Cloudera Runtime 7.1.7

Starting Apache Hive

Date published: 2019-08-21
Date modified: 2021-08-05

https://docs.cloudera.com/

Legal Notice

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.

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

Starting Hive on an insecure cluster

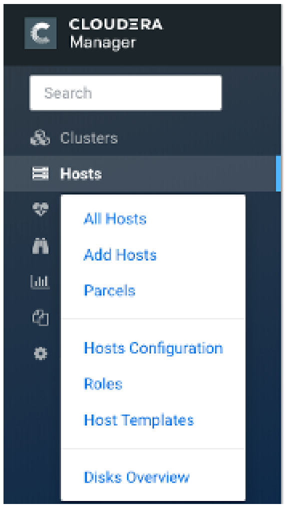
If you want to use Apache Hive for a quick test, you can do so using default Hive default authorization mode, assuming you are on an insecure cluster (no Kerberos or Ranger policies). In default authorization mode, only user hive can access Hive. Steps to start the Hive shell, not to be mistaken with the Hive CLI that CDP does not support, include how to log into a cluster.

About this task

From the cluster command line, you type hive on the command line of your cluster to start the Hive shell. In the background, Beeline launches the Hive shell.

Procedure

1. In Cloudera Manager, click Hosts All Hosts .



2. Make a note of the IP address or host name of a node in your cluster, for example myhost-vpc.cloudera.com.
3. Use ssh to log into the cluster.
   For example:

```
ssh myhost-vpc.cloudera.com
```

4. Type hive to start Hive from the command line.
5. Enter Hive queries.

```
SHOW DATABASES;
```

```
CREATE TABLE students (name VARCHAR(64), age INT, gpa DECIMAL(3,2));
```

```
INSERT INTO TABLE students VALUES ('fred flintstone', 35, 1.28), ('barney
 rubble', 32, 2.32);
```

Starting Hive using a password

You start the Hive shell using a Beeline command to query Hive as an end user authorized by Apache Ranger. As administrator, you set up the end user in the operating system and in Ranger.

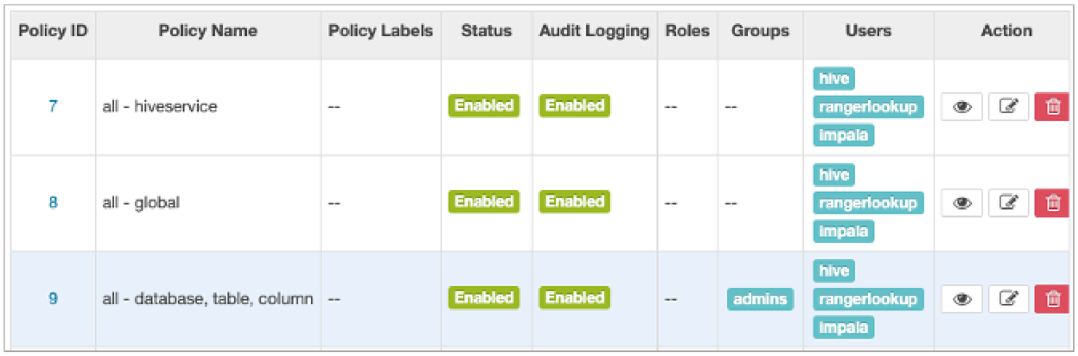
About this task

Before starting Hive for the first time, check that you are covered by Ranger policies required for basic operations as shown in the following steps. All users need to use the default database, perform operations such as listing database names, and query the information schema. The preloaded default database tables columns and information_schema database Ranger policies cover group public (all users). If these policies are disabled, you cannot use the default database, perform basic operations such as listing database names, or query the information schema. For example, if the default database tables columns policy is disabled, the following error appears if you try to use the default database:

```
hive> USE default;
Error: Error while compiling statement: FAILED: HiveAccessControlException
Permission denied: user [hive] does not have [USE] privilege on [default]
```

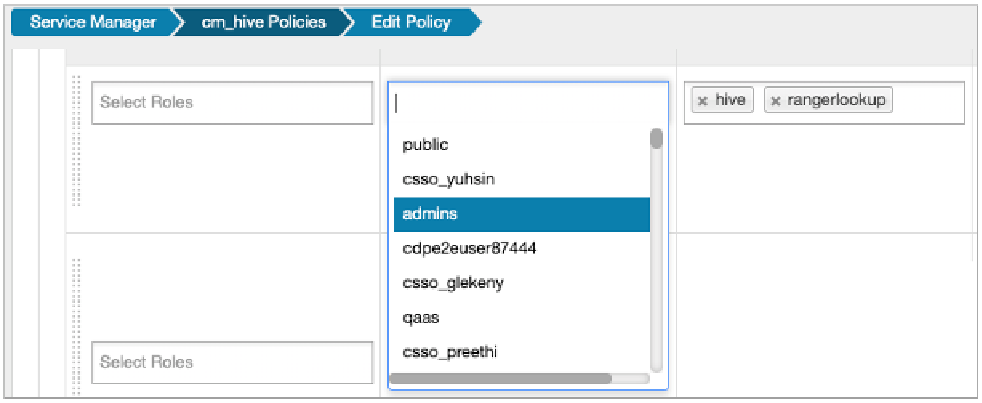
Procedure

1. Access the Ranger Console: click the Ranger Admin web UI link, enter your user name and password, then click Sign In.
2. On the far right, click Ranger Hadoop SQL , and in Allow Conditions, edit all - database, table, column.



Policy ID	Policy Name	Policy Labels	Status	Audit Logging	Roles	Groups	Users	Action
7	all - hiveservice	--	Enabled	Enabled	--	--	hive rangerlookup impala	
8	all - global	--	Enabled	Enabled	--	--	hive rangerlookup impala	
9	all - database, table, column	--	Enabled	Enabled	--	admins	hive rangerlookup impala	

3. Add your user or group name to Hive policies to grant full access to Hive.
   For example, add the admins group name to the list of groups that can access Hive.



Service Manager > cm_hive Policies > Edit Policy

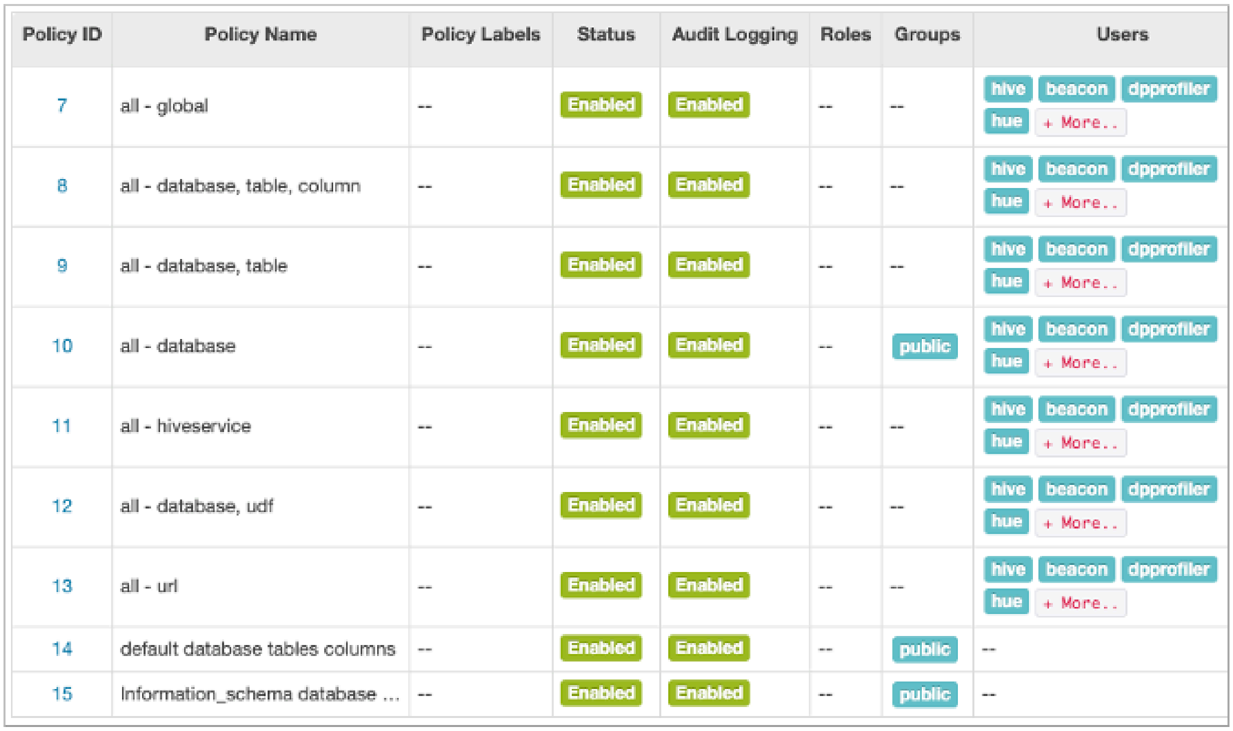
Select Roles

public
csso_yuhsin
admins
cdpe2euser87444
csso_glekeny
qaas
csso_preethi

× hive × rangerlookup

Select Roles

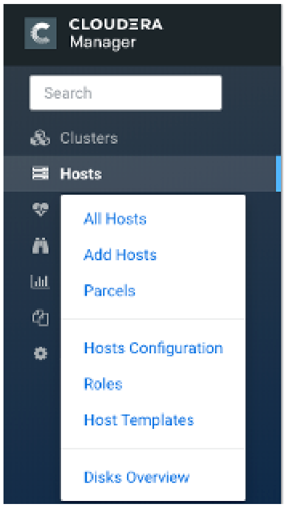
4. Check that the preloaded default database tables columns and information_schema database policies are enabled for group public.



Policy ID	Policy Name	Policy Labels	Status	Audit Logging	Roles	Groups	Users
7	all - global	--	Enabled	Enabled	--	--	hive beacon dpprofiler hue + More..
8	all - database, table, column	--	Enabled	Enabled	--	--	hive beacon dpprofiler hue + More..
9	all - database, table	--	Enabled	Enabled	--	--	hive beacon dpprofiler hue + More..
10	all - database	--	Enabled	Enabled	--	public	hive beacon dpprofiler hue + More..
11	all - hiveservice	--	Enabled	Enabled	--	--	hive beacon dpprofiler hue + More..
12	all - database, udf	--	Enabled	Enabled	--	--	hive beacon dpprofiler hue + More..
13	all - url	--	Enabled	Enabled	--	--	hive beacon dpprofiler hue + More..
14	default database tables columns	--	Enabled	Enabled	--	public	--
15	Information_schema database ...	--	Enabled	Enabled	--	public	--

5. In Environments *YOUR ENVIRONMENT YOUR CLUSTER* , click the CM-URL.

6. In Cloudera Manager, click Hosts All Hosts .



7. Make a note of the IP address or host name of a node in your cluster, for example myhost-vpc.cloudera.com.

Tip: In a Kerberized cluster, ensure the core-site.xml file is correctly configured on all nodes from which you plan to connect to beeline. Include the following property in the core-site.xml:

```
<property>
  <name>hadoop.security.authentication</name>
  <value>kerberos</value>
</property>
```

8. Use ssh to log into the cluster.
For example:

```
ssh myhost-vpc.cloudera.com
```

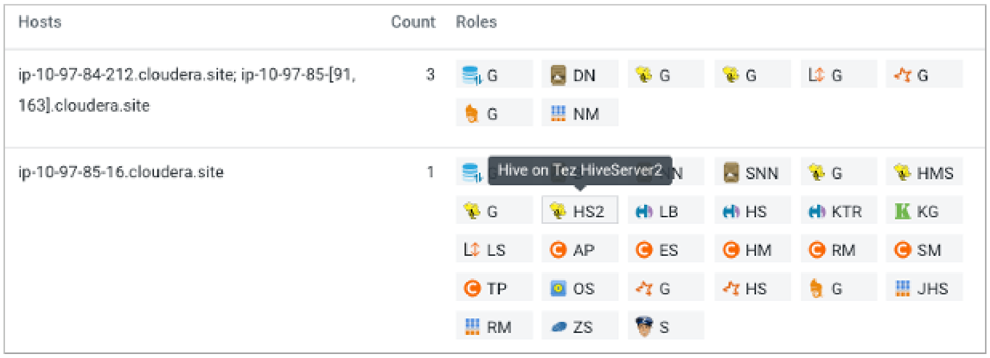
You can get help about starting the Hive shell. On the command line, type

```
hive -h
```

Output is:

```
Connect using simple authentication to HiveServer2 on localhost:10000
beeline -u jdbc:hive2://localhost:10000 username password

Connect using simple authentication to HiveServer2 on hs.local:10000 us
ing -n for username and -p for password
beeline -n username -p password -u jdbc:hive2://hs2.local:10012

Connect using Kerberos authentication with hive/localhost@mydomain.com as
 HiveServer2 principal
beeline -u "jdbc:hive2://hs2.local:10013/default;principal=hive/localho
st@mydomain.com"

Connect using SSL connection to HiveServer2 on localhost at 10000
beeline "jdbc:hive2://localhost:10000/default;ssl=true;sslTrustStore=/usr/
local/truststore;trustStorePassword=mytruststorepassword"

Connect using LDAP authentication
beeline -u jdbc:hive2://hs2.local:10013/default <ldap-username> <ldap-pas
sword>
```

9. Using the fully qualified domain name or IP address a node in your cluster from Cloudera Manager Hosts Role(s) and looking through the list of roles to find HS2 (Hive on Tez HiveServer2). For example:



Hosts	Count	Roles
ip-10-97-84-212.cloudera.site; ip-10-97-85-[91, 163].cloudera.site	3	G DN G G G G G NM
ip-10-97-85-16.cloudera.site	1	Hive on Tez HiveServer2 N SNN G HMS G HS2 LB HS KTR KG LS AP ES HM RM SM TP OS G HS G JHS RM ZS S

This node has the HiveServer role, so you can use the name or IP address in Beeline.

10. Start the Hive shell.

- Use your user name if your cluster security is set up.
- Use the user name hive and no password.

Substitute the name or IP address of your HiveServer host for 10.65.13.98.

Simple authentiation:

```
beeline -u jdbc:hive2://10.65.13.98:10000 -n <your user name> -p
```

Kerberos:

```
beeline -u "jdbc:hive2://10.65.13.98:10000/default;principal=hive/_HOST@
CLOUDERA.SITE"
```

11. Enter Hive queries.

```
SHOW DATABASES;
```

```
CREATE TABLE students (name VARCHAR(64), age INT, gpa DECIMAL(3,2));
```

```
INSERT INTO TABLE students VALUES ('fred flintstone', 35, 1.28), ('barney
 rubble', 32, 2.32);
```

Related Information

Configure a Resource-based Policy: Hive

Running a Hive command

You can run most Hive commands that push configuration variables to Hive SQL scripts from the command line of a node in your cluster. You use a keyword and options to launch these commands in Beeline.

About this task

Hive supports running Hive commands from the command line. The command you enter launches Beeline in the background. The -e flag followed by a Hive set command lists system variables.

Procedure

On the command line of a node in your CDP cluster, enter the hive command to send configuration properties to standard output.

```
> hive -e set
```

Supported commands appear. All obsolete Hive CLI commands are supported by Beeline except set key=value commands that configure Hive Metastore.

The output includes the system variable settings:

```
+----------------------------------------------------------------+
|                            set                                 |
+----------------------------------------------------------------+
| _hive.hdfs.session.path=/tmp/hive/hive/91ecb...00a             |
| _hive.local.session.path=/tmp/hive/91ecb...00a                 |
                       |
 ...
```

Converting Hive CLI scripts to Beeline

If you have legacy scripts that run Hive queries from edge nodes using the Hive CLI, you must solve potential incompatibilities with variable substitution in these scripts. CDP supports Beeline instead of Hive CLI. You can use Beeline to run legacy scripts with a few caveats.

About this task

In this task, you resolve incompatibilities in legacy Hive CLI scripts and Beeline:

- Configuration variables
    - Problem: You cannot refer to configuration parameters in scripts using the hiveconf namespace unless allowed.
    - Solution: You include the parameter in the HiveServer allowlist (whitelist).
- Namespace problems
    - Problem: Beeline does not support the system and env namespaces for variables.
    - Solution: You remove these namespace references from scripts using a conversion technique described in this task.

Procedure

1. Create a conversion script named env_to_hivevar.sh that removes env references in your SQL scripts.

```
#!/usr/bin/env bash

CMD_LINE=""

#Blank conversion of all env scoped values
for I in `env`; do
  CMD_LINE="$CMD_LINE --hivevar env:${I} "
done
echo ${CMD_LINE}
```

2. On the command line of a node in your cluster, define and export a variable named HIVEVAR, for example, and set it to run the conversion script.

```
export HIVEVAR=`./env_to_hivevar.sh`
```

3. Define and export variables to hold a few variables for testing the conversion.

```
export LOC_TIME_ZONE="US/EASTERN"
export MY_TEST_VAR="TODAY"
```

4. On the command line of a cluster node, test the conversion: Execute a command that references HIVEVAR to parse a SQL statement, remove the incompatible env namespace, and run the remaining SQL.

```
hive ${HIVEVAR} -e 'select "${env:LOC_TIME_ZONE}";'
```

```
+-------------+
|     _c0     |
+-------------+
| US/EASTERN  |
+-------------+
```

5. Create a text file named init_var.sql to simulate a legacy script that sets two configuration parameters, one in the problematic env namespace.

   ```
   set mylocal.test.var=hello;
   set mylocal.test.env.var=${env:MY_TEST_VAR};
   ```

6. Include these configuration parameters in the allowlist: In Cloudera Manager, go to Clusters HIVE_ON_TEZ-1 Configuration , and search for hive-site.
7. In HiveServer2 Advanced Configuration Snippet (Safety Valve) for hive-site.xml, add the property key: hive.security.authorization.sqlstd.confwhitelist.append.
8. Provide the property value, or values, to allowlist, for example: mylocal\..*|junk.

   This action appends mylocal.test.var and mylocal.test.env.var parameters to the allowlist.
9. Save configuration changes, and restart any components as required.
10. Run a command that references HIVEVAR to parse a SQL script, removes the incompatible env namespace, and executes the remaining SQL, including the whitelisted configuration parameters identified by hiveconf:.

    ```
    hive -i init_var.sql ${HIVEVAR} -e 'select "${hiveconf:mylocal.test.var}
    ","${hiveconf:mylocal.test.env.var}";'
    ```

    ```
    +--------+--------+
    |  _c0   |  _c1   |
    +--------+--------+
    | hello  | TODAY  |
    +--------+--------+
    ```

Related Information

Custom Configuration (about Cloudera Manager Safety Valve)

Example of using the Cloudera Manager Safety Valve