• Installing Hive:

1. Extracting the apache-hive-2.3.9-bin.tar.gz

```
ubuntu@stud:~/Downloads
ubuntu@stud:~$ cd Downloads/
ubuntu@stud:~/Downloads$ ls
apache-hive-2.3.9-bin.tar.gz files firefox.tmp fwd.zip
ubuntu@stud:~/Downloads$ tar -xvzf apache-hive-2.3.9-bin.tar.gz
apache-hive-2.3.9-bin/conf/hive-log4j2.properties.template
apache-hive-2.3.9-bin/conf/hive-exec-log4j2.properties.template
apache-hive-2.3.9-bin/conf/beeline-log4j2.properties.template
apache-hive-2.3.9-bin/conf/llap-daemon-log4j2.properties.template
apache-hive-2.3.9-bin/conf/llap-cli-log4j2.properties.template
```

2. Moving the extracted file /home/hdp123 directory:

```
hdp123@stud:~

ubuntu@stud:~/Downloads$ su hdp123
Password:
hdp123@stud:/home/ubuntu/Downloads$ sudo mv apache-hive-2.3.9-bin /home/hdp123/
[sudo] password for hdp123:
hdp123@stud:/home/ubuntu/Downloads$ ls
apache-hive-2.3.9-bin.tar.gz files firefox.tmp fwd.zip
hdp123@stud:/home/ubuntu/Downloads$ cd
hdp123@stud:~$ ls
analyselogs apache-hive-2.3.9-bin input1 snap
hdp123@stud:~$
```

3. Editing the bashrc file by adding HIVE_HOME Path:

```
hdp123@stud:~/apache-hive-2.3.9-bin/conf Q = - - ×

hdp123@stud:~$ cd apache-hive-2.3.9-bin/
hdp123@stud:~/apache-hive-2.3.9-bin$ pwd
/home/hdp123/apache-hive-2.3.9-bin$ sudo nano ~/.bashrc
hdp123@stud:~/apache-hive-2.3.9-bin$ source ~/.bashrc
```

```
. /etc/bash_completion
fi

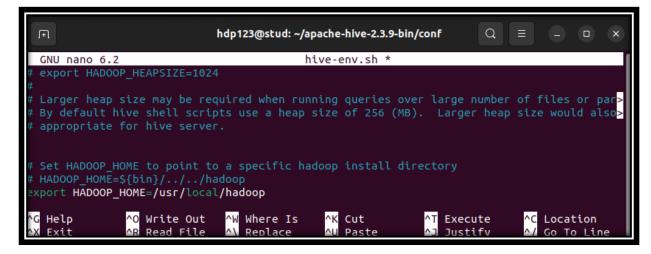
fi

export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export HADOOP_HOME=/usr/local/hadoop
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"
export HIVE_HOME=/home/hdp123/apache-hive-2.3.9-bin
```

4. Renaming the Hive-env and Hive-default files:



```
hdp123@stud:~/apache-hive-2.3.9-bin/conf$ sudo mv hive-env.sh.template hive-env.sh
hdp123@stud:~/apache-hive-2.3.9-bin/conf$ ls
beeline-log4j2.properties.template ivysettings.xml
hive-env.sh llap-cli-log4j2.properties.template
hive-exec-log4j2.properties.template llap-daemon-log4j2.properties.template
hive-log4j2.properties.template parquet-logging.properties
hive-site.xml
hdp123@stud:~/apache-hive-2.3.9-bin/conf$
```



5. Starting the Hadoop Services:

```
hdp123@stud:~/apache-hive-2.3.9-bin/conf Q = - - ×

hdp123@stud:~/apache-hive-2.3.9-bin/conf$ start-all.sh

This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh

WARNING: An illegal reflective access operation has occurred

WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.Kerbe

rosIItil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.10.2.jar) to method
```

```
hdp123@stud:~/apache-hive-2.3.9-bin/conf$ jps
10640 NameNode
11571 NodeManager
11299 SecondaryNameNode
11066 DataNode
11451 ResourceManager
11919 Jps
hdp123@stud:~/apache-hive-2.3.9-bin/conf$
```

6. Creating files and assigning access to those files:

```
hdp123@stud: /usr/local/hadoop
hdp123@stud:~/apache-hive-2.3.9-bin/conf$ cd $HADOOP_HOME
hdp123@stud:/usr/local/hadoop$ sudo chown -R hdp123 /home/hdp123/apache-hive-2.3.9-bin/
hdp123@stud:/usr/local/hadoop$ hdfs dfs -mkdir -p /user/hive/warehouse
WARNING: An illegal reflective access operation has occurred WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.Kerbe
rosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.10.2.jar) to method
sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.a
uthentication.util.KerberosUtil
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
23/05/04 20:23:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for you
r platform... using builtin-java classes where applicable hdp123@stud:/usr/local/hadoop$ hdfs dfs -chmod 777 /user/hive/warehouse
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.Kerbe
rosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.10.2.jar) to method
sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.a
uthentication.util.KerberosUtil
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
23/05/04 20:23:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for you
r platform... using builtin-java classes where applicable
hdp123@stud:/u
                       /hadoop$ hdfs dfs -chmod 777 /user/hive/warehouse
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.Kerbe
rosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.10.2.jar) to method
sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.a
uthentication.util.KerberosUtil
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
23/05/04 20:24:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for you
r platform... using builtin-java classes where applicable
hdp123@stud:/us
                        hadoop$ hdfs dfs -chmod 777 /tmp
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.Kerbe
```

7. Initializing the Schema-tool:

```
hdp123@stud:~/apache-hive-2.3.9-bin/bin Q = - - ×

hdp123@stud:~/apache-hive-2.3.9-bin/bin$ ./schematool -dbType derby -initSchema
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdp123/apache-hive-2.3.9-bin/lib/log4j-slf4j-impl-
2.6.2.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]
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 2.3.0
Initialization script hive-schema-2.3.0.derby.sql
Initialization script completed
schemaTool completed
```

8. Starting Hive:

```
hdp123@stud:~/apache-hive-2.3.9-bin/bin$ ./hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdp123/apache-hive-2.3.9-bin/lib/log4j-slf4j-impl-2.6.2.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]
Logging initialized using configuration in jar:file:/home/hdp123/apache-hive-2.3.9-bin/lib
/hive-common-2.3.9.jar!/hive-log4j2.properties Async: true
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.Kerbe
rosUtil (file:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.10.2.jar) to method
sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.a
uthentication.util.KerberosUtil
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
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consid
er using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive>
```

• Running Hive Queries:

1. Staring Hive Services:

```
hdp123@stud:~/apache-hive-2.3.9-bin/bin Q = - - ×

hdp123@stud:~/files$ ls
flight_data.csv flight_data.txt flight_ext.csv flight_fare.csv sample.txt
hdp123@stud:~/files$ cd
hdp123@stud:~/s cd apache-hive-2.3.9-bin/bin
hdp123@stud:~/apache-hive-2.3.9-bin/bin$ ./hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdp123/apache-hive-2.3.9-bin/lib/log4j-slf4j-impl-2
.6.2.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.
```

2. Creating a database and emp table:

```
hive> show databases;

OK

default

Time taken: 6.542 seconds, Fetched: 1 row(s)

hive> create database dypcoe;

OK

Time taken: 0.242 seconds

hive> create table emp(ename string,esal int) row format delimited fields terminated by '_'

stored as textfile;

OK

Time taken: 0.937 seconds

hive>
```

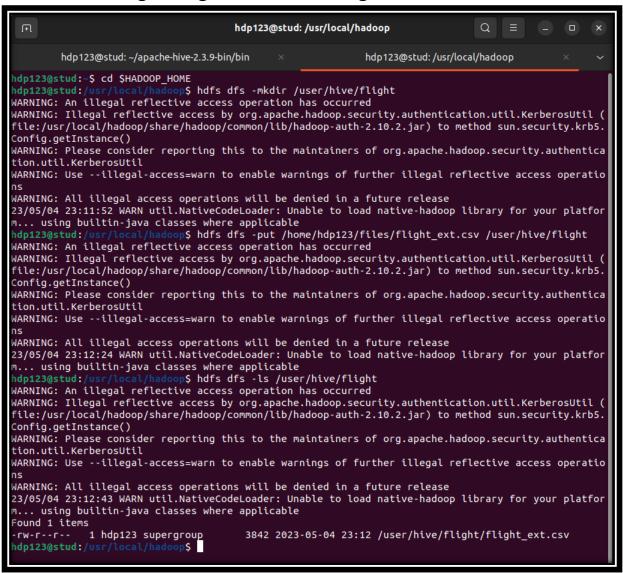
3. Loading data from sample.txt in emp table and viewing the data:

```
hive> load data local inpath '/home/hdp123/files/sample.txt' into table emp;
Loading data to table default.emp
OK
Time taken: 1.795 seconds
hive> select * from emp;
OK
Parth,15200 NULL
Aditya,58241 NULL
Yogi,14000 NULL
Tanishq,14500 NULL
Time taken: 1.823 seconds, Fetched: 4 row(s)
hive>
```

4. Creating flight table & loading data from flight_data into table :

```
hive> create table flight(Year int, Month int, DayofMonth int, DayOfWeek int, DepTime int, CRSDepT ime int, ArrTime int, CRSArrTime int, UniqueCarrier string, FlightNum int, TailNum string, ActualEl apsedTime int, CRSElapsedTime int, AirTime int, ArrDelay int, DepDelay int, Origin string, Dest string, Distance int, TaxiIn int, TaxiOut int, Cancelled int, CancellationCode string, Diverted string, CarrierDelay int, WeatherDelay int, NASDelay int, SecurityDelay int, LateAircraftDelay int) row format delimited fields terminated by ',';
OK
Time taken: 0.373 seconds
hive> load data local inpath '/home/hdp123/files/flight_data.csv' into table flight;
Loading data to table default.flight
OK
Time taken: 0.563 seconds
```

5. Creating storage files and loading data into files:



6. Creating external table(flight) and describing it:

```
hdp123@stud: ~/apache-hive-2.3.9-bin/bin
hive> create external table flight_ext(Year int, Month int, DayofMonth int, DayOfWeek int, DepTime int, CRSDepTime int, ArrTime int, CRSArrTime int, UniqueCarrier string, FlightNum int, TailNum string, ActualElapsedTime int, CRSElapsedTime int, AirTime int, ArrDelay int, DepDelay int, Origin string, Dest string, Distance int, TaxiIn int, TaxiOut int, Cancelled int, CancellationCode string, Diverted string, CarrierDelay int, WeatherDelay int, NASDelay int, SecurityDelay int, LateAircraftDelay int) row format delimited fields terminated by ',' location '/user/hive/flight';
 Time taken: 5.487 seconds
hive> describe formatted flight;
# col_name
                                            data_type
                                                                                         comment
year
                                             int
 month
                                             int
 dayofmonth
                                             int
dayofweek
                                             int
deptime
                                             int
crsdeptime
                                             int
arrtime
                                             int
crsarrtime
                                             int
uniquecarrier
                                             string
flightnum
                                             int
tailnum
                                             string
actualelapsedtime
                                             int
crselapsedtime
                                             int
airtime
                                             int
arrdelay
                                             int
depdelay
                                             int
origin
                                             string
dest
                                             string
distance
                                             int
 taxiin
                                             int
taxiout
                                             int
cancelled
                                             int
cancellationcode
                                             string
                                             string
diverted
carrierdelay
                                             int
weatherdelay
                                             int
nasdelay
                                             int
securitydelay
                                             int
 lateaircraftdelay
                                             int
```

```
# Detailed Table Information
Database:
                           default
Owner:
                           hdp123
                           Thu May 04 23:05:48 IST 2023
CreateTime:
LastAccessTime:
                           UNKNOWN
Retention:
Location:
                           hdfs://localhost:9000/user/hive/warehouse/flight
Table Type:
Table Parameters:
                           MANAGED_TABLE
         numFiles
                                    1
        numRows
                                    0
         rawDataSize
                                    0
         totalSize
                                    4674
         transient_lastDdlTime
                                    1683221778
# Storage Information
SerDe Library:
                           org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
                           org.apache.hadoop.mapred.TextInputFormat
org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
InputFormat:
OutputFormat:
Compressed:
                           No
                           -1
[]
Num Buckets:
Bucket Columns:
Sort Columns:
Storage Desc Params:
         field.delim
         serialization.format
Time t<u>a</u>ken: 0.698 seconds, Fetchéd: 58 row(s)
```

7. Counting number of rows in flight ext table:

```
hive> select count(*) from flight_ext;

WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a di or using Hive 1.X releases.

Query ID = hdp123_20230504231827_bb9d1267-d97f-452f-b01d-568d08a69292

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):
    set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
    set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
    set mapreduce.job.reduces=<number>
Starting Job = job_1683211876093_0001, Tracking URL = http://stud:8088/proxy/application_1683211876093_0001/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1683211876093_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-04 23:18:48,095 Stage-1 map = 0%, reduce = 0%
2023-05-04 23:19:05,716 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.87 sec
2023-05-04 23:19:05,716 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.29 sec
MapReduce Total cumulative CPU time: 5 seconds 290 msec
Ended Job = job_1683211876093_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.29 sec HDFS Read: 15222 HDFS Write: 102 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 290 msec

OK
39
Time taken: 42.368 seconds, Fetched: 1 row(s)
```

8. Creating internal table (flight_fare):

```
hive> create table flight_fare( FlightNum int, Fare int) row format delimited fields terminated by ',';
OK
Time taken: 0.234 seconds
hive> load data local inpath '/home/hdp123/files/flight_fare.csv' into table flight_fare;
Loading data to table default.flight_fare
OK
Time taken: 0.808 seconds
```

```
hive> select flight.Origin, flight.Dest, flight_fare.fare from flight join flight_fare on flight.FlightNum=flight_fare.FlightNum;

WARNING: 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.

Query ID = hdp123_20230504232230_d9d247a5-6726-4fe9-9f74-05d769fe831c

Total jobs = 1
```

```
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1683211876093_0002, Tracking URL = http://stud:8088/proxy/applicat
3 0002/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1683211876093_0002
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2023-05-04 23:22:51,192 Stage-3 map = 0%, reduce = 0%
2023-05-04 23:22:57,690 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.49 sec
MapReduce Total cumulative CPU time: 2 seconds 490 msec
Ended Job = job_1683211876093_0002
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 2.49 sec HDFS Read: 13095 HDFS Write: 1317
Total MapReduce CPU Time Spent: 2 seconds 490 msec
OK
SEA
        SJC
                5181
SEA
        PSP
                7905
SAN
        SEA
                8781
SEA
        GEG
                7852
                9009
TUS
        SEA
```

9. Calculating average of depdelay from flight:

```
hive> select avg(depdelay) from flight;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future vers
using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = hdp123_20230504232624_dc8110ff-e4f1-496b-9290-f9548b0828eb
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
 set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
 set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
 set mapreduce.job.reduces=<number>
Starting Job = job_1683211876093_0003, Tracking URL = http://stud:8088/proxy/applicatio
3_0003/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1683211876093_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-04 23:26:34,120 Stage-1 map = 0%, reduce = 0%
2023-05-04 23:26:40,832 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.04 sec
2023-05-04 23:26:48,252 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.28 sec
MapReduce Total cumulative CPU time: 4 seconds 280 msec
Ended Job = job_1683211876093_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1
                                     Cumulative CPU: 4.28 sec
                                                                 HDFS Read: 16632 HDFS Wri
Total MapReduce CPU Time Spent: 4 seconds 280 msec
10.021739130434783
Time taken: 26.331 seconds, Fetched: 1 row(s)
```

