORT>
```

- d.** If this fails, then the traffic to the Cloudera network is blocked on your VPC. This means that you should check the outbound rules on their VPC to make sure that the traffic to the cloudera network is allowed.

3. Collect the following logs and share them with Cloudera support:

- a.** The dump of `/var/log/messages` (which stores the global system messages) is necessary to debug the issue with the SSH tunnel setup by CCM.
- b.** The dump of the output of the `tcpdump` command on the problematic node captures the network messages with the description of contents of packets between the node and the CDP control plane. Cloudera support may ask for this if needed or when the already collected information is not sufficient to debug the issue.
- c.** The dump of `/var/log/socket_wait_cleanup.log` (which stores the logs related to the restarts of `autossh` process) Note that this log file is only present in environments created after December 12, 2020.

4. Collect the firewall logs and share them with Cloudera support:

- a.** It is useful for Cloudera support to see the logs from your firewall to see if the firewall is closing the connection to the CDP Control Plane.

5. Collect and share the environment/cluster and CCM related information:

- a.** To look up the logs, Cloudera support needs the account id of the customer, environment CRN, and the time period to scan the logs for the issue.

FreeIPA sync issues

Error or issue details	Resolution
<p>502 or "Bad Gateway" error such as:</p> <ul style="list-style-type: none"> "Bad Gateway, cause=com.cloudera.cdp.cm.ApiException: Bad Gateway" Unexpected response from FreeIPA; details: code: 502 Status: 502 Bad Gateway Response 	<p>A new parameter defining connection timeout was added to the autossh script placed on images used to provision clusters. Without it, there is no timeout for CCM connections and CCM related services won't restart automatically. If you provisioned Freeipa or Data Lake prior to the implementation of this parameter, you need to update the script by using the following steps:</p> <ol style="list-style-type: none"> Modify autossh command on FreeIPA, Data Lake and Data Hub nodes in this script: <pre>/cdp/bin/reverse-tunnel.sh</pre> <p>Add new parameter to the command:</p> <pre>-o "ConnectTimeout 30"</pre> <p>Original command:</p> <pre>exec autossh -M 0 -o "ServerAliveInterval 30" -o "ServerAliveCountMax 3" -o UserKnownHostsFile=\${CCM_PUBLIC_KEY_FILE} -N -T -R \${LOCAL_IP}:0:localhost:\${CCM_TUNNEL_SERVICE_PORT} -i \${PRIVATE_KEY} -p \${CCM_SSH_PORT} \${USER}@\${CCM_HOST} -vvv</pre> <p>New command:</p> <pre>exec autossh -M 0 -o "ConnectTimeout 30" -o "ServerAliveInterval 30" -o "ServerAliveCountMax 3" -o UserKnownHostsFile=\${CCM_PUBLIC_KEY_FILE} -N -T -R \${LOCAL_IP}:0:localhost:\${CCM_TUNNEL_SERVICE_PORT} -i \${PRIVATE_KEY} -p \${CCM_SSH_PORT} \${USER}@\${CCM_HOST} -vvv</pre> Restart ccm-tunnel@KNOX service on Data Lake and Data Hub nodes and ccm-tunnel@GATEWAY on FreeIPA, Data Lake and Data Hub nodes: <pre>systemctl restart ccm-tunnel@KNOX</pre> <pre>systemctl restart ccm-tunnel@GATEWAY</pre> Check if modification of the command was successful: <pre>ps aux grep autossh</pre>

Troubleshooting CCMv2

This page lists common issues related to troubleshooting workload connectivity using CCMv2.

Incorrect network setup

Reason for connection failure	How to fix the problem
<p>For CCM to work, the agent must be able to connect to its counterpart running inside the Control Plane. A misconfigured network often blocks this path. This can happen in following ways:</p> <ul style="list-style-type: none"> • Customer's network does not resolve CCM DNS addresses. • Customer's network does not allow outbound connections. • Customer's network does not allow outbound connections to Cloudera IPs. • Customer's network does not allow outbound connections to the required ports. • Customer expects us to use a non-transparent proxy which does not allow connection to the required Cloudera IPs / ports. • Customer expects us to use a non-transparent proxy which does deep-packet inspection to sniff underlying traffic and blocks CCM traffic. 	<p>Running the <code>cdp-telemetry doctor</code> command helps diagnose most problems of this category. See Collecting workload logs on page 16.</p> <p>Fixing the problem usually means ensuring that you set up the outbound network access destinations that are listed as CDP prerequisites. See:</p> <ul style="list-style-type: none"> • AWS outbound network access destinations • Azure outbound network access destinations • GCP outbound network access destinations

Collecting workload logs

The easiest and recommended way to collect logs is to use the `cdp-telemetry doctor` tool.

The `ccm` module of the `cdp-telemetry doctor` tools collects all the information required on a given node.

```
cdp-telemetry doctor ccm status
cdp-telemetry doctor ccm report
```



Note: Using `cdp-telemetry doctor` should report all the information CDP support requires on the cluster for debugging. However, the tool needs to be launched on every node where an agent is expected to run.

To generate the diagnostic bundle, you need to `ssh` into each node of concern and invoke the following command:

```
cdp-telemetry doctor ccm report
```

Errors related to classic clusters

See [Troubleshooting classic cluster registration errors](#).

Data Warehouse and private networking

Cloudera Data Warehouse service supports private deployments in AWS, which use private subnets.

For information on setting up an AWS environment with private subnet support, refer to [Setting up private networking in AWS environments](#) documentation. in Data Warehouse documentation.

Public Endpoint Access Gateway

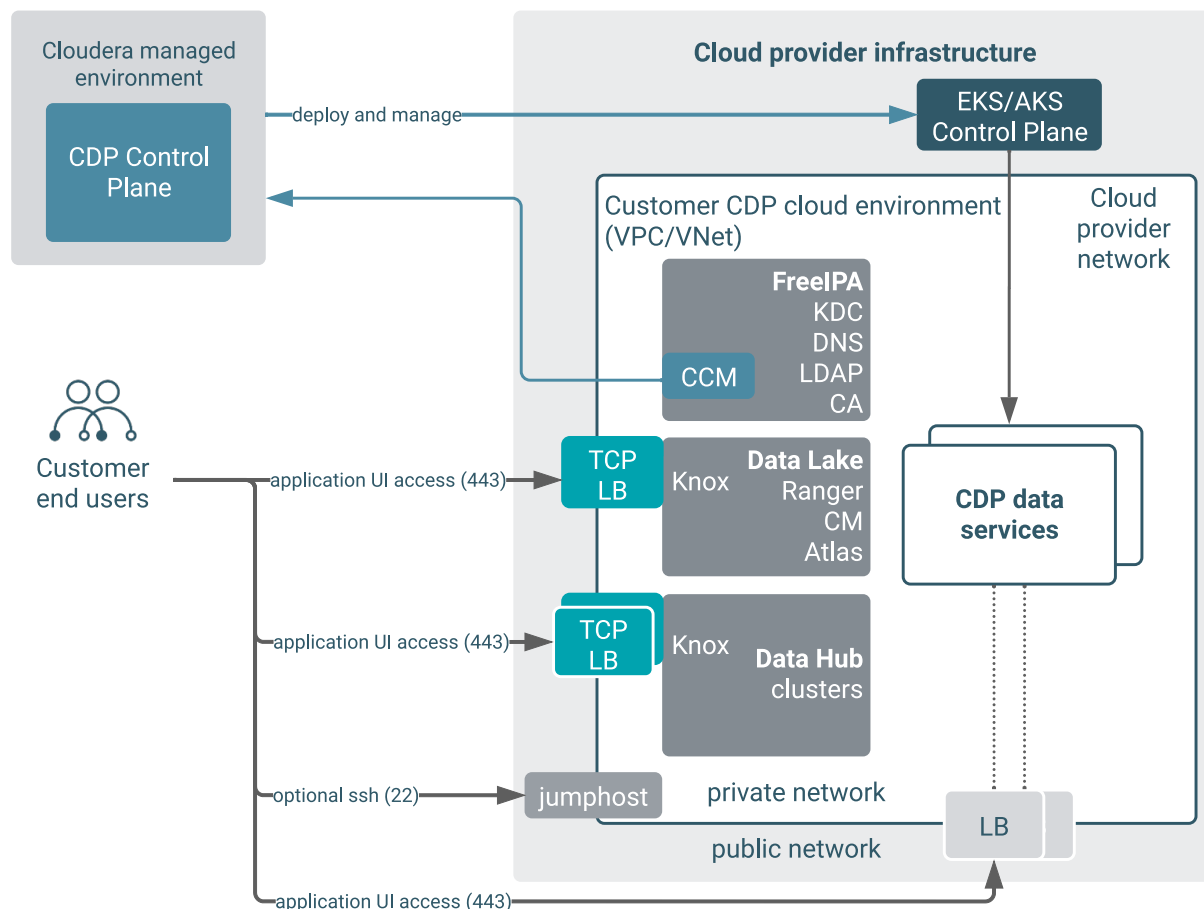
If the network into which you are deploying your CDP environment does not have pre-established connectivity with your corporate network, enabling the Public Endpoint Access Gateway can reduce the complexity users face when interacting with the CDP endpoints.

The recommended way to deploy production-ready CDP environments is to deploy them on private networks, but this additional security makes it difficult for users to access UIs and APIs without configuring complex network connectivity between users and internal cloud provider networks. The Public Endpoint Access Gateway provides secure connectivity to UIs and APIs in Data Lake and Data Hub clusters deployed using private networking, allowing

users to access these resources without complex changes to their networking or creating direct connections to cloud provider networks.

You can enable the Public Endpoint Access Gateway when registering your AWS, Azure, or GCP environment in CDP. The gateway interfaces the Knox service, which is automatically integrated with your identity provider configured in CDP, allowing you to authenticate using your SSO credentials without any additional configuration. All communication with the gateway is over TLS, so connections are secure. You can control the IP ranges from where connections to the gateway can be established by configuring your security groups.

The following diagram illustrates this setup:



Note: The gateway provides secure connectivity to UIs and APIs. All Knox-enabled endpoints are supported. The gateway does not cover SSH or data access paths (such as Kafka Broker and NiFi Site2Site endpoints). Cloudera recommends that you set up connectivity between private networks in the public cloud and internal customer networks for secure and fast Kafka and NiFi deployments.

Enable Public Endpoint Access Gateway for AWS

You can enable Public Endpoint Access Gateway during AWS environment registration.

Once activated, the gateway will be used for the Data Lake and all the Data Hubs within the environment. There is no way to activate it on a per Data Lake or per Data Hub level. Once it is enabled for an environment, there is no way to deactivate it. The gateway can be used either with an existing VPC or with a new VPC created by CDP.

Prerequisites

- If you choose to enable Public Endpoint Access Gateway, CDP will create two AWS network load balancers (AWS NLB) per cluster (that is for each Data Lake and Data Hub). Make sure that your AWS NLB limits allow for the load balancer creation.
- If you are using your existing network, you should have at least 2 public subnets in the VPC that you would like to use for CDP. The availability zones of the public and private subnets must match.

Steps

For CDP UI

When registering your AWS environment, make sure to do the following:

1. On the Region, Networking, and Security page, under Network, select your existing VPC or select to have a new VPC created.
2. If you selected an existing VPC, select at least two existing private subnets (or at least three subnets if you would like to provision Data Warehouse instances).
3. Click on Enable Public Endpoint Access Gateway to enable it. This enables UIs and APIs of the Data Lake and Data Hub clusters to be accessible over the internet.
4. If you selected an existing VPC, under Select Endpoint Access Gateway Subnets, select the public subnets for which you would like to use the gateway. The availability zones of the public subnets must be the same as the availability zones of the private subnets selected under Select Subnets.
5. Under Security Access Settings, make sure to restrict access to only be accepted from sources coming from your external network range.



Note: The security access settings do not apply to the network load balancer used by the Public Endpoint Access Gateway, but they apply to the instances that are running in private subnets and to which the Public Endpoint Access Gateway routes traffic. Therefore the security access settings should allow the users' public IP ranges to be able to connect through the public load balancer.

6. Finish registering your environment.

For CDP CLI

During environment registration via CDP CLI, you can optionally enable public endpoint access gateway using the following CLI parameters:

```
--endpoint-access-gateway-scheme PUBLIC
--endpoint-access-gateway-subnet-ids subnet-0232c7711cd864c7b subnet-05d4769d88d875cda
```

The first parameter enables the gateway and the second one allows you to specify public subnets. The availability zones of the public subnets must be the same as the availability zones of the private subnets specified under --subnet-ids. For example:

```
cdp environments create-aws-environment \
--environment-name gkldev \
--credential-name gklcred \
--region "us-west-2" \
--security-access cidr=0.0.0.0/0 \
--authentication publicKeyId="gkl" \
--log-storage storageLocationBase=s3a://gklpriv-cdp-bucket,instanceProfile=arn:aws:iam::152813717728:instance-profile/mock-idbroker-admin-role \
--vpc-id vpc-037c6d94f30017c24 \
--subnet-ids subnet-0232c7711cd864c7b subnet-05d4769d88d875cda \
--endpoint-access-gateway-scheme PUBLIC \
--endpoint-access-gateway-subnet-ids subnet-0232c7711cd864c7b subnet-05d4769d88d875cda \
--free-ipa instanceCountByGroup=1 \
```

Equivalent CLI JSON for an environment request looks like this:

```
"endpointAccessGatewayScheme": "PUBLIC",
"endpointAccessGatewaySubnetIds":
  [ "subnet-0232c7711cd864c7b",
    "subnet-05d4769d88d875cda" ],
```

Enable Public Endpoint Access Gateway for Azure

You can enable Public Endpoint Access Gateway during Azure environment registration.

Once activated, the gateway will be used for the Data Lake and all the Data Hubs within the environment. There is no way to activate it on a per Data Lake or per Data Hub level. Once it is enabled for an environment, there is no way to deactivate it. The gateway can be used either with an existing VNet or with a new VNet created by CDP.

If you choose to enable Public Endpoint Access Gateway, CDP will create two Azure load balancers per cluster (that is, two for each Data Lake and Data Hub).

Steps

For CDP UI

When registering your Azure environment, make sure to do the following:

1. On the Region, Networking, and Security page, under Network, select your existing VNet or select to have a new VNet created.
2. If you selected an existing VNet, select at least one existing private subnet (or at least three subnets if you would like to provision Data Warehouse instances).
3. Click on Enable Public Endpoint Access Gateway to enable it. This enables UIs and APIs of the Data Lake and Data Hub clusters to be accessible over the internet.
4. Under Security Access Settings, make sure to restrict access to only be accepted from sources coming from your external network range.



Note: The security access settings do not apply to the load balancer used by the Public Endpoint Access Gateway, but they apply to the instances that are running in private subnets and to which the Public Endpoint Access Gateway routes traffic. Therefore the security access settings should allow the users' public IP ranges to be able to connect through the public load balancer.

5. Finish registering your environment.

For CDP CLI

During Azure environment registration via CDP CLI, you can optionally enable public endpoint access gateway using the `--endpoint-access-gateway-scheme` CLI parameter. For example:

```
cdp environments create-azure-environment
...
--endpoint-access-gateway-scheme PUBLIC
```

Equivalent CLI JSON for an environment request looks like this:

```
cdp environments create-azure-environment
...
"endpointAccessGatewayScheme": "PUBLIC"
```

Enable Public Endpoint Access Gateway for GCP

You can enable Public Endpoint Access Gateway during GCP environment registration.

Once activated, the gateway will be used for the Data Lake and all the Data Hubs within the environment. There is no way to activate it on a per Data Lake or per Data Hub level. Once it is enabled for an environment, there is no way to deactivate it.

If you choose to enable Public Endpoint Access Gateway, CDP will create two Google Cloud Load Balancers (GCLB) per cluster (that is, two for each Data Lake and two for each Data Hub).

Prerequisites

If you would like to use this feature, make sure that "Private Google Access" is disabled on at least one subnet in the VPC.

Steps

For CDP UI

When registering your GCP environment, make sure to do the following:

1. On the Region, Networking, and Security page, under Network, select your existing VPC network.
2. Select at least one existing private subnet.
3. Click on Enable Public Endpoint Access Gateway to enable it. This enables UIs and APIs of the Data Lake and Data Hub clusters to be accessible over the internet.



Enable Endpoint Access Gateway

Select Subnets for the Endpoint Access Gateway*

These subnets must match the availability zones of the selected subnets above.

Please select subnet(s) *



4. If you selected an existing VPC, under Select Endpoint Access Gateway Subnets, select the public subnets for which you would like to use the gateway. The availability zones of the public subnets must be the same as the availability zones of the private subnets selected under Select Subnets.
5. Under Security Access Settings, make sure to restrict access to only be accepted from sources coming from your external network range.



Note: The security access settings do not apply to the load balancer used by the Public Endpoint Access Gateway, but they apply to the instances that are running in private subnets and to which the Public Endpoint Access Gateway routes traffic. Therefore the security access settings should allow the users' public IP ranges to be able to connect through the public load balancer.

6. Finish registering your environment.

For CDP CLI

During GCP environment registration via CDP CLI, you can optionally enable Public Endpoint Access Gateway using the following CLI parameter:

```
--endpoint-access-gateway-scheme PUBLIC
```

For example:

```
cdp environments create-gcp-environment
...
--endpoint-access-gateway-scheme PUBLIC
```

Equivalent CLI JSON for an environment request looks like this:

```
cdp environments create-gcp-environment
...
"endpointAccessGatewayScheme": "PUBLIC"
```

Azure Load Balancers in Data Lakes and Data Hubs

The [Azure Load Balancer](#) is used in multiple places in CDP Data Lakes and Data Hubs. It is used as a frontend for Knox in both Data Lakes and Data Hubs, and for Oozie HA in HA Data Hubs.

The following table describes all use cases where an Azure Load Balancer is used in Data Lakes and Data Hubs running on Azure:

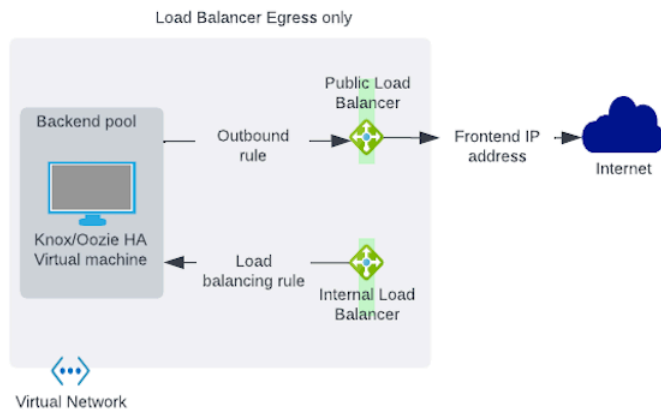
CDP component	Azure Load Balancer use case
Data Lake	A load balancer is configured in front of Knox for Data Lakes of all shapes.
HA Data Hub	A load balancer is configured for all Data Hubs created from a default template where Knox and/or Oozie is running in HA mode. This can be overridden by setting the “enableLoadBalancer” setting in a custom template to “false”.
An environment with Public Endpoint Access Gateway enabled	When the Endpoint Gateway is enabled, Load balancers are created in front of Knox for all Data Lakes and Data Hubs attached to the environment.

In the event that a Data Lake or Data Hub uses private networks (meaning the “Create Public IPs” option is disabled during environment creation and the Public Endpoint Access Gateway is not enabled), an internal load balancer is created for ingress traffic to Knox in all Data Lakes and in Knox HA Data Hubs running in that environment.

Because CDP uses a Standard SKU Azure Load Balancer, the private load balancer does not allow for public egress, and a public load balancer is required for public outbound destinations as described in the [Azure network reference architecture](#). If you are creating an environment with an existing network of your own configuration, which is assumed to be fully private (the “Create Public IPs” option and Public Endpoint Access Gateway are disabled), it is your responsibility to create egress connectivity for the required subnets in your VNet. This can be accomplished through a [NAT gateway setup](#) or [user-defined routing](#).

If you want CDP to create a secondary load balancer for egress in an existing network of your own configuration, be aware that it requires certain public IP permissions that are granted as part of the [required Azure permissions](#). However, you can create this secondary public egress load balancer when you create an Azure environment through the CLI with the options `--no-enable-outbound-load-balancer` and `--enable-outbound-load-balancer`. The secondary public egress load balancer created by CDP has only outbound rules defined, and does not handle ingress traffic.

This is illustrated in the following diagram:



If you are creating a new network during environment registration, CDP ensures that egress connectivity is available. If the "Create Public IPs" option and Public Endpoint Access Gateway are disabled in your network, a separate load balancer is created for egress, though this load balancer requires certain public IP permissions that are granted as part of the [required Azure permissions](#). If either "Create Public IPs" or Public Endpoint Access Gateway is enabled, then a public load balancer is created to handle both public ingress to port 443 and public egress.



Note: When an Azure VNet has both a NAT Gateway and a public load balancer for outbound connectivity, the NAT Gateway takes precedence. For more information, see [Design virtual networks with NAT gateway](#).