Cloudera Data Hub

Accessing Clusters

Date published: 2019-12-17
Date modified: 2025-08-18

https://docs.cloudera.com/

Legal Notice

Contents

Enabling admin and user access to Cloudera Data Hub clusters

Cloudera Data Hub resource roles can be assigned on the scope of a specific Cloudera Data Hub cluster.

When you grant access to admins and users of a Cloudera Data Hub, consider the following guidelines:

- Any user or group that needs to access a specific Cloudera Data Hub needs the EnvironmentUser role at the scope of the environment where that Cloudera Data Hub cluster is running.
- A user with the DataHubCreator (or higher) account role can create Cloudera Data Hub clusters.
- The user who creates a Cloudera Data Hub cluster gets the Owner role for that Cloudera Data Hub cluster.
- The Owner of the Cloudera Data Hub cluster can grant others access to the cluster. The following roles can be assigned:
  - Owner - Grants the permission to mange the Cloudera Data Hub cluster in Cloudera and delete it. It does not grant any cluster-level access (such as access to Cloudera Manager).
  - DataHubAdmin (Technical Preview) - Grants administrative rights over the Cloudera Data Hub cluster.

The roles are described in detail in *Resource roles*. The steps for assigning the roles are described in *Assigning resource roles to users* and *Assigning resource roles to groups*.

Related Information

Resource roles

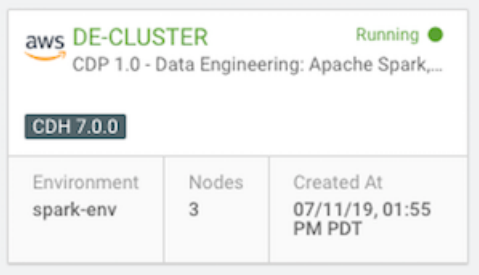
Assigning resource roles to users

Assigning resource roles to groups

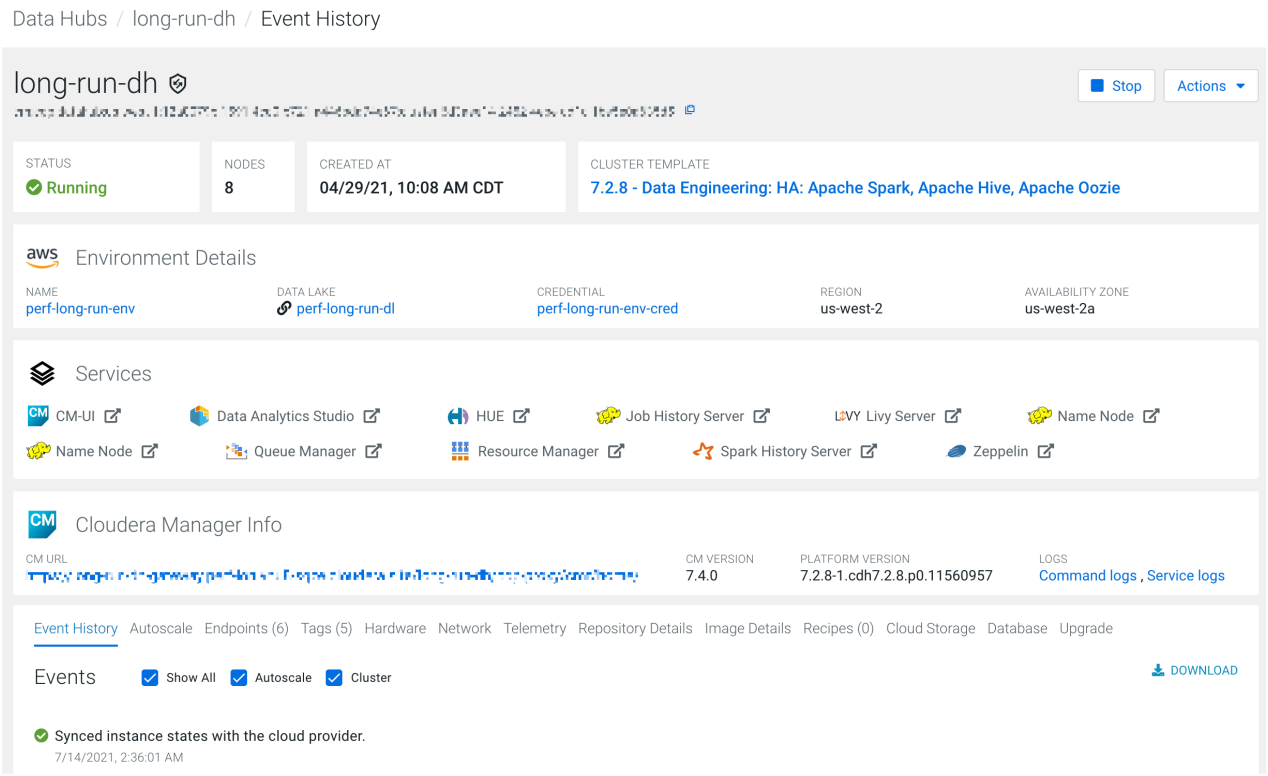
Understanding Cloudera Data Hub cluster details

To access information about Data Hub clusters, navigate to the Data Hub Clusters service or to the Cloudera Management Console service > Data Hub Clusters.

Each cluster is represented by a tile that includes basic information about the cluster:



Click the cluster tile to display more information:



The summary bar includes the following information:

Item	Description
Status	Current cluster status. When a cluster is healthy, the status is Running.
Nodes	The current number of cluster nodes.
Created at	The date and time when the cluster was created. The format is MM/DD/YY, Time AM/PM Timezone. For example: 06/20/19, 7:56 AM PDT.
Cluster Template	The name of the cluster template used to create the cluster.

Environment Details

You can find information about the cluster cloud provider environment under Environment Details:

Item	Description
Cloud Provider	The logo of the cloud provider on which the cluster is running.
Name	The name of the environment used to create the cluster.
Data Lake	The name of a Data Lake to which the cluster is attached.
Credential	The name of the credential used to create the cluster.
Region	The region in which the cluster is running in the cloud provider infrastructure.
Availability Zone	The availability zone within the region in which the cluster is running.

Services

In the Services section, you will find links to cluster UIs. The exact content depends on what components are running on your cluster.

Cloudera Manager Info

The Cloudera Manager Info section provides the following information:

Item	Description
Cloudera Manager URL	Link to the Cloudera Managerr web UI.
Cloudera Manager Version	The Cloudera Manager version which the cluster is currently running.
Platform Version	The Cloudera Runtime version which the cluster is currently running.

Event History and other tabs

Under Cloudera Manager, the Event History tab is displayed by default. You can also click the other tabs to view additional cluster information.

Item	Description
Event History	Shows events logged for the cluster, with the most recent event at the top. The Download option allows you to download the event history.
Hardware	This section includes information about your cluster instances: instance names, instance IDs, instance types, their status, fully qualified domain names (FQDNs), and private and public IPs. If you click on the », you can access more information about the instance, storage, image, and packages installed on the image.
Tags	This section lists keys and values of the user-defined tags, in the same order as you added them.
Endpoints	This section includes the endpoints for various cluster servces.
Recipes	This section includes recipe-related information. For each recipe, you can see the host group on which a recipe was executed, recipe name, and recipe type.
Repository Details	This section includes Cloudera Manager and Cloudera Runtime repository information, as you provided when creating a cluster.
Image Details	This section includes information about the prewarmed or base image that was used for the cluster nodes.
Network	This section includes information about the names of the network and subnet in which the cluster is running and the links to related cloud provider console.
Cloud Storage	This section provides information about the base storage locations used for YARN and Zeppelin.
Database	This section provides information about any external managed database you might have created for the cluster.
Telemetry	This section provides information about logging, metering, cluster log collection, and other analytics.

Actions Menu

Click Show Cluster Template on the Actions menu to review the template used in cluster creation. Click Show CLI Command to review the CDP CLI command used to create the cluster (which you can copy to create similar clusters via the CDP CLI). Select Manage Access to manage access to the cluster.

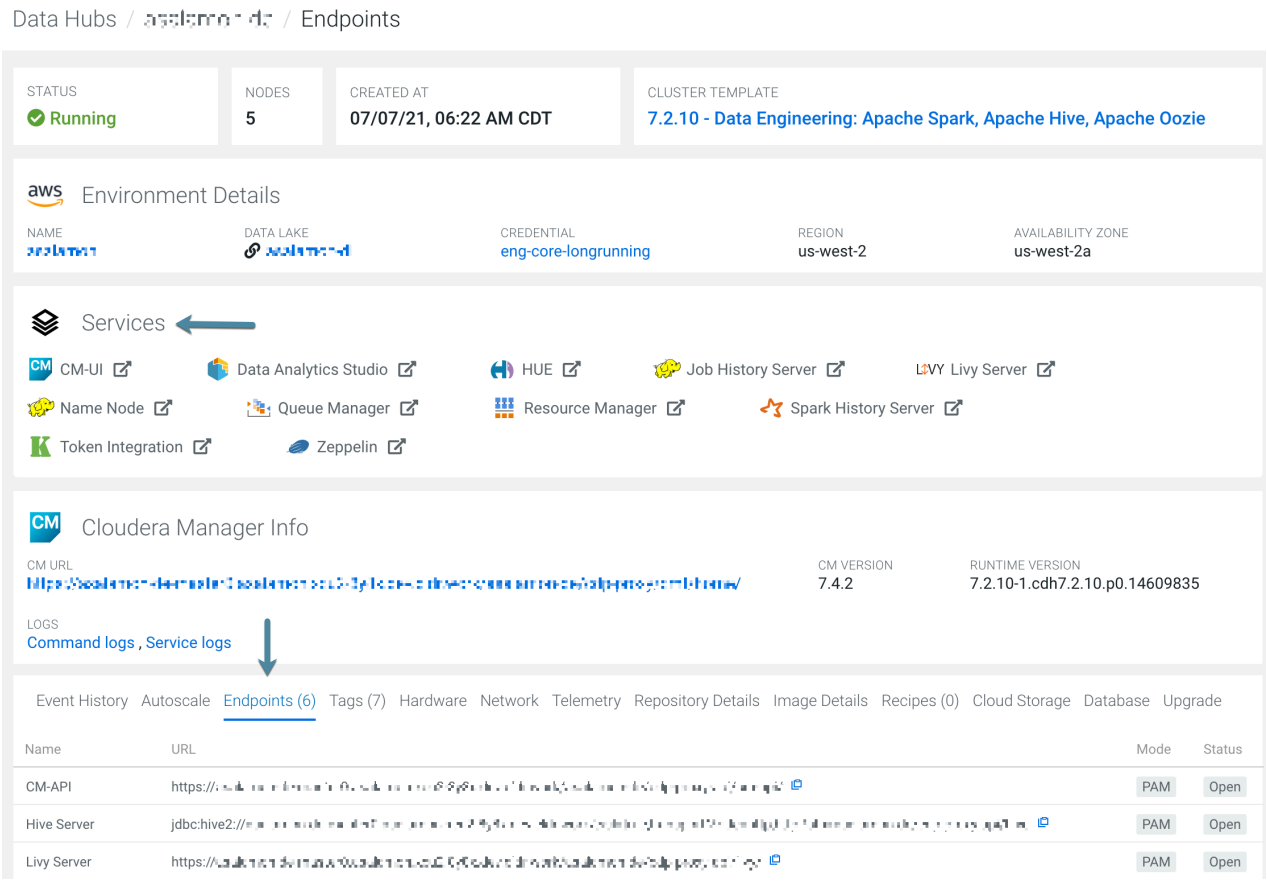
You can also perform some basic Cloudera Manager management functions from the Actions menu, such as resizing, retrying, and repairing the cluster, as well renewing the host certificate.

Accessing Cloudera Manager, cluster UIs, and endpoints

Cluster UIs and endpoints can be accessed from cluster details.

Required role: EnvironmentUser at the scope of the environment where the Data Hub is running, but Cloudera Manager access is read-only. EnvironmentAdmin grants a limited administrator role in Cloudera Manager. DatahubAdmin or the Owner of the Cloudera Data Hub can access cluster details, but access to Cloudera Manager is read-only.

To access cluster UIs and endpoints navigate to the to the Data Hub Clusters service and click the tile for your cluster. This opens the cluster details page, which lists the URLs for the cluster UIs and endpoints:



Click the URL for the service that you would like to access and you will be logged in automatically with your Cloudera credentials. All of the UIs and endpoints are accessible via the Knox gateway. The URLs listed connect you to a chosen service via Knox, and Knox securely passes your Cloudera credentials.

Credentials to use for logging in

The following table lists the credentials to use to access clusters:

Method	URL	Credentials
Cloudera Manager web UI	Access from the URL listed in cluster details > Services section.	You do not need to provide any credentials. You are automatically logged in with your Cloudera credentials. When accessing a Cloudera Data Hub cluster via Cloudera Manager, you assume the Configurator role.

Method	URL	Credentials
All cluster web UIs	Access from the URLs listed in cluster details.	You do not need to provide any credentials. You are automatically logged in with your Cloudera credentials.
Data Analytics Studio (DAS)	Access from the URLs listed in cluster details.	Access DAS with your workload user name and workload password. When accessing Cloudera for the first time, you must set a workload password. For more instructions on how to set your workload password, refer to Set or reset workload password. For instructions on how to find your workload user name, refer to Finding your workload user name.
All cluster endpoints	Access by using the API endpoint listed in cluster details > Endpoints tab.	Access all cluster API endpoints (such as JDBC and ODBC) with your workload user name and workload password. When accessing Cloudera for the first time, you must set a workload password. For more instructions on how to set your workload password, refer to Set or reset workload password. For instructions on how to find your workload user name, refer to Finding your workload user name. For information on how to set up a connection from a business intelligence tool such as Tableau, refer to Configuring JDBC for Impala and Configuring ODBC for Impala.

Security exception

The first time you access the UIs, your browser will attempt to confirm that the SSL Certificate is valid. Since CDP automatically generates a certificate with self-signed CA, your browser will warn you about an untrusted connection and ask you to confirm a security exception. Depending on your browser, perform the steps below to proceed:

Browser	Steps
Firefox	Click Advanced > Click Add Exception… > Click Confirm Security Exception
Safari	Click Continue
Chrome	Click Advanced > Click Proceed…

Accessing a Cloudera Data Hub cluster via SSH

You can use SSH to access cluster nodes via a command line client.

Non-root SSH access to cluster VMs

Required role: Any user who has access to the environment (EnvironmentUser, DataSteward, and EnvironmentAdmin) can access Cloudera Data Hub clusters via SSH.

To execute privileged operating system actions on a Cloudera Data Hub, the EnvironmentPrivilegedUser role is required.

All authorized users can access cluster nodes via SSH using either a private key that matches the user's public key previously provided in Cloudera, or their workload user name and workload password.

For SSH access through a workload user name and password:

When accessing Cloudera for the first time, you must set a workload password. The password also needs to be reset each time you are added to a new environment.

For more information about workload passwords and instructions for setting/resetting it, refer to Set or Reset Workload Password.

For instructions on how to find your workload user name, refer to Finding Your Workload User Name.

On Mac OS, you can use the following syntax to SSH to the VM:

```
$ ssh <workloaduser>@<nodeIPaddress>
```

For example:

```
$ ssh jsmith@190.101.0.132
```

To SSH to a cluster using the private key file that pairs with the public key associated with a user, use the ssh utility:

```
$ ssh -i <path-to-private-key-file> <cdpusername>@<nodeIPaddress>
```

For example:

```
$ ssh -i ~/.ssh/my-private-key jsmith@192.12.141.12
```

On Windows, you can access your cluster via SSH by using an SSH client such as PuTTY. For more information, refer to How to use PuTTY on Windows.

Root SSH access to cluster VMs

Required role: None

Cloudera administrators can access cluster nodes as cloudbreak user with the SSH key provided during environment creation.

On Mac OS, you can use the following syntax to SSH to the VM:

```
ssh -i <path-to-cloudbreak-private-key-file> cloudbreak@<nodeIPaddress>
```

For example:

```
ssh -i ~/.ssh/cloudbreak-private-key cloudbreak@p90.101.0.132
```

On Windows, you can access your cluster via SSH by using an SSH client such as PuTTY. For more information, refer to How to use PuTTY on Windows.

Setting workload password

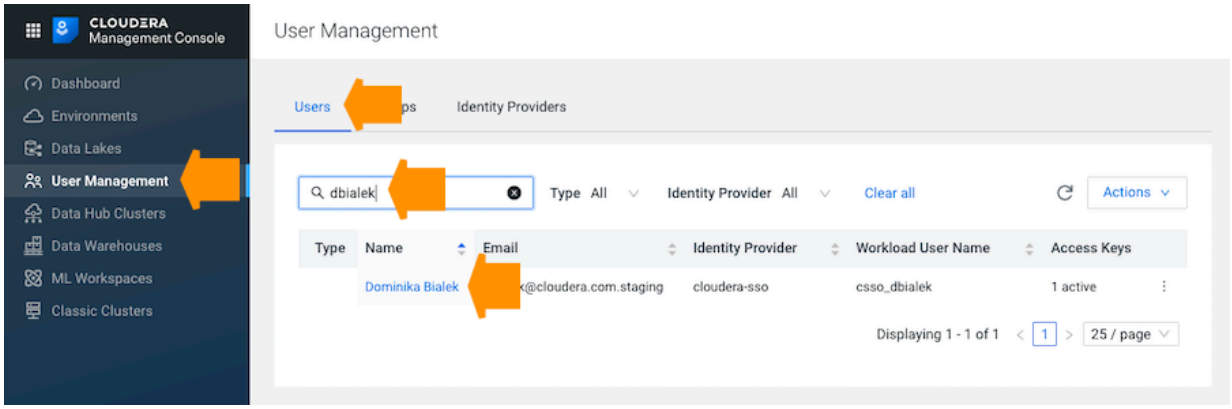
A workload password is used to access Cloudera Data Hub clusters via SSH, endpoints such as JDBC/ODBC, and some UIs such as DAS.

Required role: All users can manage their workload passwords from the account management page. All users can manage their workload password from CDP CLI, but this action requires an API access key, which can only be generated by users with the IAMUser role. As a Cloudera administrator or PowerUser, you can manage the workload password for all user accounts.

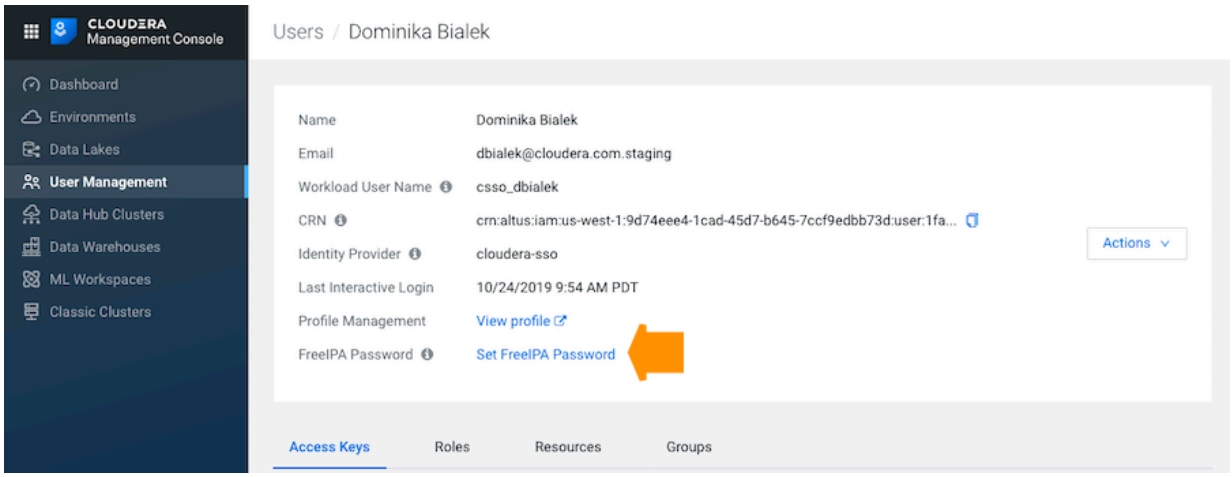
The workload password is independent from your SSO password.

To set or reset the password:

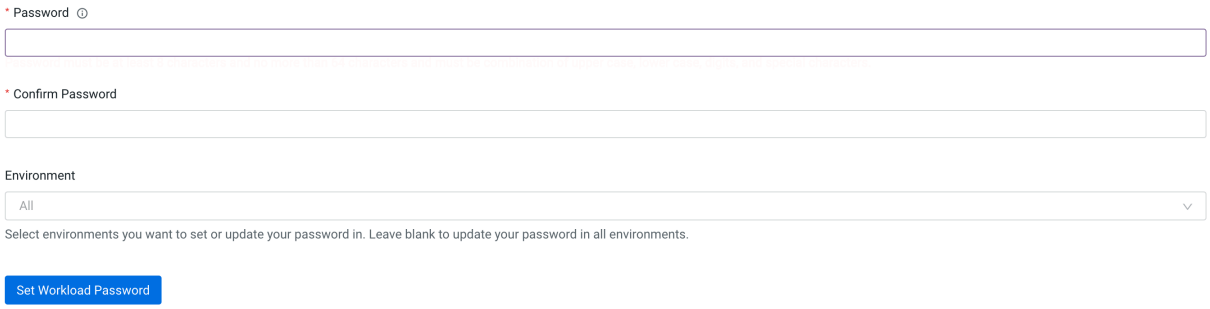
1. Sign in to the Cloudera web interface.
2. Navigate to Cloudera Management Console > User Management.
3. Search for your user name, and then click your user name:



4. Click Set Password for User:



5. In the dialog box that appears, enter the workload password twice:

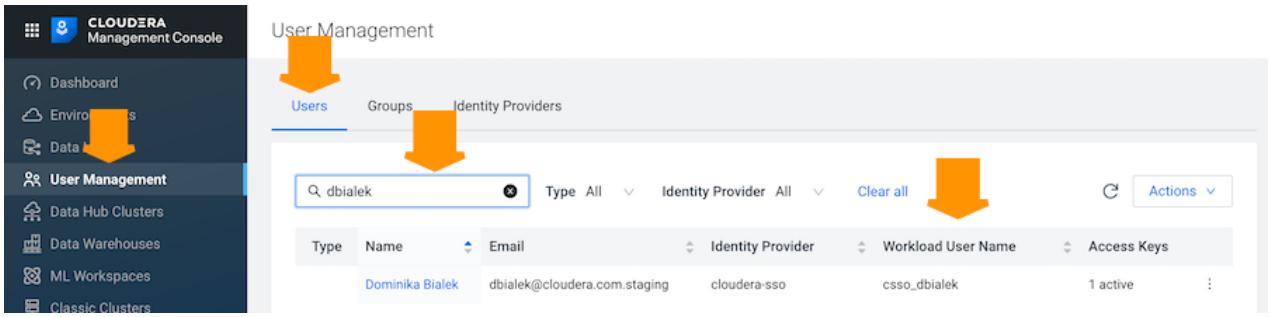


6. Click Set Workload Password.

Finding your workload user name

Once you have reset your workload password, locate your workload user name.

To check your workload user name, select Cloudera Management Console > User Management > Users, find your user name, and then find your Workload User Name:



Retrieving keytabs for workload users

A keytab file stores long-term keys for a principal in Kerberos. You can generate a keytab either through the Cloudera Management Console user interface or the CDP CLI.

About this task

Required roles: All users can retrieve their keytabs from the account management page. All users can retrieve their keytabs from CDP CLI, but this action requires an API access key, which can only be generated by users with the IAMUser role. As a Cloudera administrator or PowerUser, you can retrieve the keytab for all user accounts.

You may need to generate a keytab for a workload user in certain Cloudera Data Hub use cases, for example long-running Spark streaming jobs, which require a keytab as a long-lived credential; or NiFi flows requiring a keytab to write data into HBase.

Note: Keytabs are scoped to an environment, whereas workload passwords are the same for every environment. A keytab is, however, tied to the workload password. If you change the workload password, you must retrieve a new keytab. When you change a workload password, retrieve the keytab only after the user sync operation is complete. For more information on user sync, see *Assigning resources to users*.

Procedure

You can retrieve a keytab either in the Cloudera Management Console or in the CDP CLI:

- Cloudera Management Console:
  a. Click User Management Users and then search for and select the Name of the user that you want to get a keytab for.
  b. Click Actions Get Keytab .
  c. Select the environment in which the Cloudera Data Hub cluster is running and then click Download.
  d. Save the keytab file in a location of your choice.

     Once you have downloaded the keytab file, you can copy it to the machine on which the cluster runs and use the keytab to authenticate as the workload user principal, or point to the keytab file when running a Spark job or other job that requires a keytab.
- CDP CLI:
  a. Keytab retrieval (get-keytab) is within the environments module. Run cdp environments get-keytab help for more information. You will need to pass the environment name and an actor CRN:

```
cdp environments get-keytab \
--environment-name=EnvironmentName \
--actor-crn=ActorCrn
```

b. The output of the command is a base64-encoded representation of a keytab. The contents of the output must be base64 decoded and saved to a file for it to work as a keytab.

Note: There are ways to generate keytabs with utilities outside of Cloudera, such as ipa-getkeytab or ktutil. Cloudera recommends against using these methods as they may not work as expected. For example, ipa-getk eytab creates a keytab that may work but only temporarily.

Related Information

Assigning resources to users
CLI client setup

Running workloads

Once your cluster is running, refer to the Cloudera Runtime and Cloudera Data Hub documentation for information on how to run workloads using services such as Hive, Spark, Impala, Hue, or Kafka.

Related Information

Data Access
Data Science
Streaming
Flow Management
Operational Database