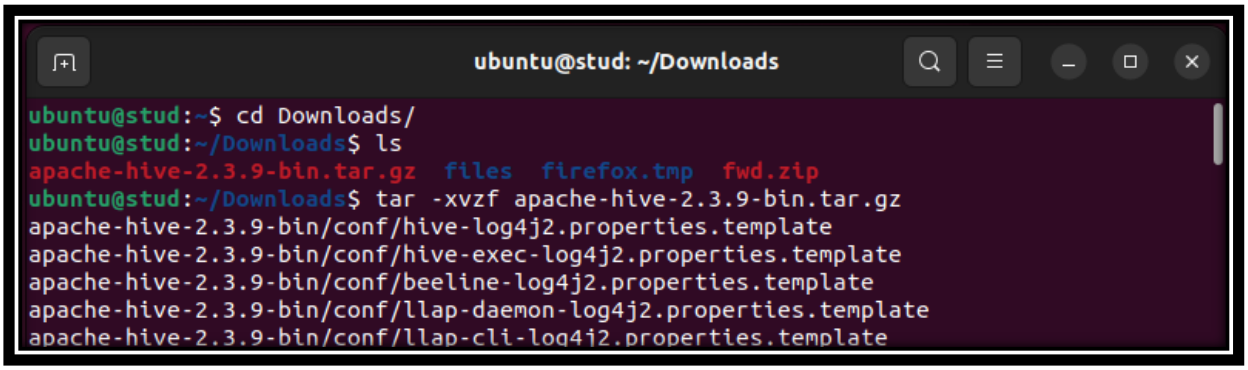
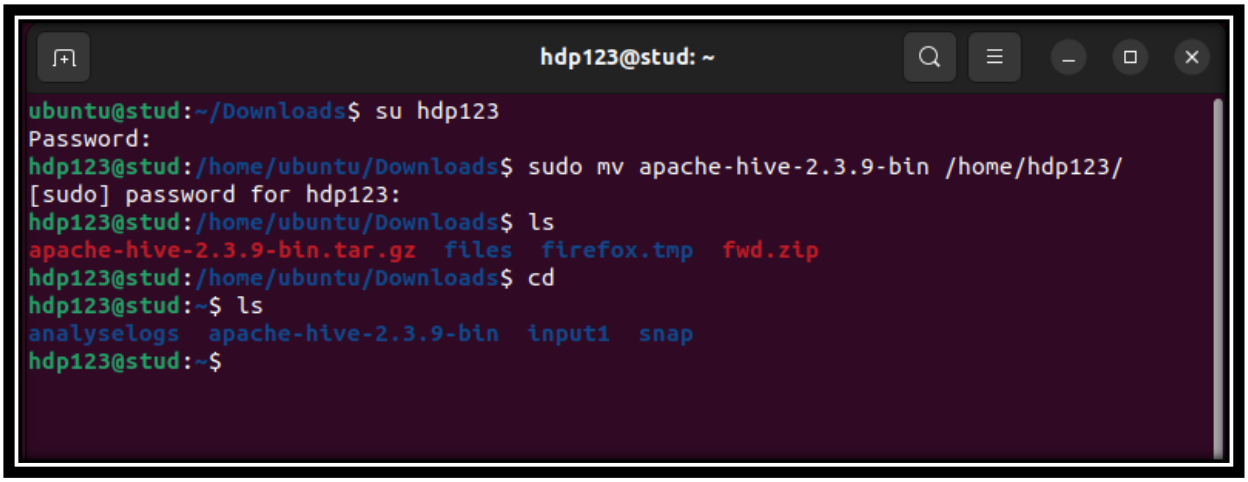
- **Installing Hive :**

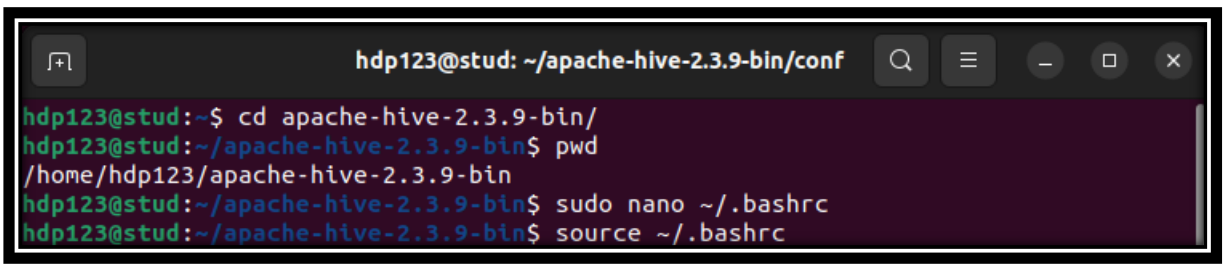
1. **Extracting the apache-hive-2.3.9-bin.tar.gz**



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



3. **Editing the bashrc file by adding HIVE_HOME Path :**



```
    . /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=/home/hdp123/apache-hive-2.3.9-bin
```

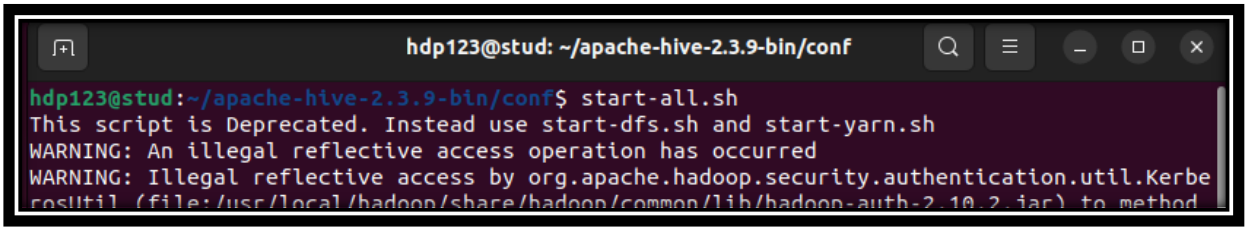
4. Renaming the Hive-env and Hive-default files :

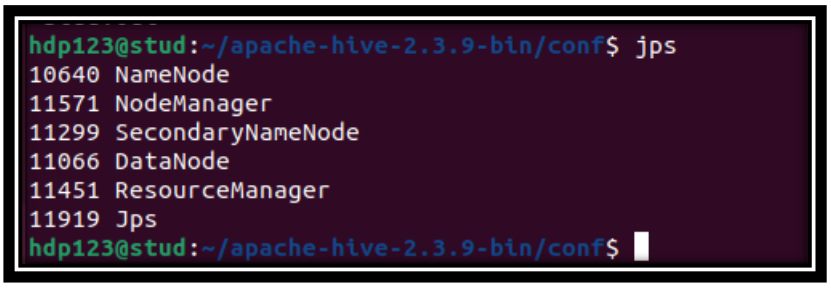
```
hdp123@stud: ~/apache-hive-2.3.9-bin/conf
hdp123@stud:~/apache-hive-2.3.9-bin$ cd conf/
hdp123@stud:~/apache-hive-2.3.9-bin/conf$ sudo mv hive-default.xml.template hive-site.xml
```

```
hdp123@stud: ~/apache-hive-2.3.9-bin/conf
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$
```

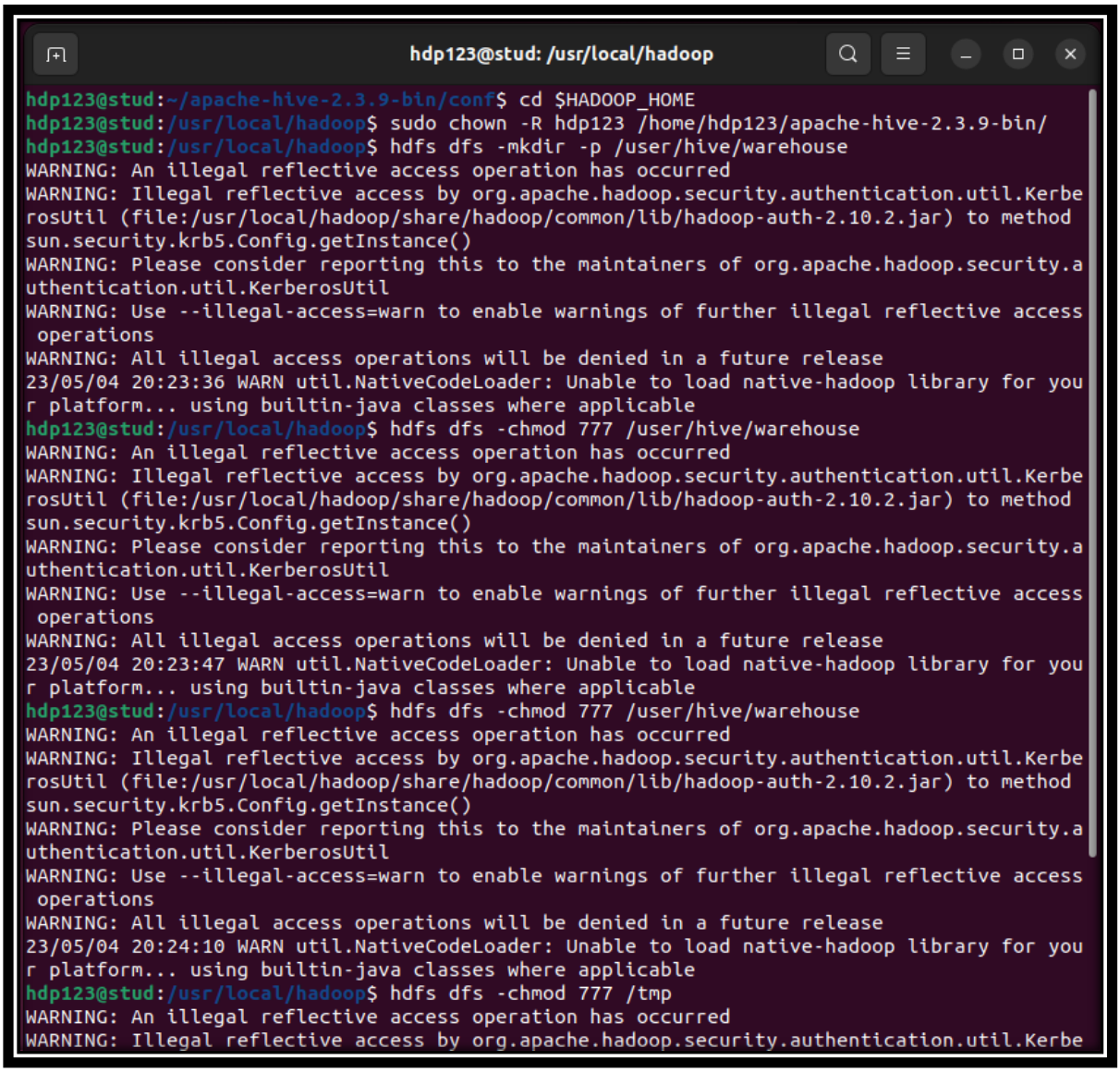
```
hdp123@stud: ~/apache-hive-2.3.9-bin/conf
 GNU nano 6.2                          hive-env.sh *
# export HADOOP_HEAPSIZE=1024
#
# Larger heap size may be required when running queries over large number of files or par>
# By default hive shell scripts use a heap size of 256 (MB).  Larger heap size would also>
# appropriate for hive server.


# Set HADOOP_HOME to point to a specific hadoop install directory
# HADOOP_HOME=${bin}/../../hadoop
export HADOOP_HOME=/usr/local/hadoop

^G Help      ^O Write Out  ^W Where Is   ^K Cut       ^T Execute    ^C Location
^X Exit      ^R Read File  ^\ Replace    ^U Paste     ^J Justify    ^/ Go To Line
```

5. Starting the Hadoop Services :

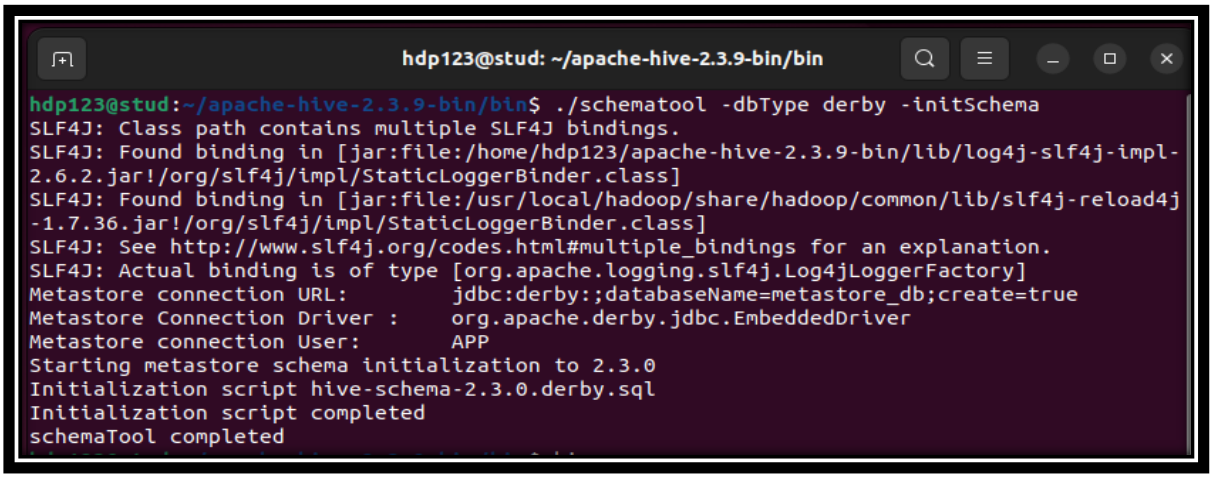




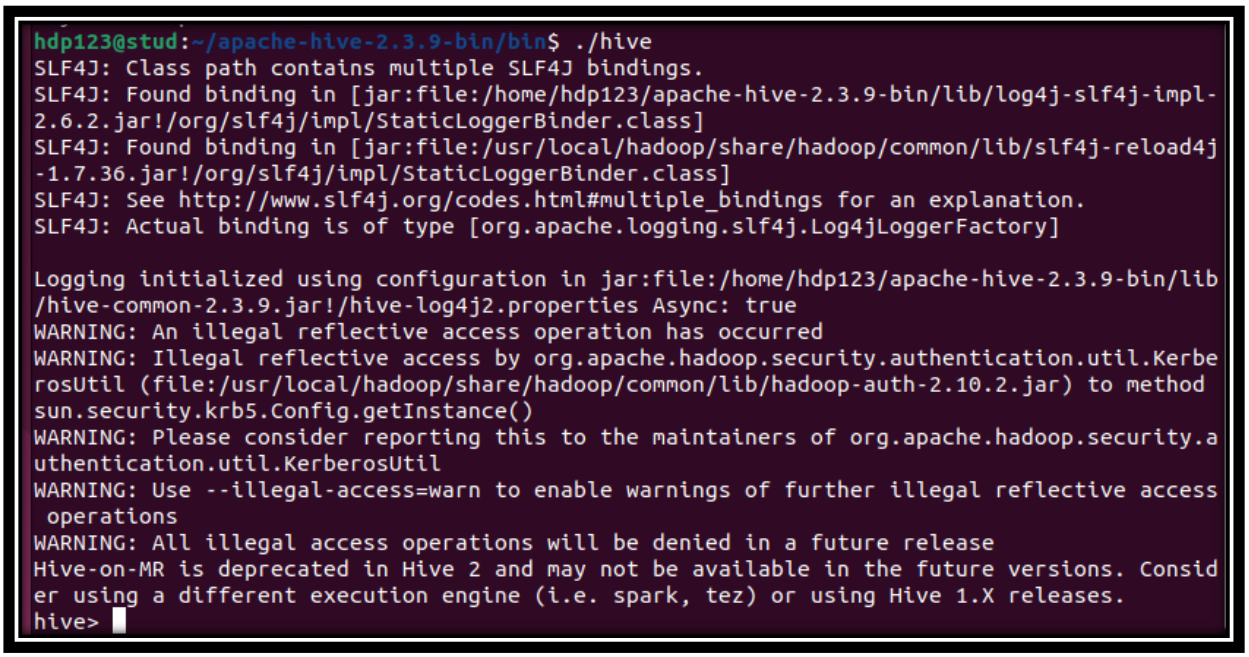
6. Creating files and assigning access to those files :



7. Initializing the Schema-tool :

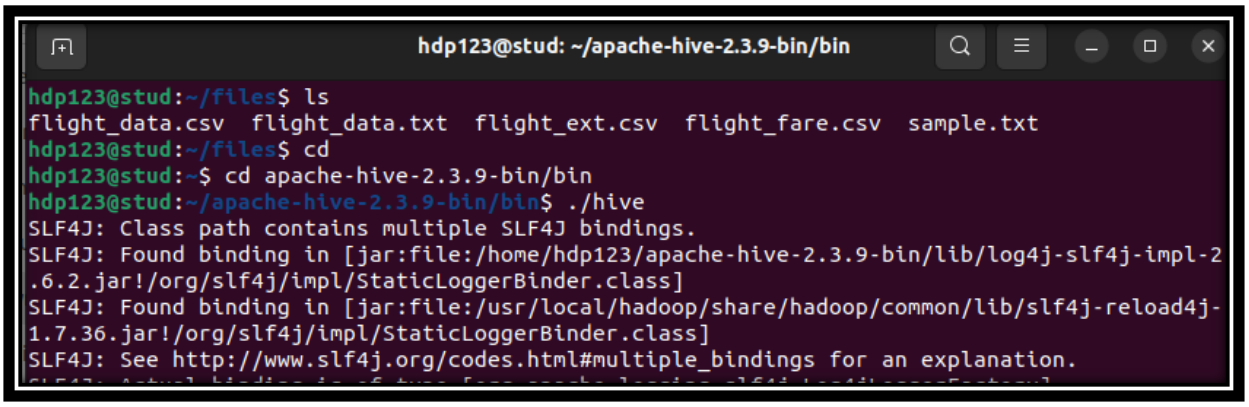


8. Starting Hive :

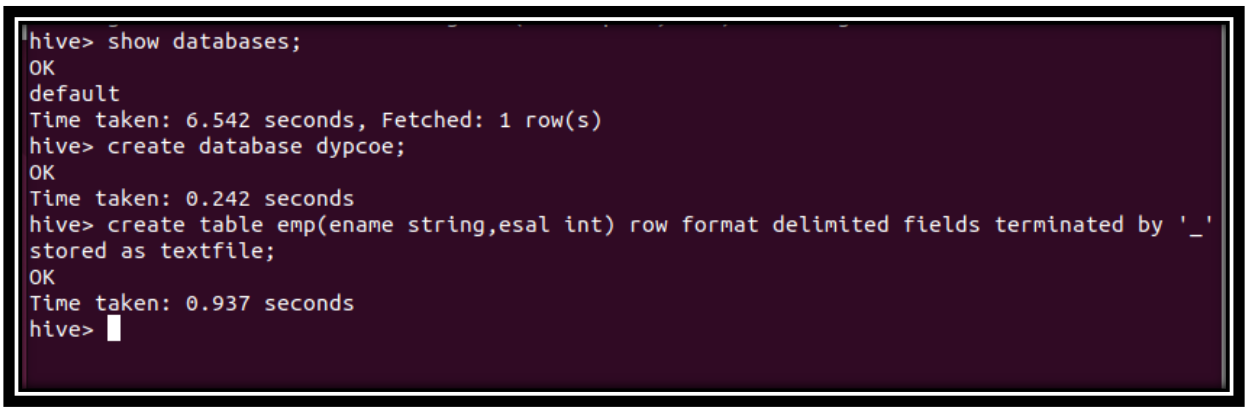


- # Running Hive Queries :

1. Staring Hive Services :



2. Creating a database and emp table :



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

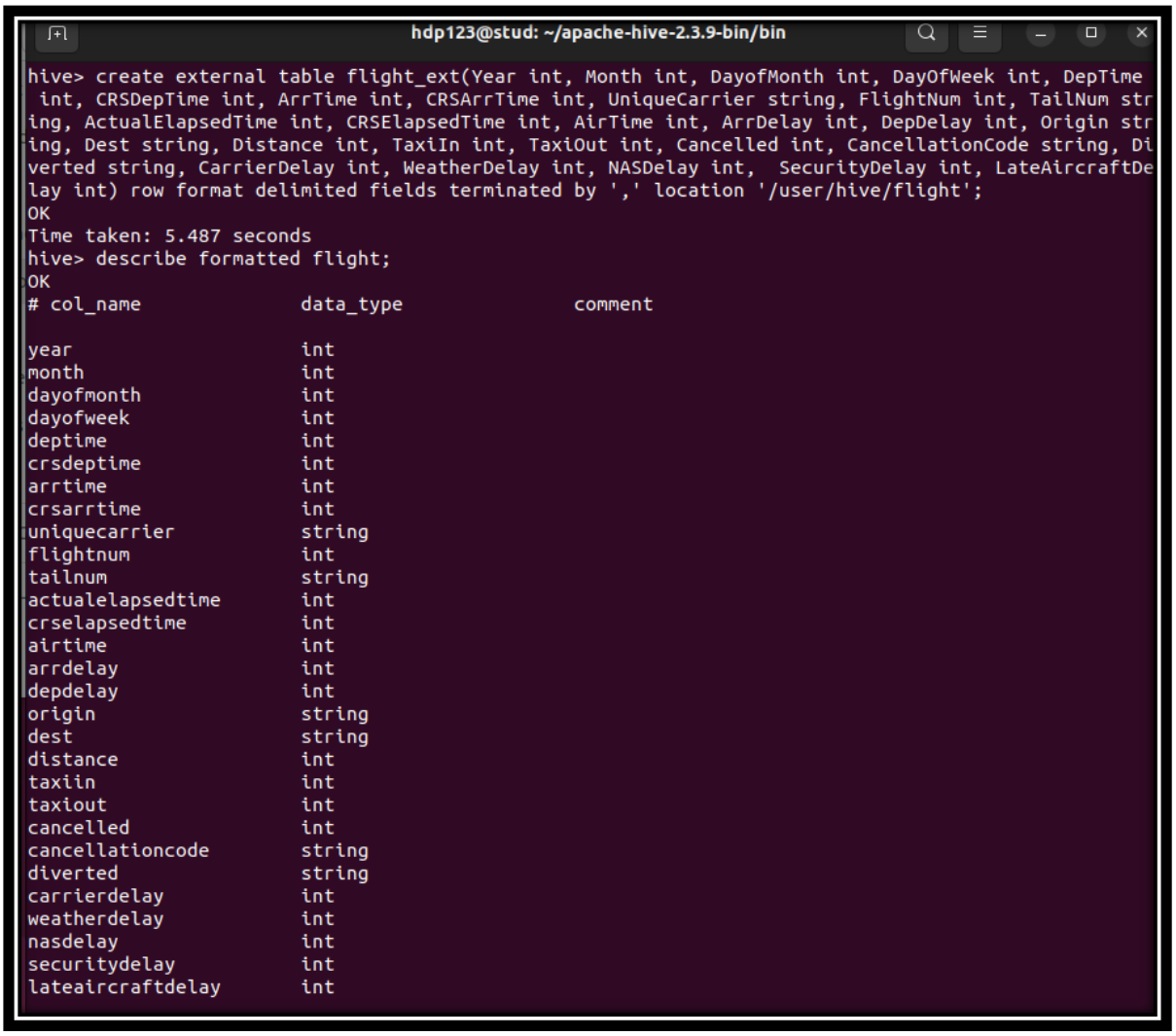
4. **Creating flight table & loading data from flight_data into table :**

```
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 str
ing, 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 :**

```
hdp123@stud: /usr/local/hadoop
hdp123@stud: ~/apache-hive-2.3.9-bin/bin   hdp123@stud: /usr/local/hadoop
hdp123@stud:~$ cd $HADOOP_HOME
hdp123@stud:/usr/local/hadoop$ hdfs dfs -mkdir /user/hive/flight
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-2.10.2.jar) to method sun.security.krb5.
Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentica
tion.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operatio
ns
WARNING: All illegal access operations will be denied in a future release
23/05/04 23:11:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfor
m... using builtin-java classes where applicable
hdp123@stud:/usr/local/hadoop$ hdfs dfs -put /home/hdp123/files/flight_ext.csv /user/hive/flight
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-2.10.2.jar) to method sun.security.krb5.
Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentica
tion.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operatio
ns
WARNING: All illegal access operations will be denied in a future release
23/05/04 23:12:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfor
m... using builtin-java classes where applicable
hdp123@stud:/usr/local/hadoop$ hdfs dfs -ls /user/hive/flight
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-2.10.2.jar) to method sun.security.krb5.
Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentica
tion.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operatio
ns
WARNING: All illegal access operations will be denied in a future release
23/05/04 23:12:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfor
m... using builtin-java classes where applicable
Found 1 items
-rw-r--r--   1 hdp123 supergroup       3842 2023-05-04 23:12 /user/hive/flight/flight_ext.csv
hdp123@stud:/usr/local/hadoop$
```

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



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 23:18:57,180 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.87 sec
2023-05-04 23:19:05,716 Stage-1 map = 100%,  reduce = 100%, 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.Flig
htNum=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/applica
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
TUS     SEA     9009
```

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
S
Total MapReduce CPU Time Spent: 4 seconds 280 msec
OK
10.021739130434783
Time taken: 26.331 seconds, Fetched: 1 row(s)
```