Manual to Install Apache HIVE in Hadoop

1. login (or ssh) through hadoop user to the hadoop server.

2. Sit in the home directory:

cd ~

3. Download the HIVE file:

wget https://dlcdn.apache.org/hive/hive-4.0.0/apache-hive-4.0.0-bin.tar.gz

4. Untar the HIVE file:

tar –xvzf apache-hive-4.0.0-bin.tar.gz

5. Open the bashrc file

sudo vi ~/.bashrc

6. Add the lines listed below into the bashrc file, and save the file:

SET HIVE HOME

export HIVE_HOME=/home/hadoop/apache-hive-4.0.0-bin

export PATH=$PATH:/home/hadoop/apache-hive-4.0.0-bin/bin

export CLASSPATH=$CLASSPATH:$HADOOP_HOME/lib/*:$HIVE_HOME/lib/*

```
  GNU nano 4.8                                                    /home/hadoop/.bashrc

# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi


export HADOOP_HOME=/usr/local/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"


# SET HIVE HOME

export HIVE_HOME=/home/hadoop/apache-hive-4.0.0-bin
export PATH=$PATH:/home/hadoop/apache-hive-4.0.0-bin/bin
export CLASSPATH=$CLASSPATH:$HADOOP_HOME/lib/*:$HIVE_HOME/lib/*
```

7. Activate the environment variables

source ~/.bashrc

8. Go to the conf directory inside the hive home directory

cd $HIVE_HOME/conf

9. Copy the file hive-default.xml.template to hive-site.xml

cp hive-default.xml.template hive-site.xml

10. Edit hive-site.xml file:

vi hive-site.xml

and do the following:

a. Replace all occurrences of ${system:java.io.tmpdir} to /tmp/hive

This is the location Hive stores all it's temporary files.

b. Replace all occurrences of ${system:user.name} to username, the username should be the one you log in with, i.e. hadoop

11. Create the HIVE data warehouse directory on HDFS, and the temporary tmp directory

hdfs dfs -mkdir /user/hive/warehouse

hdfs dfs -mkdir /user/tmp

12. Give necessary permissions the directories.

hdfs dfs -chmod g+w /user/tmp

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

13. Hive uses an RDBMS like Derby for efficient management, retrieval, and updating of metadata, which is essential for query planning, optimization, transaction management, concurrency control, and maintaining data integrity.

Now initialize the derby database:

cd $HIVE_HOME

schematool -initSchema -dbType derby

```
hadoop@hadoop-iims:~/apache-hive-4.0.0-bin/conf$ schematool -initSchema -dbType derby
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-4.0.0-bin/lib/log4j-slf4j-impl-2.18.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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]
Initializing the schema to: 4.0.0
Metastore connection URL:        jdbc:derby:;databaseName=metastore_db;create=true
Metastore connection Driver :    org.apache.derby.jdbc.EmbeddedDriver
Metastore connection User:       APP
Starting metastore schema initialization to 4.0.0
Initialization script hive-schema-4.0.0.derby.sql
```

```
Initialization script completed
hadoop@hadoop-iims:~/apache-hive-4.0.0-bin/conf$
```

14. Go to $HIVE_HOME

cd $HIVE_HOME

15. Check if hive is installed correctly.

bin/hive --version

```
hadoop@hadoop-iims:~/apache-hive-4.0.0-bin$ bin/hive --version
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-4.0.0-bin/lib/log4j-slf4j-impl-2.18.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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]
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-4.0.0-bin/lib/log4j-slf4j-impl-2.18.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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]
Hive 4.0.0
Git git://MacBook-Pro.local/Users/denyskuzmenko/projects/hive/fork/hive -r 183f8cb41d3dbed961ffd27999876468ff06690c
Compiled by denyskuzmenko on Mon Mar 25 12:44:09 CET 2024
From source with checksum e3c64bec52632c61cf7214c8b545b564
hadoop@hadoop-iims:~/apache-hive-4.0.0-bin$
```

16. Launch HIVE Query Shell

There are several limitations using Hive CLI hence in the new version its been deprecated and introduced Beeline to connect to Hive. Now you can connect to Hive from a remote server either using Beeline or from Java, Scala, Python applications using Hive JDBC Connection string

bin/beeline -u jdbc:hive2:// -n scott -p tiger

```
hadoop@hadoop-iims:~/apache-hive-4.0.0-bin$ bin/beeline -u jdbc:hive2:// -n scott -p tiger
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-4.0.0-bin/lib/log4j-slf4j-impl-2.18.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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]
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-4.0.0-bin/lib/log4j-slf4j-impl-2.18.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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]
Connecting to jdbc:hive2://
Hive Session ID = be7e52a2-0da8-405a-844d-31be2055cd67
```

```
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.hive.common.StringInternUtils
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
24/06/24 07:25:16 [main]: WARN hikari.HikariConfig: objectstore - leakDetectionThreshold is less than 2000ms or more than maxLifetime, disabling it.
24/06/24 07:25:17 [main]: WARN hikari.HikariConfig: objectstore-secondary - leakDetectionThreshold is less than 2000ms or more than maxLifetime, disabling it.
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:19 [main]: WARN DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
24/06/24 07:25:21 [main]: WARN exec.FunctionRegistry: UDF Class org.apache.hive.org.apache.datasketches.hive.cpc.UnionSketchUDF does not have description. Please annota
te the class with the org.apache.hadoop.hive.ql.exec.Description annotation and provide the description of the function.
24/06/24 07:25:21 [main]: WARN exec.FunctionRegistry: UDF Class org.apache.hive.org.apache.datasketches.hive.hll.UnionSketchUDF does not have description. Please annota
te the class with the org.apache.hadoop.hive.ql.exec.Description annotation and provide the description of the function.
24/06/24 07:25:21 [main]: WARN exec.FunctionRegistry: UDF Class org.apache.hive.org.apache.datasketches.hive.theta.IntersectSketchUDF does not have description. Please
annotate the class with the org.apache.hadoop.hive.ql.exec.Description annotation and provide the description of the function.
24/06/24 07:25:21 [main]: WARN exec.FunctionRegistry: UDF Class org.apache.hive.org.apache.datasketches.hive.theta.EstimateSketchUDF does not have description. Please a
nnotate the class with the org.apache.hadoop.hive.ql.exec.Description annotation and provide the description of the function.
24/06/24 07:25:21 [main]: WARN exec.FunctionRegistry: UDF Class org.apache.hive.org.apache.datasketches.hive.theta.ExcludeSketchUDF does not have description. Please an
notate the class with the org.apache.hadoop.hive.ql.exec.Description annotation and provide the description of the function.
24/06/24 07:25:21 [main]: WARN exec.FunctionRegistry: UDF Class org.apache.hive.org.apache.datasketches.hive.theta.UnionSketchUDF does not have description. Please anno
tate the class with the org.apache.hadoop.hive.ql.exec.Description annotation and provide the description of the function.
24/06/24 07:25:21 [main]: WARN exec.FunctionRegistry: UDF Class org.apache.hive.org.apache.datasketches.hive.tuple.ArrayOfDoublesSketchToValuesUDTF does not have descri
ption. Please annotate the class with the org.apache.hadoop.hive.ql.exec.Description annotation and provide the description of the function.
24/06/24 07:25:22 [main]: WARN session.SessionState: Configuration hive.reloadable.aux.jars.path not specified
Connected to: Apache Hive (version 4.0.0)
Driver: Hive JDBC (version 4.0.0)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 4.0.0 by Apache Hive
0: jdbc:hive2://> show databases;
+----------------+
| database_name  |
+----------------+
| default        |
+----------------+
1 row selected (1.79 seconds)
```

17. Run a sample command to check HIVE QL

```
ption. Please annotate the class with the org.apache
24/06/24 07:25:22 [main]: WARN session.SessionState:
Connected to: Apache Hive (version 4.0.0)
Driver: Hive JDBC (version 4.0.0)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 4.0.0 by Apache Hive
0: jdbc:hive2://> show databases;
+----------------+
| database_name  |
+----------------+
| default        |
+----------------+
1 row selected (1.79 seconds)
0: jdbc:hive2://>
```

This confirms that HIVE has been installed, and we can perform other Hive related tasks here.

*** END OF MANUAL***

Additional:

TO RUN HIVE USING AN IP ADDRESS:

1. Stop the running hive session by hitting Ctrl+C

2. Go to the Hive home directory

cd $HIVE_HOME

3. Edit the hive-site.xml file:

vi conf/hive-site.xml

4. Set the below property to false

```
<property>
 <name>hive.server2.enable.doAs</name>
 <value>false</value>
 </property>
```

5. In the same file, find the below property:

```
<property>
    <name>hive.conf.restricted.list</name>
```

Remove the value “hive.users.in.admin.role” and save the file.

6. Now run HiveServer 2.

$HIVE_SERVER/bin/hiveserver2

7. Open a new terminal window.

8. Sit in the Hive Home Directory

cd $HIVE_HOME

9. Start beeline hive

bin/beeline -u jdbc:hive2://10.4.47.55:10000 hadoop

10. The Hive Command Line is now available.