..

# Creating hybrid environments

**Date published: 2025-11-14**
**Date modified:**

**https://docs.cloudera.com/**

# Legal Notice

# Contents

# Prerequisites

## Security requirements

Learn how Kerberos authentication and bidirectional cross-realm trust ensure security.

- You need to ensure that the Kerberos authentication method is enabled. It ensures secure communication with the services, regardless of whether they reside in public cloud or on-premises. Both public cloud and on-premises clusters must be Kerberized.
- Kerberos bidirectional cross-realm trust setup will be established so that on-premise KDC and public cloud KDCs trust each other.

## Network requirements

Learn the address, domain name service, and network mechanism requirements for the hybrid cloud architecture.

### Concepts

The following concepts need to be considered when reviewing the network requirements for hybrid cloud:

- CIDR

  Classless Inter-Domain Routing is a notation to represent an IPv4 address range, for example: 10.0.0.0/8. The bitmask, the number after the slash, represents the number of fixed bits in the IP address range. Higher bitmask values indicate a lower number of addresses in the range.
- Private IP ranges

  IP ranges used for cloud networks and on-premise networks.

  - 10.0.0.0-10.255.255.255
  - 172.16.0.0-172.31.255.255
  - 192.168.0.0-192.168.255.255
- Overlapping and non-overlapping networks
  - Overlapping
    - 10.0.0.0/8 and 10.0.0.0/16
  - Not overlapping
    - 10.0.0.0/16 and 10.1.0.0/16
    - 10.0.0.0/8 and 172.16.0.0/16
    - 10.0.0.0/16 and 192.168.0.0./16
    - 192.0.2.128/28 and 192.0.2.144/28
- DNS (Domain Name System)

  A hierarchical, distributed naming system used to translate human-readable domain names into IP addresses. DNS enables resources in cloud and on-premise networks to communicate using stable, memorable names instead of numerical IP addresses. DNS resolution can occur through public resolvers, private/internal DNS servers, or a hybrid approach where conditional forwarding rules direct specific domains to their correct authoritative resolver. Consistent DNS configuration across cloud and on-premise networks is essential to ensure service discovery, Kerberos authentication, and cross-environment workload communication.
- Direct Connect / Dedicated Interconnect + Site-to-Site VPN

  Mechanisms used to establish secure, private connectivity between on-premise data centers and cloud networks. A **Direct Connect** (or equivalent dedicated interconnect service) provides a high-bandwidth, low-latency physical

link that bypasses the public internet. A **Site-to-Site VPN** creates an encrypted IPsec tunnel over the internet to connect the on-premise network to the cloud virtual network. These two connectivity options are often used together, with the VPN serving as a failover path when the dedicated link becomes unavailable. When configuring hybrid connectivity, all participating networks must use non-overlapping CIDR ranges, and routing must be set up so on-premise subnets and cloud subnets can reach each other consistently.

### Requirements

The following requirements must be met when creating hybrid environments:

- The on-premise network and the public cloud network must be peered.
- No overlapping CIDRs between the on-premise network and the public cloud network.
- Fully connected routing between the on-premise network and the public cloud network with a bidirectional line of sight based on the available ports.
- High-bandwidth and lower-latency public cloud to private cloud network to transfer the data required for cloud bursting.
- No network address translation between the on-premise network and the public cloud network.
- Firewalls are permitted and supported, but they must allow the communication channels detailed in Security policies and firewall rules.
- Egress connectivity must be enabled on both the on-premise data lake and on the public cloud FreeIPA instance.
  - For outbound destinations regarding the on-premise network, please refer to our documentation:
    - If AWS is the provider, refer to AWS outbound network access destinations
    - If Google Cloud is the provider, refer to GCP outbound network access destinations
    - If Microsoft Azure is the provider, refer to Azure outbound network access destinations
  - For outbound destinations regarding the public cloud network, please refer to Outbound network access destinations for Cluster Connectivity Manager v2 in our documentation.

## Security policies and firewall rules

Learn about hybrid environments' security policies and firewall rules.

- Ensure that security policies and firewall rules allow traffic from all nodes in the public cloud Cloudera on cloud network to access the on-premise ports used by runtime services. Please refer to Ports Used by Cloudera Runtime Components for the list of ports typically used by runtime components.
- Ensure that security policies and firewall rules allow traffic between your FreeIPA instance and your Active Directory Domain Controller. Direct communication between FreeIPA servers and Active Directory Domain Controllers is required over a range of ports for various protocols.

The following table details the essential ports that must be opened between all **FreeIPA Trust Controllers and all Active Directory Domain Controllers**.

| Port | Protocol | Service Name | Direction | Purpose |
|---|---|---|---|---|
| 53 | TCP/UDP | DNS | Bidirectional | Name and Service Record Resolution |
| 88 | TCP/UDP | Kerberos | Bidirectional | Authentication, Ticket Granting |
| 123 | UDP | NTP | Bidirectional | Time Synchronization |
| 135 | TCP | RPC Endpoint Mapper (EPMAP) | Bidirectional | LSA RPC for Trust Management |
| 138 | UDP | NetBIOS Datagram Service | Bidirectional | NetBIOS Communication |
| 139 | TCP | NetBIOS Session Service | Bidirectional | NetBIOS Communication |
| 389 | TCP/UDP | LDAP | Bidirectional | User/Group Resolution, Trust Validation |
| 445 | TCP | SMB/CIFS | Bidirectional | Trust Creation (IPC$ share access) |
| 464 | TCP/UDP | Kerberos Password Change | Bidirectional | Kerberos kpasswd service |
| 636 | TCP | LDAPS | IPA -> AD | Secure User/Group Resolution (if configured) |

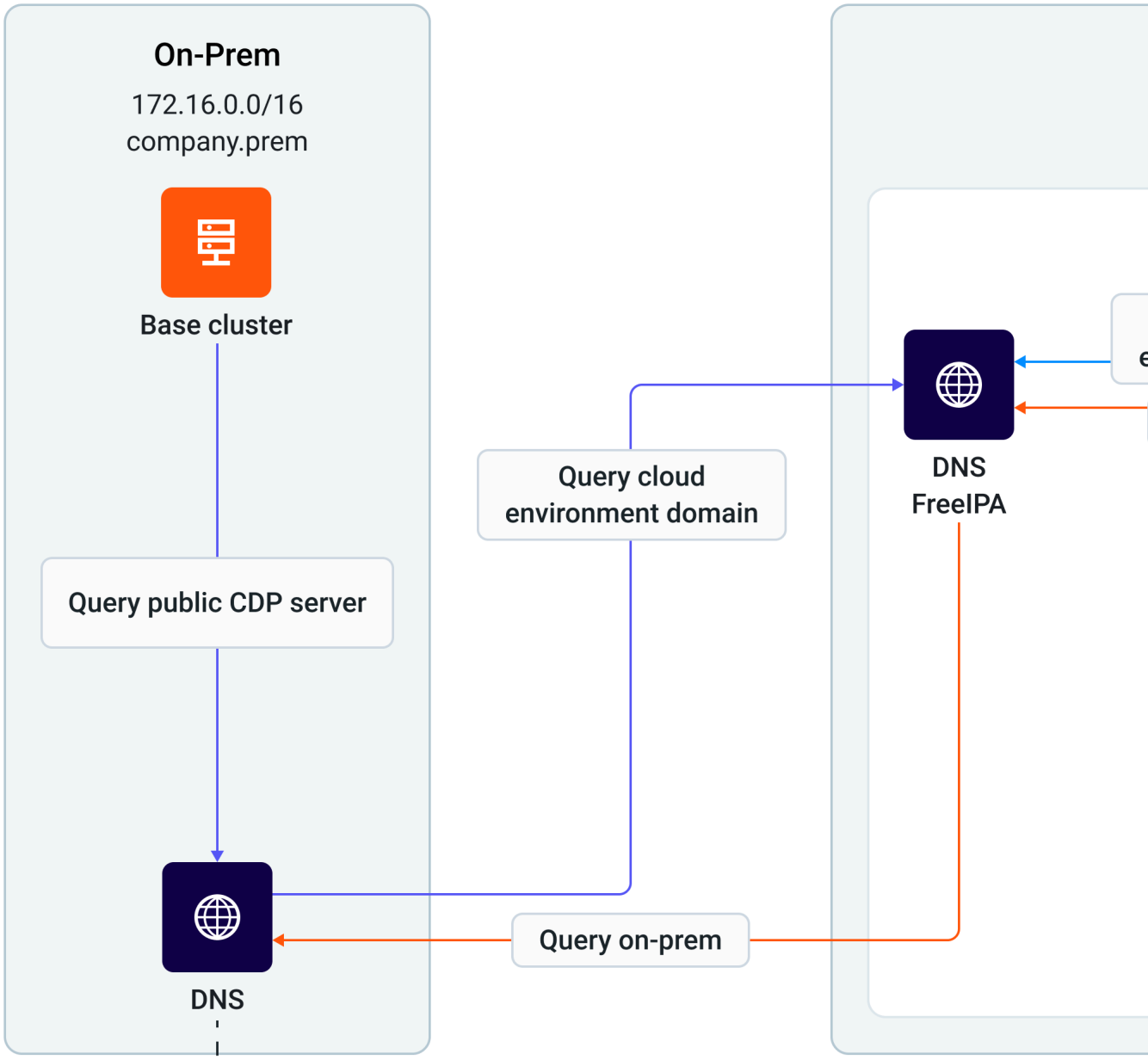
| Port | Protocol | Service Name | Direction | Purpose |
|---|---|---|---|---|
| 3268 | TCP | LDAP Global Catalog (GC) | IPA -> AD | Forest-wide User/Group lookups |
| 3269 | TCP | LDAPS Global Catalog (GC) | IPA -> AD | Secure Forest-wide lookups (if configured) |
| 49152-65535 | TCP | RPC Dynamic Ports | Bidirectional | High ports for RPC communication initiated via EPMAP |

## Hybrid Domain Name Resolution Architecture

Learn more about the Hybrid Domain Name Resolution architecture.

The following diagram outlines the DNS resolution strategy for a hybrid environment, connecting the **On-Premises** network (company.prem) and the **Cloud** network (company.cloud).

The design ensures that resources in either environment can discover and communicate with each other using their local domain names, while also retaining the ability to resolve external internet addresses.




| Domain | Type |
|---|---|
| company.prem | local |
| 16.172.in-addr.arpa | local |
| company.cloud | forward |

The table in the diagram details the core forwarding rules configured on the on-premises DNS server that enable this hybrid connectivity.

- company.prem (local): The server is authoritative for this domain.
- 16.172.in-addr.arpa (local): The server is authoritative for the on-premises reverse DNS lookup zone.
- company.cloud (forward): All queries for this domain are forwarded to the FreeIPA server's load balancer.

### Architecture Overview

Learn more about the Hybrid Domain Name Resolution architecture.

The architecture includes two primary environments, each with its own DNS server that collaborates to provide seamless name resolution.

The customer's on-premises corporate DNS server will contain records that delegate forward DNS resolution to the public cloud FreeIPA DNS service for all public cloud node zones.

- **On-Premises Environment (company.prem)**
  - **Network:** 172.16.0.0/16
  - **Components:**
    - **Base Cluster:** A collection of services and workloads that need to communicate with both on-premises and cloud resources.
    - **Local DNS Server:** The primary DNS for the on-premises environment (e.g., Active Directory). It is the authority for the company.prem domain and the reverse lookup zone 16.172.in-addr.arpa.
- **Cloud Environment (company.cloud)**
  - **Network:** 10.2.0.0/16
  - **Components:**
    - **FreeIPA:** The central DNS server for the public cloud environment. It is the authority for the company.cloud domain and acts as the primary resolver for all cloud-based services.
    - **Data Hubs:** A collection of services and workloads running in the public cloud that need to communicate with FreeIPA and with the on-premises environment.
    - **Network Default Nameserver:** The public or provider-supplied DNS resolver used for all external internet queries.

### DNS Resolution Flows

Learn more about the three primary query scenarios.

#### On-Prem-to-Cloud (Blue Arrow)

This flow allows on-premise applications to resolve cloud resources.

1. The **Base Cluster** needs to resolve an address in the cloud (e.g., datahub.company.cloud).
2. It sends the query to its local **On-premise DNS** server.
3. The **On-premise DNS** has a conditional forwarding rule (as shown in the configuration table) for the company.cloud domain.
4. The query is forwarded to the **FreeIPA** server in the cloud.
5. **FreeIPA** resolves the name and returns the IP address to the **On-premise DNS server**, which then sends it to the **Base Cluster**.

#### Cloud-to-On-Prem (Orange Arrows)

This flow allows cloud applications to resolve on-premises resources.

1. A cloud application (like **Data Hub**) needs to resolve an address on-premises (e.g., basecluster.company.prem).
2. It sends the query to its local resolver, **FreeIPA**.
3. **FreeIPA** has a forwarding rule configured to send all queries for company.prem to the **On-premise DNS** server.

4. The **On-premise DNS** resolves the name and returns the IP address to **FreeIPA**, which then sends it to the **Data Hub**.

#### Cloud-to-External (Purple Arrow)

This flow allows cloud applications to resolve public internet addresses.

1. A cloud application (or **FreeIPA** itself) needs to resolve an external address (e.g., google.com).
2. It sends the quesry to **FreeIPA**.
3. **FreeIPA** determines the domain is not local (company.cloud) and not part of any specific forwarding rule (company.prem).
4. It forwards the query to the **Network Default Nameserver**.
5. The **Network Default Nameserver** resolves the public address and returns the result.

### Network bandwidth considerations for performance

Network bandwidth and latency, file format, and compression settings impact performance in hybrid cloud environments, where compute resources run in the cloud and data remains on-premises.

- **Remote data access** is a practical model for bursty workloads; however, performance is heavily impacted by available network bandwidth.
- **Columnar file formats** (e.g., Parquet, ORC) drastically reduce execution time and data transfer compared to CSV, making them a prerequisite for hybrid setups.
- **Bandwidth constraints** (e.g., 5 Gbit/s) increase execution time and reduce CPU efficiency for I/O-intensive queries.
- **Gzip Compression** significantly reduces data transfer volume, improving performance under limited bandwidth. **Snappy** offers minor gains with lower CPU overhead.
- **Not all queries are impacted equally**. CPU-bound queries run efficiently even under constrained bandwidth, while I/O-bound queries degrade sharply.

The strategic use of columnar formats and compression enables many workloads to run efficiently in hybrid environments, even with limited network capacity. For CPU-intensive Spark jobs, this setup is a viable architecture for burst-to-cloud use cases. In contrast, I/O-intensive jobs remain highly sensitive to network limits, making this approach less suitable for data-heavy pipelines without further optimization.

## Identity Provider Requirements

Learn more about the supported identity providers and their configuration requirements.

### Supported Identity Providers

Learn more about the supported identity providers.

Cloudera Management Console supports **Microsoft Active Directory** as the on-premises Identity Provider. You must log in as a user with **Domain Administrator**, **Enterprise Administrator**, or equivalent privileges to make all required changes on the Active Directory instance.

### Active Directory Encryption Settings

For enhanced security, Cloudera configures the hybrid environment’s FreeIPA to only permit strong encryption for Kerberos. Therefore, you have to make sure that Active Directory is configured accordingly.

#### About this task

Starting from Active Directory 2022, AES-128 and AES-256 Kerberos encryption types are enabled by default for all newly created users, therefore additional configuration is not needed there.

### Before you begin

For Active Directory versions prior to 2022, you have to configure your on-premises Cloudera Manager to set the encryption type in Active Directory when creating users.

### Procedure

1. Select  Administration Settings .
2. Set krb_enc_types to either **aes128-cts** or **aes256-cts**.
3. Enable ad_set_encryption_types.

   These settings only apply to newly created users, so you have to regenerate all keytabs.
4. Select  Administration Security Kerberos Credentials .
5. Select all credentials and click Regenerate Selected.

   **Note:**

   All services have to be stopped to perform these steps.
6. The Kerberos configuration also has to be edited on the on-premise cluster's nodes.

   Following is a sample configuration for aes128 compatibility:

```
default_tgs_enctypes = aes128-cts-hmac-sha256-128 aes128-cts-hmac-sha1-96
default_tkt_enctypes = aes128-cts-hmac-sha256-128 aes128-cts-hmac-sha1-96
permitted_enctypes = aes256-cts-hmac-sha384-192 aes128-cts-hmac-sha256-128
 aes256-cts-hmac-sha1-96 aes128-cts-hmac-sha1-96 camellia256-cts-cmac came
llia128-cts-cmac
```

7. Locate the latest keytab used by the Cloudera Manager process and check its encryption:

```
KEYTAB=$(find /run/cloudera-scm-agent/process/ -name "*.keytab" -printf
"%T@ %p\n" | sort -n | tail -n1 | awk '{print $2}')
klist -kte $KEYTAB

Keytab name: FILE:/run/cloudera-scm-agent/process/241-impala-IMPALAD/imp
ala.keytab
KVNO Timestamp           Principal
---- ------------------- ------------------------------------------------
-----
   1 11/13/2025 13:36:34 HTTP/ccycloud-3.ad-dbajzath.root.comops.site@QE-I
NFRA-AD.CLOUDERA.COM (aes128-cts-hmac-sha1-96)
   1 11/13/2025 13:36:34 impala/ccycloud-3.ad-dbajzath.root.comops.site@Q
E-INFRA-AD.CLOUDERA.COM (aes128-cts-hmac-sha1-96)
```

## User Identity Requirements

Since **Kerberos authentication** is used to access Cloudera Hybrid Data Hubs and on-premises data lake services, usernames must align across both identity providers to ensure proper authentication and authorization.

For seamless access between Cloudera Hybrid Data Hubs and on-premises data lake services, user identities must be consistent across both environments. This requirement exists because a **trust relationship** is established between the **FreeIPA** cluster deployed in the cloud (as part of the hybrid environment) and the Identity Provider (e.g., **Active Directory**) used by the on-premise data lake.

### Example

A user **user1@env.cloudera.site** from the hybrid environment attempts to access **HDFS** running on the on-premises data lake. FreeIPA manages the public cloud hybrid environment domain (env.cloudera.site).

When this occurs:

1. **FreeIPA** issues a **Kerberos token** for **user1@env.cloudera.site**.

2. The on-premises HDFS service attempts to validate this token with the **Active Directory KDC**.

   If **user1** exists in the Active Directory but under the on-premises domain (e.g., user1@acme.ad.com), the token can still be accepted. This is possible because, during the Hybrid trust setup, the Kerberos auth_to_local user mappings are configured to map both identities to a common local user (**user1**).
3. Access is granted when the mapping resolves successfully, allowing the public cloud user to securely interact with on-premises HDFS resources.

This configuration enables cross-realm authentication between cloud-based and on-premises environments while maintaining Kerberos security guarantees.

## Time synchronization

Learn more about how both FreeIPA and Active Directory environments are synchronized to an authoritative time source.

The Kerberos protocol's security model is critically dependent on synchronized time across all participating systems. Timestamps are embedded within Kerberos tickets to prevent replay attacks, where an attacker might capture a ticket and attempt to reuse it later. If the time difference (clock skew) between a client, a server, and the KDC exceeds a configured tolerance, authentication requests will be rejected with errors similar to Clock skew too great.

Because of this sensitivity, both FreeIPA and Active Directory environments must be synchronized to an authoritative time source.

FreeIPA uses the `chrony` service to synchronize its time with an NTP server. You can verify time synchronization settings with the following commands:

- `less /etc/chrony.conf`
- `chronyc sources`

```
#####################################
##### FreeIPA commands      #####
#####################################
# check for "server" in chrony configuration
less /etc/chrony.conf
server 169.254.169.123 prefer iburst trust

chronyc sources
MS Name/IP address         Stratum Poll Reach LastRx Last sample
===============================================================================
======
^* 169.254.169.123               3   8   377   223  -7697ns[  -11us] +/-
 324us

# The IP address (169.254.169.123) is the address for the Amazon Time Sync
Service that provides highly accurate time synchronization for instances run
ning using the Network Time Protocol (NTP)
```

In an Active Directory environment, the server holding the **PDC Emulator (PDCe) FSMO** role acts as the primary time source for the entire domain.

Log in to any Active Directory Domain Controller and run the `netdom query fsmo` command to find the PDC.

```
rem # Find your PDC Emulator, the server name listed for "PDC"
netdom query fsmo
Schema master               EC2AMAZ-UVREDMU.hybrid.cloudera.org
Domain naming master        EC2AMAZ-UVREDMU.hybrid.cloudera.org
PDC                         EC2AMAZ-UVREDMU.hybrid.cloudera.org
RID pool manager            EC2AMAZ-UVREDMU.hybrid.cloudera.org
Infrastructure master       EC2AMAZ-UVREDMU.hybrid.cloudera.org
The command completed successfully.
```

Log in to the PDC Emulator instance and run the following commands:

```
#######################################
##### AD commands             #####
#######################################


rem # Check the presence of the NTP server
w32tm /query /status
Leap Indicator: 0(no warning)
Stratum: 4 (secondary reference - syncd by (S)NTP)
Precision: -23 (119.209ns per tick)
Root Delay: 0.0006070s
Root Dispersion: 0.0249495s
ReferenceId: 0xA9FEA97B (source IP:  169.254.169.123)
Last Successful Sync Time: 11/5/2025 9:07:38 AM
Source: 169.254.169.123,0x9
Poll Interval: 6 (64s)


w32tm /query /peers
#Peers: 3

Peer: time.windows.com,0x8
State: Active
Time Remaining: 22.4656548s
Mode: 3 (Client)
Stratum: 3 (secondary reference - syncd by (S)NTP)
PeerPoll Interval: 6 (64s)
HostPoll Interval: 6 (64s)

Peer: 169.254.169.123,0x9
State: Active
Time Remaining: 636.8803514s
Mode: 3 (Client)
Stratum: 3 (secondary reference - syncd by (S)NTP)
PeerPoll Interval: 6 (64s)
HostPoll Interval: 6 (64s)
```

Both instances synced with an NTP server should be sufficient, but the safest choice is to configure both your FreeIPA server and your Active Directory PDC Emulator to use the same authoritative time source.

Active Directory reconfiguration is less viable, so the preferred method is to configure the new server and add a new FreeIPA recipe to persist the change on each subsequently upscaled FreeIPA node.

```
#!/bin/bash
# ---
# This script updates /etc/chrony.conf to use a new, single time server.
# It comments out all existing 'server' or 'pool' lines and adds the
# new server (provided as an argument) with 'prefer iburst'.
#
# Usage:
#   sudo ./update_chrony.sh <new-server-fqdn-or-ip>
#
# Example:
#   sudo ./update_chrony.sh ad-pdc.your-domain.com
# ---

# Exit immediately if a command fails
set -e

# --- Configuration ---
NEW_SERVER="$1"
```

```
CONFIG_FILE="/etc/chrony.conf"
SERVICE_NAME="chronyd"
# --- 1. Validation ---

# Check if running as root
if [ "$(id -u)" -ne 0 ]; then
   echo "This script must be run as root or with sudo."
   exit 1
fi
# Check if a server argument was provided
if [ -z "$NEW_SERVER" ]; then
    echo "Error: No server provided."
    echo "Usage: sudo $0 <new-server-fqdn-or-ip>"
    exit 1
fi

echo "--- Starting Chrony Update ---"

# --- 2. Backup ---
BACKUP_FILE="${CONFIG_FILE}.bak-$(date +%Y%m%d-%H%M%S)"
echo "Creating backup at $BACKUP_FILE..."
cp "$CONFIG_FILE" "$BACKUP_FILE"

# --- 3. Update Config ---
echo "Commenting out existing 'server' and 'pool' directives in $CONFIG_FI
LE..."
# This sed command finds lines starting with 'server ' or 'pool ' and puts
 a # in front.
sed -i -E 's/^(server|pool)\s/#&/' "$CONFIG_FILE"

echo "Adding new server: $NEW_SERVER..."
# This sed command inserts the new server line right before the 'driftfile'
line.
sed -i "/^driftfile/i server $NEW_SERVER prefer iburst\n" "$CONFIG_FILE"

echo "Config file updated."
# --- 4. Restart Service ---
echo "Restarting $SERVICE_NAME service to apply changes..."
systemctl restart "$SERVICE_NAME"
echo "Service restarted."

# --- 5. Verify ---
echo
echo "Verifying $SERVICE_NAME status (should be 'active'):"
# Show just the 'Active:' line from the status
systemctl status "$SERVICE_NAME" --no-pager | grep 'Active:'
echo
echo "Checking new sources... (it may take a moment to connect and sync)"
# Give chrony a second to start, then check sources
sleep 1
chronyc sources
echo
echo "--- Chrony Update Complete ---"
```

# Registering Cloudera Hybrid Environments

Learn how to register a hybrid environment.

### Before you begin

Ensure that you have fulfilled the following Cloud Provider requirements based on the Cloud Service Provider of your choice: AWS, Azure, or GCP.

**For AWS**

- Ensure that your AWS account has the right permissions as described in AWS account permissions.
- Create the Cross-account Role ARN corresponding to the Cross Account Role as described in Creating cross-account access IAM role.
- Create your credential to provision the environment as described in Creating a provisioning credential for AWS.
- Ensure that the VPC and subnet requirements are met as described in VPC and subnets.
- Create or use your existing security groups as described in Security groups.
  - If you plan to create new Security Groups, make sure you have the required IP address range information.
  - If you plan to use existing Security Groups, you need to open all the required ports.
- If you plan to use Customer Managed Encryption Keys (CMEK), configured them as described in Customer managed encryption keys.
- Create or use your existing SSH public key as described in SSH key pair.
  - If you plan to create a new SSH key, make sure you are using an RSA or ED25519 public key. This will create a new EC2 key pair on the AWS side, and all cloud resources will use it for SSH authentication.
  - If you plan to use an existing SSH key, you need to refer to an existing AWS EC2 key pair. The Cloudera Control Plane will validate your key existence
- Create an S3 bucket and set up the Logs location as described in AWS cloud storage prerequisites..

**For Azure**

- Ensure that your Azure account has the right permissions as described in Azure subscription requirements.
- Create a custom role with the required set of permissions as described in Azure credential prerequisites.
- Create your credential to provision the environment as described in Creating a provisioning credential for Azure.
- Create or let Cloudera create resource groups for the environment as descirbed in Resource groups.
- Ensure that the VNET and subnet requirements are met as described in VNet and subnets.
- Configure Azure Flexible Server as described in Private setup for Azure Flexible Server.
- If you plan to use Customer Managed Encryption Keys (CMEK), configured them as described in Encrypting Azure resources with customer managed keys.
- Create or use your existing security groups as described in Network security groups.
  - If you are planning to create new Security Groups, make sure you have the required IP address range information.
  - If you are planning to use existing Security Groups, you need to open all the required ports.
- Create or use your existing SSH public key as described in SSH key pair.
- Create an ADLS Gen2 storage and set up the Logs location as described in Azure cloud storage prerequisites.

**For GCP**

- Ensure that your Google account has the right permissions as described in GCP permissions.
- Create a Google project as described in GCP project.
- Create a service account with the required set of permissions as described in Service account for credential.
- Create your credential to provision the environment as described in Creating a GCP credential.
- Ensure that the VPC and subnet requirements are met as described in VPC network and subnet.
- If you plan to use Customer Managed Encryption Keys (CMEK), configured them as described in Customer managed encryption keys.
- Create or use your existing SSH public key as described in SSH key pair.

- Create a Google storage bucket and set up the Logs location as described in GCP cloud storage prerequisites.

- **EnvironmentCreator**

## Procedure

1. Navigate to Cloudera Management Console.
2. Select Environments.
3. Click Register environment.
4. In the Purpose section, select Hybrid Cloud Environment.
5. Enter the following information for the new hybrid environment:

| Environments Page | |
|---|---|
| General Information | |
| Environment Name | Enter a name for the new hybrid environment. |
| Description (optional) | Enter a short description for the new hybrid environment. |
| Select Cloud Provider | Select the cloud provider of your choice. |

6. If you already have a credential set up, select it from the dropdown list.
7. If you need to create new credentials, enter or select the following information:

### For AWS

| Register Environment page | |
|---|---|
| Amazon Web Services Credential section | |
| Name | Enter a name for the new credential. |
| Description (optional) | Enter a short description for the new credential. |
| Enable Permission Verification | Use this toggle to have Cloudera check permissions for your credential. Cloudera will verify that you have the required permissions for your environment. |
| Default \| Minimal | Select whether to use Default or Minimal role.<br>Use the provided JSON to create the AWS IAM policy.<br>Use Minimal role for a general Hybrid environment. Use Default if you plan to use Data Services. |
| Service Manager Account ID<br>External ID | Use the provided IDs to create the AWS IAM role. |
| Cross-account Role ARN | Enter the cross-account ARN role. |
| Show CLI Command (optional) | Click this button to display the command required to create the credential from the CLI. |
| Create Credential | Click this button to create the new credential. |

Click Next to proceed to the Region, Networking and Security page.

| Region, Networking and Security page | |
|---|---|
| Region, Location section | |
| Select Region | Select the region for the new environment. |
| Network section | |
| Select Network | Select the existing virtual network where you would like to provision all Cloudera resources. For more information, refer to VPC and subnet. |

| Region, Networking and Security page | |
|---|---|
| Select Subnets | Select existing subnets within the selected VPC. For more information, refer to VPC and subnet. |
| Enable Cluster Connectivity Manager (CCM) | The Cluster Connectivity Manager allows Cloudera to communicate with Cloudera Data Hub clusters and on-prem classic clusters that are on private subnets. For more information, refer to the Cluster Connectivity Manager documentation. |
| Enable Endpoint Access Gateway | When the Cluster Connectivity Manager is enabled, you can optionally enable Endpoint Access Gateway to provide secure connectivity to UIs and APIs in Cloudera Data Hub clusters deployed using private networking.<br>Under Select Subnets for the Endpoint Access Gateway, select the public subnets for which you would like to use the gateway. The number of subnets must be the same as selected under Select Subnets and the availability zones must match.<br>For more information, refer to the Public Endpoint Access Gateway section. |
| Encryption section | |
| Enable Customer Managed Keys | Enable this if you would like to provide a Customer-Managed Key (CMK) to encrypt the environment's disks and databases. For more information, refer to Customer managed encryption keys. |
| Proxies section | |
| Select Proxy Configuration | Select one of the following options:<br>• Do Not Use Proxy Configuration<br>• Create New Proxy Configuration<br>If you would like Cloudera to automatically create security groups for you and open them to the CIDR range specified.<br>Enter the following information for the new proxy configuration:<br>• Name<br>• Description (optional)<br>• Protocol<br>• Server Host<br>• Server Port<br>• No Proxy Hosts<br>• Inbound Proxy CIDR<br>• Username<br>• Password<br>• Existing proxy configuration<br>You need to open all the required ports if you would like to use your existing security groups.<br>For more information, refer to Setting up a proxy server. |

| Region, Networking and Security page | |
|---|---|
| Security Access Settings | Select one of the following options to determine inbound security group settings that allow connections to the Cloudera Data Hub clusters from your organization's computers:<br>• Create New Security Groups<br>If you would like Cloudera to automatically create security groups for you and open them to the CIDR range specified.<br>• Access CIDR<br>Enter a custom CIDR IP range for all new security groups that will be created for the Cloudera Data Hub clusters.<br>• Select Existing Security Groups<br>If you would like to use your existing security groups. In this case, you need to open all the required ports. Refer to Security groups to ensure that you open all ports required for your users to access environment resources.<br>• Select Existing Security Group for Gateway Nodes.<br>• Select Existing Security Group as default. |
| SSH Settings section | |
| New SSH public key | Enter a new SSH public key. |
| Existing SSH public key | Enter the name of an existing EC2 key pair name with your desired SSH key pair. |
| Add tags section | |
| Add (optional) | You can optionally add tags to be created for your resources on AWS. For more information, refer to Defining custom tags. |
| Advanced options section | |
| Network And Availability | Enable Multiple Availability Zones for FreeIPA. For more information, refer to Deploying Cloudera in multiple AWS availability zones. |
| Hardware And Storage | Enter FreeIPA nodes instance types. Click the edit icon in the top right corner and select the instance type from the drop-down list. For more information on instance types, refer to Amazon EC2 instance types. |
| Cluster Extensions | You can optionally select and attach previously registered recipes to run on a specific FreeIPA host group. For more information, see Recipes. |
| Security | Select the SELinux mode based on your requirements:<br>• Permissive<br>• Enforcing |

Click Next to proceed to the Storage page.

| Storage page | |
|---|---|
| Logs section | |
| Logger Instance Profile | Select the IAM instance profile (or IAM role) that provides Cloudera with write access to the S3 logs data location. |
| Logs Location Base | Provide a path to an existing S3 bucket or a directory within an existing S3 bucket where log data will be stored. |
| Backup Location Base (optional) | Provide a path to an existing S3 bucket or a directory within an existing S3 bucket where IPA backups will be stored.<br>If none is provided, the log location will be used. |
| Telemetry section | |

| Storage page | |
|---|---|
| Enable Cloudera Observability (optional) | When this is enabled, diagnostic information about job and query execution is sent to Cloudera Observability for Cloudera Data Hub clusters. For more information, refer to Enabling workload analytics and logs collection. |

## For Azure

| Register Environment page | |
|---|---|
| Microsoft Azure Credential section | |
| Name | Enter a name for the new credential. |
| Description (optional) | Enter a short description for the new credential. |
| Default \| Minimal | Select whether to use Default or Minimal role.<br>Use the provided JSON to create the AWS IAM policy.<br>Use Minimal role for a general Hybrid environment. Use Default if you plan to use Data Services. |
| Command 1 | Use the provided command in the Azure Shell to associate the new certificate with the service principal. |
| Command 2 | Use the provided command in the Azure Shell to identify your Subscription ID and Tenant ID. |
| Show CLI Command (optional) | Click this button to display the command required to create the credential from the CLI. |
| Create Credential | Click this button to create the credential. |

Click Next to proceed to the Region, Networking and Security page.

| Region, Networking and Security page | |
|---|---|
| Region, Location section | |
| Select Region | Select the region for the new environment. |
| Resource Group section | |
| Select Resource Group | Select one of the following:<br>• Select an existing resource group to have all Cloudera resources provisioned into that resource group.<br>• Select Create new resource groups to have Cloudera create multiple resource groups. |
| Network section | |
| Select Network | Select the existing virtual network where you would like to provision all Cloudera resources. Refer to VPC and subnet. |
| Select Subnets | This option is only available if you choose to use an existing network. Multiple subnets must be selected and Cloudera distributes resources evenly within the subnets. |
| Enable Cluster Connectivity Manager (CCM) | The Cluster Connectivity Manager allows Cloudera to communicate with Cloudera Data Hub clusters and on-prem classic clusters that are on private subnets. For more information, refer to the Cluster Connectivity Manager documentation. |

| Region, Networking and Security page | |
|---|---|
| Enable Endpoint Access Gateway | When Cluster Connectivity Manager is enabled, you can optionally enable Public Endpoint Access Gateway to provide secure connectivity to UIs and APIs in Cloudera Data Hub clusters deployed using private networking.<br><br>If you are using your existing VNET, under **Select Endpoint Access Gateway Subnets**, select the public subnets for which you would like to use the gateway. The number of subnets must match that set under **Select Subnets**, and the availability zones must match. For more information, refer to Public Endpoint Access Gateway. |
| Create Public IPs | This option is disabled by default when Cluster Connectivity Manager is enabled. It is enabled by default when Cluster Connectivity Manager is disabled. |
| Database section | |
| Database | Select one of the following:<br><br>• Flexible Server<br>• Flexible Servier with Private Link<br><br>You must select the Private DNS Zone for the database from the drop-down menu.<br>• Flexible Server with Delegated Subnet (deprecated)<br><br>For more information on Flexible Servers, refer to Using Azure Database for PostgreSQL Flexible Server. |
| Encryption section | |
| Enable Encryption at Host | ? |
| Enable Customer Managed Keys | Enable this option if you would like to provide a Customer-Managed Key (CMK) to encrypt the environment's disks and databases. For more information, refer to Customer managed encryption keys. |
| Proxies section | |
| Select Proxy Configuration | Select one of the following:<br><br>• Do Not Use Proxy Configuration<br>• Create New Proxy Configuration<br><br>If you would like Cloudera to automatically create security groups for you and open them to the CIDR range specified.<br><br>Enter the following information for the new proxy configuration:<br><br>• Name<br>• Description (optional)<br>• Protocol<br>• Server Host<br>• Server Port<br>• No Proxy Hosts<br>• Inbound Proxy CIDR<br>• Username<br>• Password<br>• Select existing proxy configuration<br><br>You need to open all the required ports if you would like to use your existing security groups.<br><br>For more information, refer to Setting up a proxy server. |

| Region, Networking and Security page | |
|---|---|
| Security Access Settings | Select one of the following options to determine inbound security group settings that allow connections to the Cloudera Data Hub clusters from your organization's computers:<br>• Create New Security Groups<br>If you would like Cloudera to automatically create security groups for you and open them to the CIDR range specified.<br>• Access CIDR<br>Enter a custom CIDR IP range for all new security groups that will be created for the Cloudera Data Hub clusters.<br>• Select Existing Security Groups<br>If you would like to use your existing security groups. In this case, you need to open all the required ports. Refer to Security groups to ensure that you open all ports required for your users to access environment resources.<br>• Select Existing Security Group for Gateway Nodes.<br>• Select Existing Security Group as default. |
| SSH Settings section | |
| New SSH public key | Enter a new SSH public key. |
| Existing SSH public key | Enter the name of an existing SSH key pair. |
| Add tags section | |
| Add (optional) | You can optionally add tags to be created for your resources on Azure. For more information, refer to Defining custom tags. |
| Advanced options section | |
| Network And Availability | Enable Multiple Availability Zones for FreeIPA. For more information, refer to Deploying Cloudera in multiple Azure availability zones. |
| Hardware And Storage | You can specify an instance type for each host group. For more information on instance types, refer to Sizes for virtual machines in Azure. |
| Cluster Extensions | You can optionally select and attach previously registered recipes to run on FreeIPA nodes. For more information, see Recipes. |
| Security | Select the SELinux mode based on your requirements:<br>• Permissive<br>• Enforcing |

Click Next to proceed to the Storage page.

| Storage page | |
|---|---|
| Logs section | |
| Logger Instance Profile | The logger requires Storage Blob Data Contributor role on the provided storage account. |
| Logs Location Base | Provide your filesystem and storage account name in a filesystem@storageaccountname.dfs.core.windows.net[/subfolders] format where data will be stored.<br>• Filesystem must already exist.<br>• The storage account must be Storage V2.<br>• Subfolders are optional. |

| Storage page | |
|---|---|
| Backup Location Base (optional) | Provide your filesystem and storage account name in a filesystem@storageaccountname.dfs.core.windows.net[/subfolders] format where IPA backups will be stored.<br>• Filesystem must already exist.<br>• The storage account must be Storage V2.<br>• Subfolders are optional. |
| Telemetry section | |
| Enable Cloudera Observability (optional) | When this is enabled, diagnostic information about job and query execution is sent to Cloudera Observability for Data Hub clusters. For more information, refer to Enabling workload analytics and logs collection. |

## For GCP

| Register Environment page | |
|---|---|
| Google Cloud Platform Credential section | |
| Name | Enter a name for the new credential. |
| Description (optional) | Enter a short description for the new credential. |
| Default \| Minimal | Select whether to use Default or Minimal role.<br>Use the provided commands to create a service account through the Google Cloud SDK or Google Cloud Shell.<br>Use Minimal role for a general Hybrid environment. Use Default if you plan to use Data Services. |
| Upload file | Use the Upload file button to upload a service account private key in JSON format. |
| Show CLI Command (optional) | Click this button to display the command required to create the credential from the CLI. |
| Create Credential | Click this button to create the credential. |

Click Next to proceed to the Region, Networking and Security page.

| Region, Networking and Security page | |
|---|---|
| Region, Location section | |
| Select Region | Select the region for the new environment. |
| Select Zone | Select the zone within the selected region. |
| Network section | |
| Use Shared VPC | Shared VPC allows an organization to connect resources from multiple projects to a common Virtual Private Cloud (VPC) network, so that they can communicate with each other securely and efficiently using internal IPs from that network. When you use Shared VPC, you designate a project as a host project and attach one or more service projects to it. The VPC networks in the host project are called Shared VPC networks. Eligible resources from service projects can use subnets in the Shared VPC network. For more information, see https://cloud.google.com/vpc/docs/shared-vpc |
| Select Network | Select the existing VPC where you would like to provision all Cloudera resources. Refer to VPC and subnet. |
| Select Subnets | Select at least one subnet within the selected VPC. Refer to VPC and subnet. |

| Region, Networking and Security page | |
|---|---|
| Create Private Subnets | This option is only available if you select to have a new network and subnets created. It is turned on by default so that private subnets are created in addition to public subnets. If you disable it, only public subnets will be created.<br><br>For production deployments, Cloudera recommends that you use private subnets. Work with your internal IT teams to ensure that users can access the browser interfaces for cluster services. |
| Enable Cluster Connectivity Manager (CCM) | The Cluster Connectivity Manager allows Cloudera to communicate with Cloudera Data Hub clusters and on-prem classic clusters that are on private subnets. For more information, refer to the Cluster Connectivity Manager documentation. |
| Enable Endpoint Access Gateway | When Cluster Connectivity Manager is enabled, you can optionally enable Public Endpoint Access Gateway to provide secure connectivity to UIs and APIs in Cloudera Data Hub clusters deployed using private networking.<br><br>If you are using your existing VPC, under **Select Endpoint Access Gateway Subnets**, select the public subnets for which you would like to use the gateway. The number of subnets must match that set under **Select Subnets**, and the availability zones must match. For more information, refer to Public Endpoint Access Gateway. |
| Create Public IPs | This option is disabled by default when Cluster Connectivity Manager is enabled. It is enabled by default when Cluster Connectivity Manager is disabled. |
| Encryption section | |
| Enable Customer Managed Keys | Enable this if you would like to provide a Customer-Managed Key (CMK) to encrypt the environment's disks and databases. For more information, refer to Customer managed encryption keys. |
| Proxies section | |
| Select Proxy Configuration | Select one of the following:<br>• Do Not Use Proxy Configuration<br>• Create New Proxy Configuration<br>If you would like Cloudera to automatically create security groups for you and open them to the CIDR range specified.<br>Enter the following information for the new proxy configuration:<br>• Name<br>• Description (optional)<br>• Protocol<br>• Server Host<br>• Server Port<br>• No Proxy Hosts<br>• Inbound Proxy CIDR<br>• Username<br>• Password<br>• Existing proxy configuration<br>If you would like to use your existing security groups. In this case, you need to open all required ports.<br><br>For more information, refer to Setting up a proxy server. |

| Region, Networking and Security page | |
|---|---|
| Security Access Settings | You have two options:<br>• Do not create firewall rule<br>Select this option if you are using a shared VPC and have already set the firewall rules directly on the VPC.<br>• Provide existing firewall rules<br>If not all of your firewall rules are set directly on the VPC, provide the previously created firewall rules for SSH and UI access. You should select two existing firewall rules, one for Knox gateway-installed nodes and another for all other nodes. You may select the same firewall rule in both places if needed.<br>For information on required ports, refer to Firewall rules. |
| SSH Settings section | |
| New SSH public key | Enter a new SSH public key. |
| Existing SSH public key | Enter the name of an existing SSH key pair. |
| Add tags section | |
| Add (optional) | You can optionally add tags to be created for your resources on GCP. For more information, refer to Defining custom tags. |
| Advanced options section | |
| Network And Availability | Enable Multiple Availability Zones for FreeIPA. For more information, refer to Deploying Cloudera In Multiple GCP Availability Zones. |
| Hardware And Storage | You can specify an instance type for each host group. For more information on instance types, refer to Sizes for virtual machines in Azure. |
| Cluster Extensions | You can optionally select and attach previously registered recipes to run on FreeIPA nodes. |
| Security | Select the SELinux mode based on your requirements:<br>• Permissive<br>• Enforcing |

Click Next to proceed to the Storage page.

| Storage page | |
|---|---|
| Logs section | |
| Logger Service Profile | Select the Service Account that provides Cloudera with write access to the Google Cloud Storage location where logs will be stored. |
| Logs Location Base | Provide a path to an existing GCS bucket or a directory within an existing GCS bucket where data will be stored. For more information, refer to Minimum setup for cloud storage. |
| Backup Location Base (optional) | Provide a path to an existing GCS bucket or a directory within an existing GCS bucket where FreeIPA backups will be stored. For more information, refer to Minimum setup for cloud storage. |
| Telemetry section | |
| Enable Cloudera Observability (optional) | When this is enabled, diagnostic information about job and query execution is sent to Cloudera Observability for Cloudera Data Hub clusters. For more information, refer to Enabling workload analytics and logs collection. |

8. Click Register Environment to finish the hybrid environment registration process.

### Results
You have created the Hybrid Environment.

### What to do next

After your environment is running, perform the following steps:

- You must assign roles to users and groups to grant them access to the environment, and perform user sync. For steps, refer to Enabling admin and user access to environments.
- You must onboard your users and groups for cloud storage. For steps, refer to Onboarding Cloudera users and groups for cloud storage.

# Connecting Cloudera on premises with Cloudera on cloud

You need to connect your on-premises and cloud environments using your Cloudera Base on premises cluster as a Classic Cluster. This means that the Cloudera Base on premises cluster is used as the Data Lake in your cloud environment.

## Adding a Cloudera Base on premises cluster as a Classic Cluster

Register a Cloudera Base on premises cluster as a classic cluster using Cloudera Manager so that you can use the Base cluster in your hybrid environment as the Data Lake.

### Before you begin

- All the clusters must meet the requirements identified in Prerequisites for adding classic clusters.

### Procedure

1. Log in to Cloudera Management Console.
2. Click Classic Clusters.
3. On the **Classic Clusters** page, click ADD CLUSTER.
4. In the **Add Cluster** dialog box, navigate to the Cloudera Base on premises tab and enter the following details:
   a) If your cluster is not reachable by a public network, click “My cluster is accessible only in my private network”.
   b) Cloudera Manager IP address - Enter the IP address of the Cloudera Manager of the Cloudera Base on premises cluster. The Cloudera Management Console uses this IP address to identify the cluster for registration purposes.
   c) Cloudera Manager Port - Enter the port of the Cloudera Manager of the Cloudera Base on premises cluster.
   d) Data center - Enter a unique data center name for the Cloudera Base on premises cluster.
   e) Select the My cluster runs on HTTPS option if the Cloudera Base on premises cluster uses HTTPS.
   f) Clear the Register KNOX endpoint (Optional) option, if selected.
   g) Click CONNECT.

   The Cloudera Management Console acquires the configuration details from Cluster Connectivity Manager service. After Cloudera successfully connects to your new cluster (which should take no more than 5 minutes), it will highlight Step 2.
5. On the Classic Clusters page, click Files in the Step 2 pane.

6. Follow the instructions in the **Setup Connectivity Client** dialog box. You need to download the jumpgate-agent rpm file and the cluster_connectivity_setup_files zip file onto Cloudera Manager host in your new cluster:
   a) In the command line interface, copy the jumpgate-agent RPM and cluster_connectivity_setup_files.zip to the Cloudera Manager host.
   b) Unzip the cluster_connectivity_setup_files file. Inside this zip file there is a script install.sh.
   c) SSH to the Cloudera Manager host.
   d) Install the jumpgate-agent rpm using:

```
yum --nogpgcheck localinstall <
downloaded-jumpgate-agent-rpm >
```

   e) Run install.sh by using ./install.sh command.
   f) Check service status to see if the agent has been connected:

```
systemctl status jumpgate-agent.service
```

   **Note:** If you regenerate the script files, you cannot use the previously downloaded cluster_connecti vity_setup_files.zip file because the file is no longer valid.
7. On the **Classic Clusters** page, click Test Connection in the Step 2 pane to verify whether the connection is successful.
8. Click Register in the Step 3 pane.
9. In the **Cluster Details** dialog box, enter the Cloudera Manager credentials that have Admin access to Cloudera Manager and the cluster services.
10. Click CONNECT.
11. To complete the registration, enter the following details on the **Classic Clusters** page:
    a) Cluster Location - Enter the geographical location of the Data Lake.
    b) Data Center - Ensure that the data center name is the name that you provided for the Cloudera Base on premises cluster during registration.
    c) Tags - Optionally, enter the tags for the cluster.
    d) Description - Optionally, enter a description.
12. Click Add.

### What to do next

You can use the registered Classic Cluster in your hybrid environment.

# Setting up trust for hybrid environments

Learn more about how to set up trust for hybrid environments.

### About this task

Once the hybrid environment has been created, its status changes to Trust Setup Required. This requires setting up a cross-realm trust between FreeIPA in Cloudera on cloud and Active Directory in Cloudera on premises.

### Before you begin

- You have the following options to connect the cloud and on-premises environments:
  - Connect the cloud and on-premises Control Plane.
  - Register the Cloudera on premises cluster as a Classic cluster in Cloudera on cloud.
- The on-premises Control Plane provides additional functionalities over the Classic clusters, such as Data Services and more controlled user-level access to the on-premises data lake services, but it is not mandatory for the Hybrid burst to cloud use cases.

- Creating a connection between the cloud and on-premises platform, the Cloudera Base cluster acts as the data lake for the hybrid cloud environment.

  The Cloudera Data Hub cluster created in the Cloudera Hybrid Environments can directly access data and metadata in the Cloudera on premises Data Lake cluster.
- To have a seamless connection between Cloudera on cloud and Cloudera on premises, you also have to connect the Cloudera on cloud Kerberos KDC and DNS server to the same one used by the Cloudera on premises account.
- Register your Cloudera on premises control plane in the Cloudera Hybrid Environments.

## Procedure

1. Navigate to the Cloudera Management Console for Cloudera on cloud.
2. Select Environments.
3. Select the environment created in Registering Hybrid Environments.
4. Select the Data Lake tab.
5. Click Next.
6. Select the on-premises environment from the drop-down list.
7. The wizard will fetch the required information and complete the necessary items.

   **Realm**

   The Active Directory REALM is typically your domain name, but entered in all capital letters (e.g., CORP.EXAMPLE.COM). It is used for Kerberos authentication to identify the specific security domain.

   **IP Address**

   This is the specific IP address of the Domain Controller responsible for authentication and other directory services. You can find this by running `nslookup` on your Active Directory FQDN.

   **FQDN**

   The Fully Qualified Domain Name (FQDN) is the complete domain name for your Active Directory, such as corp.example.com. It uniquely identifies your domain on the Internet.
8. Click Validate and Configure.
9. In the Authorize On-Premises Components section under Active Directory, click Show commands.
10. Copy the commands and run them on the Active Directory instance in a command prompt with administrative rights. You must log in with a user who has **Domain Administrator**, **Enterprise Administrator**, or equivalent elevated privileges.
11. Under Data Lake, click Show Instructions.
12. Copy the code.
13. Create a new file in the /etc/krb5.conf.d/ folder with a custom name and .conf extension on Cloudera on premises.
14. Paste the code into the file and save it. It is automatically processed by the Kerberos service.
15. Configure the *AUTH_TO_LOCAL* service setting in Cloudera Manager on-premises as explained in the Adding trusted realms to the cluster section of the Cloudera Base on premises documentation.
16. Click Test connection.

## Results

When the connection is successful, the data lake services are listed, and the cross-realm setup is finished.

## What to do next

You can start creating Cloudera Data Hub clusters in your hybrid cloud environment.