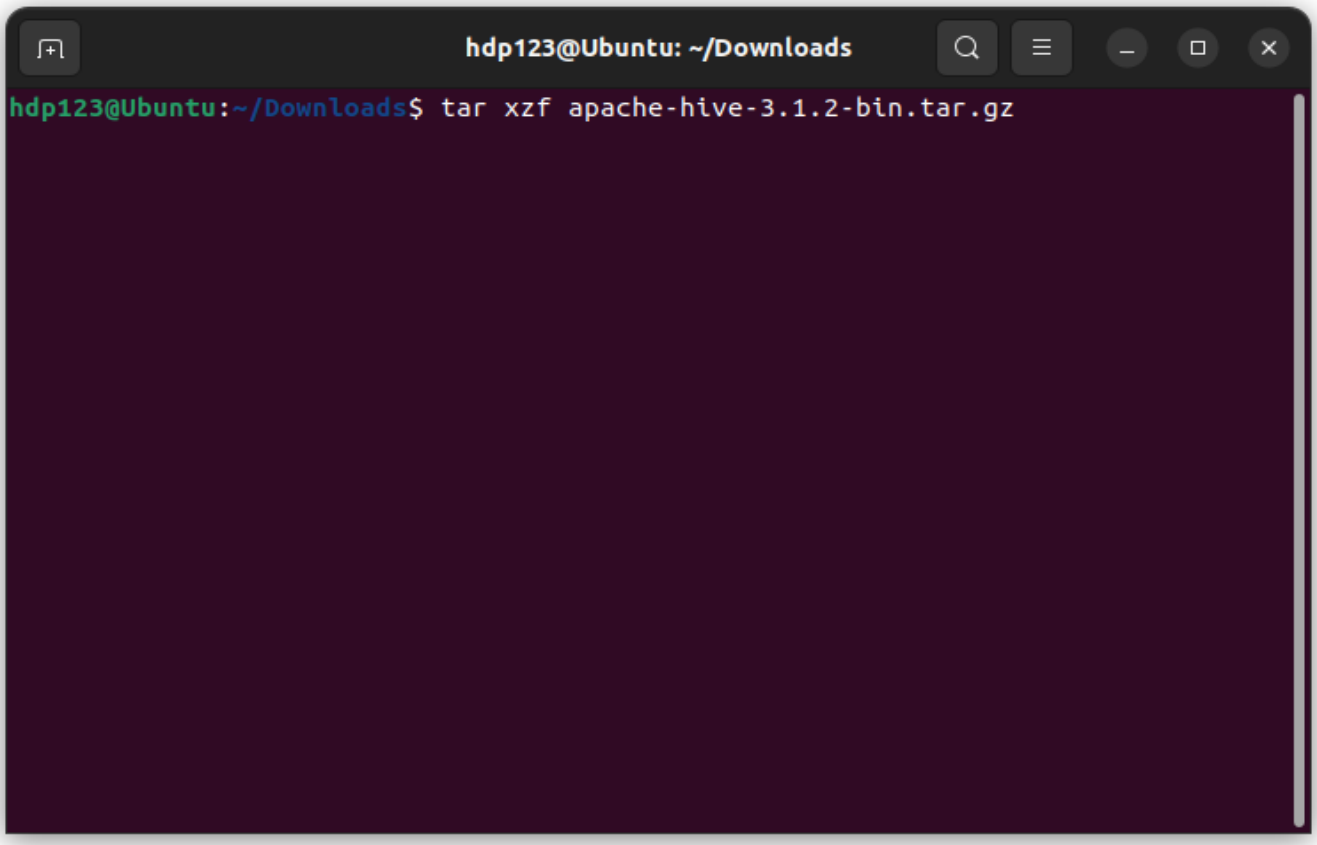
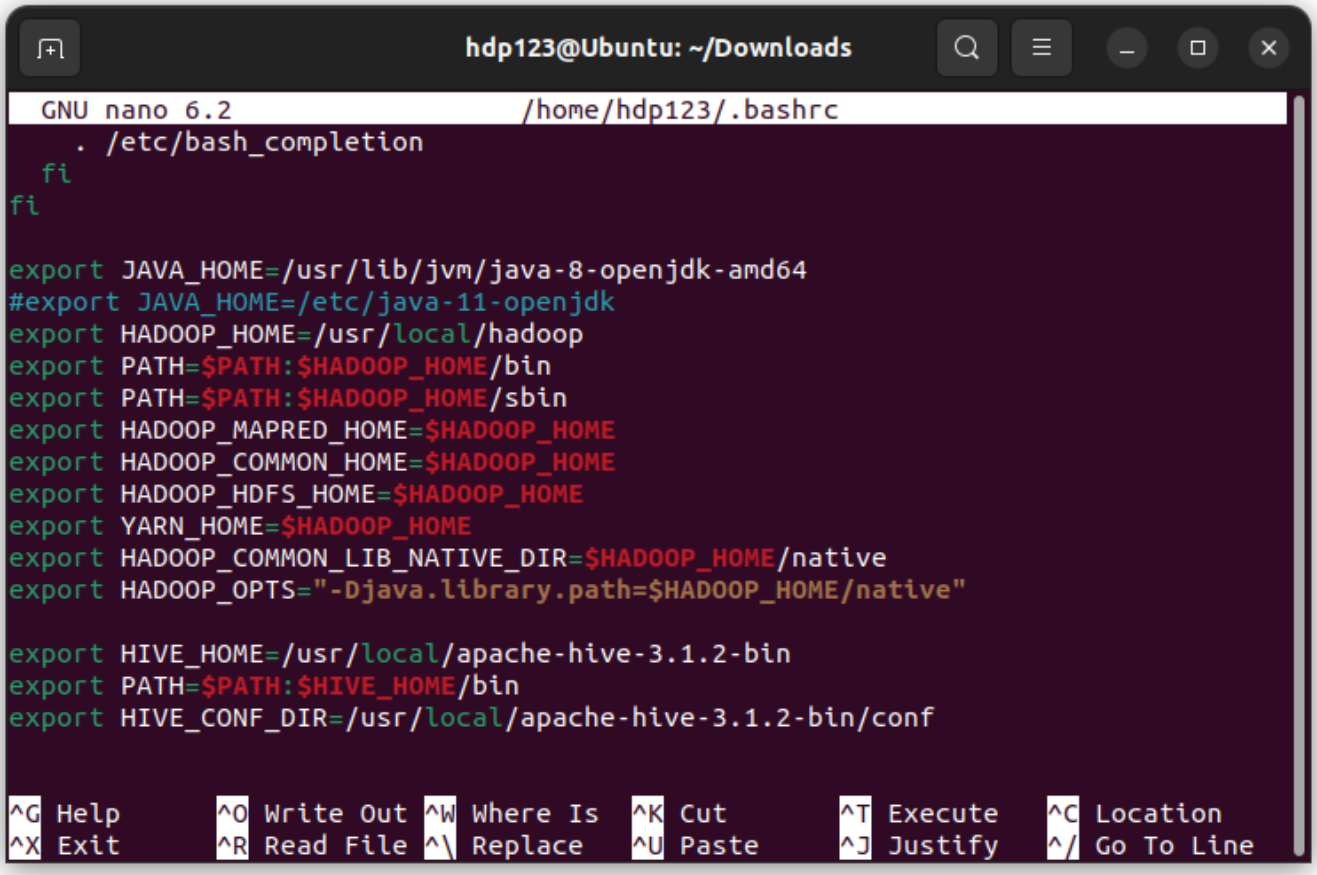
Output –

Downloading Hive -

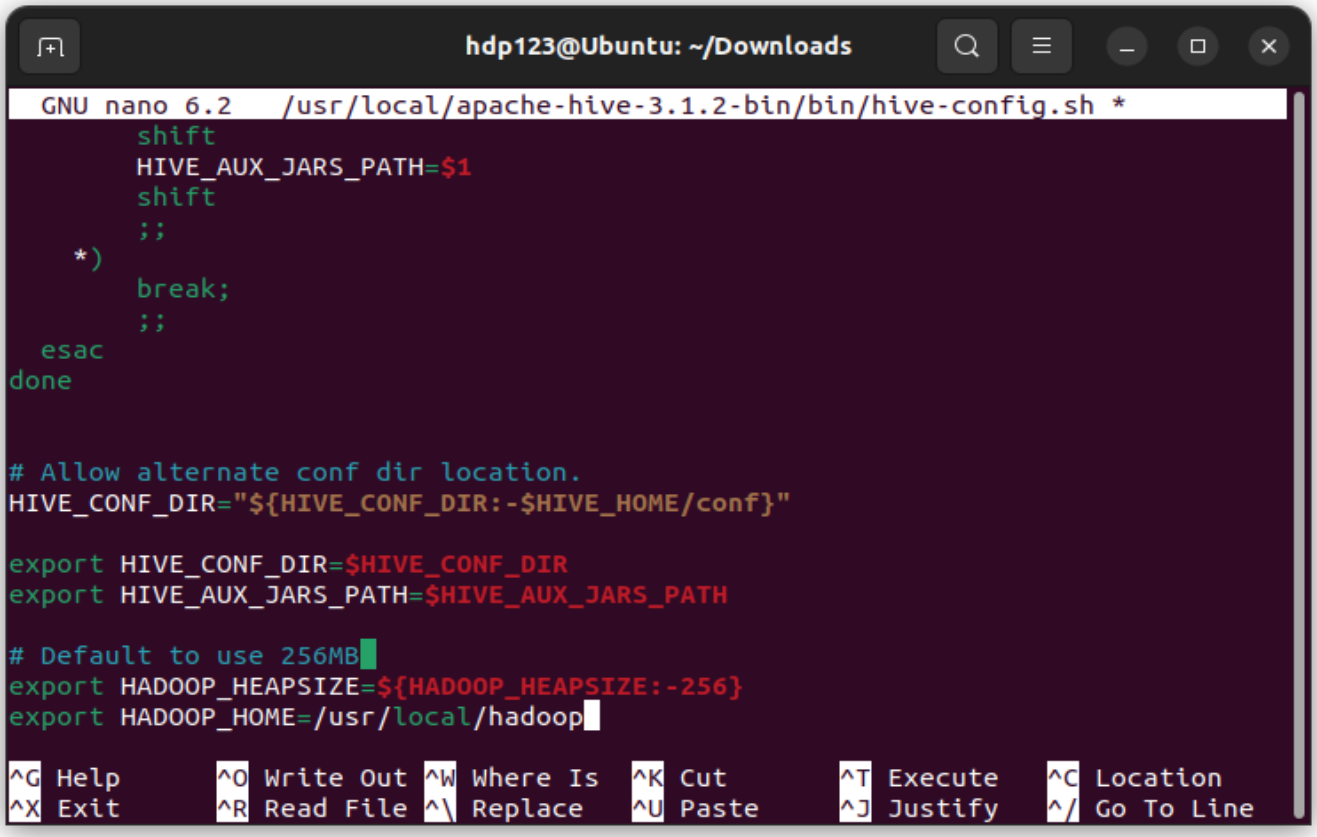
Extracting Hive -



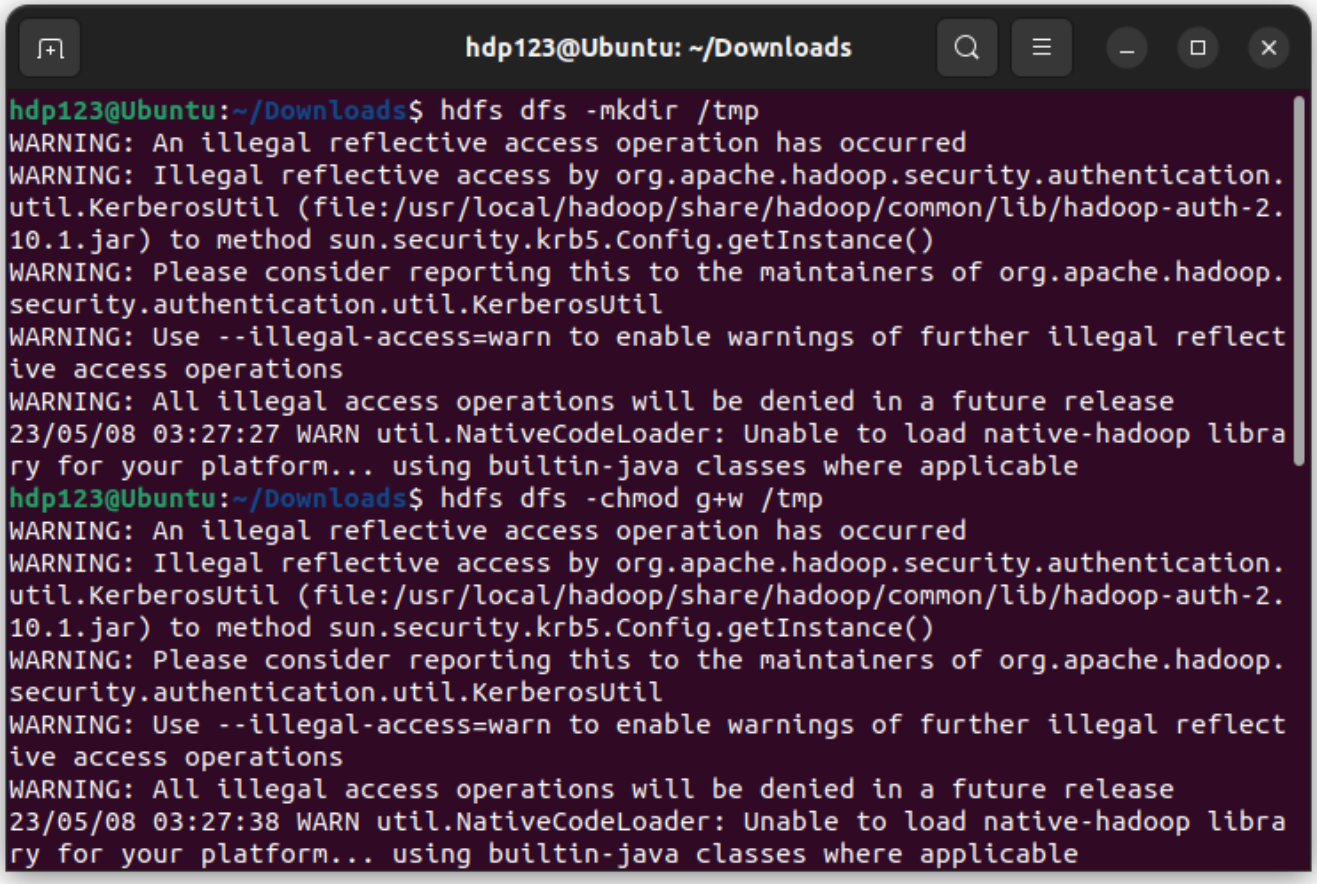
sudo nano ~/.bashrc

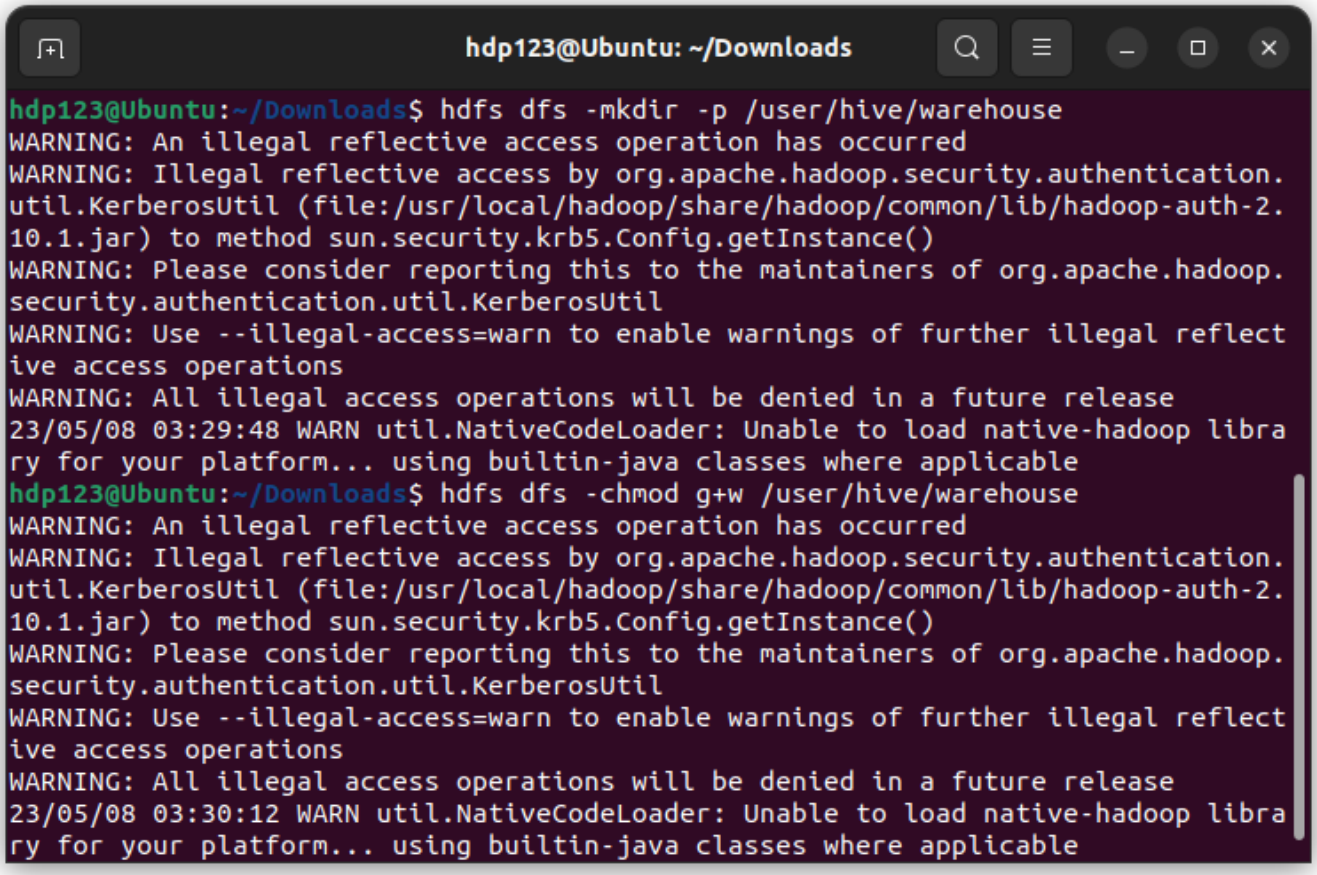


sudo nano $HIVE_HOME/bin/hive-config.sh

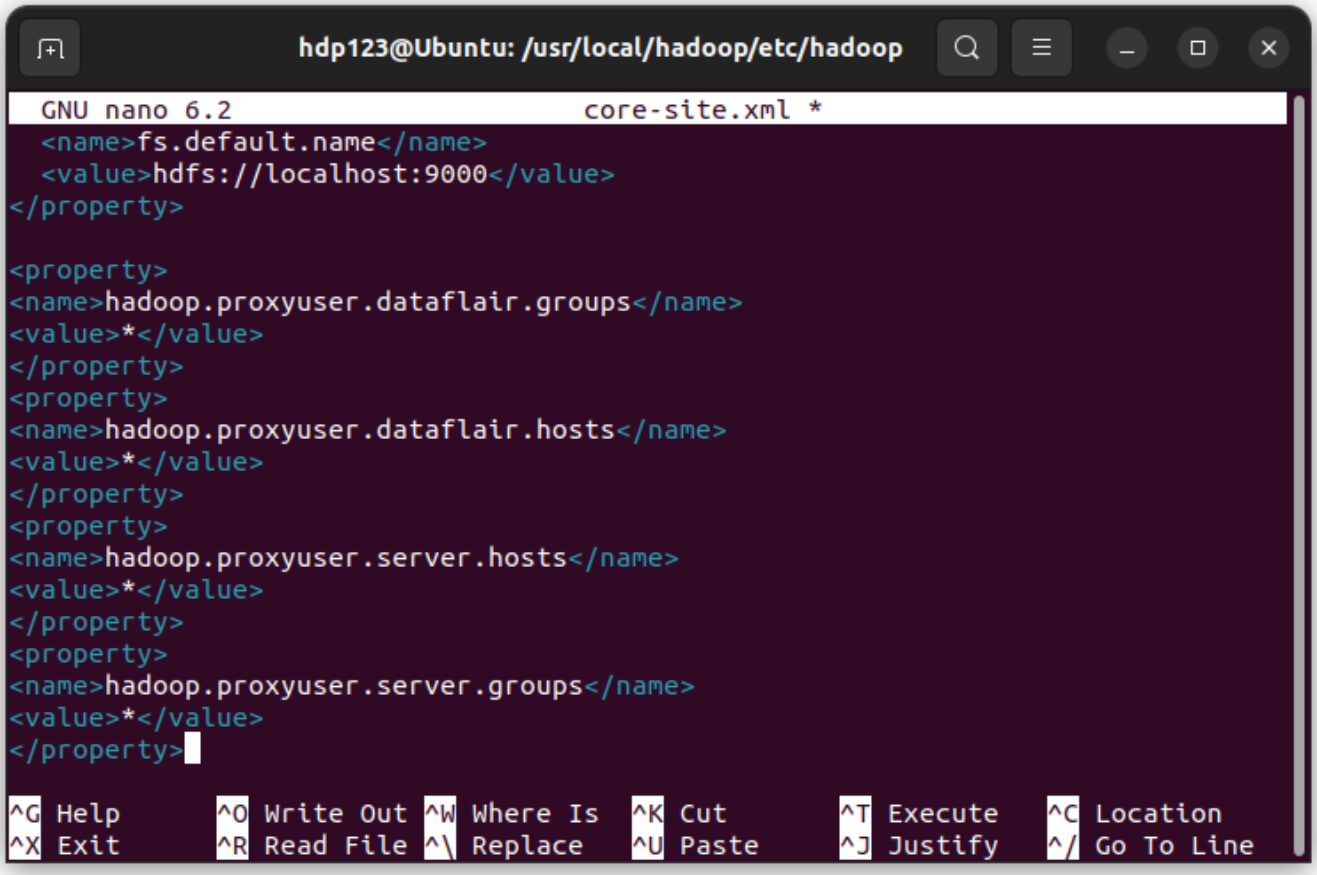


Making files for Hive in HDFS -

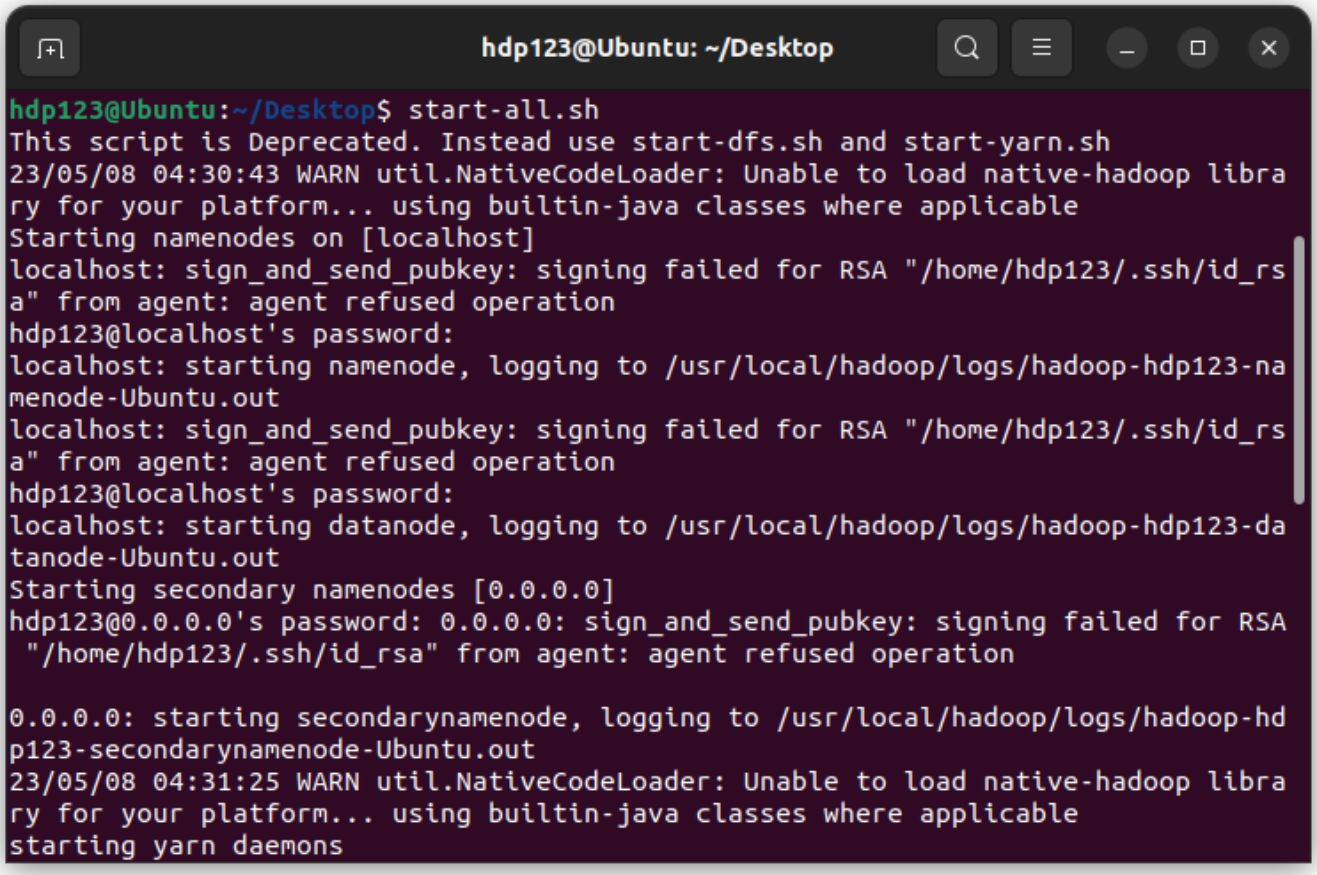




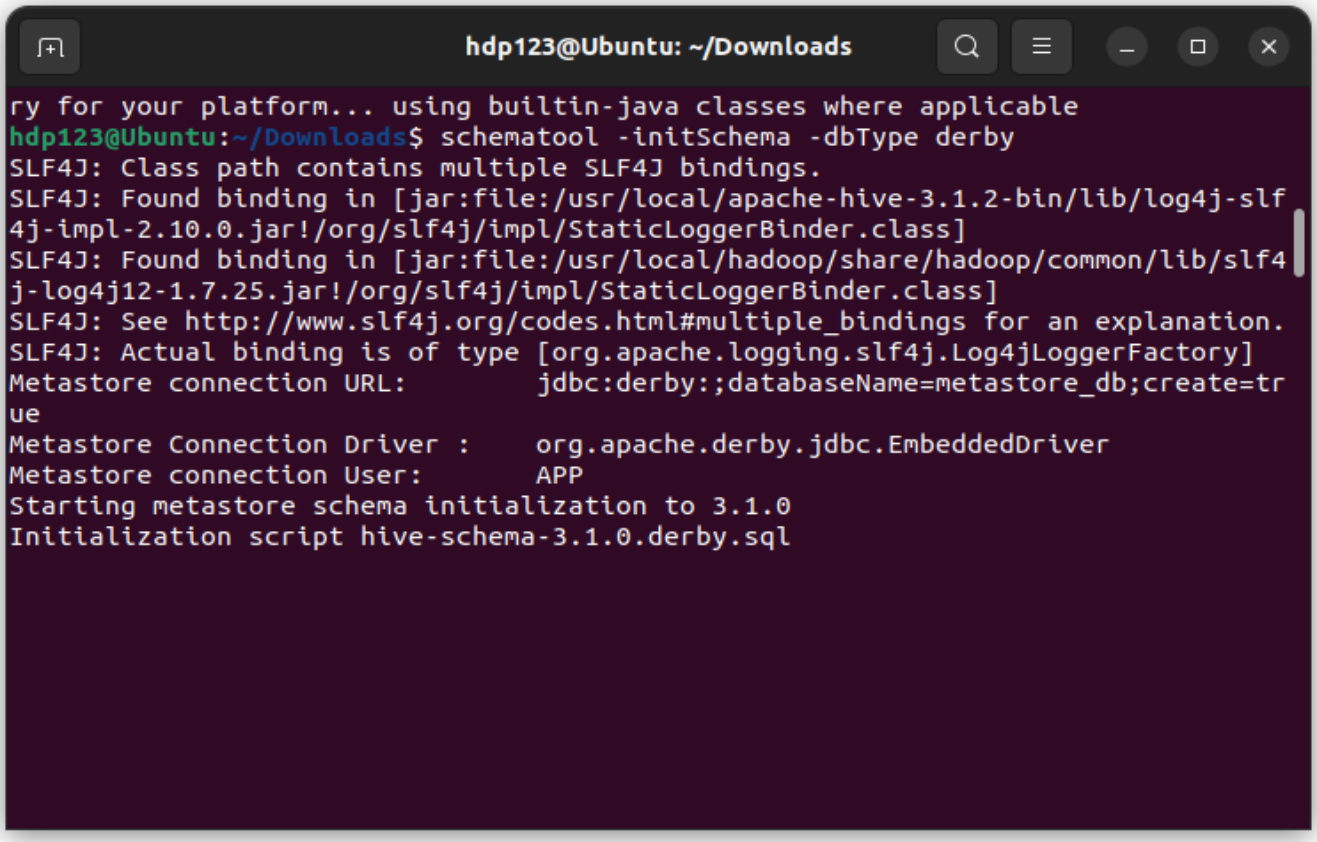
sudo nano core-site.xml

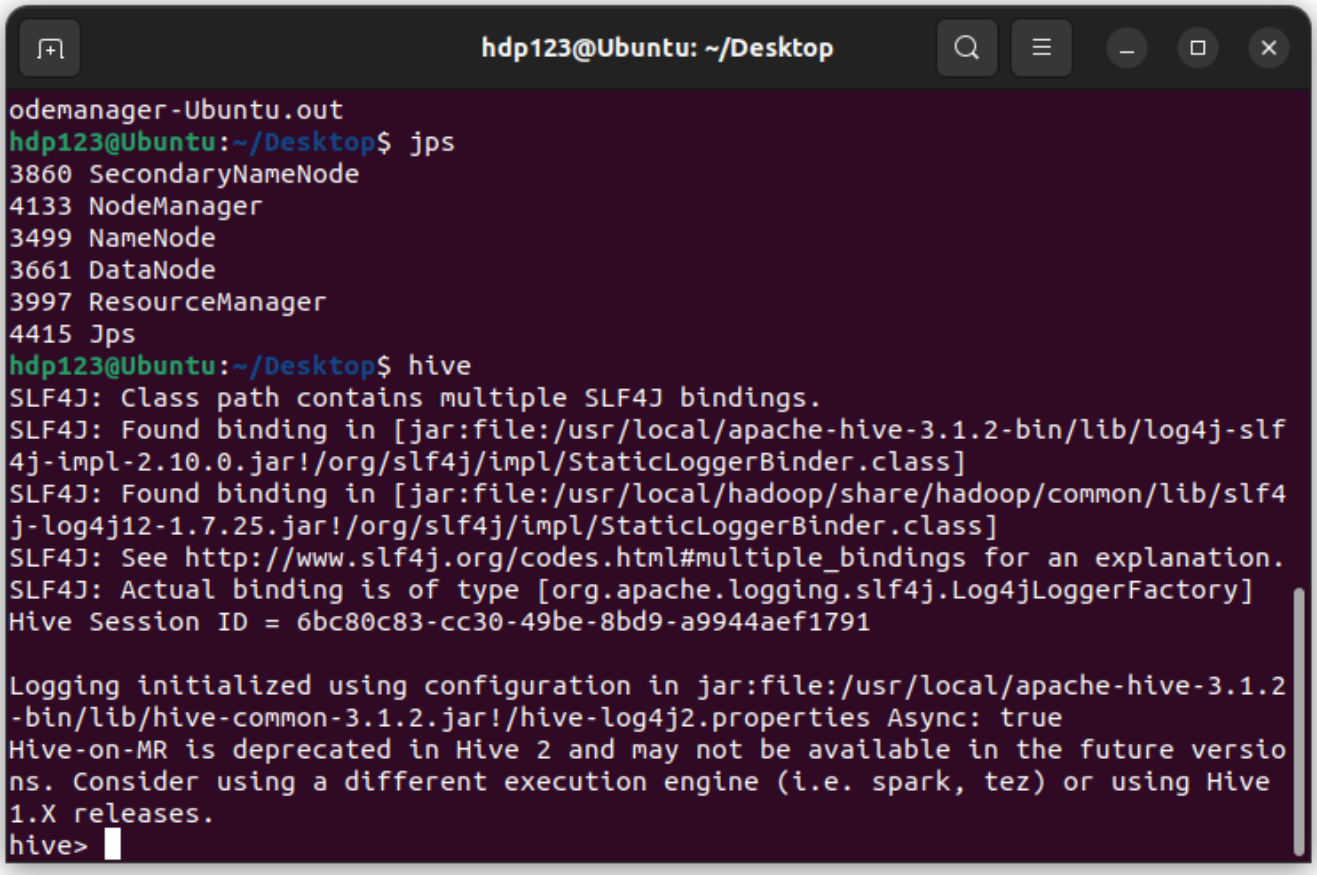


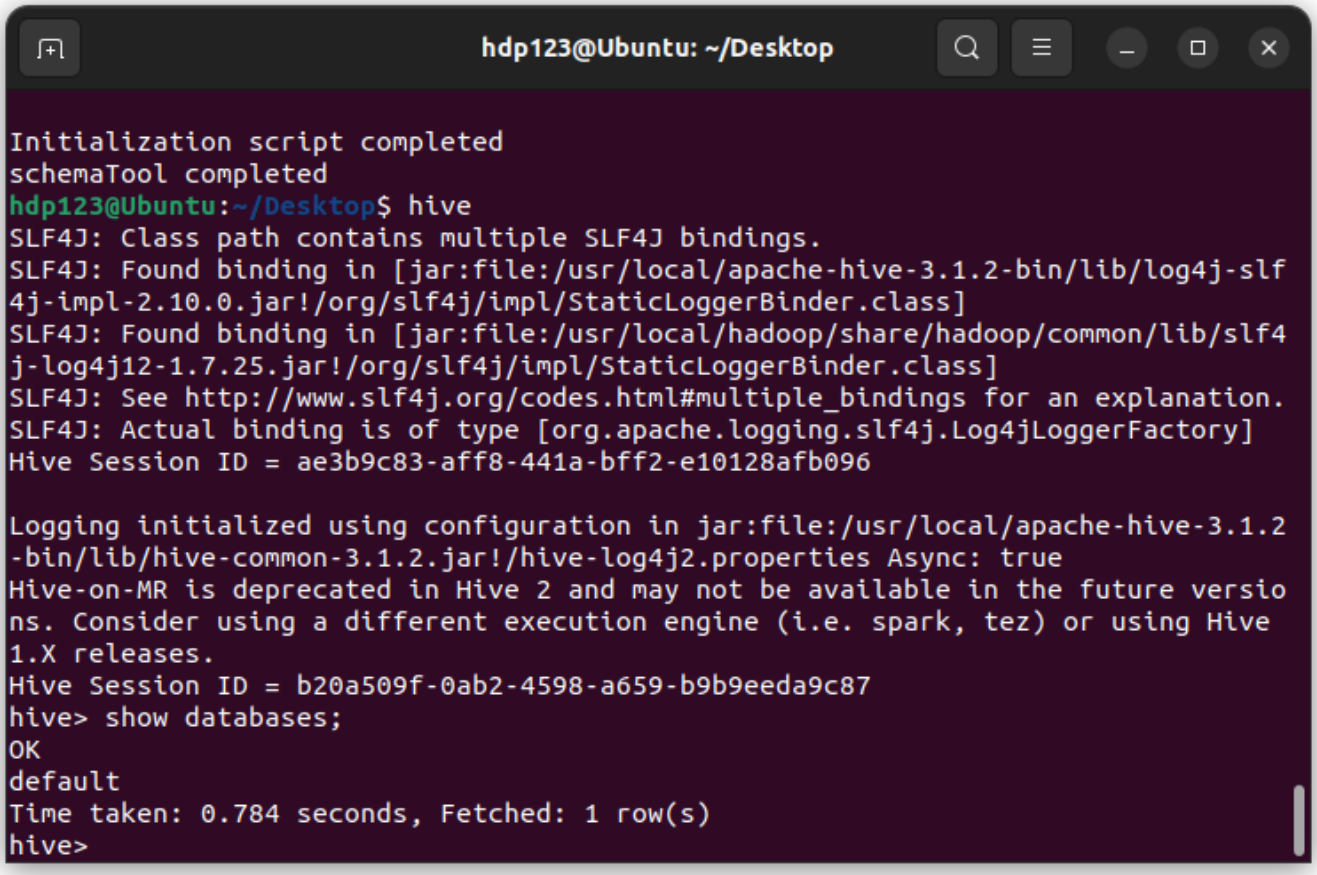
Starting Hadoop -



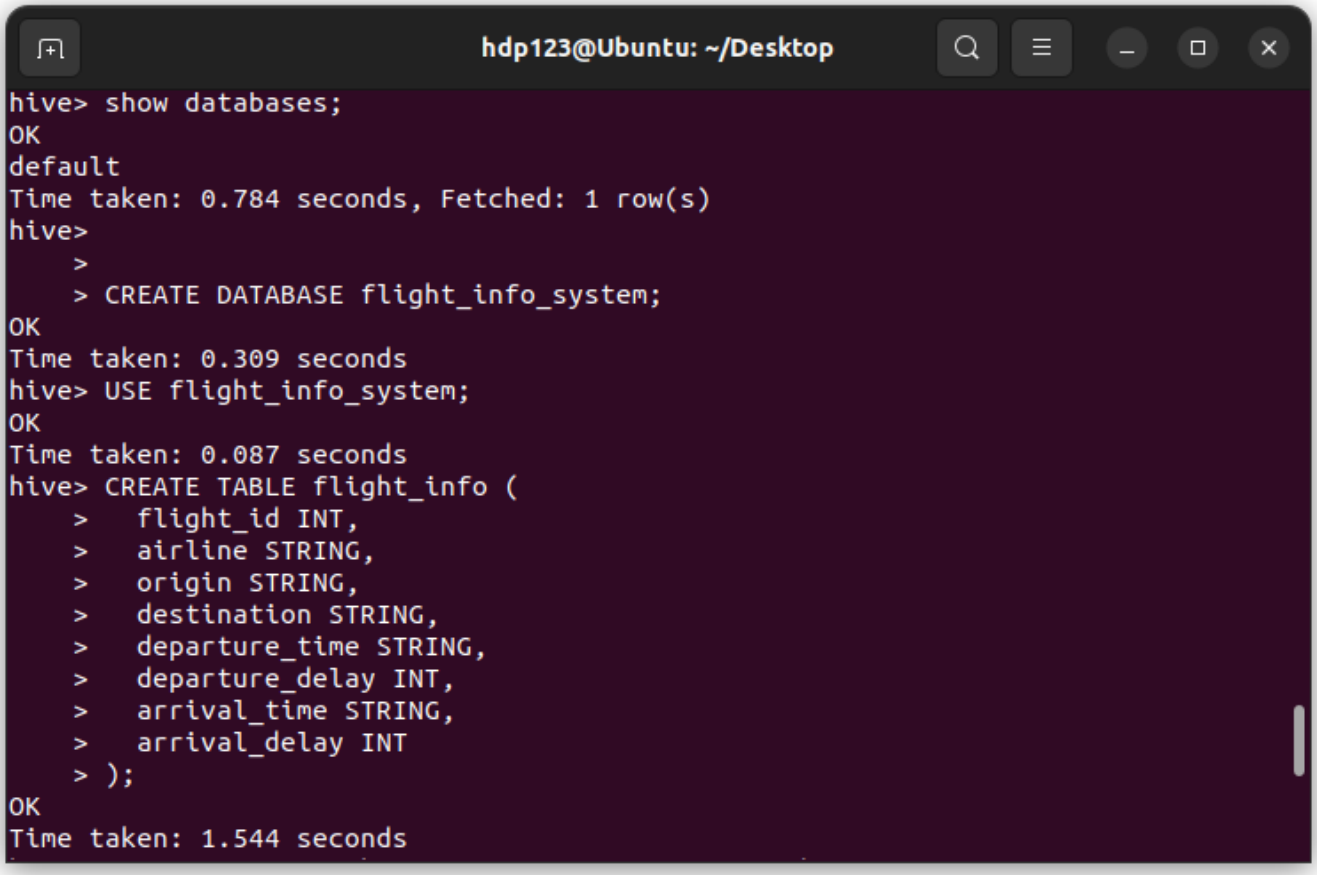
Starting schematool and Hive -

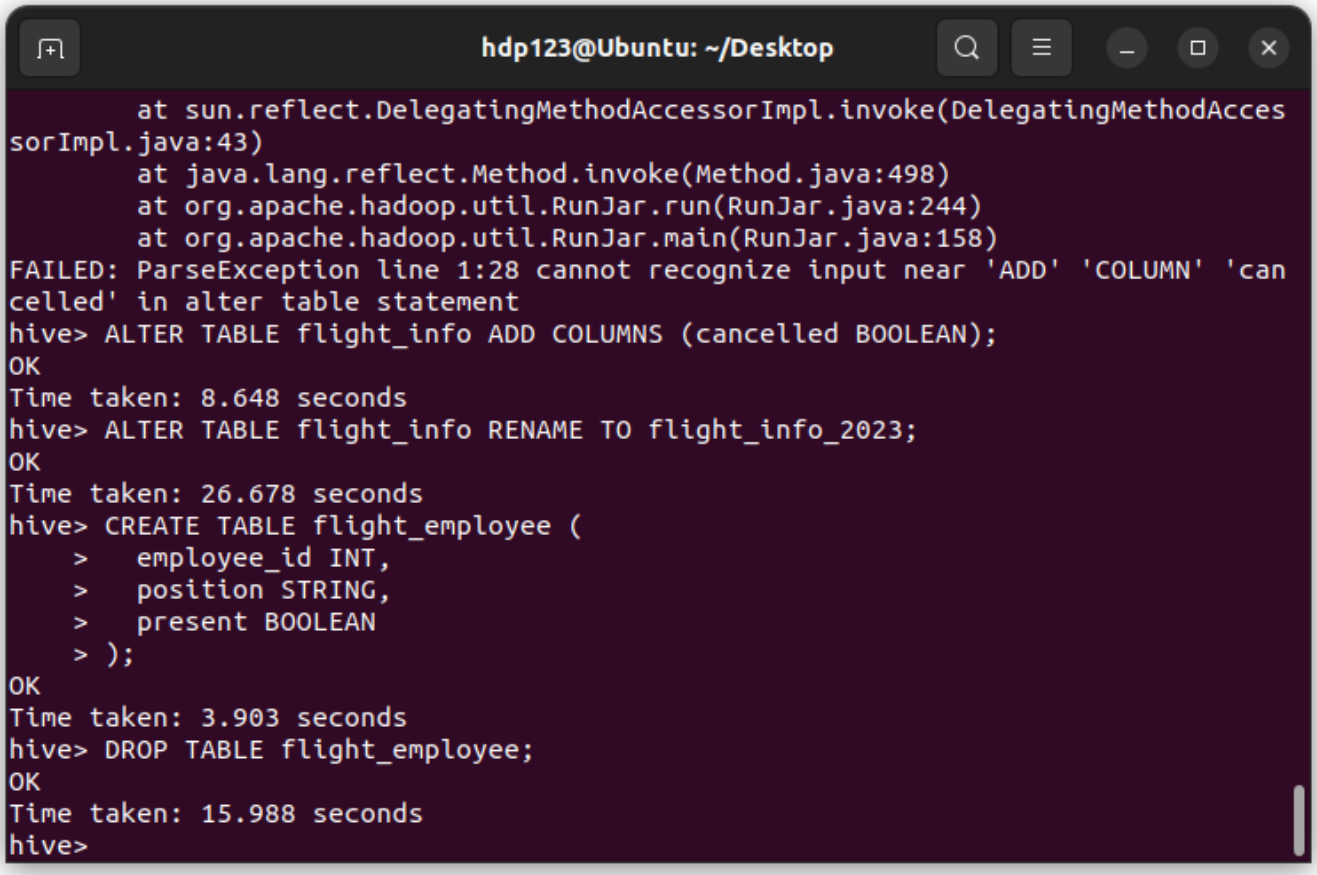



hdp123@Ubuntu: ~/Desktop
odemanager-Ubuntu.out
hdp123@Ubuntu:~/Desktop$ jps
3860 SecondaryNameNode
4133 NodeManager
3499 NameNode
3661 DataNode
3997 ResourceManager
4415 Jps
hdp123@Ubuntu:~/Desktop$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/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]
Hive Session ID = 6bc80c83-cc30-49be-8bd9-a9944aef1791
Logging initialized using configuration in jar:file:/usr/local/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>


hdp123@Ubuntu: ~/Desktop
Initialization script completed
schemaTool completed
hdp123@Ubuntu:~/Desktop$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/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:/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]
Hive Session ID = ae3b9c83-aff8-441a-bff2-e10128afb096
Logging initialized using configuration in jar:file:/usr/local/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 = b20a509f-0ab2-4598-a659-b9b9eeda9c87
hive> show databases;
OK
default
Time taken: 0.784 seconds, Fetched: 1 row(s)
hive>

A. Creating, Dropping and Altering Database tables -





B. Creating an external Hive table -

```
hdp123@Ubuntu: ~/Desktop
    > );
OK
Time taken: 3.903 seconds
hive> DROP TABLE flight_employee;
OK
Time taken: 15.988 seconds
hive> CREATE EXTERNAL TABLE flight_info_external (
    >   flight_number INT,
    >   origin_airport_id INT,
    >   origin STRING,
    >   destination_airport_id INT,
    >   destination STRING,
    >   departure_time INT,
    >   departure_delay INT,
    >   arrival_time INT,
    >   arrival_delay INT
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > STORED AS TEXTFILE
    > LOCATION '/home/hdp123/Downloads/flightdata';
OK
Time taken: 1.226 seconds
hive>
```

flightdata.txt -

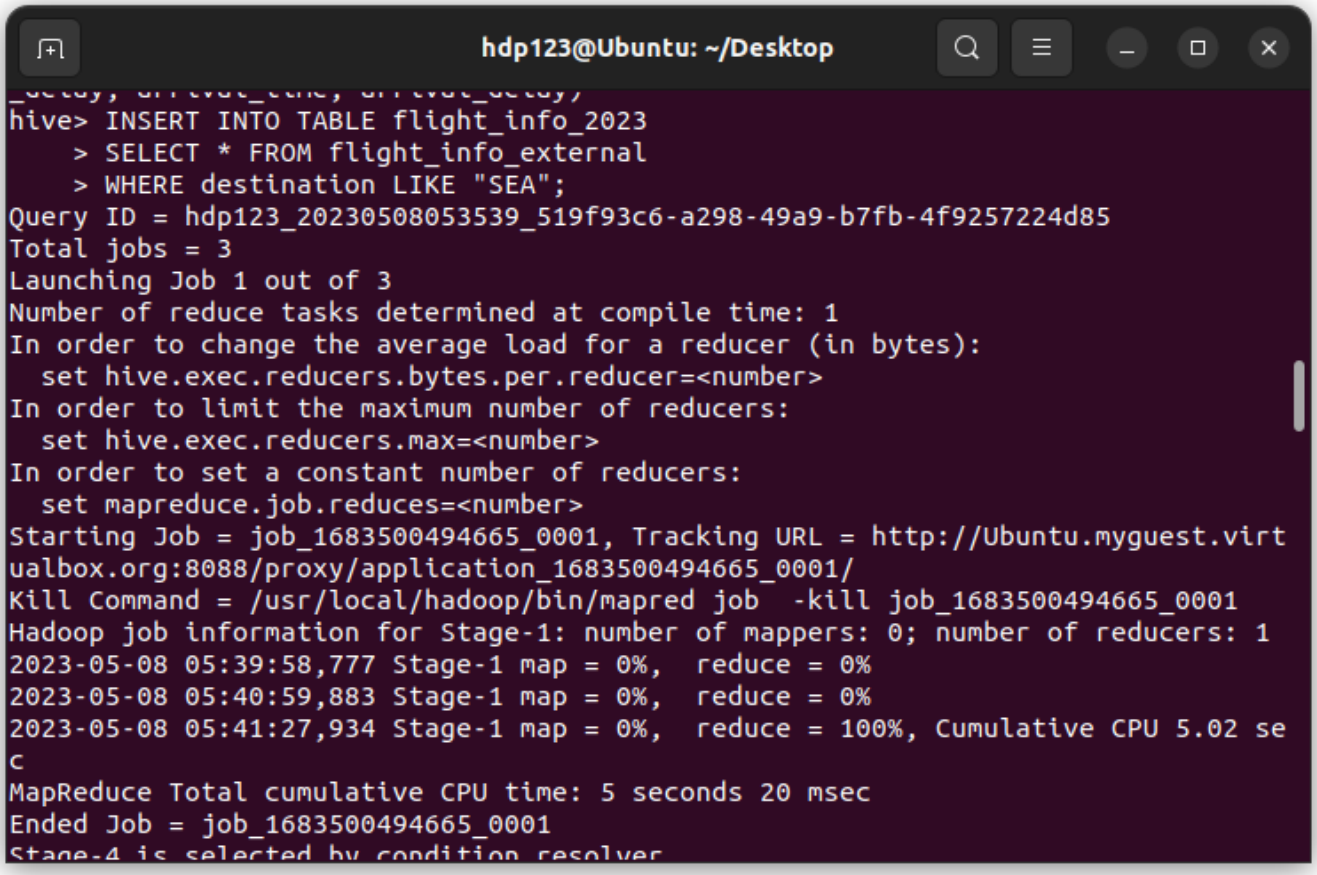
flightdata.txt - Notepad

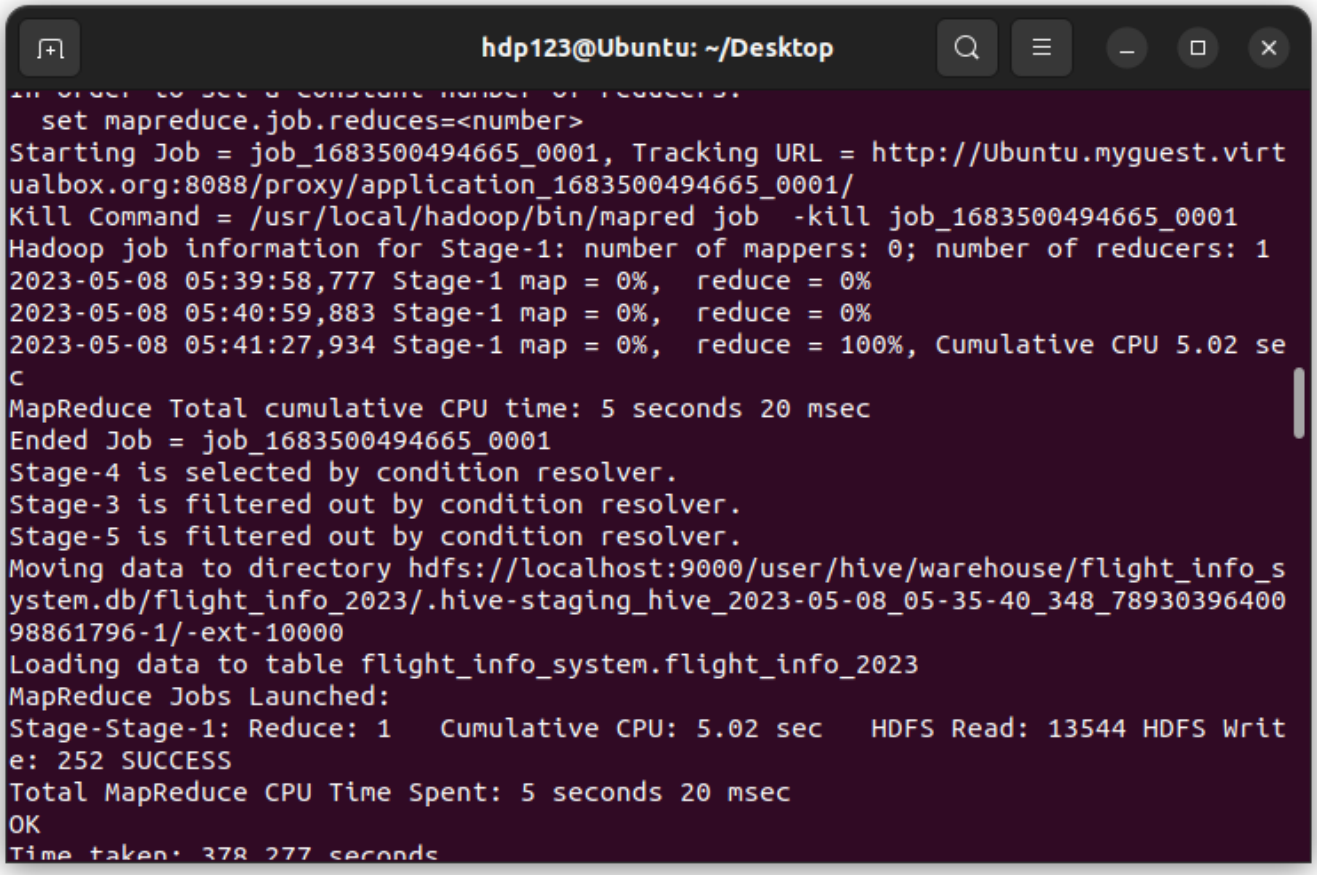
File Edit Format View Help

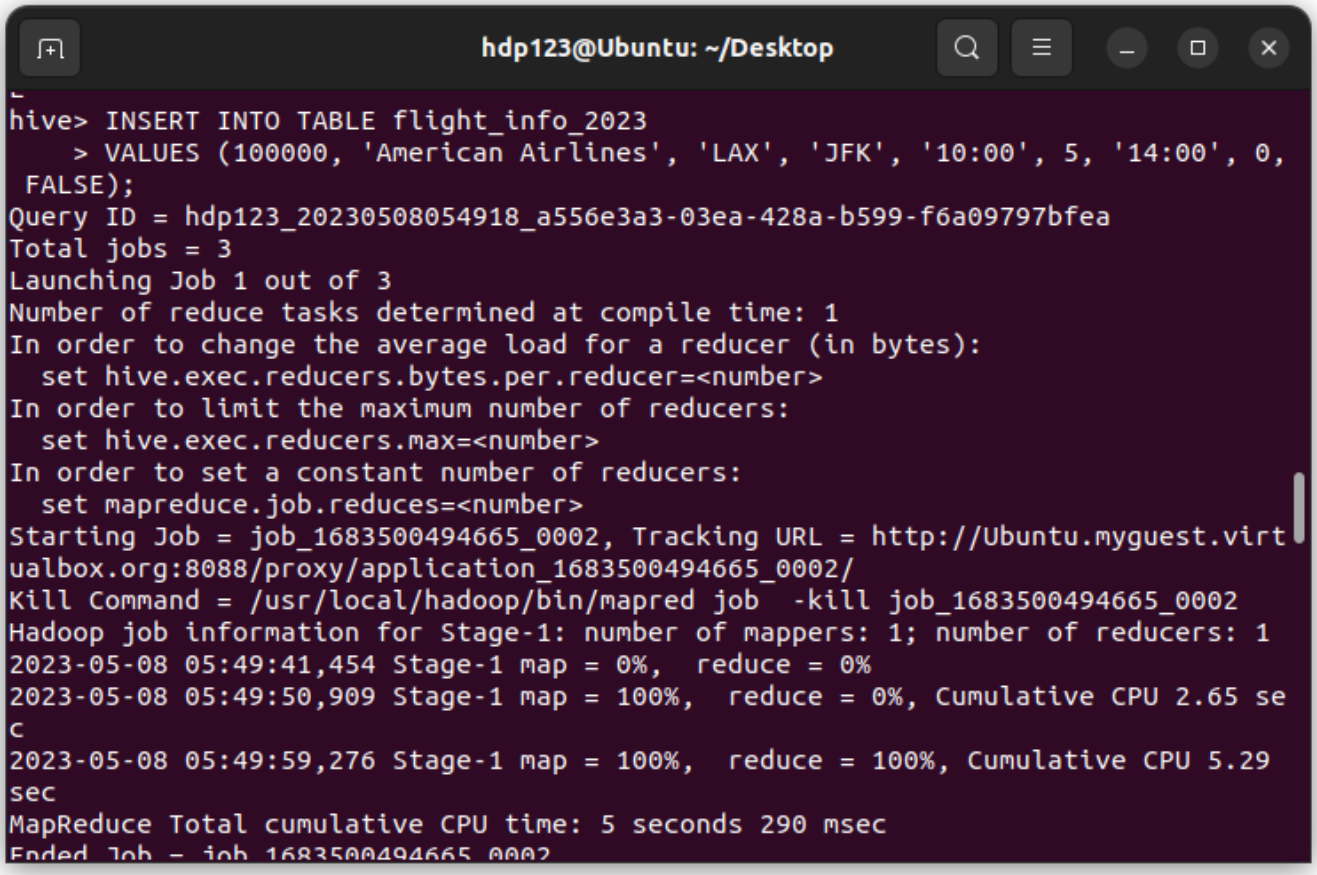
```
FL_NUM,ORIGIN_AIRPORT_ID,ORIGIN,DEST_AIRPORT_ID,DEST,DEP_TIME,DEP_DELAY,ARR_TIME,ARR_DELAY
1399,10397,ATL,14747,SEA,1907,2,2102,-41
1476,11433,DTW,13487,MSP,1344,-1,1439,4
1597,10397,ATL,14747,SEA,942,2,1142,-33
1768,14747,SEA,13487,MSP,820,1,1345,10
1823,14747,SEA,11433,DTW,2256,-4,615,8
1975,13487,MSP,10397,ATL,1127,-2,1441,-18
2074,10397,ATL,13487,MSP,1745,0,1920,-11
2151,13487,MSP,14747,SEA,1751,11,1908,-21
2221,13487,MSP,14747,SEA,1115,0,1255,-10
2291,13487,MSP,10397,ATL,1443,13,1800,-1
2350,10397,ATL,12478,JFK,828,3,1029,-9
2444,10397,ATL,14747,SEA,1355,10,1605,-16
2610,10397,ATL,13487,MSP,721,-4,903,-1
2826,11433,DTW,14747,SEA,841,6,1023,-24
2845,11433,DTW,10397,ATL,1622,-2,1805,-25
86,13487,MSP,11433,DTW,1337,-8,1616,-4
423,12478,JFK,10397,ATL,1258,-2,1519,-19
1823,11433,DTW,14747,SEA,1724,-4,1905,-24
1972,14747,SEA,12478,JFK,715,0,1519,-11
2005,10397,ATL,11433,DTW,723,-2,920,0
2039,10397,ATL,14747,SEA,847,23,1101,2
902,10397,ATL,11433,DTW,1454,-2,1643,-12
997,13487,MSP,11433,DTW,832,-1,1105,-11
1108,13487,MSP,10397,ATL,1510,2,1818,-17
1231,10397,ATL,11433,DTW,1339,-1,1524,-13
1291,14747,SEA,10397,ATL,750,5,1528,-1
1397,13487,MSP,14747,SEA,1439,-1,1604,-21
1444,11433,DTW,10397,ATL,2021,1,2211,-16
1491,13487,MSP,10397,ATL,608,-2,922,-18
1573,14747,SEA,10397,ATL,4,9,736,4
1646,10397,ATL,14747,SEA,1621,1,1838,-21
1734,13487,MSP,11433,DTW,1151,-1,1425,-8
1587,11433,DTW,10397,ATL,1350,5,1532,-17
1699,11433,DTW,14747,SEA,1236,21,1437,17
1770,14747,SEA,13487,MSP,1521,1,2048,16
991,10397,ATL,11433,DTW,1205,0,1405,1
1104,10397,ATL,13487,MSP,1219,14,1355,10
```

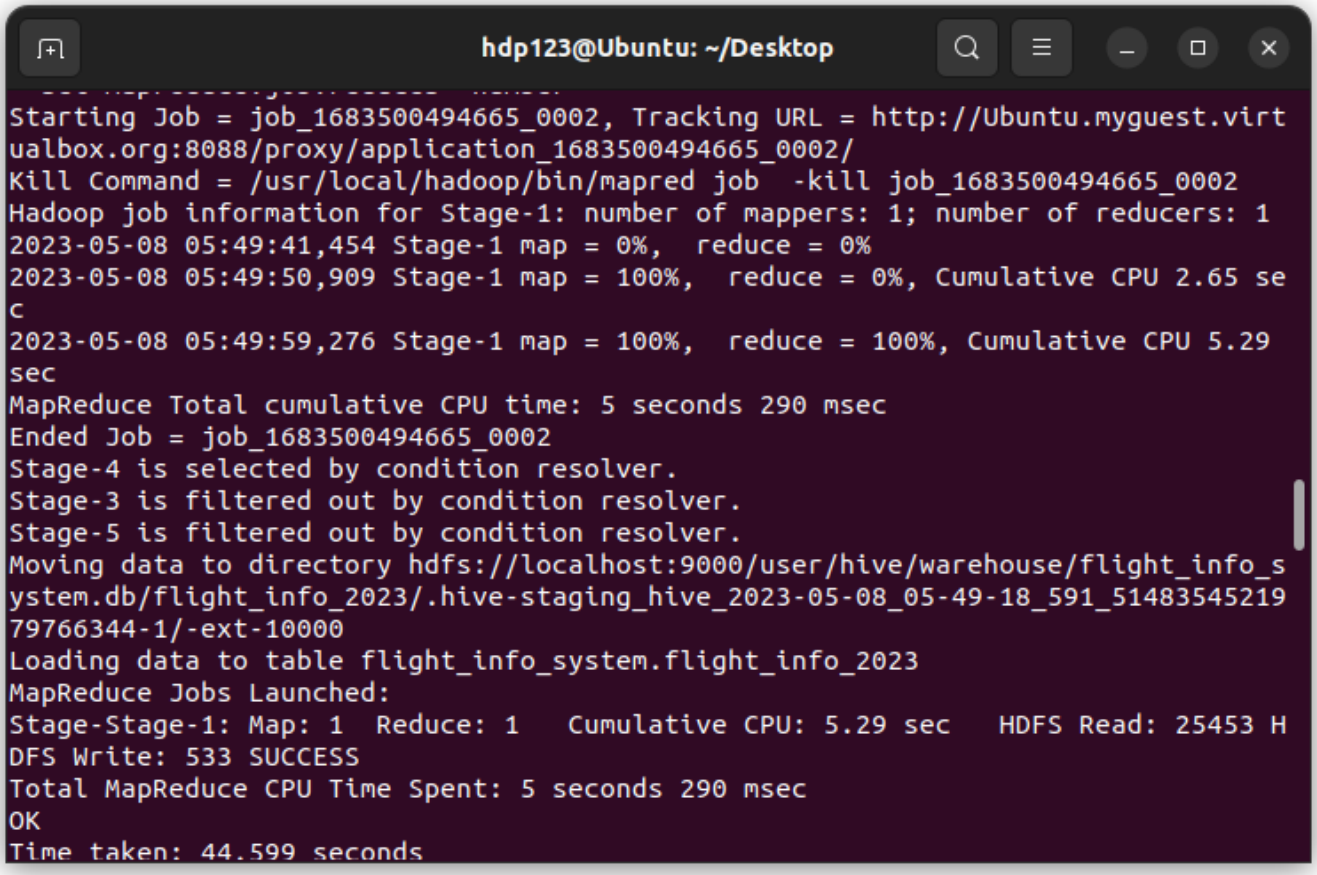
Ln 1, Col 1 | 100% | Windows (CRLF) | UTF-8

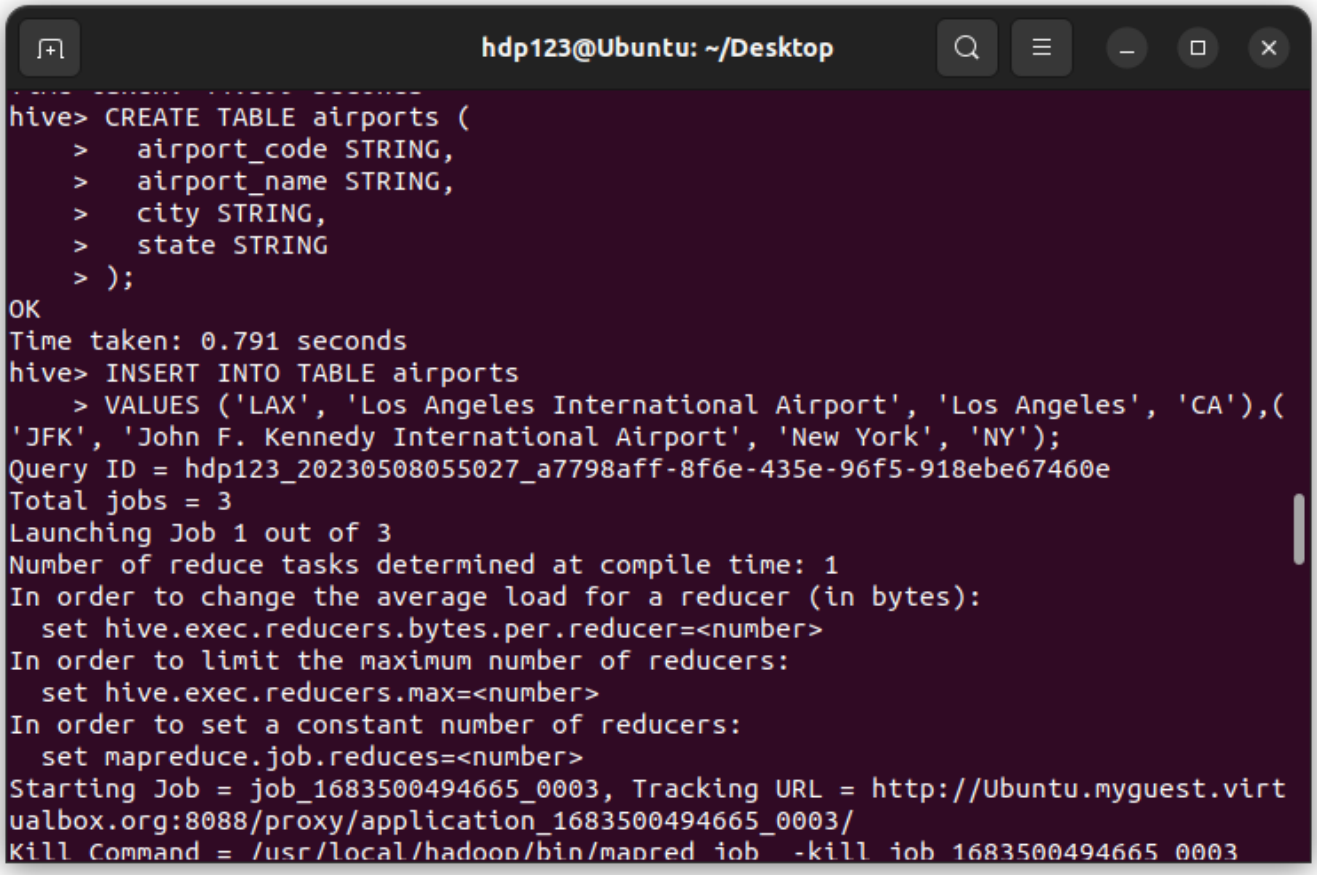
C. Load table with data, insert new values and field in the table, join tables with Hive -

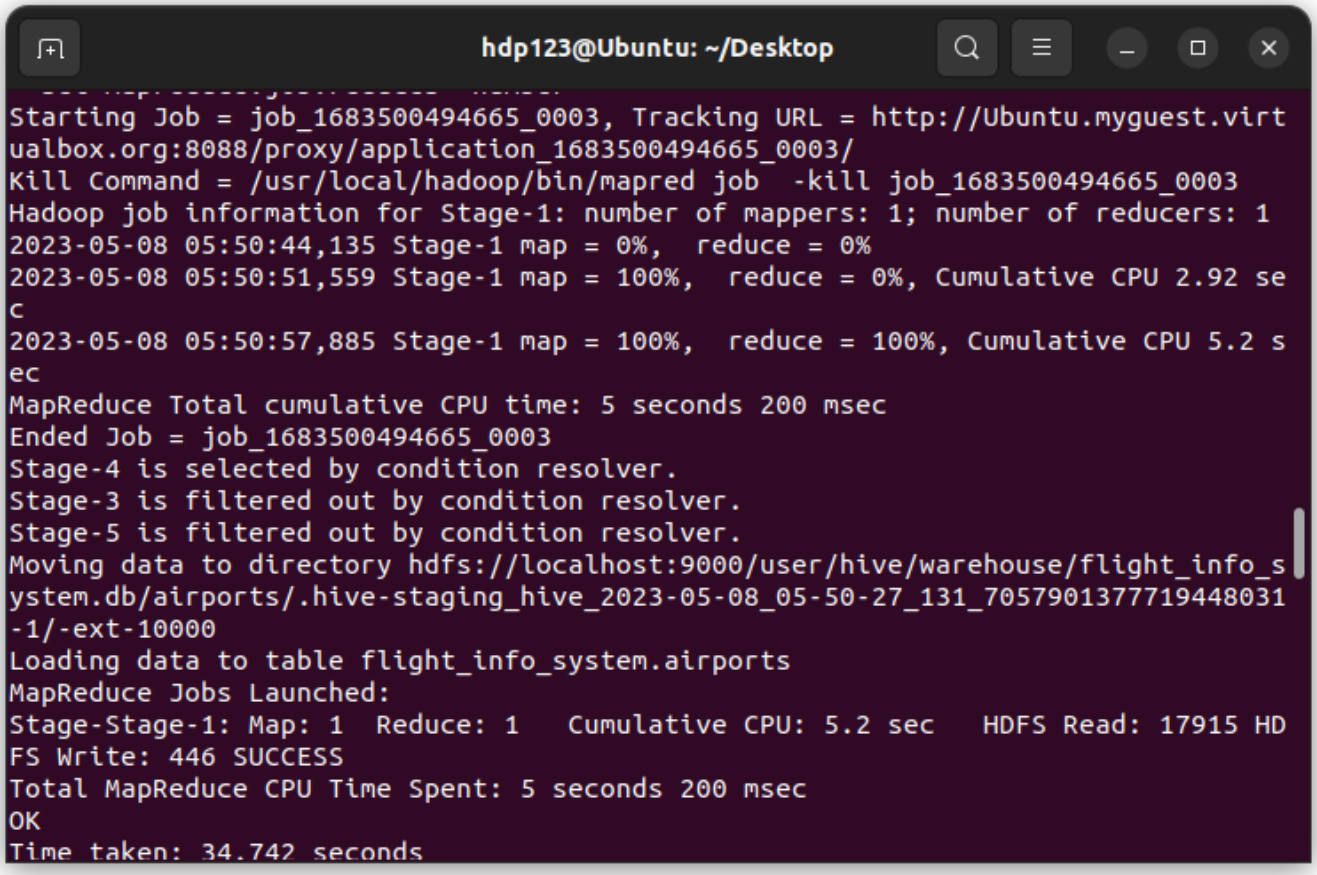


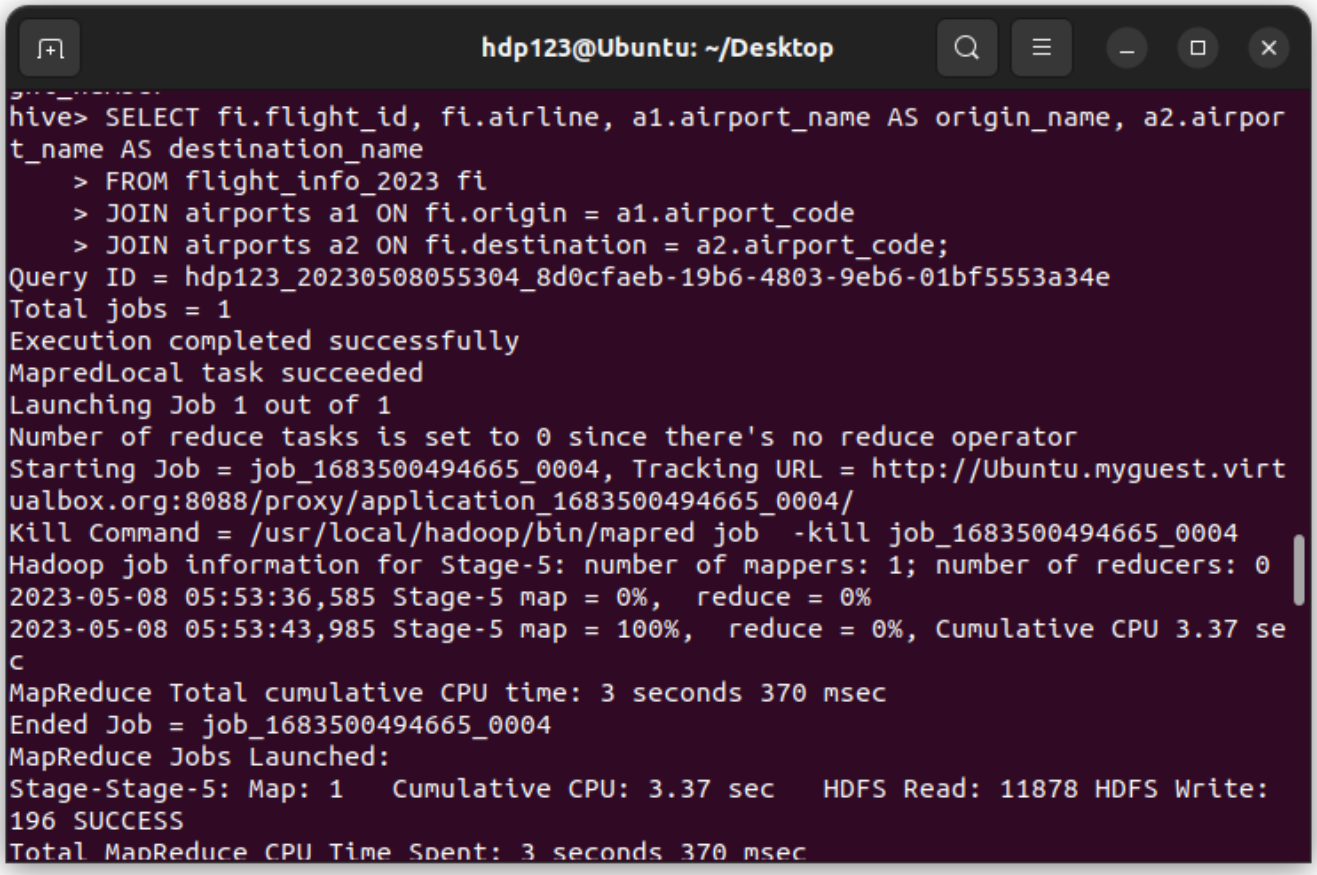


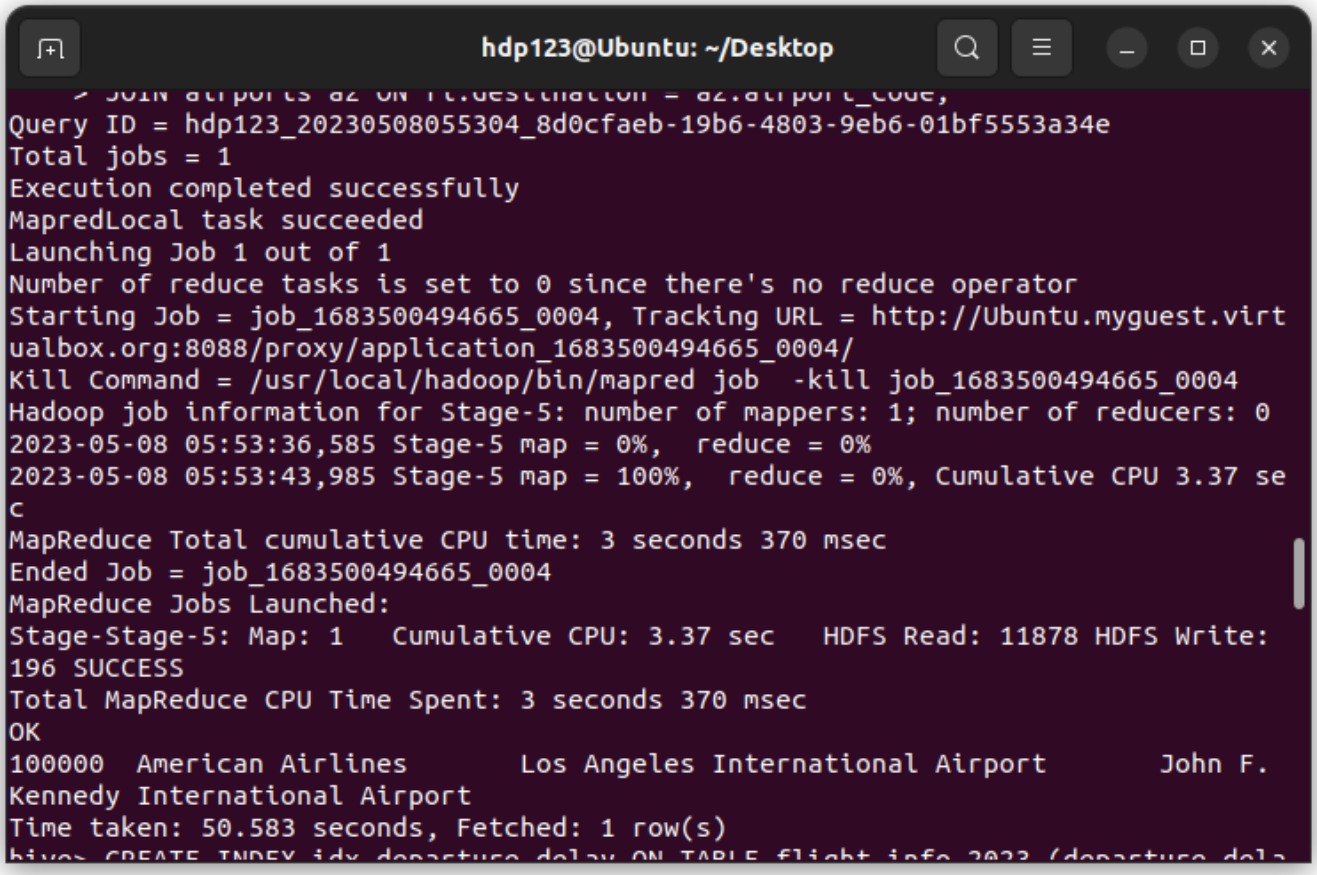

hdp123@Ubuntu: ~/Desktop
hive> INSERT INTO TABLE flight_info_2023
> VALUES (100000, 'American Airlines', 'LAX', 'JFK', '10:00', 5, '14:00', 0,
FALSE);
Query ID = hdp123_20230508054918_a556e3a3-03ea-428a-b599-f6a09797bfea
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile 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_1683500494665_0002, Tracking URL = http://Ubuntu.myguest.virtualbox.org:8088/proxy/application_1683500494665_0002/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1683500494665_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-08 05:49:41,454 Stage-1 map = 0%, reduce = 0%
2023-05-08 05:49:50,909 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.65 sec
2023-05-08 05:49:59,276 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.29 sec
MapReduce Total cumulative CPU time: 5 seconds 290 msec


hdp123@Ubuntu: ~/Desktop
Starting Job = job_1683500494665_0002, Tracking URL = http://Ubuntu.myguest.virtualbox.org:8088/proxy/application_1683500494665_0002/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1683500494665_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-08 05:49:41,454 Stage-1 map = 0%, reduce = 0%
2023-05-08 05:49:50,909 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.65 sec
2023-05-08 05:49:59,276 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.29 sec
MapReduce Total cumulative CPU time: 5 seconds 290 msec
Ended Job = job_1683500494665_0002
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://localhost:9000/user/hive/warehouse/flight_info_system.db/flight_info_2023/.hive-staging_hive_2023-05-08_05-49-18_591_5148354521979766344-1/-ext-10000
Loading data to table flight_info_system.flight_info_2023
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.29 sec HDFS Read: 25453 HDFS Write: 533 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 290 msec
OK
Time taken: 44.599 seconds


hdp123@Ubuntu: ~/Desktop
hive> CREATE TABLE airports (
> airport_code STRING,
> airport_name STRING,
> city STRING,
> state STRING
>);
OK
Time taken: 0.791 seconds
hive> INSERT INTO TABLE airports
> VALUES ('LAX', 'Los Angeles International Airport', 'Los Angeles', 'CA'),(
'JFK', 'John F. Kennedy International Airport', 'New York', 'NY');
Query ID = hdp123_20230508055027_a7798aff-8f6e-435e-96f5-918ebe67460e
Total jobs = 3
Launching Job 1 out of 3
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_1683500494665_0003, Tracking URL = http://Ubuntu.myguest.virt
ualbox.org:8088/proxy/application_1683500494665_0003/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1683500494665_0003


hdp123@Ubuntu: ~/Desktop
Starting Job = job_1683500494665_0003, Tracking URL = http://Ubuntu.myguest.virt
ualbox.org:8088/proxy/application_1683500494665_0003/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1683500494665_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-08 05:50:44,135 Stage-1 map = 0%, reduce = 0%
2023-05-08 05:50:51,559 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.92 se
c
2023-05-08 05:50:57,885 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.2 s
ec
MapReduce Total cumulative CPU time: 5 seconds 200 msec
Ended Job = job_1683500494665_0003
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://localhost:9000/user/hive/warehouse/flight_info_s
ystem.db/airports/.hive-staging_hive_2023-05-08_05-50-27_131_7057901377719448031
-1/-ext-10000
Loading data to table flight_info_system.airports
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.2 sec HDFS Read: 17915 HD
FS Write: 446 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 200 msec
OK
Time taken: 34.742 seconds


hdp123@Ubuntu: ~/Desktop
hive> SELECT fi.flight_id, fi.airline, a1.airport_name AS origin_name, a2.airport_name AS destination_name
> FROM flight_info_2023 fi
> JOIN airports a1 ON fi.origin = a1.airport_code
> JOIN airports a2 ON fi.destination = a2.airport_code;
Query ID = hdp123_20230508055304_8d0cfaeb-19b6-4803-9eb6-01bf5553a34e
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_1683500494665_0004, Tracking URL = http://Ubuntu.myguest.virtualbox.org:8088/proxy/application_1683500494665_0004/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1683500494665_0004
Hadoop job information for Stage-5: number of mappers: 1; number of reducers: 0
2023-05-08 05:53:36,585 Stage-5 map = 0%, reduce = 0%
2023-05-08 05:53:43,985 Stage-5 map = 100%, reduce = 0%, Cumulative CPU 3.37 sec
MapReduce Total cumulative CPU time: 3 seconds 370 msec
Ended Job = job_1683500494665_0004
MapReduce Jobs Launched:
Stage-Stage-5: Map: 1 Cumulative CPU: 3.37 sec HDFS Read: 11878 HDFS Write: 196 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 370 msec


hdp123@Ubuntu: ~/Desktop
Query ID = hdp123_20230508055304_8d0cfaeb-19b6-4803-9eb6-01bf5553a34e
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_1683500494665_0004, Tracking URL = http://Ubuntu.myguest.virtualbox.org:8088/proxy/application_1683500494665_0004/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1683500494665_0004
Hadoop job information for Stage-5: number of mappers: 1; number of reducers: 0
2023-05-08 05:53:36,585 Stage-5 map = 0%, reduce = 0%
2023-05-08 05:53:43,985 Stage-5 map = 100%, reduce = 0%, Cumulative CPU 3.37 sec
MapReduce Total cumulative CPU time: 3 seconds 370 msec
Ended Job = job_1683500494665_0004
MapReduce Jobs Launched:
Stage-Stage-5: Map: 1 Cumulative CPU: 3.37 sec HDFS Read: 11878 HDFS Write: 196 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 370 msec
OK
100000 American Airlines Los Angeles International Airport John F. Kennedy International Airport
Time taken: 50.583 seconds, Fetched: 1 row(s)

E. Find the average departure delay per day in 2008/2023 -