Cloudera Data Engineering 1.5.4

Configuring users to create jobs

Date published: 2020-07-30
Date modified: 2024-05-30

https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

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

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

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

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

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

Contents

Configuring users to create jobs

You must complete the manual steps to prepare the cluster for each user that needs to submit jobs. You can also configure a service account Kerberos key tab file to a machine user id in order to submit CDE jobs. The machine user will now have access to the hdfs storage through jobs.

Note: After you create a cluster, you must initialize the cluster, and configue users before creating jobs.

Configuring LDAP users

You must complete the following manual steps to prepare the cluster for each user that needs to submit jobs. Perform these steps for each user that needs to submit jobs to the virtual cluster.

Before you begin

In Cloudera Data Engineering (CDE), jobs are associated with virtual clusters. Before you can create a job, you must create a virtual cluster that can run it. For more information, see Creating virtual clusters.

Procedure

1. If you already downloaded the utility script and uploaded it to an ECS or HDFS gateway cluster host as documented in Creating virtual clusters, you can skip to step 8.
2. Download cde-utils.sh to your local machine.
3. Create a directory to store the files, and change to that directory:

   ```
   mkdir -p /tmp/cde-1.3.4 && cd /tmp/cde-1.3.4
   ```

4. Embedded Container Service (ECS)

   Copy the extracted utility script (cde-utils.sh) to one of the Embedded Container Service (ECS) cluster hosts. To identify the ECS cluster hosts:

   a. Log in to the Cloudera Manager web interface.
   b. Go to Clusters Experience Cluster ECS Hosts .
   c. Select one of the listed hosts, and copy the script to that host.

   Red Hat OpenShift Container Platform (OCP)

   Copy the extracted utility script (cde-utils.sh) and the OpenShift kubeconfig file to one of the HDFS service gateway hosts, and install the kubectl utility:

   a. Log in to the Cloudera Manager web interface.
   b. Go to Clusters Base Cluster HDFS Instances .
   c. Select one of the Gateway hosts, and copy the script to that host.
   d. Copy the OCP kubeconfig file to the same host.
   e. On that host, install the kubectl utility following the instructions in the Kubernetes documentation.
5. On the cluster host that you copied the script to, set the script permissions to be executable:

   ```
   chmod +x /path/to/cde-utils.sh
   ```

6. Identify the virtual cluster endpoint:

   a. In the Cloudera Manager web UI, go to the Experiences page, and then click Open CDP Private Cloud Experiences.

   b. Click the Data Engineering tile.

   c. Select the CDE service containing the virtual cluster you want to activate.

   d. Click ✎ Cluster Details.

   e. Click JOBS API URL to copy the URL to your clipboard.

   f. Paste the URL into a text editor to identify the endpoint host. For example, the URL is similar to the following:

   ```
   https://dfdj6kgx.cde-2cdxw5x5.apps.ecs-demo.example.com/dex/api/v1
   ```

   The endpoint host is dfdj6kgx.cde-2cdxw5x5.apps.ecs-demo.example.com.

7. On the ECS or HDFS gateway host, create a filename containing the user principal, and generate a keytab. If you do not have the ktutil utility, you might need to install the krb5-workstation package. The following example commands assume the user principal is psherman@EXAMPLE.COM

   a. Create a file named <*username*>.principal (for example, psherman.principal) containing the user principal:

   ```
   psherman@EXAMPLE.COM
   ```

   b. Generate a keytab named <*username*>.keytab for the user using ktutil:

   ```
   sudo ktutil
   ktutil:  addent -password -p psherman@EXAMPLE.COM -k 1 -e aes256-cts
   Password for psherman@EXAMPLE.COM:
   ktutil:  addent -password -p psherman@EXAMPLE.COM -k 2 -e aes128-cts
   Password for psherman@EXAMPLE.COM:
   ktutil:  wkt psherman.keytab
   ktutil:  q
   ```

8. Validate the keytab using klist and kinit:

   ```
   klist -ekt psherman.keytab
   Keytab name: FILE:psherman.keytab
   KVNO Timestamp           Principal
   ---- ------------------- ----------------------------------------------
   --------
      1 08/01/2021 10:29:47 psherman@EXAMPLE.COM (aes256-cts-hmac-sha1-96)
      1 08/01/2021 10:29:47 psherman@EXAMPLE.COM (aes128-cts-hmac-sha1-96)

   kinit -kt psherman.keytab psherman@EXAMPLE.COM
   ```

   Make sure that the keytab is valid before continuing. If the kinit command fails, the user will not be able to run jobs in the virtual cluster. After verifying that the kinit command succeeds, you can destroy the Kerberos ticket by running kdestroy.

9. Use the cde-utils.sh script to copy the user keytab to the virtual cluster hosts:

   ```
   ./cde-utils.sh init-user-in-virtual-cluster -h <endpoint_host> -u <user> -
   p <principal_file> -k <keytab_file>
   ```

   For example, using the psherman user, for the dfdj6kgx.cde-2cdxw5x5.apps.ecs-demo.example.com endpoint host:

   ```
   ./cde-utils.sh init-user-in-virtual-cluster -h dfdj6kgx.cde-2cdxw5x5.app
   s.ecs-demo.example.com -u psherman -p psherman.principal -k psherman.key
   tab
   ```

10. Repeat these steps for all users that need to submit jobs to the virtual cluster.

Configuring service account key tab to the machine user

You can configure a service account Kerberos key tab file to a machine user id in order to submit CDE jobs. The machine user will now have access to the hdfs storage through jobs.

Procedure

1. *Create a CDP machine user*. For example, `mu_cde`.
2. *Generate access key* and secret key for the machine user and download the credential information.
3. On the ECS or HDFS gateway host, create a filename containing the user principal, and generate a keytab. If you do not have the `ktutil` utility, you might need to install the krb5-workstation package. The following example commands assume the user principal is `mu_cde@example.com`.

   a. Create a file named <username>.principal (for example, *mu_cde.principal*) containing the user principal:

      ```
      mu_cde@example.com
      ```

   b. Generate a key tab file for the actual service account. For example, `svc_acc`. Name this keytab file as `<username>.keytab`, example: `mu_cde.keytab`. The generation of keytab can be done for the user using the `ktutil` command:

      ```
      sudo ktutil
      ktutil:  addent -password -p mu_cde@example.com -k 1 -e aes256-cts
      Password for pmu_cde@example.com:
      ktutil:  addent -password -p mu_cde@example.com -k 2 -e aes128-cts
      Password for mu_cde@example.com:
      ktutil:  wkt mu_cde.keytab

      ktutil:  q
      ```

4. Validate the keytab using `klist` and `kinit` commands:

   ```
   klist -ekt mu_cde.keytab

   Keytab name: FILE:mu_cde.keytab
   KVNO Timestamp           Principal
   ---- ------------------- ------------------------------------------------------
   -----
      1 08/01/2021 10:29:47 mu_cde@example.com (aes256-cts-hmac-sha1-96)
      1 08/01/2021 10:29:47 mu_cde@example.com (aes128-cts-hmac-sha1-96)
   kinit -kt mu_cde.keytab mu_cde@example.com
   ```

   Make sure that the keytab is valid before continuing. If the `kinit` command fails, the user will not be able to run jobs in the virtual cluster. After verifying that the `kinit` command succeeds, you can destroy the Kerberos ticket by running `kdestroy`.

5. Upload the keytab file for the user id as `mu_cde` and use the principal and keytab file of the actual service account (`svc_acc`).

   ```
   ./cde-utils.sh init-user-in-virtual-cluster -h <endpoint_host> -u <user> -p <principal_file> -k <keytab_file>
   ```

   For example, using the `mu-cde` user, for the dfdj6kgx.cde-2cdxw5x5.apps.ecs-demo.example.com endpoint host:

   ```
   ./cde-utils.sh init-user-in-virtual-cluster -h dfdj6kgx.cde-2cdxw5x5.apps.ecs-demo.example.com -u mu-cde -p mu-cde.principal -k mu-cde.keytab
   ```
6. Set the *Ranger authorization policy* for the `svc_acc account`.

Related Information

Ranger authorization policy

Creating a machine user in CDP

Generate Access Key