Cloudera on Premises Data Services 1.5.5

Cloudera Data Services on premises Environments

Date published: 2023-12-16
Date modified: 2025-11-08

https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2025. 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

Environments

In Cloudera, a private cloud environment is a logical entity that represents the association of your Cloudera on premises user account with multiple compute resources using which you can provision and manage workloads such as Cloudera Data Warehouse, Cloudera Data Engineering, and Cloudera AI. You can register as many environments as you require.

Registering an environment provides Cloudera with access to your user account and identifies the resources in your user account that Cloudera services can access or provision. For Cloudera on premises environments, resources include compute clusters such as Kubernetes as well as Data Lake clusters in Cloudera. Compute workloads are deployed within these environments.

A workload receives access to a Kubernetes cluster for compute purposes and a Data Lake cluster for storage, metadata, and security purposes within the environment in which it is deployed. Administrators can define user permissions and set resource quotas in each environment.

Cloudera on premises environment prerequisites

Cloudera on premises requires that you set up specific account criteria prior to registering an environment.

Set up the following before registering an environment:

- The trust store file that is configured in Cloudera Manager must meet the following requirements:
  - The file must not be empty
  - The file must be in JKS format (not PKCS12)
  - The file must contain the Certificate Authorities that have signed the following certificates:
    - The certificates of all Cloudera Manager servers that the user plans to use as base clusters in Cloudera on premises
    - The certificates of all external Postgres servers used by Hive Metastore services in all base clusters

  You can access the trust store file in Cloudera Manager at Administration > Settings > Security > "Cloudera Manager TLS/SSL Client Trust Store File".
- Your base cluster must meet the following requirements:
  - Ranger configured and running
  - Atlas configured and running
  - HMS configured and running
  - HDFS available and warehouse root configured
- A Kubeconfig file with cluster admin privileges.
- The associated Cloudera Manager user name and password.
- The Apache Ozone service must be enabled on your Cloudera Base on premises cluster that you are using for the Cloudera Data Engineering service.

Registering a Cloudera on premises environment

After you have set up the Cloudera on premises requirements, you can register the environment.

Steps

Management Console UI

1. Sign into the Cloudera console.
2. Click Environments.
3. On the Environments page, click Register Environment.

4. On the Register Environment page, provide the required information.

   Environment

   Environments in Cloudera on premises provide shared data, security, and governance (metadata) for your Cloudera AI and Cloudera Data Warehouse applications.

Property	Description
Environment Name	Enter a name for your environment. This name will be used to refer to this environment in Cloudera. Note: Cloudera Data Warehouse service requires that you specify the environment name less than 45 characters long. This is because Cloudera Data Warehouse uses a deterministic namespace and adds a prefix to the environment name. The length of the namespace ID after Cloudera Data Warehouse applies a prefix to the Environment name, including the hyphen (-), should not exceed 63 characters.

   Resource Quota

   Optionally you can set quota for CPU, Memory, and GPU resources for this environment. When setting a quota for the environment, ensure that it satisfies the resource requirement of the data services to be created within the environment. However, if you do not specify a quota for a resource, then no quota will be set for that resource at the time of the environment creation.

   Note: Ensure that at least 4 CPU Cores and 30 GB of RAM are reserved for Monitoring.

Property	Description
CPU (Cores)	You can specify a CPU quota (in Cores) for the environment. The quota must be a positive number.
Memory (GB)	You can specify a Memory quota (in GigaBytes) for the environment. The quota must be a positive number.
GPU (Cores)	You can specify a GPU quota (in Cores) for the environment. The quota must be a positive number.

   Compute Cluster Resources

   To run workloads, you must specify a Kubeconfig file to register a Kubernetes cluster with Cloudera on premises.

Property	Description
Kubernetes Configurations	Click Upload Files, then select a Kubeconfig file to enable Cloudera to access a Kubernetes cluster.
Storage class	The storage class on the OpenShift cluster. If you do not specify this value, the default storage class is used.
Domain	The default domain suffix for workload applications.

   Data Lake

   A Data Lake refers to the shared security and governance services in a Cloudera Data Center cluster linked to a Cloudera on premises environment, and managed by Cloudera Manager. To register an environment, Cloudera on premises needs to access Cloudera Manager and its Data Lake services.

   Note: When registering an environment in Cloudera Data Services on premises - Data Lake services - Cloudera Manager, you need to use FQDN (Fully Qualified Domain Name) rather than IP address.

Parameter	Description
Cloudera Manager URL	The Cloudera Manager URL.
Cloudera Manager Admin Username	The Cloudera Manager administrator user name.
Cloudera Manager Admin Password	The Cloudera Manager administrator password.

5. Under Data Lake, click Connect.

   When Cloudera on premises has successfully connected to Cloudera Manager, a confirmation message appears, along with the Data Lake cluster services.
6. Click Register.

   The environment page appears. The new environment is also listed on the Environments page.

   Note: You can repeat steps 2 through 5 to register more environments. If required, you can select a different Data Lake cluster while registering the additional environment. Cloudera Data Services on premises has multiple base cluster support, and users can create multiple environments each pointing to a different base or data lake cluster.

CLI

You can use the following command to create a new environment:

```
cdp environments create-private-environment \
--environment-name <value> \
--address <value> \
--authentication-token <value> \
----cluster-names <value>
```

For a detailed description of the command properties, use cdp environments create-private-environment --help

Creating multiple environments with different base or Data Lake clusters

To register an environment with a data lake cluster managed by a Cloudera Manager that is different from your existing Cloudera Manager, you need to add the certificates of the new Cloudera Manager to the Cloudera Management Console. If the existing Cloudera Manager and the new Cloudera Manager share the same root CA, and the root CA is already uploaded as the data lake certificate, then no additional certificate needs to be added.

Cloudera Manager certificates can be accessed from the Cloudera Manager server.

1. To get the certificate path on the Cloudera Manager server, navigate to the Administration tab on your Cloudera Manager UI.
2. Select Settings, and within the Settings, select Security as the category.
3. In the list of settings, Cloudera Manager TLS/SSL Server Keystore File Location points to the Cloudera Manager server certificate. This certificate needs to be combined with the certificates of your existing Cloudera Manager server(s).

   Note: The combined certificate needs to be in X.509 PEM format.

   For information on combining certificates, see Configuring multiple base clusters with ECS .
4. The combined certificate then needs to be uploaded as a Datalake certificate to Cloudera Management Console > Administration > CA Certificates . This ensures that the Cloudera Data Services on premises cluster is configured to support both the new and the existing data lake clusters.

Accessing a Cloudera on premises environment

Once an environment exists, you can access it from the Cloudera Management Console.

About this task

From the details available on the Environments page, you can access the Data Lake cluster, the Data Hub clusters running within the environment, and the Summary page listing the information.

Procedure

1. To access an existing environment, navigate to Cloudera Management Console > Environments and click on your environment.

2. On the environment details page, you can access the shared Data Lake cluster services and Cloudera Manager.

Managing the compute cluster

Compute Cluster provides information about the Kubernetes version, number of nodes in the cluster, and cluster registration time. You can access the RedHat OpenShift dashboard from here. You can download and update the Kubernetes configuration file.

Updating the Kubernetes configuration

You must update your kubeconfig file before it expires to ensure continued access to the cluster. The Kubeconfig file contains the cluster access information and authentication information for the admin user.

Procedure

1. Navigate to the Cloudera Management Console > Environments.
2. Click the title of your environment.
3. Click the Compute Cluster tab.
4. Click Actions > Update Kubernetes Configuration. This option is available only in OpenShift Container Platform.
5. Upload the new Kubeconfig file using the Choose File option and click Save.

What to do next

You can see the updated date and time in the Registered section.

Downloading the Kubernetes Configuration (OpenShift Container Platform deployments only)

On OpenShift Container Platform deployments, you can download and view the kubeconfig file used to register your environment.

Procedure

1. Navigate to the Cloudera Management Console > Environments.
2. Click the title of your environment.
3. Click the Compute Cluster tab.
4. Click Actions > Download Kubernetes Configuration.

What to do next

The Kubernetes configuration file is downloaded to your system.

Handling ingress controller certificate rotation

This section provides guidance on managing ingress controller certificate rotation for Cloudera environments. It addresses the process for rotating certificates in Cloudera Control Plane and ensuring updates are reflected in the Monitoring namespace.

Currently, Cloudera Control Plane supports certificate rotation through a Cloudera Manager command, which rotates the certificate and recreates the necessary secrets within the Cloudera Control Plane namespace. However, the Monitoring namespace only copies the ingress secret during cluster creation, and subsequent updates, such as certificate rotation, are not automatically propagated.

Updating Monitoring namespace for certificate rotation

You must manually update the Monitoring namespace to ensure seamless certificate rotation and synchronization across namespaces.

Before you begin

Verify that you have kubectl installed and configured to point to the cluster that requires fixing.

Procedure

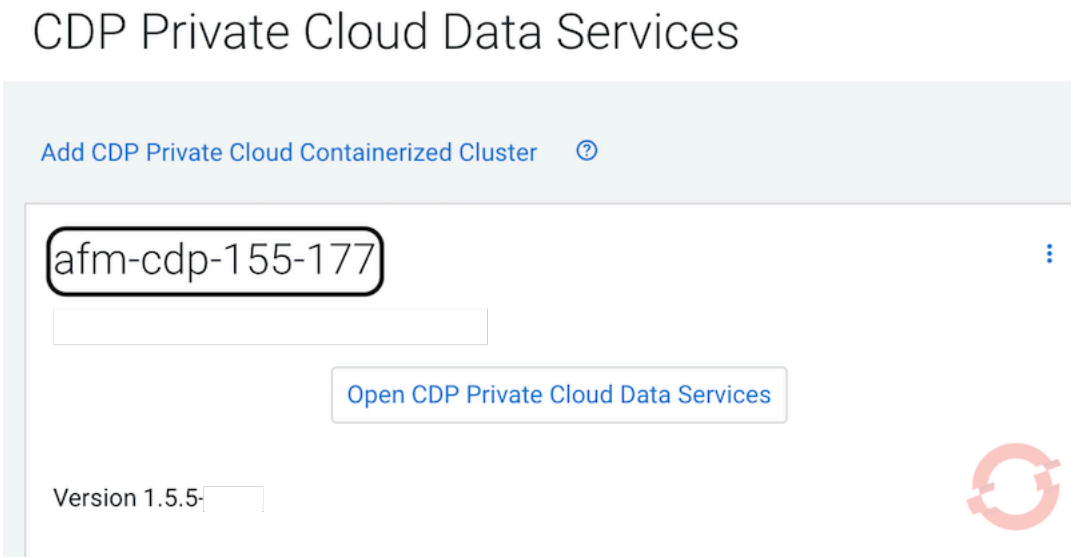
1. Copy the below script to a new file and save it as update-monitoring-namespace-cacert.sh

```
export SOURCE_NAMESPACE="<CONTROL_PLANE_NAMESPACE>"
export SOURCE_CONFIGMAP="cdp-pvc-truststore"
export SOURCE_KEY="JKS"
export TARGET_NAMESPACE="<MONITORING_PLATFORM_NAMESPACE>"
export TARGET_CONFIGMAP="monitoring-sdx-ca-certs"
export TARGET_KEY="cacerts"

# Store the data in a variable
DATA=$(kubectl get configmap $SOURCE_CONFIGMAP -n
$SOURCE_NAMESPACE -o jsonpath="{.binaryData.$SOURCE_KEY}")

# Update the target configmap using the variable
kubectl patch configmap $TARGET_CONFIGMAP -n $TARGET_NAMESPACE \
--type='json' -p="[{\"op\": \"replace\", \"path\": \"/binaryData/$TARGET_K
EY\",
\"value\":\"$DATA\"}]"
```

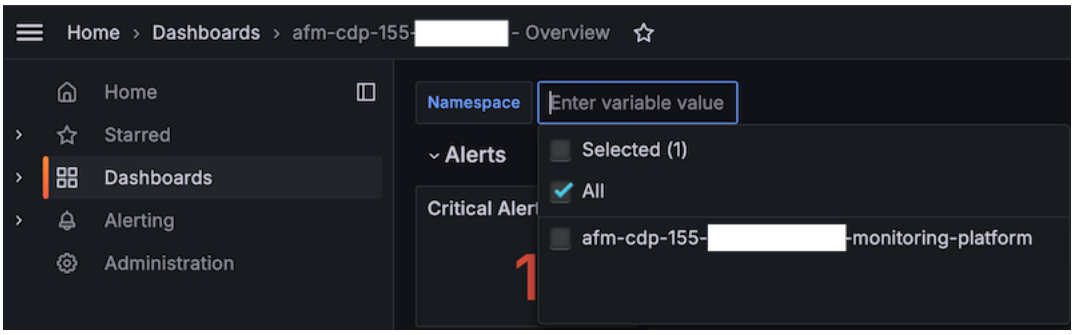
2. Edit the script by replacing <CONTROL_PLANE_NAMESPACE> with the name of the Cloudera Control Plane namespace.



3. Edit the script by replacing <MONITORING_PLATFORM_NAMESPACE> with the Monitoring platform namespace for the problematic environment. This is the namespace where the monitoring-cm-health-exporter is located.

   Note: Use the full name. For example, afm-cdp-155-<hash>-monitoring-platform

   a) Log in to the Monitoring dashboard
   b) Click the Namespace drop-down
   c) Copy the Monitoring platform namespace



4. Save and run the script `update-monitoring-namespace-cacert.sh`

What to do next

Verify recovery:

1. The monitoring-cm-health-exporter should recover the next time it restarts.
2. Alternatively, you can terminate the problematic pod to force it to recover immediately. Run the command `kubectl rollout restart deployment monitoring-cm-health-exporter -n <monitoring_namespace>`

Deleting a Cloudera on premises environment

Deleting an environment also removes all the resources associated with that environment.

Deleting a Cloudera on premises

Cloudera Management Console

1. On the Environments page, click the environment that you want to remove.
2. Select Actions > Delete Environment.
3. Click Delete to confirm the deletion of the selected environment.

CLI

You can use the following command to delete an environment:

```
cdp environments delete-environment \
--environment-name <value> \
```

For a detailed description of the command properties, use cdp environments delete-environment --help