Cloudera Runtime 7.2.18

Integrating Apache Hive with BI tools

Date published: 2019-08-21
Date modified: 2024-03-20

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

Connecting Hive to BI tools using a JDBC/ODBC driver in Data Hub

To query, analyze, and visualize data stored in CDP, you use drivers provided by Cloudera to connect Apache Hive to Business Intelligence (BI) tools.

About this task
How you connect to Hive depends on a number of factors: the location of Hive inside or outside the cluster, the HiveServer deployment, the type of transport, transport-layer security, and authentication. HiveServer is the server interface that enables remote clients to run queries against Hive and retrieve the results using a JDBC or ODBC connection.

Before you begin
- Choose a Hive authorization model.
- Configure authenticated users for querying Hive through JDBC or ODBC driver. For example, set up a Ranger policy.

Getting the JDBC or ODBC driver in Data Hub

You need to download the Cloudera Hive JDBC or ODBC driver to your Data Hub cluster and provide it to your client. In the BI tool on the client's cluster, you need to provide the HiveServer (HS2) host name portion of the JDBC URL. The BI tool requires this information to connect to Hive in CDP.

Procedure
1. Download the JDBC or ODBC driver from the Cloudera Downloads page.
2. Follow instructions in the documentation to install the driver on the same downloads page.

Configuring the BI tool

After you download an ODBC or JDBC driver and give it to your client, you need to provide the information that the BI tool requires to connect to Hive in CDP.
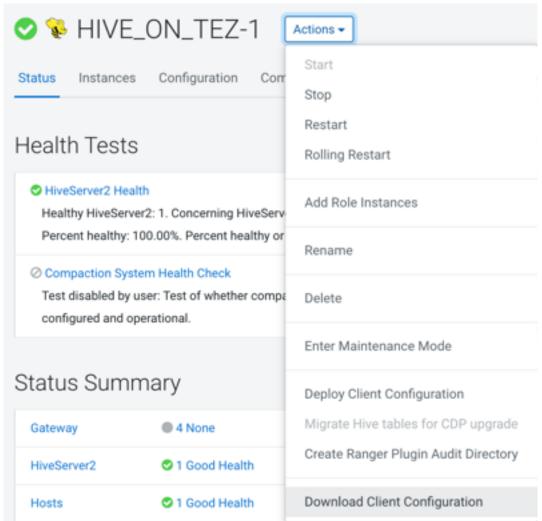
Before you begin
You have downloaded the JDBC or ODBC driver.

Procedure
1. Depending on the type of driver you obtain, proceed as follows:
   - ODBC driver: follow instructions on the ODBC driver download site, and skip the rest of the steps in this procedure.
   - JDBC driver: add the driver to the classpath of your JDBC client. For example, check the client documentation about where to put the driver.
2. In Cloudera Manager (CM), click Clusters Hive on Tez to go to the Hive on Tez service page.

3. From the Hive on Tez service page, click Actions and select Download Client Configuration.

4. Unpack hive_on_tez-clientconfig.zip, open the beeline-site.xml file, and copy the value of beeline.hs2.jdbc.url.hiv
e_on_tez. This value is the JDBC URL.
For example

```
jdbc:hive2://my_hiveserver.com:2181/;serviceDiscoveryMode=zooKeeper; \
zooKeeperNamespace=hiveserver2
```

5. In the BI tool, configure the JDBC connection using the JDBC URL and driver class name, com.cloudera.hive.jd
bc.HS2Driver.