Installation of HIVE Instructions (for wsl/Ubuntu systems)

Note: This is for those who have installed Hadoop version 3.x.x

1. Start ssh service and check if it is working

$ bash (Open cmd and type this to get into the wsl environment (only for wsl users))

$ sudo service ssh start

$ ssh localhost (to check if ssh is working properly)

2. Prerequisites:

Note that HIVE is incompatible with jdk11.

So, we will have to install jdk8. If you are running Hadoop using jdk8, you can ignore this step, just run the last step of this section.

$ sudo apt install openjdk-8-jre-headless

$ sudo apt-get install openjdk-8-jdk

$ sudo apt-get update (to update just incase it gives an error during installation of jdk8)

3. Once ssh is working properly, go to Hadoop and get the applications running.

$ cd ~/hadoop/hadoop-3.3.0 (In my case, it was 3.3.0. It can be 3.1.2 or even something else depending on your version of Hadoop installation)

$ sbin/start-dfs.sh

$ sbin/start-yarn.sh

$ jps (Check if all applications are up and running over here)

4. Installation:

Type "cd" to get back to home directory. All installations should be done in home itself.

$ wget https://mirrors.estointernet.in/apache/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz

$ mkdir ~/hive-home

$ tar -xvzf apache-hive-3.1.2-bin.tar.gz -C ~/hive-home

5. Opening the .bashrc file:

For Ubuntu users:

$ gedit ~/.bashrc (Ubuntu users, move to next step)

For wsl users, I would suggest you to have Visual Studio Code for editing this as editing it on shell gets very difficult and open bashrc file located in your root/home/username folder.

If you are unable to find this folder, type this into the address bar of your file system:

C:\Users\{Your_Windows_Username}\AppData\Local\Packages\CanonicalGroupLimited.Ubuntu18.04onWindows_79rhkp1fndgsc\LocalState\rootfs\home\{Your_Ubuntu_Username}

Replace {Your_Windows_Username} by your windows username and {Your_Ubuntu_Username} by your ubuntu username. You can find windows username by going to C: drive and then Users if you are not too sure about it. You can find ubuntu username in the cmd itself.

Also, if you are not using 18.04 Ubuntu in wsl, replace that by your version number in CanonicalGroupLimited.Ubuntu{version}onWindows_....:

For eg: If your version of Ubuntu on wsl is 20.04, then type this in address bar:

C:\Users\{Your_Windows_Username}\AppData\Local\Packages\CanonicalGroupLimited.Ubuntu20.04onWindows_79rhkp1fndgsc\LocalState\rootfs\home\{Your_Ubuntu_Username}

If you still can't find your home directory, type this in your cmd, you will reach there.:

$ explorer.exe . (Don't forget the "." at the end)

(Once you have reached your home directory, "Pin to Quick Access" to reach there easily the next time.)

Now, open this directory using VS Code and open .bashrc file under it.

6. Editing the .bashrc file:

Add the following code to the end of .bashrc file:

export HIVE_HOME=<HIVE_HOME_DIR>

(Replace <HIVE_HOME_DIR> by your hive-home dir.

For me, it was ~/hive-home/apache-hive-3.1.2-bin (it usually is this only))

export PATH=$PATH:$HIVE_HOME/bin

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

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

Also confirm that your bash profile includes:

export HADOOP_HOME=<HADOOP_HOME>

export PATH=$PATH:$HADOOP_HOME/bin

Also, make sure you JAVA_HOME is set to to jdk8 and not jdk11.

The next step is only for those who installed Hadoop using jdk11 and not jdk8. You may skip this if you have installed Hadoop using jdk8 itself.

For that, in .bashrc, there should be:

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

Open hadoop-env.sh that we edited during Hadoop installation and insert the same line as give above, or if it already exists, change it to the line mentioned above.

7. Create hive-site.xml

Go back to terminal and go to home in terminal using $ cd (if you are not already there)

$ cd ~/hive-home/apache-hive-3.1.2-bin/conf

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

8. Open the hive-site.xml file and insert the following in between the <configuration> tags

(Preferable after the <configuration> tag at the start)

This file is present in hive-home/apache-hive-3.1.2-bin/conf folder

Insert these:

```
 <property>
   <name>system:java.io.tmpdir</name>
   <value>/tmp/hive/java</value>
 </property>
 <property>
   <name>system:user.name</name>
   <value>{Your_Ubuntu_Username}</value>
 </property>
 <property>
   <name>javax.jdo.option.ConnectionURL</name>
   <value>jdbc:derby:;databaseName=metastore_db;create=true </value>
   <description>JDBC connect string for a JDBC metastore </description>
 </property>
```

9. Remove unwanted characters:

In the hive-site.xml file press Ctrl+F and find "hive.txn.xlock.iow".

You can see that it is between the <name> tags.

For the same property, check the description, there are some unwanted characters.

Remove those four extra characters "&#&;" between the words "for" and "transactional".

In the end, the description should be:

  <description>

    Ensures commands with OVERWRITE (such as INSERT OVERWRITE) acquire Exclusive locks for transactional tables. This ensures that inserts (w/o overwrite) running concurrently are not hidden by the INSERT OVERWRITE.

  </description>

10. Replace jar files:

Come back to cmd and type "cd" to get to the home-directory.

I hope you know your Ubuntu Username by now since we used it before, else it's in cmd.

Since, our versions of Hive and Hadoop aren't compatible, we run these two commands:

$ rm /home/{Your_Ubuntu_Username}/hive-home/apache-hive-3.1.2-bin/lib/guava-19.0.jar

$ cp /home/{Your_Ubuntu_Username}/hadoop/hadoop-3.3.0/share/hadoop/hdfs/lib/guava-27.0-jre.jar /home/{Your_Ubuntu_Username}/hive-home/apache-hive-3.1.2-bin/lib/

(Replace Hadoop-3.3.0 by your version of Hadoop in the second(cp) command)

11. Setting up the environment:

Come back to cmd and type "cd" to get to home-directory

$ cd ~/hive-home/apache-hive-3.1.2-bin

Now execute these statements:

$ hadoop fs -mkdir /tmp

$ hadoop fs -mkdir /user

$ hadoop fs -mkdir /user/hive

$ hadoop fs -mkdir /user/hive/warehouse

$ hadoop fs -chmod g+w /tmp

$ hadoop fs -chmod g+w /user/hive/warehouse

12. Run hive:

$ hive

(It will take about a minute or two, but it will run giving you a Session-ID).

13. Exit hive:

To exit hive after finishing the work, type:

exit; (inside hive prompt)

14. Stop everything:

Come back to home directory using "cd"

$ cd ~/hadoop/hadoop-3.3.0 (Replace 3.3.0 by your version number)

Run these three commands:

$ sbin/stop-yarn.sh

$ sbin/stop-dfs.sh

$ sudo service ssh stop