

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

© 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

Connecting Hive to BI tools using a JDBC/ODBC driver in Data Hub.....	4
Getting the JDBC or ODBC driver in Data Hub.....	4
Configuring the BI tool.....	4

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.

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

The screenshot shows the Cloudera Hive on Tez service page for a service named 'HIVE_ON_TEZ-1'. The page has tabs for 'Status', 'Instances', 'Configuration', and 'Components'. The 'Status' tab is active. Under 'Health Tests', there are two tests: 'HiveServer2 Health' (Healthy) and 'Compaction System Health Check' (disabled). The 'Status Summary' section shows: Gateway (4 None), HiveServer2 (1 Good Health), and Hosts (1 Good Health). An 'Actions' dropdown menu is open, listing various actions, with 'Download Client Configuration' highlighted at the bottom.

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

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

- In the BI tool, configure the JDBC connection using the JDBC URL and driver class name, `com.cloudera.hive.jdbc.HS2Driver`.