Cloudera Data Warehouse on premises 1.5.5

Using BI tools with Cloudera Data Warehouse

Date published: 2020-08-17
Date modified: 2025-11-08



https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2026. 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

Built-in BI clients and drivers 4
Download root certificates 4
Download root cert from Cloudera Data Warehouse UI 4
Download root cert using CDP CLI 4
Download Beeline tarball 5
Connect to Hive from Tableau 6
Connecting to a Trino Virtual Warehouse from a BI client 9
Download JDBC JAR 9
Upload additional JARs 10
Use Impala shell 11
Using the Trino command-line interface 13

Built-in BI clients and drivers in Cloudera Data Warehouse

You can connect BI tools and SQL clients such as Beeline, impala-shell, impyla, Tableau, and so on to Cloudera Data Warehouse and use them to explore and query data in the Data Lakehouse.

Cloudera Data Warehouse provides built-in downloadables such as Hive JDBC driver for connecting JDBC-compliant tools to the Virtual Warehouses, Beeline CLI, Impala JDBC and ODBC drivers for connecting to Impala Virtual Warehouses, and a JDBC driver for using with Unified Analytics. You can download these from the Resources and Downloads tile present on the **Overview** page.

Cloudera recommends that you use the latest versions of Beeline, JDBC or ODBC drivers, impala-shell, or Impyla to connect to Cloudera Data Warehouse. The versions of these BI client tools that are installed on the Cloudera Base on premises nodes could be old and may not have newer features that Cloudera Data Warehouse supports.

Downloading root certificates for Kubernetes environments from Cloudera Data Warehouse on premises

You can download the signed certificates for Hive in the Java KeyStore (JKS) format and for Impala in the PEM format from the Cloudera Data Warehouse user interface or using the Beta CDP CLI, and use them to connect to Impala or Hive Virtual Warehouses from the BI clients such as Impala-shell, Impyla, or Beeline.

Suppose your organization uses a custom Certificate Authority (CA) to sign security certificates for Embedded Container Service (ECS) or OpenShift Container Platform (OCP) environments. In that case, you can obtain the TLS certificates for the Kubernetes environment without depending on your Administrators (DWAdmin users).

Downloading root certificates from Cloudera Data Warehouse web UI

You can download the root certificates for establishing a secure connection from BI clients to a Virtual Warehouse from the Virtual Warehouse more options menu in Cloudera Data Warehouse on premises.

Procedure

1. Log in to the Cloudera Data Warehouse service.
2. Go to the Virtual Warehouses tab, locate the Virtual Warehouse you want to connect to, and click Download Kubernetes cluster certificates .
3. (Impala) On the **Kubernetes certificate** modal, click **Copy certificate**.

   This copies the TLS certificate to the clipboard. Save the copied content in a text file.

   (Hive) On the **Kubernetes certificate** modal, click Download truststore and note the Java truststore password either by clicking  or Copy password.

Downloading root certificates using CDP CLI

Using the CDP CLI, you can download the root certificates for establishing a secure connection from BI clients to a Virtual Warehouse in Cloudera Data Warehouse on premises

About this task

Note: This feature is available only with Beta CDP CLI.

Procedure

1. SSH in to the cluster host on which you have installed Beta CDP CLI and can access the Cloudera Data Services on premises cluster.
2. Run the respective commands to download the root certificates in PEM or JCEKS formats:

   PEM

   ```
   cdp dw get-k8s-cert-pem
   ```

   Java KeyStore (JKS)

   ```
   cdp dw get-k8s-cert-jks
   ```
3. Save the output to a file for future use.

Downloading the Beeline CLI tarball

Download the Beeline CLI tarball from Cloudera Data Warehouse to your local system and use the Beeline client to connect to a Hive Virtual Warehouse and run queries. The archive file contains all the dependent JARs and libraries that are required to run the Beeline script.

Before you begin

From the Cloudera Management Console user profile, note the Workload User Name and Workload Password.

Attention: Ensure that Java is installed on the node on which you want to download and use the Beeline CLI, and you have set the JAVA_HOME environment variable correctly while installing the JDK.

Procedure

1. Log in to the CDP web interface and navigate to the Cloudera Data Warehouse service.
2. In the **Overview** page of the Cloudera Data Warehouse, click See More in the Resources and Downloads tile.
3. Select Beeline CLI and click to download the file.
4. Save the apache-hive-beeline-x.x.xxxx.tar.gz file in your local system and extract the tarball.
5. In the Data Warehouse service **Overview** page, for the Virtual Warehouse you want to connect to the client, click and select Copy JDBC URL.
6. Paste the copied JDBC URL in a text file, to be used in later steps.

   ```
   jdbc:hive2://<your-virtual-warehouse>.<your-environment>.<dwx.company.com>/default;transportMode=http;httpPath=cliservice;ssl=true;retries=3
   ```
7. Open a terminal window and go to the folder where the tarball is extracted to start Beeline.
   bin/beeline

   This starts an interactive Beeline shell where you can connect to Hive and run SQL queries.
8. Run the connect command to connect to Hive using the JDBC URL that you copied earlier.

   ```
   beeline> !connect [***JDBC URL***]
   ```

Note: If the root certificate is untrusted, set the value of ssl to false.

```
Connecting to jdbc:hive2://<your-virtual-warehouse>.<your-environment>.<
dwx.company.com>/default;transportMode=http;httpPath=cliservice;ssl=true
;retries=3
```

9. Enter the Workload User Name and Workload Password when you are prompted for the user credentials.

```
Enter username for jdbc:hive2://<your-virtual-warehouse>.<your-environme
nt>.<dwx.company.com>/default: [***WORKLOAD USERNAME***]
Enter password for jdbc:hive2://<your-virtual-warehouse>.<your-environm
ent>.<dwx.company.com>/default: [***WORKLOAD PASSWORD***]
Connected to: Apache Hive (version 3.1.2000.7.0.2.2-24)
Driver: Hive JDBC (version 3.1.2000.7.0.2.2-24)
Transaction isolation: TRANSACTION_REPEATABLE_READ
```

10. To verify if you are connected to HiveServer2 on the Virtual Warehouse, run the following SQL command: SHOW TABLES;

```
INFO : Compiling command(queryId=hive_20200214014428_182d2b63-a510-421f-
8bbc-65a4ae24d1d6): show tables
INFO : Semantic Analysis Completed (retrial = false)
INFO : Completed compiling command(queryId=hive_20200214014428_182d2b63-
a510-421f-8bbc-65a4ae24d1d6); Time taken: 0.054 seconds
INFO : Executing command(queryId=hive_20200214014428_182d2b63-a510-421f
-8bbc-65a4ae24d1d6): show tables
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20200214014428_182d2b63-a
510-421f-8bbc-65a4ae24d1d6); Time taken: 0.018 seconds
INFO : OK
-----------

table_name
-----------
-----------
No rows selected (0.311 seconds)
```

Connecting to Hive Virtual Warehouses from Tableau

This topic describes how to connect to Tableau with Hive Virtual Warehouses on Cloudera Data Warehouse service.

About this task

Required role: DWUser

Before you begin

Before you can use Tableau with Hive Virtual Warehouses, you must have populated your Database Catalog with sample data when you create it. You must also create a Hive Virtual Warehouse, which is configured to connect to the Database Catalog that is populated with data.

Procedure

1. Download the latest version of the Hive ODBC driver from Cloudera Downloads page.
2. Install the driver on the local host where you intend to use Tableau Desktop.
3. Log in to the Cloudera Data Warehouse service as DWUser.

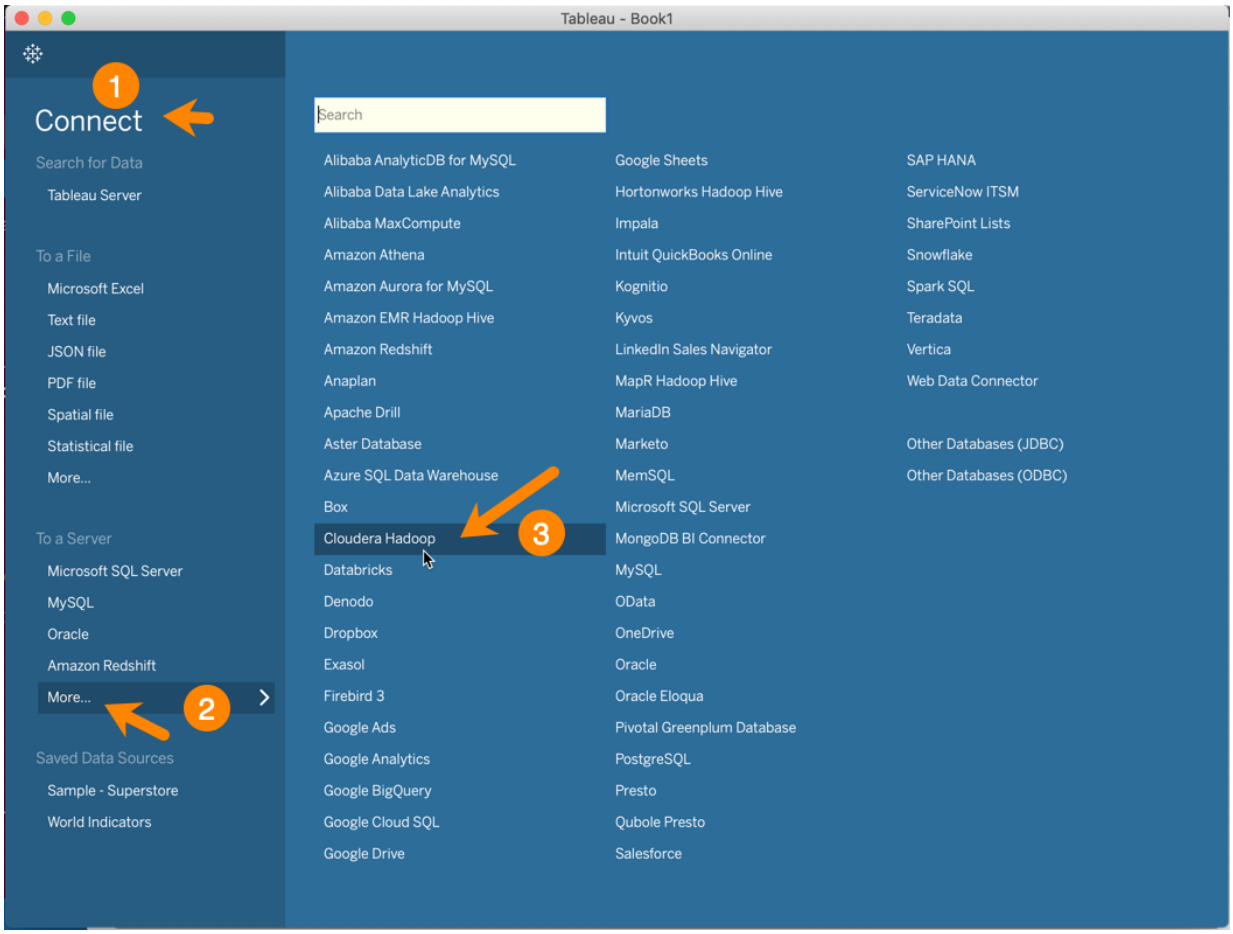
4. Go to the **Virtual Warehouses** tab, locate the Hive Virtual Warehouse you want to connect to, and click ⋮ Copy JDBC URL .

   This copies the JDBC URL to your system's clipboard.
5. Paste the copied JDBC URL into a text file. It should look similar to the following:

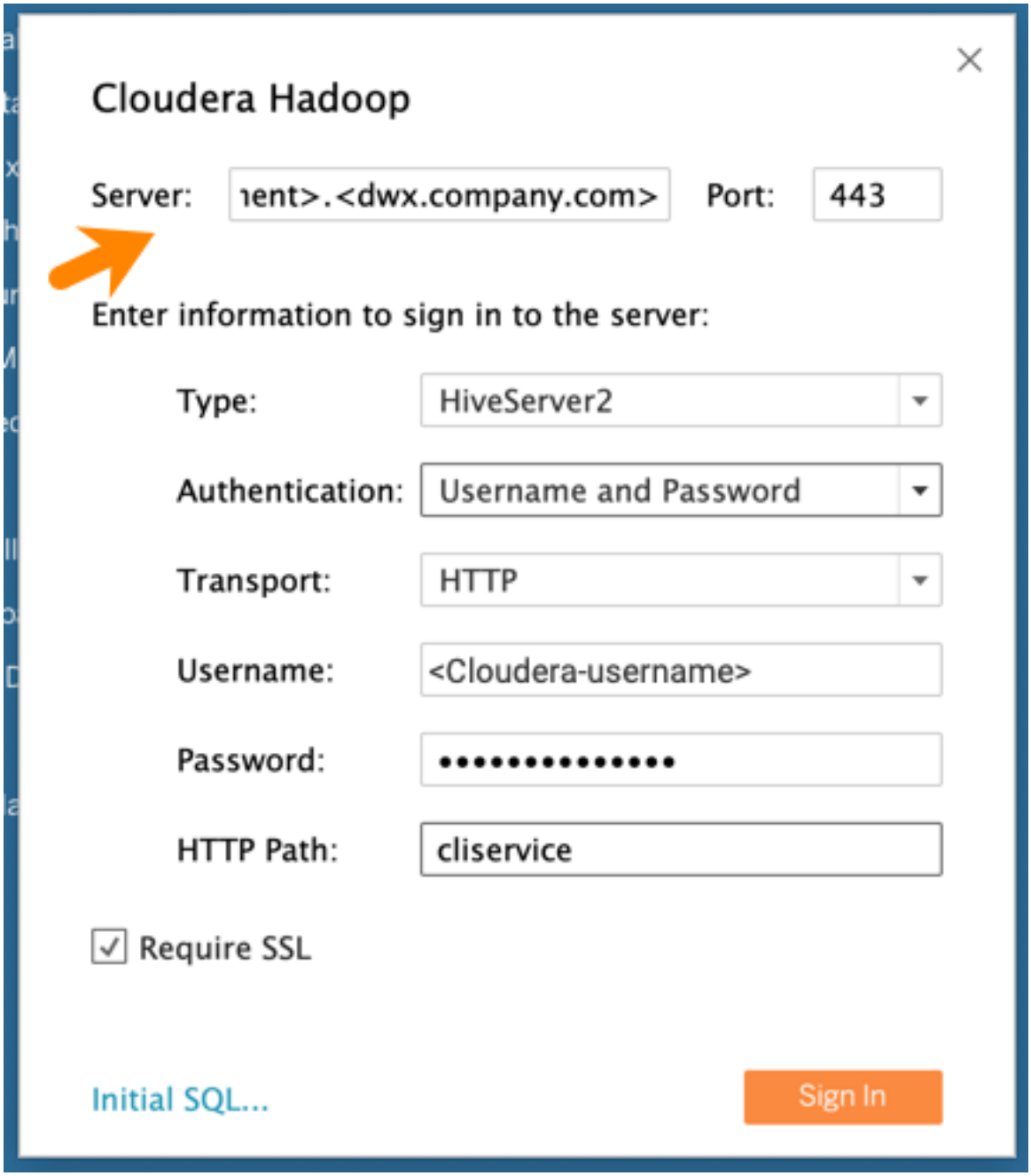
   ```
   jdbc:hive2://<your-virtual-warehouse>.<your-environment>.<dwx.company.com>/default;transportMode=http;httpPath=cliservice;ssl=true;retries=3
   ```

   Note: If the root certificate is untrusted, set the value of ssl to false.
6. From the text file where you just pasted the URL, copy the host name from the JDBC URL to your system's clipboard. For example, the host name in the URL is:

   ```
   <your-virtual-warehouse>.<your-environment>.<dwx.company.com>
   ```
7. Start Tableau and navigate to ConnectMore…Cloudera Hadoop :



   This launches the Cloudera Hadoop dialog box.
8. In the Tableau Cloudera Hadoop dialog box, paste the host name you copied to your clipboard in Step 7 into the Server field:



9. Then in the **Tableau Cloudera Hadoop** dialog box, set the following other options:
   - Port: 443
   - Type: HiveServer2
   - Authentication: Username and Password
   - Transport: HTTP
   - Username: Username you use to connect to the CDP Data Warehouse service.
   - Password: Password you use to connect to the CDP Data Warehouse service.
   - HTTP Path: cliservice
   - Require SSL: Make sure this is checked.
10. Click Sign In.

Related Information

Cloudera Hadoop connection option described in the Tableau documentation

Connecting to a Trino Virtual Warehouse from a BI client

You can connect BI tools and SQL clients to a Trino Virtual Warehouse to explore and query data. Cloudera Data Warehouse provides built-in downloadables such as Trino JDBC/ODBC driver for connecting JDBC/ODBC-compliant tools to the Virtual Warehouses.

About this task

Note: Trino is available as a Technical Preview offering starting from the Cloudera Data Warehouse on premises 1.5.5 SP1 release. It is available as a GA offering in the Cloudera Data Warehouse on premises 1.5.5 SP2 release.

Procedure

1. Log in to the Cloudera Data Warehouse service as DWUser.
2. In the **Overview** page of the Cloudera Data Warehouse service, click See More in the **Resources and Downloads** tile.
3. Click Trino JDBC/ODBC Driver to download the Trino JDBC and ODBC drivers zip file.
4. Provide the JAR file that you downloaded to your JDBC or ODBC client.

   On most clients, add the JAR file under the Libraries folder. Refer to your client documentation for information on the location to add the JAR file.
5. In the Cloudera Data Warehouse service **Overview** page, for the Trino Virtual Warehouse you want to connect to the client, click ⋮ and select Copy Trino JDBC URL.

   A URL is copied to your system clipboard in the following format:

   ```
   jdbc:trino://<your_virtual_warehouse>.<your_environment>.<dwx.company.com>:<port>
   ```
6. Configure your client BI tool to connect to the Virtual Warehouse using the copied JDBC URL.

Downloading a JDBC driver from Cloudera Data Warehouse

To use third-party BI tools, your client users need a JDBC JAR to connect your BI tool and the service. You learn how to download the JDBC JAR to give to your client, and general instructions about how to use the JDBC JAR.

Before you begin

Before you can use your BI tool with the Cloudera Data Warehouse service:

- You created a Database Catalog.

  You have the option to populate your Database Catalog with sample data when you create it.
- You created a Virtual Warehouse and configured it to connect to the Database Catalog.

Of course, to query tables in the Virtual Warehouse from a client, you must have populated the Virtual Warehouse with some data.

Procedure

1. Log in to the Cloudera Data Warehouse service as DWUser.

2. In the **Overview** page of the Cloudera Data Warehouse service, click See More in the Resources and Downloads tile.
3. Click Hive JDBC Jar to download the Apache Hive JDBC JAR file.
4. Provide the JAR file you downloaded to your JDBC client.

   On most clients, add the JAR file under the Libraries folder. Refer to your client documentation for information on the location to add the JAR file.
5. In the Cloudera Data Warehouse service **Overview** page, for the Virtual Warehouse you want to connect to the client, click ⋮ and select Copy JDBC URL.

   A URL is copied to your system clipboard in the following format:

   ```
   jdbc:hive2://<your_virtual_warehouse>.<your_environment>.<dwx.company.com>/default;transportMode=http;httpPath=cliservice;ssl=true;retries=3
   ```

   Note: If the root certificate is untrusted, set the value of ssl to false.
6. Paste the URL into a text editor and configure your client BI tool to connect to the Virtual Warehouse using the following portion of the URL, represents the server name of the Virtual Warehouse:

   ```
   <your_virtual_warehouse>.<your_environment>.<dwx.company.com>
   ```
7. In your JDBC client, set the following other options:

   Authentication: Username and Password

   Username: Username you use to connect to the Cloudera Data Warehouse service.

   Password: Password you use to connect to the Cloudera Data Warehouse service.

Uploading additional JARs to CDW

You add additional Java Archive (JAR) files to the Cloudera Data Warehouse Hive classpath that might be required to support dependency JARs, third-party Serde, or any Hive extensions.

About this task

- The JARs are added to the end of the Hive classpath and do not override the Hive JARs.
- Cloudera recommends that you do not use this procedure to add User-Defined Function (UDF) JARs. If you do, then you must restart HiveServer2 or reload the UDF. For more information about reloading functions, see the Hive Data Definition Language (DDL) manual.

Before you begin

You have the EnvironmentAdmin role permissions to upload the JAR to your object storage.

Procedure

1. Build the archive file.

   The archive file can be either a .jar file or a tar.gz file. For a tar.gz archive file, only JARs present in the top level are considered.

   For example, if the tar.gz file contains these files — test1.jar, test2.jar, and deps/test3.jar, only test1.jar and test2.jar are considered; deps/test3.jar is excluded.

   Note: There is no defined priority to use a particular file. If there are multiple files with the same name or same class in multiple jars, any file can be in effect.

2. Upload the archive file to the Hive Virtual Warehouse on Cloudera Data Warehouse object storage, such as HDFS.
3. Log in to the Cloudera Data Warehouse service and from the Overview page, locate the Hive Virtual Warehouse that uses the bucket or container where you placed the archive file, and click ⋮ and select Edit.
4. In the **Virtual Warehouse Details** page, click Configurations Hiveserver2 .
5. From the Configuration files drop-down list, select env.
6. Search for CDW_HIVE_AUX_JARS_PATH and add the archive file to the environment variable.

   If you add a directory, the .jar or tar.gz files within the directory are copied and extracted. For a tar.gz file, only the JARs present in the top level are copied.

   Consider the following JAR path - /common-jars/common-jars.tar.gz:/common-jars/single-jar.jar:/serde-specific-jar/serde.jar. In this example, common-jars.tar.gz is extracted and single-jar.jar and serde.jar files are copied.
7. Repeat the previous step and add the archive file or directory for Query coordinator and Query executor.

   If the CDW_HIVE_AUX_JARS_PATH environment variable is not present, click + and add the following custom configuration:

```
CDW_HIVE_AUX_JARS_PATH=[***VALUE***]
```

Results

On applying the configuration changes, Hive Virtual Warehouse restarts and the archive files are available and added to the end of the Hive classpath.

Related Information

Hive Data Definition Language manual

Using Impala shell

This topic describes how to download and install the Impala shell to query Impala Virtual Warehouses in the Cloudera Data Warehouse service.

About this task

Required role: DWUser

You can install the Impala shell on a local computer and use it as a client to connect with an Impala Virtual Warehouse instance. If you are connecting from a node that is already a part of a CDH or Cloudera cluster, you already have Impala shell and do not need to install it.

Before you begin

Make sure that you have the latest stable version of Python 3.0 or above and a pip installer associated with that build of Python installed on the computer where you want to run the Impala shell.



Note: The following procedure cannot be used on a Windows computer.

Procedure

1. Open a terminal window on the computer where you want to install the Impala shell, and run the following pip installer command to install the shell on your local computer:

```
pip install impala-shell
```

Tip: To ensure you're using the correct version of Impala shell, always download it directly from the UI. To do this, go to your Impala Virtual Warehouse, click ⋮ Copy Impala shell Download Command .

After you run this command, if your installation was successful, you receive success messages that are similar to the following messages:

```
Successfully built impala-shell bitarray prettytable sasl sqlparse thrift
 thrift-sasl
Installing collected packages: bitarray, prettytable, six, sasl, sqlparse,
 thrift, thrift-sasl, impala-shell
Successfully installed bitarray-1.0.1 impala-shell-3.3.0.dev20190730101121
prettytable-0.7.1 sasl-0.2.1 six-1.11.0 sqlparse-0.1.19 thrift-0.11.0 thri
ft-sasl-0.2.1
```

2. To confirm that the Impala shell has installed correctly, run the following command which displays the help for the tool:

```
impala-shell --help
```

If the tool help displays, the Impala shell is installed properly on your computer.

3. To connect to your Impala Virtual Warehouse instance using this installation of Impala shell:
   a) Log in to the CDP web interface and navigate to the Data Warehouse service.
   b) Go to the **Virtual Warehouses** tab, locate the Impala Virtual Warehouse you want to connect to, and select ⋮ Copy Impala shell command .

   This copies the shell command to your computer's clipboard. This command enables you to connect to the Virtual Warehouse instance in Cloudera Data Warehouse service using the Impala shell that is installed on your local computer.
4. In the terminal window on your local computer, at the command prompt, paste the command you just copied from your clipboard. The command might look something like this:

```
impala-shell --protocol='hs2-http' --ssl -i "tpcds-impala.your_company.c
om:443"
```

5. Press return and you are connected to the Impala Virtual Warehouse instance. A "Starting Impala Shell..." message similar to the following displays:

```
Starting Impala Shell without Kerberos authentication
SSL is enabled. Impala server certificates will NOT be verified (set --ca_
cert to change)
Warning: --connect_timeout_ms is currently ignored with HTTP transport.
Opened TCP connection to
tpcds-impala.your_company.com:443
Connected to tpcds-impala.your_company.com:443
Server version: impalad version 3.4.0-SNAPSHOT RELEASE (build
d133a7140a3b97508ec77b1c73bb4f55f5dcb928)
************************************************************************
*******
Welcome to the Impala shell.
(Impala Shell v3.3.0-SNAPSHOT (a509cff) built on Mon Jul 29 18:37:09 PDT
 2019)
Every command must be terminated by a ';'.
************************************************************************
********
```

6. Run the following SQL command to confirm that you are connected properly to the Impala Virtual Warehouse instance:

```
SHOW DATABASES;
```

If you are connected properly, this SQL command should return the following type of information:

```
Query: show databases
+------------------------+----------------------------------------------+
| name                   | comment                                      |
+------------------------+----------------------------------------------+
| _impala_builtins       | System database for Impala builtin functions |

| default                | Default Hive database                        |
                  |                          |
                |
+------------------------+----------------------------------------------+
Fetched 2 row(s) in 0.30s
```

If you see a listing of databases similar to the above example, your installation is successful and you can use the shell to query the Impala Virtual Warehouse instance from your local computer.

Related Information

Download the latest stable version of Python 2
Configuring client access to Impala
Impala shell tool
Impala shell configuration options
Impala shell configuration file

Using the Trino command-line interface

The Trino command-line interface (CLI) provides a terminal-based, interactive shell for running queries. The CLI is a self-executing JAR file and acts like a normal UNIX executable.

Procedure

1. Log in to the Cloudera web interface and navigate to the Cloudera Data Warehouse service.
2. In the **Overview** page of Cloudera Data Warehouse, click See More in the **Resources and Downloads** tile.
3. Select Trino CLI and click to download the trino-cli-executable.jar file.
4. Save the trino-cli-executable.jar file and rename it to trino.
5. Make trino as executable file by running the following command:
   chmod +x trino.jar
6. In the Cloudera Data Warehouse service **Overview** page, for the Virtual Warehouse you want to connect to the client, click and select Copy Trino URL.
7. Paste the copied Trino URL in a text file, to be used in later steps.
8. Open a terminal window to start Trino CLI. You can run the following command to check the version of the CLI:

```
./trino --version
```

Run the CLI with --help or -h to see all available options.

9. To start the Trino CLI in an interactive mode, specify the URL of the Trino coordinator that you copied earlier.

```
./trino http://trino.example.com:8080
```

This starts an interactive shell where you can connect to Trino and run SQL queries.

Related Information

Trino CLI