**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*](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*

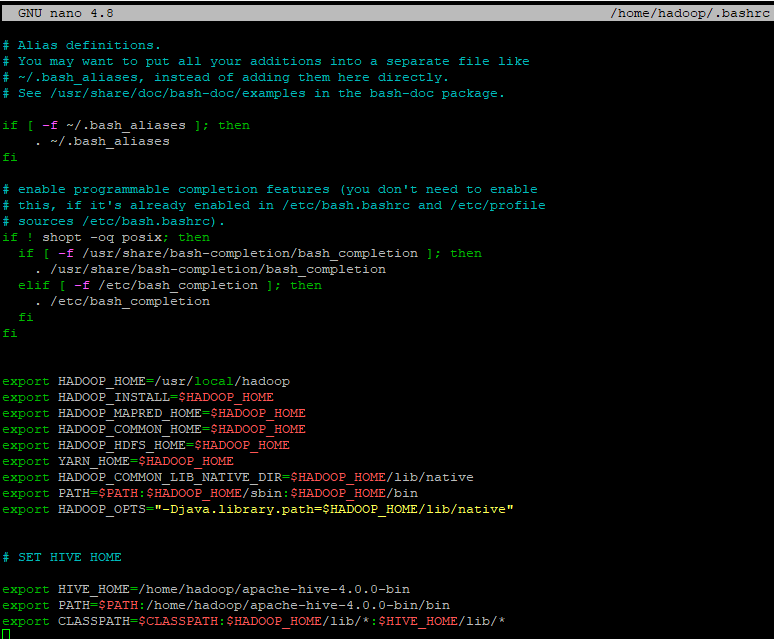
**6. Add the lines listed below into the bashrc file, and save the file:**

# 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/\*



**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*

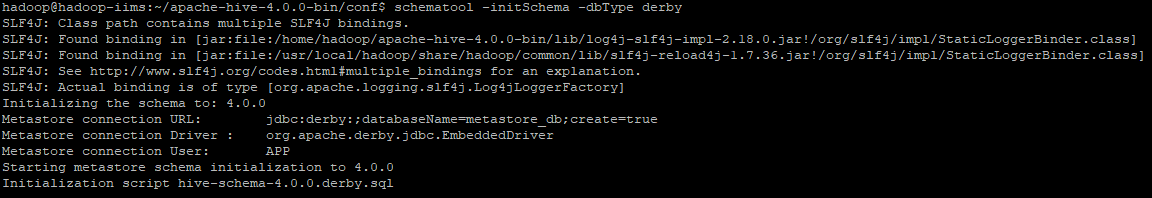
*hdfs dfs -chmod g+w /user/hive/warehouse*

**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*

**



**14. Go to $HIVE\_HOME**

*cd $HIVE\_HOME*

**15. Check if hive is installed correctly.**

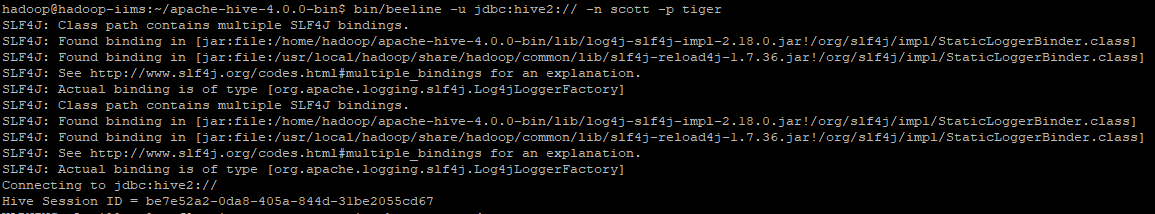
*bin/hive --version*

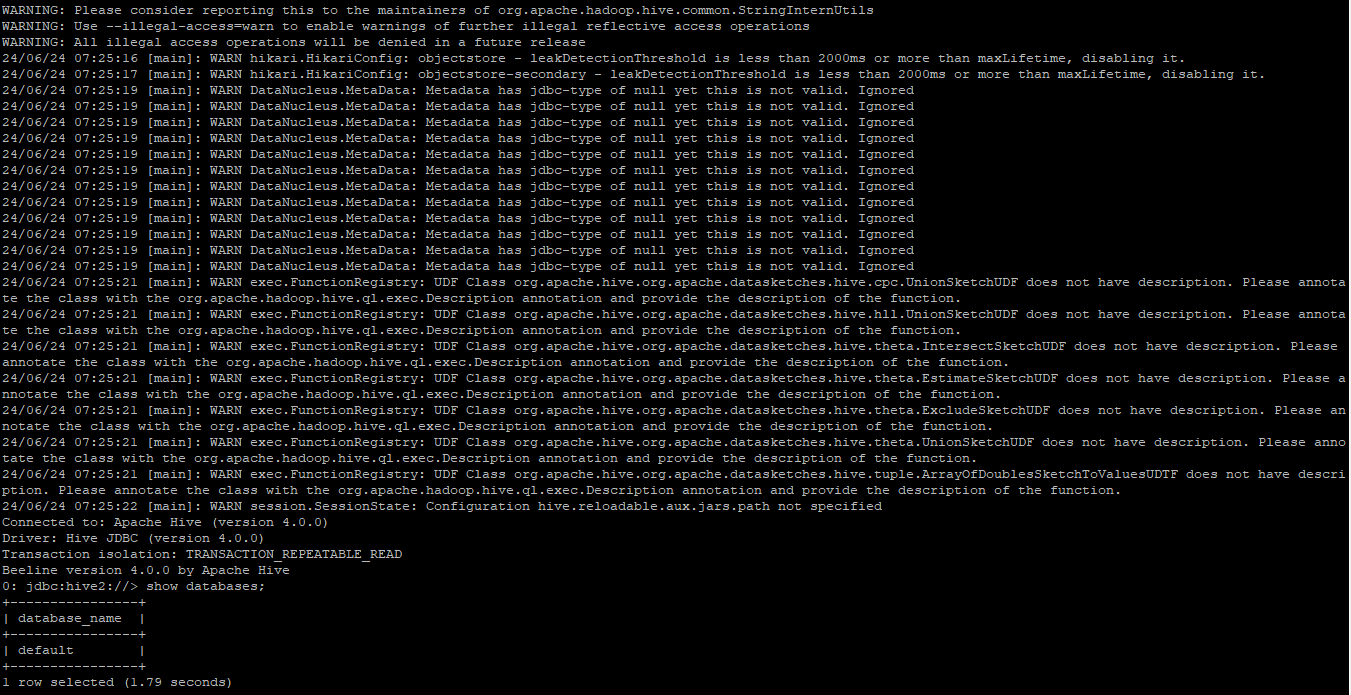


**16. Launch HIVE Query Shell**

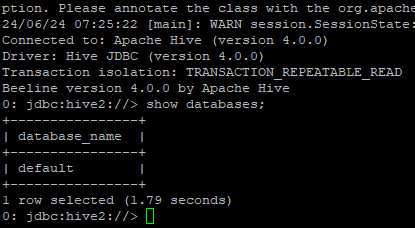
There are several limitations using Hive CLI hence in the new version its been deprecated and introduced Beeline to connect to Hive. Now you can connect to Hive from a remote server either using Beeline or from Java, Scala, Python applications using Hive JDBC Connection string

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

**



**17. Run a sample command to check HIVE QL**



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.**