Cloudera Data Science Workbench

Hadoop Authentication

Date published: 2020-02-28
Date modified:

CLOUDƎRA

Legal Notice

Contents

Hadoop Authentication with Kerberos for Cloudera Data Science Workbench

Cloudera Data Science Workbench users can authenticate themselves using Kerberos against the cluster KDC defined in the host's /etc/krb5.conf file.

Cloudera Data Science Workbench does not assume that your Kerberos principal is always the same as your login information. Therefore, you will need to make sure Cloudera Data Science Workbench knows your Kerberos identity when you sign in.

To authenticate against your cluster's Kerberos KDC, go to the top-right dropdown menu, click Account settings Hadoop Authentication, and enter your Kerberos principal. To authenticate, either enter your password or click Upload Keytab to upload the keytab file directly to Cloudera Data Science Workbench. Once successfully authenticated, Cloudera Data Science Workbench uses your stored credentials to ensure that you are secure when running your workloads.

When you authenticate with Kerberos, Cloudera Data Science Workbench will store your keytab in an internal database. When you subsequently run an engine, the keytab is used by a Cloudera Data Science Workbench sidecar container to generate ticket-granting tickets for use by your code. Ticket-granting tickets allow you to access resources such as Spark, Hive, and Impala, on Kerberized CDH clusters.

While you can view your current ticket-granting ticket by typing klist in an engine terminal, there is no way for you or your code to view your keytab. This prevents malicious code and users from stealing your keytab.

⚠️ **Important:**

- (New in 1.6.1) CDSW 1.6.1 fixes an issue where setting HADOOP_USER_NAME to the CDSW username had certain unintended consequences. This fix now sets HADOOP_USER_NAME to the first part of the Kerberos principal in kerberized environments. In non-kerberized environments, it is still set to the CDSW username.
- If the /etc/krb5.conf file is not available on all Cloudera Data Science Workbench hosts, authentication will fail.
- If you do not see the Hadoop Authentication tab, make sure you are accessing your personal account's settings from the top right menu. If you have selected a team account, the tab will not be visible when accessing the Team Settings from the left sidebar.
- When you upload a Kerberos keytab to authenticate yourself to the CDH cluster, Cloudera Data Science Workbench might display a fleeting error message ('cancelled') in the bottom right corner of the screen, even if authentication was successful. This error message can be ignored.

UI Behavior for Non-Kerberized Clusters

The contents of the Hadoop Authentication tab change depending on whether the cluster is kerberized.

For a secure cluster with Kerberos enabled, the Hadoop Authentication tab displays a Kerberos section with fields to enter your Kerberos principal and username. However, if Cloudera Data Science Workbench cannot detect a krb5 .conf file on the host, it will assume the cluster is not kerberized, and the Hadoop Authentication tab will display Hadoop Username Override configuration instead.

For a non-kerberized cluster, by default, your Hadoop username will be set to your Cloudera Data Science Workbench username. To override this default and set an alternative HADOOP_USER_NAME, go to the Hadoop Username Override setting at Account settingsHadoop Authentication.

If the Hadoop Authentication tab is incorrectly displaying Kerberos configuration fields for a non-kerberized cluster, make sure the krb5.conf file is not present on the host running Cloudera Data Science Workbench. If you do find any instances of krb5.conf on the host, depending on your deployment, perform one of the following sets of actions:

- On CSD deployments, go to Cloudera Manager and stop the CDSW service. Remove the krb5.conf file(s) from the Cloudera Data Science Workbench gateway host, and then start the CDSW in Cloudera Manager.

  OR
- On RPM deployments, run cdsw stop, remove the krb5.conf file(s) from the Cloudera Data Science Workbench gateway host, and run cdsw start.

You should now see the expected Hadoop Username Override configuration field.

Limitations

- Cloudera Data Science Workbench does not support the use of Kerberos plugin modules in krb5.conf.

Limitations

Provides limitations for Cloudera Data Science Workbench support for Kerberos.

Cloudera Data Science Workbench does not support the use of Kerberos plugin modules in krb5.conf.

Configure FreeIPA

In addition to MIT Kerberos and Active Directory, Cloudera Data Science Workbench also supports FreeIPA as an identity management system. However, this support comes with one major caveat: if your Kerberos configuration file (/etc/krb5.conf) contains references to any external files that reside on the host operating system, Kerberos authentication could fail. This is because those files will not automatically be mounted into the engines where Cloudera Data Science Workbench runs workloads. As a result, any utilities or plugins referenced in this manner will not work.

About this task

To enable FreeIPA support you must perform the following steps.

Procedure

1. Modify krb5.conf to remove references to external files

   You do not need to edit the krb5.conf file on the host operating system. Instead, make a copy of the file, and make your changes there. Points to note:
   include and includedir directives

   > While the include and includedir directives do typically reference external files, CDSW does account for those directives. Therefore, they are safe to use and no changes need to be made here.

   [plugins] directives

   > The [plugins] will always refer to a shared library on the host, which will not be available inside engines. An example of this is:

   ```
   [plugins]
   localauth = {
   module = sssd:/usr/lib64/sssd/modules/sssd_krb5_localauth_plugi
   n.so
   ```

```
    }
```

You must remove the entire [plugins] section from krb5.conf. It is not needed for the commands used by Cloudera Data Science Workbench.

PKINIT

The PKINIT option can also point to external files which will not exist by default in CDSW engines. An example of such a configuration is:

```
[libdefaults]
    EXAMPLE.COM = {
        pkinit_anchors = FILE:/usr/local/example.com.crt
    }
```

If the realm that uses PKINIT is not one that CDSW users will need a keytab for, it can be removed from the krb5.conf file. Otherwise, users will need to create a keytab outside of CDSW and upload it to the Settings Hadoop Authentication page.

default_ccache_name directive

A default_ccache_name using the Linux-specific KEYRING directive does not work with Cloudera Data Science Workbench. An example of this line is:

```
 default_ccache_name = KEYRING:persistent:%
```

You must remove this line from the krb5.conf file; the default value will work properly inside CDSW engines.

Note: If you require other configurations of krb5.conf that have not been discussed here, the recommended alternative method is to manually upload a keytab on the User SettingsHadoop Authentication page in the Cloudera Data Science Workbench web UI.

2. Copy the contents of krb5.conf to the Site Administration panel

a. Log into Cloudera Data Science Workbench as a site administrator.

b. Click Admin Security.

c. Copy the contents of the modified krb5.conf from Step 1 to the Kerberos Configuration text box. Click Update.

The contents of this text box will now be used as the krb5.conf file in engines launched for user workloads.