

Hive installation in Linux (ubuntu)

#login to hdoop user

\$ su - hdoop

```
mdtaha@NuvobookV1:~$ su - hdoop
Password:
hdoop@NuvobookV1:~$
```

#start Hadoop server

\$ start-all.sh

\$ jps

```
hdoop@NuvobookV1:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hdoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [172.18.132.51]
Starting datanodes
Starting secondary namenodes [NuvobookV1]
Starting resourcemanager
Starting nodemanagers
hdoop@NuvobookV1:~$ jps
13749 NodeManager
13627 ResourceManager
13163 DataNode
13405 SecondaryNameNode
14158 Jps
13039 NameNode
hdoop@NuvobookV1:~$ |
```

#download hive 3.1.3 and extract it

\$ wget <http://download.nust.na/pub2/apache/hive/hive-3.1.3/apache-hive-3.1.3-bin.tar.gz>

```
hdoop@phoenixNAP:~$ wget https://downloads.apache.org/hive/hive-4.0.0/apache-hive-4.0.0-bin.tar.gz
--2024-09-02 07:57:53-- https://downloads.apache.org/hive/hive-4.0.0/apache-hive-4.0.0-bin.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 88.99.208.237, 135.181.214.104, 2a01:4f8:10a:39da::2, ...
Connecting to downloads.apache.org (downloads.apache.org)|88.99.208.237|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 458782861 (438M) [application/x-gzip]
Saving to: 'apache-hive-4.0.0-bin.tar.gz'

apache-hive-4.0.0-b 100%[=====>] 437.53M 17.6MB/s in 26s

2024-09-02 07:58:19 (16.9 MB/s) - 'apache-hive-4.0.0-bin.tar.gz' saved [458782861/458782861]
```

\$ tar xzf apache-hive-3.1.3-bin.tar.gz

\$ ls

```
hdoop@NuvobookV1:~$ tar xzf apache-hive-3.1.3-bin.tar.gz
hdoop@NuvobookV1:~$ ls
apache-hive-3.1.3-bin      hadoop-3.4.0             pig-0.17.0
apache-hive-3.1.3-bin.tar.gz  hadoop-3.4.0.tar.gz      pig-0.17.0.tar.gz
derby.log                 hadoop-streaming-2.7.3.jar sample.txt
dev                        lab                       test.txt
dfsdata                   metastore_db              tmpdata
hdoop@NuvobookV1:~$ |
```

#configur system variables

\$ nano .bashrc

-insert this:

```
export HIVE_HOME=/home/hdoop/apache-hive-3.1.3-bin
```

```
export PATH=$PATH:$HIVE_HOME/bin
```

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

```
export PATH=$PATH:$JAVA_HOME/bin
```

```
export HADOOP_CLASSPATH=$(hadoop classpath)
#Hive sitting
export HIVE_HOME=/home/hdoop/apache-hive-3.1.3-bin
export PATH=$PATH:$HIVE_HOME/bin
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:$JAVA_HOME/bin
```

-click Ctrl+s and Ctrl+x to go out the editor

\$ source ~/.bashrc

```
hadoop@NuvobookV1:~$ nano .bashrc
hadoop@NuvobookV1:~$ source .bashrc
hadoop@NuvobookV1:~$ |
```

#Make directories inside Hadoop and change their permissions:

\$hadoop fs -mkdir /tmp

\$hadoop fs -chmod g+w /tmp

```
hadoop@NuvobookV1:~$ hadoop fs -mkdir /tmp
hadoop@NuvobookV1:~$ hadoop fs -chmod g+w /tmp
hadoop@NuvobookV1:~$ |
```

\$hadoop fs -mkdir /user

\$hadoop fs -mkdir /user/hive

\$hadoop fs -mkdir /user/hive/warehouse

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

```
hadoop@NuvobookV1:~$ hadoop fs -mkdir /user
hadoop@NuvobookV1:~$ hadoop fs -mkdir /user/hive
hadoop@NuvobookV1:~$ hadoop fs -mkdir /user/hive/warehouse
hadoop@NuvobookV1:~$ hadoop fs -chmod g+w /user/hive/warehous
```

Edit hive-config.sh file

export HADOOP_HOME=/home/hadoop/hadoop-3.4.0

```
# Disable the JNDI. This feature has critical RCE vulnerability.
# when 2.x <= log4j.version <= 2.14.1
export HADOOP_CLIENT_OPTS="$HADOOP_CLIENT_OPTS -Dlog4j2.formatMsgNoLookups=>
export HADOOP_HOME=/home/hadoop/hadoop-3.4.0
```

-click Ctrl+s and Ctrl+x to go out the editor

#replace guava file

\$ cd sudo \$HIVE_HOME/lib/

\$ ls -l guava*

\$ rm -rf guava-19.0.jar

\$ cp \$HADOOP_HOME/share/hadoop/hdfs/lib/guava-27.0-jre.jar

\$HIVE_HOME/lib

```
hadoop@NuvobookV1:~$ cd $HIVE_HOME/lib/
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$ ls -l guava*
-rw-r--r-- 1 hadoop hadoop 2308517 Oct 18 2019 guava-19.0.jar
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$ rm -rf guava-19.0.jar
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$ ls -l guava*
ls: cannot access 'guava*': No such file or directory
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$ cp $HADOOP_HOME/share/hadoop/
hdfs/lib/guava-27.0-jre.jar $HIVE_HOME/lib
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$ ls -l guava*
-rw-r--r-- 1 hadoop hadoop 2747878 Oct 17 08:00 guava-27.0-jre.jar
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$ |
```

Configure hive-site.xml File

\$ cd \$HIVE_HOME/conf

\$ ls -l

```
hadoop@phoenixNAP:~/apache-hive-4.0.0-bin/conf$ ls -l
total 844
-rw-r--r-- 1 hadoop hadoop 1775 Jan 22 2020 beeline-log4j2.properties.template
-rw-r--r-- 1 hadoop hadoop 413104 Jan 22 2020 hive-default.xml.template
-rw-r--r-- 1 hadoop hadoop 2365 Jan 22 2020 hive-env.sh.template
-rw-r--r-- 1 hadoop hadoop 2274 Jan 22 2020 hive-exec-log4j2.properties.template
-rw-r--r-- 1 hadoop hadoop 3086 Jan 22 2020 hive-log4j2.properties.template
-rw-r--r-- 1 hadoop hadoop 413104 Sep 3 04:15 hive-site.xml
-rw-r--r-- 1 hadoop hadoop 2060 Jan 22 2020 ivysettings.xml
-rw-r--r-- 1 hadoop hadoop 3558 Jan 22 2020 llap-cli-log4j2.properties.template
-rw-r--r-- 1 hadoop hadoop 7093 Jan 22 2020 llap-daemon-log4j2.properties.template
-rw-r--r-- 1 hadoop hadoop 2662 Jan 22 2020 parquet-logging.properties
```

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

\$ nano hive-site.xml

-click Ctrl+w to search for “hive.metastore.warehouse.dir”, do it again and again until you find this

```
GNU nano 7.2                               hive-site.xml
Type of database used by the metastore. Information schema & JDBCStor>
</description>
</property>
<property>
  <name>hive.metastore.warehouse.dir</name>
  <value>/user/hive/warehouse</value>
  <description>location of default database for the warehouse</description>
</property>
```

-add “/home/hadoop”, replace hadoop with your system username.

```
hadoop@NuvobookV1:~$ |
```

-click Ctrl+w and search for “for” and remove it

```
<description>
<!-- Exclusive locks for transactional tables. This ensures that inserts
are not hidden by the INSERT OVERWRITE.
</description>
</property>
<property>
  <name>hive.txn.timeout</name>
  <value>300s</value>
  <description>
    Expects a time value with unit (d/day, h/hour, m/min, s/sec, ms/msec,
    time after which transactions are declared aborted if the client has
  </description>
</property>
<property>
  <name>hive.txn.heartbeat.threadpool.size</name>
  <value>5</value>
  <description>The number of threads to use for heartbeating. For Hive CL
</property>
<property>
  <name>hive.txn.manager.dump.lock.state.on.acquire.timeout</name>
  <value>false</value>
Search [for]: for
^G Help      M-C Case Sens M-B Backwards ^P Older      ^T Go To Line
^C Cancel    M-R Reg.exp.  ^R Replace    ^N Newer
```

-click Ctrl+s and Ctrl+x to go out the editor

\$ nano hive-site.xml

-add these under “<configuration>”

```
<property>  
  <name>system:java.io.tmpdir</name>  
  <value>/tmp/hive/java</value>  
</property>
```

```
<property>  
  <name>system:user.name</name>  
  <value>${user.name}</value>  
</property>
```

```
Limitations under the License.  
--><configuration>  
  <!-- WARNING!!! This file is auto generated for documentation purposes ON>  
  <!-- WARNING!!! Any changes you make to this file will be ignored by Hive>  
  <!-- WARNING!!! You must make your changes in hive-site.xml instead.    >  
  <!-- Hive Execution Parameters -->  
  
  <property>  
    <name>system:java.io.tmpdir</name>  
    <value>/tmp/hive/java</value>  
  </property>  
  
  <property>  
    <name>system:user.name</name>  
    <value>${user.name}</value>  
  </property>
```

-click Ctrl+s and Ctrl+x to go out the editor

#Remove log4j file:

```
$cd $HIVE_HOME/lib
```

```
$ ls -l log4j*.*
```

```
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/conf$ cd $HIVE_HOME/lib
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$ ls -l log4j*.*
-rw-r--r-- 1 hadoop hadoop 208006 Jan 6 2022 log4j-1.2-api-2.17.1.jar
-rw-r--r-- 1 hadoop hadoop 301872 Jan 6 2022 log4j-api-2.17.1.jar
-rw-r--r-- 1 hadoop hadoop 1790452 Jan 6 2022 log4j-core-2.17.1.jar
-rw-r--r-- 1 hadoop hadoop 24279 Jan 6 2022 log4j-slf4j-impl-2.17.1.jar
-rw-r--r-- 1 hadoop hadoop 35962 Jan 6 2022 log4j-web-2.17.1.jar
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$
```

```
$ rm -rf log4j-slf4j-impl-2.17.1.jar
```

```
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$ rm -rf log4j-slf4j-impl-2.17.1
.jar
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$ ls -l log4j*.*
-rw-r--r-- 1 hadoop hadoop 208006 Jan 6 2022 log4j-1.2-api-2.17.1.jar
-rw-r--r-- 1 hadoop hadoop 301872 Jan 6 2022 log4j-api-2.17.1.jar
-rw-r--r-- 1 hadoop hadoop 1790452 Jan 6 2022 log4j-core-2.17.1.jar
-rw-r--r-- 1 hadoop hadoop 35962 Jan 6 2022 log4j-web-2.17.1.jar
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/lib$ |
```

#insiate the local database

```
$ cd $HIVE_HOME/bin
```

```
$ ./schematool -initSchema -dbType derby
```

```
Initialization script completed
schemaTool completed
hadoop@NuvobookV1:~/apache-hive-3.1.3-bin/bin$ |
```

#Start hive

```
$ cd $HIVE_HOME/bin
```

```
$ ./hive
```



```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin/bin
hadoop@hadoop-VirtualBox:~$ cd apache-hive-3.1.2-bin/bin
hadoop@hadoop-VirtualBox:~/apache-hive-3.1.2-bin/bin$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoop/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.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 Session ID = 8d24955e-52bf-43df-92b0-8f14b2f1247c

Logging initialized using configuration in jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.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 Session ID = 902ae3da-f1a7-4759-8c4f-cc4ef18a99c6
hive> show databases;
OK
bda
default
students
students1
students2
Time taken: 0.437 seconds, Fetched: 5 row(s)
hive> █
```

>show databases;

>use default;

>show tables;

```
hadoop@hadoop-VirtualBox: ~/apache-hive-3.1.2-bin/bin
hadoop@hadoop-VirtualBox:~$ cd apache-hive-3.1.2-bin/bin
hadoop@hadoop-VirtualBox:~/apache-hive-3.1.2-bin/bin$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hadoop/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.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 Session ID = 8d24955e-52bf-43df-92b0-8f14b2f1247c

Logging initialized using configuration in jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.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 Session ID = 902ae3da-f1a7-4759-8c4f-cc4ef18a99c6
hive> show databases;
OK
bda
default
students
students1
students2
Time taken: 0.437 seconds, Fetched: 5 row(s)
hive> █
```


Note: if hive shell command does not look like above write “exit;” to exit hive and follow the next processor:

\$ bin/hiveserver2

```
hadoop@phoenixNAP:~/apache-hive-4.0.0-bin$ bin/hiveserver2
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:/home/hadoop/hadoop-3.4.0/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:/home/hadoop/hadoop-3.4.0/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]
2024-09-04 05:41:10: Starting HiveServer2
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:/home/hadoop/hadoop-3.4.0/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 Session ID = d88ad44d-0012-4809-871e-742c81e2d119
```

-Wait for the server to start and show the Hive Session ID.


- In another terminal tab, switch to the Hadoop user using the su command:

\$ su – hadoop

\$cd \$HIVE_HOME

\$ bin/beeline -n db_user -u jdbc:hive2://localhost:10000

```

hadoop@phoenixNAP:~/apache-hive-4.0.0-bin$ bin/beeline -n db_user -u jdbc:hive2://localhost:10000
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:/home/hadoop/hadoop-3.4.0/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:/home/hadoop/hadoop-3.4.0/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://localhost:10000
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://localhost:10000> 

```

>show databases;

```

0: jdbc:hive2://localhost:10000> show databases;
INFO : Compiling command(queryId=hadoop_20240904061926_a842bdbb-180a-4342-93f1-3ddc8366db4c): show databases
INFO : Semantic Analysis Completed (retrial = false)
INFO : Created Hive schema: Schema(fieldSchemas:[FieldSchema(name:database_name, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hadoop_20240904061926_a842bdbb-180a-4342-93f1-3ddc8366db4c); Time taken: 1.715 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hadoop_20240904061926_a842bdbb-180a-4342-93f1-3ddc8366db4c): show databases
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hadoop_20240904061926_a842bdbb-180a-4342-93f1-3ddc8366db4c); Time taken: 0.165 seconds
+-----+
| database_name |
+-----+
| default      |
+-----+
1 row selected (2.474 seconds)
0: jdbc:hive2://localhost:10000>

```