HIVE INSTALLATION:


```
stud@stud-OptiPlex-3060:~$ source ~/.bashrc
stud@stud-OptiPlex-3060:~$ echo $JAVA_HOME
/usr/lib/jvm/java-1.8.0-openjdk-amd64
stud@stud-OptiPlex-3060:~$ su - hdp123
Password:
hdp123@stud-OptiPlex-3060:~$ start-dfs.sh
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/local/hadoop
/share/hadoop/common/lib/hadoop-auth-2.10.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.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
23/04/27 16:48:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hdp123-namenode-stud-OptiPlex-3060.out
localhost: WARNING: An illegal reflective access operation has occurred
localhost: WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/usr/l
ocal/hadoop/share/hadoop/common/lib/hadoop-auth-2.10.1.jar) to method sun.security.krb5.Config.getInstance()
localhost: WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.K
erberosUtil
localhost: WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
localhost: WARNING: All illegal access operations will be denied in a future release
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hdp123-datanode-stud-OptiPlex-3060.out
```



```
hdp123@stud-OptiPlex-3060:/usr/local$ jps
8480 NameNode
9062 ResourceManager
9195 NodeManager
8683 DataNode
8891 SecondaryNameNode
10075 Jps
hdp123@stud-OptiPlex-3060:/usr/local$ cd hadoop
hdp123@stud-OptiPlex-3060:/usr/local/hadoop$ ls
bin    include   libexec      logs          README.txt   share
etc    lib       LICENSE.txt   NOTICE.txt   sbin
hdp123@stud-OptiPlex-3060:/usr/local/hadoop$ cd bin
hdp123@stud-OptiPlex-3060:/usr/local/hadoop/bin$ hdfs dfs -mkdir -p /usr/hive/warehouse
23/04/27 16:58:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
hdp123@stud-OptiPlex-3060:/usr/local/hadoop/bin$ hdfs dfs -mkdir -p /tmp/hive
23/04/27 16:59:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
hdp123@stud-OptiPlex-3060:/usr/local/hadoop/bin$ hdfs dfs -chmod 777 /tmp
23/04/27 17:00:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
hdp123@stud-OptiPlex-3060:/usr/local/hadoop/bin$ hdfs dfs -chmod 777 /usr/hive/warehouse
23/04/27 17:00:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
```



```
hdp123@stud-OptiPlex-3060:/usr/local/hadoop/bin$ hdfs dfs -chmod 777 /temp/hive
23/04/27 17:01:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
chmod: `/temp/hive': No such file or directory
hdp123@stud-OptiPlex-3060:/usr/local/hadoop/bin$ hdfs dfs -chmod 777 /tmp/hive
23/04/27 17:01:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
a classes where applicable
hdp123@stud-OptiPlex-3060:/usr/local/hadoop/bin$ cd ..
hdp123@stud-OptiPlex-3060:/usr/local/hadoop$ cd ..
hdp123@stud-OptiPlex-3060:/usr/local$ cd hive
hdp123@stud-OptiPlex-3060:/usr/local/hive$ ls
bin   binary-package-licenses   conf   examples   hcatalog   jdbc   lib   LICENSE   NOTICE   RELEASE_NOTES.txt   scripts
hdp123@stud-OptiPlex-3060:/usr/local/hive$ cd bin
hdp123@stud-OptiPlex-3060:/usr/local/hive/bin$ ls
beeline   ext   hive   hive-config.sh   hiveserver2   hplsql   metatool   schematool
hdp123@stud-OptiPlex-3060:/usr/local/hive/bin$ $HIVE_HOME/bin/schematool -dbType derby -initSchema
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/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-log4j12-1.7.25.jar!/org/slf4j/impl/Stat
icLoggerBinder.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: 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
hdp123@stud-OptiPlex-3060:/usr/local/hive/bin$ ls
beeline  derby.log  ext  hive  hive-config.sh  hiveserver2  hplsql  metastore_db  metatool  schematool
hdp123@stud-OptiPlex-3060:/usr/local/hive/bin$ ./hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/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-log4j12-1.7.25.jar!/org/slf4j/impl/Stati
cLoggerBinder.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:/usr/local/hive/lib/hive-common-2.3.9.jar!/hive-log4j2.properties Asyn
c: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution e
ngine (i.e. spark, tez) or using Hive 1.X releases.
hive> show databases;
```

Adding path of hive in bashrc:


```
  GNU nano 6.2                          /home/hdp123/.bashrc
# 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 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_HDFS_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=/usr/local/hive


^G Help        ^O Write Out   ^W Where Is    ^K Cut         ^T Execute     ^C Location    M-U Undo       M-A Set Mark
^X Exit        ^R Read File   ^\ Replace     ^U Paste       ^J Justify     ^/ Go To Line  M-E Redo       M-6 Copy
```

Hive Terminal:

```
hive> show tables;
OK
emp
Time taken: 5.405 seconds, Fetched: 1 row(s)
hive> drop table emp;
OK
Time taken: 0.399 seconds
hive> create table emp(ename string,esal int) row format delimited fields terminated by ',' stored as textfile;
OK
Time taken: 0.487 seconds
hive> load data local inpath '/home/stud/Desktop/emp_details.txt' into table emp;
Loading data to table default.emp
OK
Time taken: 0.365 seconds
hive> select * from emp;
OK
sachin    10000
ayush     50000
yuvraj    70000
yogesh    90000
rahul     11000
mark      20000
stocks    40000
Time taken: 0.74 seconds, Fetched: 7 row(s)
hive>
```

Creating an external hive table:

```
hdp123@stud-OptiPlex-3060:/usr/local$ hdfs dfs -mkdir /user/hive/flight
23/05/02 14:31:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where appli
cable
hdp123@stud-OptiPlex-3060:/usr/local$ hdfs dfs -put /home/stud/Downloads/flight_ext.csv /user/hive/flight
23/05/02 14:34:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where appli
cable
hdp123@stud-OptiPlex-3060:/usr/local$ hdfs dfs -ls /user/hive/flight
23/05/02 14:34:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where appli
cable
Found 1 items
-rw-r--r--   1 hdp123 supergroup      3842 2023-05-02 14:34 /user/hive/flight/flight_ext.csv
```

```
hdp123@stud-OptiPlex-3060:/usr/local/hive/bin$ ./hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/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-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]

Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-2.3.9.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, t
ez) or using Hive 1.X releases.
hive> create external table flight_ext(Year int, Month int, DayofMonth int, DayOfWeek int, DepTime  int, CRSDepTime int, ArrTime int, CRSArrTi
me int, UniqueCarrier string, FlightNum int, TailNum string, ActualElapsedTime int, CRSElapsedTime int, AirTime int, ArrDelay int, DepDelay in
t, Origin string, Dest string, Distance int, TaxiIn int, TaxiOut int, Cancelled int, CancellationCode string, Diverted string, CarrierDelay in
t, WeatherDelay int, NASDelay int,  SecurityDelay int, LateAircraftDelay int) row format delimited fields terminated by ',' location '/user/hi
ve/flight';
OK
Time taken: 20.014 seconds
hive> show tables;
OK
emp
flight
flight_ext
flight_fare
Time taken: 0.078 seconds, Fetched: 4 row(s)
```

To find the details of table in hive:

```
hive> describe formatted flight_ext;
OK
# 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
diverted                    string
carrierdelay                int
weatherdelay                int
nasdelay                    int
securitydelay               int
lateaircraftdelay           int
```

Count the number of rows in 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 different execution engine (i.e.
 spark, tez) or using Hive 1.X releases.
Query ID = hdp123_20230502143924_c7d1e875-d69a-4964-8cba-18477918dc69
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_1683014056661_0002, Tracking URL = http://stud-OptiPlex-3060:8088/proxy/application_1683014056661_0002/
Kill Command = /usr/local/hadoop/bin/hadoop job  -kill job_1683014056661_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-02 14:39:30,423 Stage-1 map = 0%,  reduce = 0%
2023-05-02 14:39:34,513 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.36 sec
2023-05-02 14:39:39,592 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.79 sec
MapReduce Total cumulative CPU time: 2 seconds 790 msec
Ended Job = job_1683014056661_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.79 sec   HDFS Read: 15222 HDFS Write: 102 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 790 msec
OK
39
Time taken: 15.681 seconds, Fetched: 1 row(s)
hive>
```

Creating an internal table in hive:

Creating flight table:

Creating flight_fare table:

Join operation on flight and flight_fare table:


```
hive> select flight.Origin, flight.Dest, flight_fare.fare, flight_fare.Distance from flight join flight_fare on flight.FlightNum=flight_fare.F
lightNum;
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_20230502141316_5a35aed2-d5d5-4c9c-a5fe-681ec37204ea
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/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-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]
2023-05-02 14:13:24     Starting to launch local task to process map join;       maximum memory = 477626368
2023-05-02 14:13:25     Dump the side-table for tag: 1 with group count: 47 into file: file:/tmp/hdp123/639da944-04a1-4db5-9ced-63b7468f8b13/h
ive_2023-05-02_14-13-16_922_5956281314495289346-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
2023-05-02 14:13:25     Uploaded 1 File to: file:/tmp/hdp123/639da944-04a1-4db5-9ced-63b7468f8b13/hive_2023-05-02_14-13-16_922_595628131449528
9346-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (1522 bytes)
2023-05-02 14:13:25     End of local task; Time Taken: 0.96 sec.
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_1683014056661_0001, Tracking URL = http://stud-OptiPlex-3060:8088/proxy/application_1683014056661_0001/
Kill Command = /usr/local/hadoop/bin/hadoop job  -kill job_1683014056661_0001
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2023-05-02 14:13:33,983 Stage-3 map = 0%,  reduce = 0%
2023-05-02 14:13:38,085 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.96 sec
MapReduce Total cumulative CPU time: 1 seconds 960 msec
Ended Job = job_1683014056661_0001
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.96 sec   HDFS Read: 13263 HDFS Write: 1550 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 960 msec
OK
SEA     SJC     5181    697
SEA     PSP     7905    987
SAN     SEA     8781    1050
SEA     GEG     7852    224
TUS     SEA     9009    1216
LAX     SEA     6250    954
```

To find average departure delay per day in 2008:


```
hive> select avg(depdelay) from flight;
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_20230502152931_7721f663-a16b-4624-8b5f-63ad67c70ddf
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_1683021424703_0001, Tracking URL = http://stud-OptiPlex-3060:8088/proxy/application_1683021424703_0001/
Kill Command = /usr/local/hadoop/bin/hadoop job  -kill job_1683021424703_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-02 15:29:39,834 Stage-1 map = 0%,  reduce = 0%
2023-05-02 15:29:43,943 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.15 sec
2023-05-02 15:29:49,020 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.63 sec
MapReduce Total cumulative CPU time: 2 seconds 630 msec
Ended Job = job_1683021424703_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.63 sec   HDFS Read: 16632 HDFS Write: 118 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 630 msec
OK
10.021739130434783
Time taken: 20.024 seconds, Fetched: 1 row(s)
hive>
```