Connection to Private Subnets with CCMv2 (Preview)

Date published: 2021-08-10
Date modified: 2021-08-10

Legal Notice

Contents

CCMv2 vs CCMv1

CCMv2 replaces CCMv1. All new environments created after enabling CCMv2 on your tenant use CCMv2, but existing environments created prior to that continue to use CCMv1.

While CCMv1 establishes and uses a tunnel based on the SSH protocol, with CCMv2 the connection is via HTTP. The steps to register an environment with CCMv2 are similar to CCMv1 configuration steps. The main differences are:

- If you are deploying in an environment with restricted outbound network access, a new port needs to be added to the allow list. See Outbound network access destinations for CCMv2.
- If you are registering a Classic Cluster, the steps have changed. See *Registering Classic Clusters with CCMv2* posted in the Preview Features.

Cluster Connectivity Manager overview

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 Cluster Connectivity Manager (CCM) enables the CDP Control Plane to communicate with workload clusters that do not expose public IPs. This functionality is available for CDP deployments on all supported cloud providers. 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.

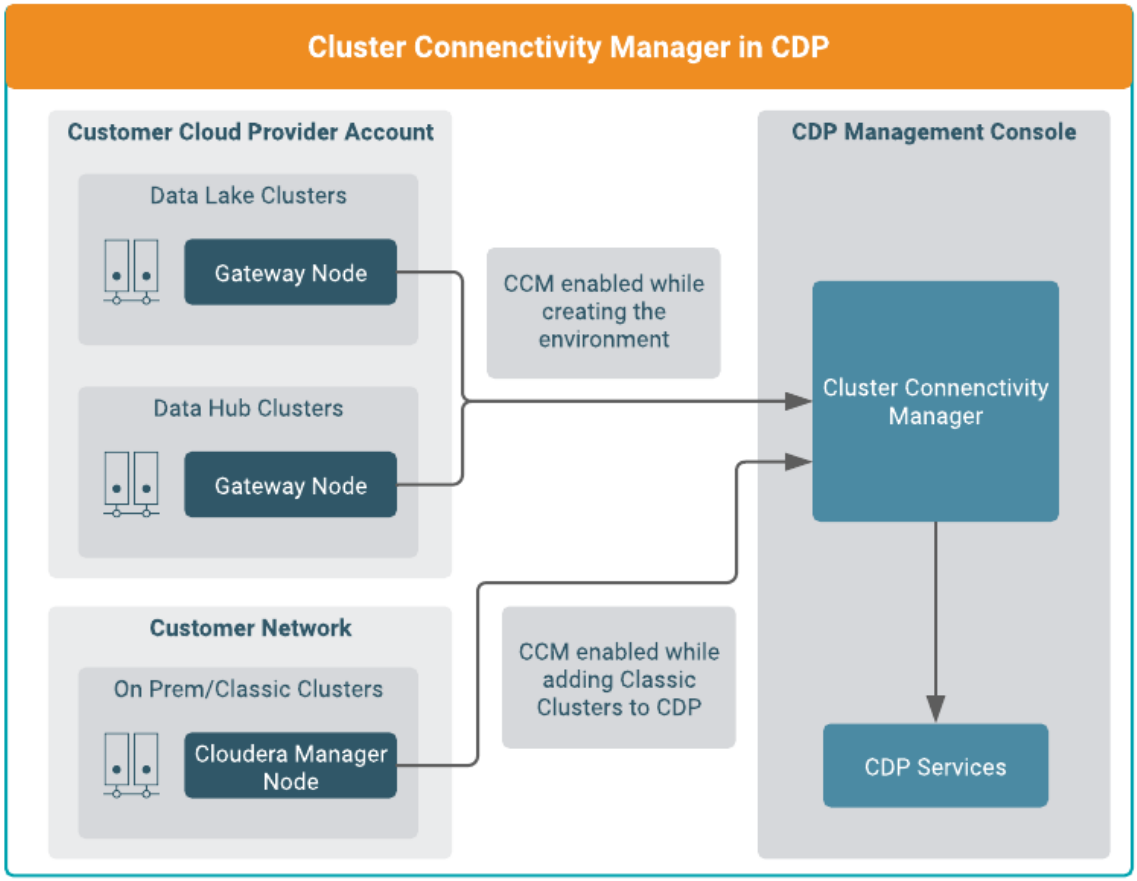
CCMv2

CCMv2 agents deployed on FreeIPA nodes initiate an HTTP 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.

The following diagram illustrates which CDP services are supported by CCMv2 and when the connection is enabled for each service:

Note: CCMv2 supports Data Lake, Data Hub, and classic clusters running on AWS, Azure, and GCP. It is also possible to contact Cloudera and enable CCMv2 for Data Warehouse deployments on Azure (Preview), but not on AWS. Other services are currently not supported.

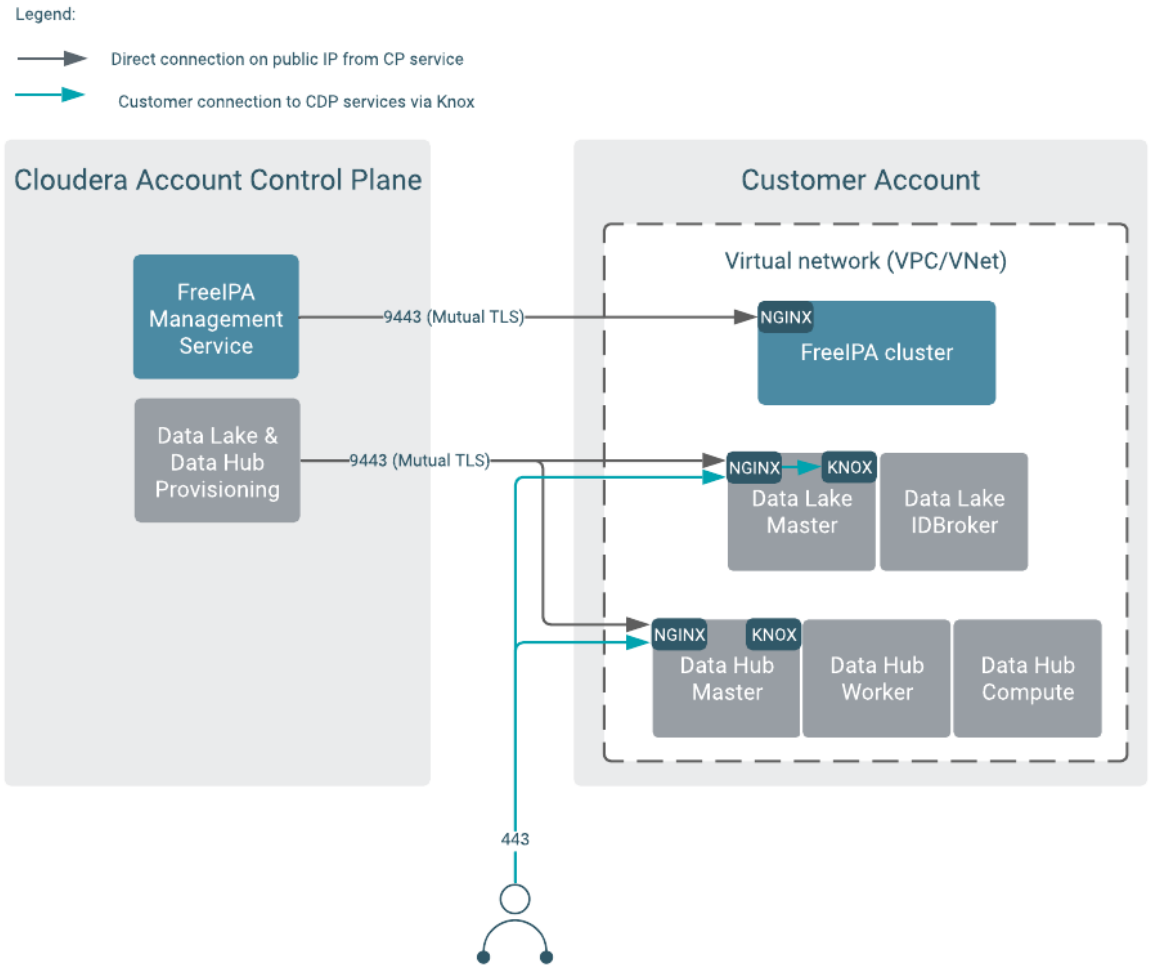


Important: CCM establishes connectivity with the CDP Control Plane. You must ensure you have connectivity to the private network you set up. To ensure connectivity, you can either enable Public Endpoint Access Gateway (available for AWS only) or work with your network IT department to set up this connectivity. See Outbound network access destinations for CCMv2 documentation.

The following three diagrams illustrate CDP connectivity to a customer account without using CCM, using CCMv2, and using CCMv1.

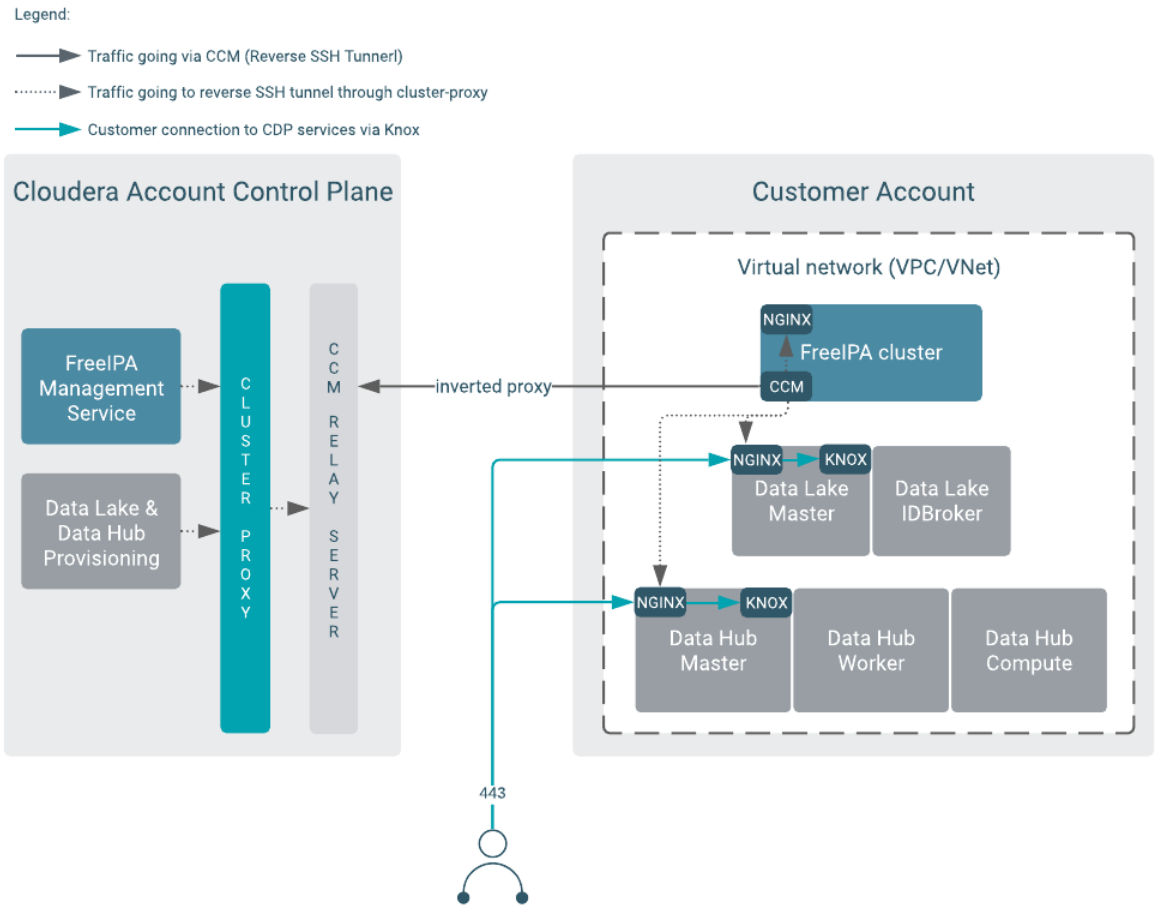
The first diagram illustrates the CDP connectivity to a customer account without CCM. When CDP is deployed in public mode, security groups (called firewall rules in Google Cloud) must be configured to allow inbound access to the environment from the CDP Control Plane exit IP range, in addition to end-user access rules restricting traffic to only originate from the customer's own network.

- This is done automatically for new networks created by CDP, so the only CIDRs required during deployment are from the customer's own network.
- Customer-provided security groups must be configured to whitelist the Cloudera Control Plane CIDRs in addition to the customer's own network CIDR.



The second diagram illustrates the CDP connectivity to a customer account with CCMv2 enabled. 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

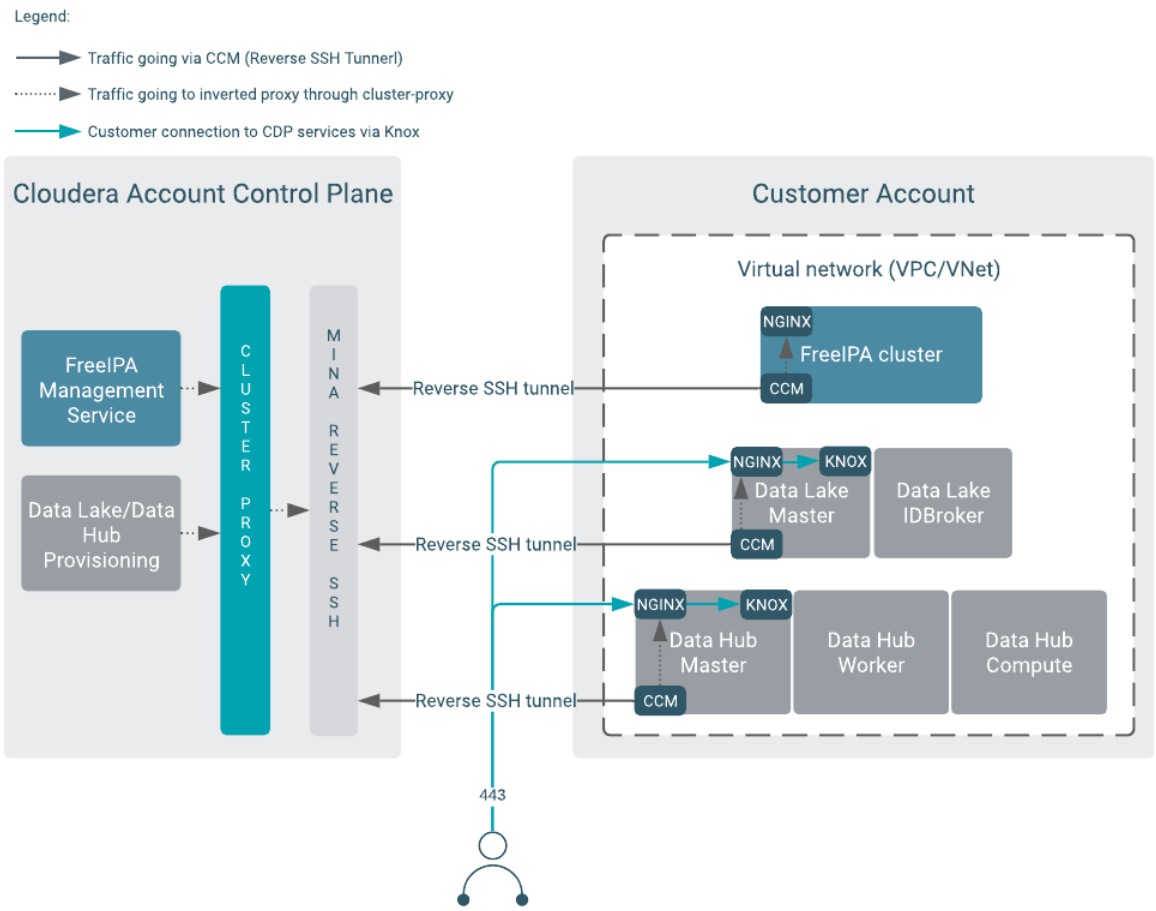
restrict access to a certain range of private IPs.



CCMv1

The third 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 HTTP. 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.

Note: CCMv1 supports Data Lake, Data Hub, and classic clusters running on AWS, Azure, and GCP. No other services are supported.



Outbound network access destinations for CCMv2

When using CCMv2 in an environment with limited internet access or proxy, make sure to refer to the following two documentation sets:

- Outbound network access destinations documentation for AWS, Azure, and GCP.
- In addition, additional endpoints listed below must be added to the allow list.

Additional endpoints for CCMv2

Description/ Usage	CDP Service	Destination	Protocol & Authentication	IP Protocol / Port	Comments

Cloudera CCMv2 Persistent Control Plane connection	All services	IP: 35.80.24.128/27, 3.65.246.128/27 Hostname pattern: *.ccm.cdp.cloudera.com	HTTPS with mutual authentication	TCP/443	Multiple long-lived/persistent connections

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

Configuring a CCM-compatible VPC on 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 your private 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 VPC 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. See Outbound network access destinations for CCMv2.

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.
8. Create your security groups as described in Security groups.

Configuring a CCM-compatible VNets on 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 your private 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 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 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 CCM-compatible VPC 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.

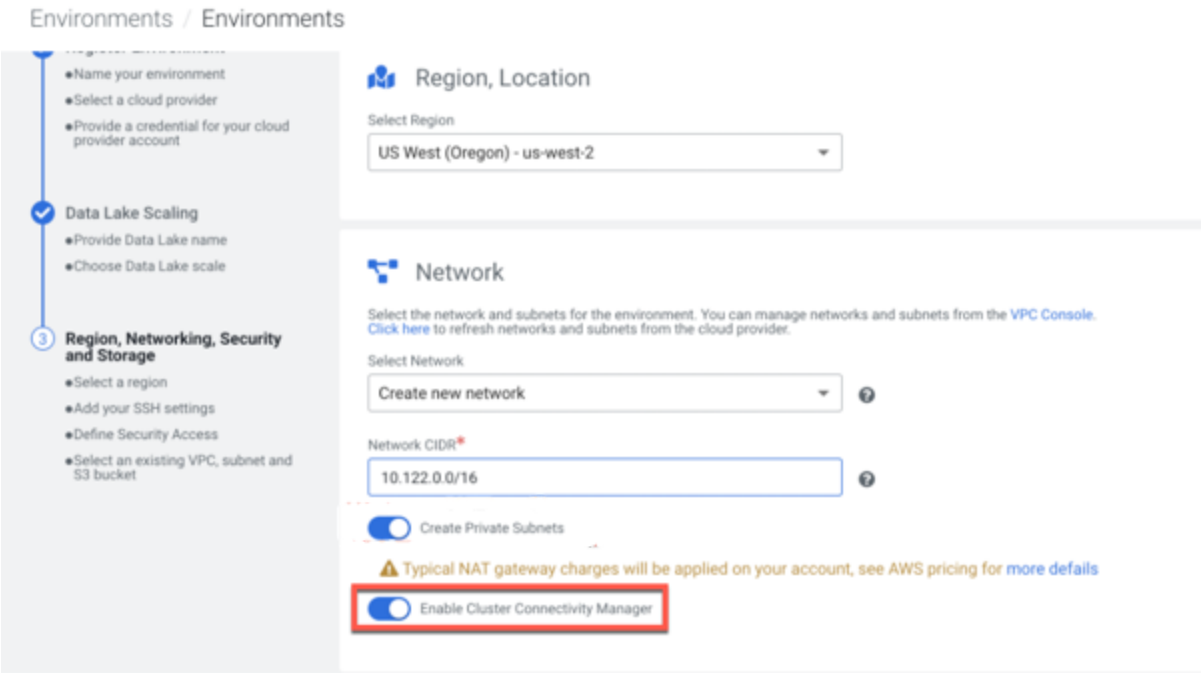
In the Google Cloud console, configure the following:

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 using the Management Console

You can enable the Cluster Connectivity Manager while creating the environment using the CDP Management Console.

While registering an environment in CDP on the **Region, Networking and Storage** page, under **Security Access Settings**, make sure that **Enable Cluster Connectivity Manager** is enabled. This option is enabled by default.



For detailed documentation related to environment registration, see:

- Register an AWS environment
- Register an Azure environment
- Register a GCP Environment

Note: CCM is used by default for environments registered via CDP web interface, but it is not used by default for environments registered via CDP CLI. Therefore, if you are registering your environment via CDP CLI and you would like to use CCM, you must explicitly enable it during environment registration.

Enabling CCM using the CDP CLI

You can enable the Cluster Connectivity Manager while creating the environment using the CLI.

While creating an environment using CDP CLI, specify the parameter to enable CCM as part of the create environment command as follows:

`--enable-tunnel`
Add this parameter to enable CCM for an environment.

`--no-enable-tunnel`
Use this parameter to disable CCM for an environment.

Note: CCM is used by default for environments registered via CDP web interface, but it is not used by default for environments registered via CDP CLI. Therefore, if you are registering your

environment via CDP CLI and you would like to use CCM, you must explicitly enable it during environment registration.