

## Task-4: Install Hive

- **Step 1: Download Hive tar.**

*Command: wget http://archive.apache.org/dist/hive/hive-2.1.0/apache-hive-2.1.0-bin.tar.gz*

- **Step 2: Extract the tar file.**

*Command: tar -xzf apache-hive-2.1.0-bin.tar.gz*

*Command: ls*

- **Step 3: Edit the “.bashrc” file to update the environment variables for user.**

*Command: sudo gedit .bashrc*

Add the following at the end of the file:

*# Set HIVE\_HOME*

*export HIVE\_HOME=/home/mks/apache-hive-2.1.0-bin*

*export PATH=\$PATH:/home/mks/apache-hive-2.1.0-bin/bin*

Run below command to make the changes work in same terminal.

*Command: source .bashrc*

- **Step 4: Check hive version.**

*Command: hive --version*

- **Step 5: Create Hive directories within HDFS. The directory ‘warehouse’ is the location to store the table or data related to hive.**

*Command:*

*hdfs dfs -mkdir -p /user/hive/warehouse*

*hdfs dfs -mkdir /tmp*

- **Step 6: Set read/write permissions for table.**

*Command:*

*In this command, we are giving write permission to the group:*

*hdfs dfs -chmod g+w /user/hive/warehouse*

*hdfs dfs -chmod g+w /tmp*

- **Step 7: Set Hadoop path in hive-env.sh**

*Command: cd apache-hive-2.1.0-bin/*

*Command: gedit conf/hive-env.sh*

*Copy following lines*

*export HADOOP\_HOME=/home/mks/hadoop-2.7.3*

*export HADOOP\_HEAPSIZE=512*

*export HIVE\_CONF\_DIR=/home/mks/apache-hive-2.1.0-bin/conf*

*Command: gedit conf/hive-env.sh*

- **Step 8: Edit hive-site.xml**

*Command: gedit conf/hive-site.xml*

*Copy following lines*

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?><!--
Licensed to the Apache Software Foundation (ASF) under one or more
contributor license agreements. See the NOTICE file distributed with
this work for additional information regarding copyright ownership.
The ASF licenses this file to You under the Apache License, Version
2.0
(the "License"); you may not use this file except in compliance with
the License. You may obtain a copy of the License at
```

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

See the License for the specific language governing permissions and limitations under the License.

-->

```
<configuration>
```

```
<property>
```

```
<name>javax.jdo.option.ConnectionURL</name>
```

```
<value>jdbc:derby;;databaseName=/home/edureka/apache-hive-2.1.0-
bin/metastore_db;create=true</value>
```

```
<description>
```

JDBC connect string for a JDBC metastore.

To use SSL to encrypt/authenticate the connection, provide database-specific SSL flag in the connection URL.

For example, jdbc:postgresql://myhost/db?ssl=true for postgres database.

```
</description>
```

```
</property>
```

```
<property>
```

```

<name>hive.metastore.warehouse.dir</name>
<value>/user/hive/warehouse</value>
<description>location of default database for the
warehouse</description>
</property>
<property>
<name>hive.metastore.uris</name>
<value/>
<description>Thrift URI for the remote metastore. Used by metastore
client to connect to remote metastore.</description>
</property>
<property>
<name>javax.jdo.option.ConnectionDriverName</name>
<value>org.apache.derby.jdbc.EmbeddedDriver</value>
<description>Driver class name for a JDBC metastore</description>
</property>
<property>
<name>javax.jdo.PersistenceManagerFactoryClass</name>
<value>org.datanucleus.api.jdo.JDOPersistenceManagerFactory</value>
<description>class implementing the jdo persistence</description>
</property>
</configuration>

```

- **Step 9: By default, Hive uses Derby database. Initialize Derby database.**

*Command: bin/schematool -initSchema -dbType derby*

```

edureka@localhost:~$ cd apache-hive-2.1.0-bin/
edureka@localhost:~/apache-hive-2.1.0-bin$ bin/schematool -initSchema -d
bType derby
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/edureka/apache-hive-2.1.0-bin/li
b/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/edureka/hadoop-2.7.3/share/hadoo
p/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder
.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an expl
anation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFa
ctory]
Metastore connection URL:          jdbc:derby:;databaseName=/home/edureka/
apache-hive-2.1.0-bin/metastore_db;create=true
Metastore Connection Driver :      org.apache.derby.jdbc.EmbeddedDriver
Metastore connection User:         APP
Starting metastore schema initialization to 2.1.0
Initialization script hive-schema-2.1.0.derby.sql
Initialization script completed
schemaTool completed
edureka@localhost:~/apache-hive-2.1.0-bin$

```

- **Step 10: Launch Hive.**

Command: *hive*

Following should be the output.

```
edureka@localhost:~/apache-hive-2.1.0-bin$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/edureka/apache-hive-2.1.0-bin/lib/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/edureka/hadoop-2.7.3/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/home/edureka/apache-hive-2.1.0-bin/lib/hive-common-2.1.0.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive>
```

*If you are getting any error related to java then that may be due to incompatible java version with current hive. Apache hive2.0 or 3.0 does not work with jdk-11.*

*In that case it is suggested to remove jdk-11 and install jdk-8.*

- Command to remove jdk: *sudo rm -r path-to-jdk*
  - How to know the path?: *echo \$JAVA\_HOME*
  - To install java: *sudo apt install openjdk-8-jdk openjdk-8-jre*
  - Update the JAVA path at two places: *.bashrc* and *hadoop-env.sh*
- 
- **Step 11: Run few queries in Hive shell.**  
Command: *show databases;*  
  
Command: *create table employee (id string, name string, dept string);*  
  
Command: *show tables;*
  - **Step 12: To exit from Hive.**  
Command: *exit;*