Database Assignment Part-2

Given That:

Field Description - flights.csv

Date

AirlineCode

FlightNum

Origin

Destination

DepartureTime

DepartureDelay

ArrivalTime

ArrivalDelay

Airtime

Distance

Field Description - airlines.csv

AirlineCode

Description

Use Cases:

1. Create Hive tables for the above datasets by identifying the right datatypes

2. Solve the following use cases

    Find count of flights that had arrival delay

    Find count of flights that had departure delay

    find the average distance travelled by a flight

    List the data that belong to the airline - American Airlines Inc

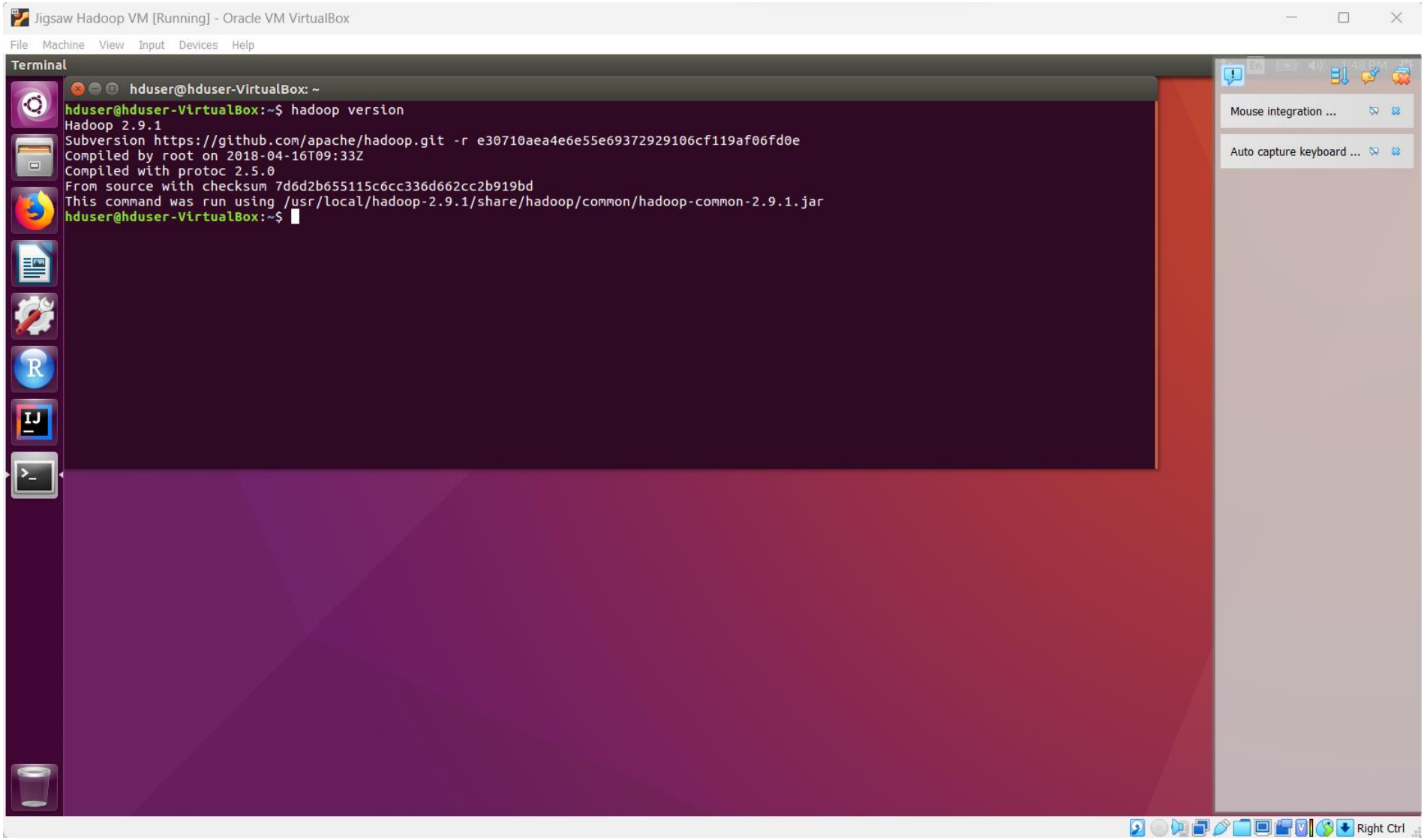
---

Attach screenshots of the outputs along with the query executed in the solutions file

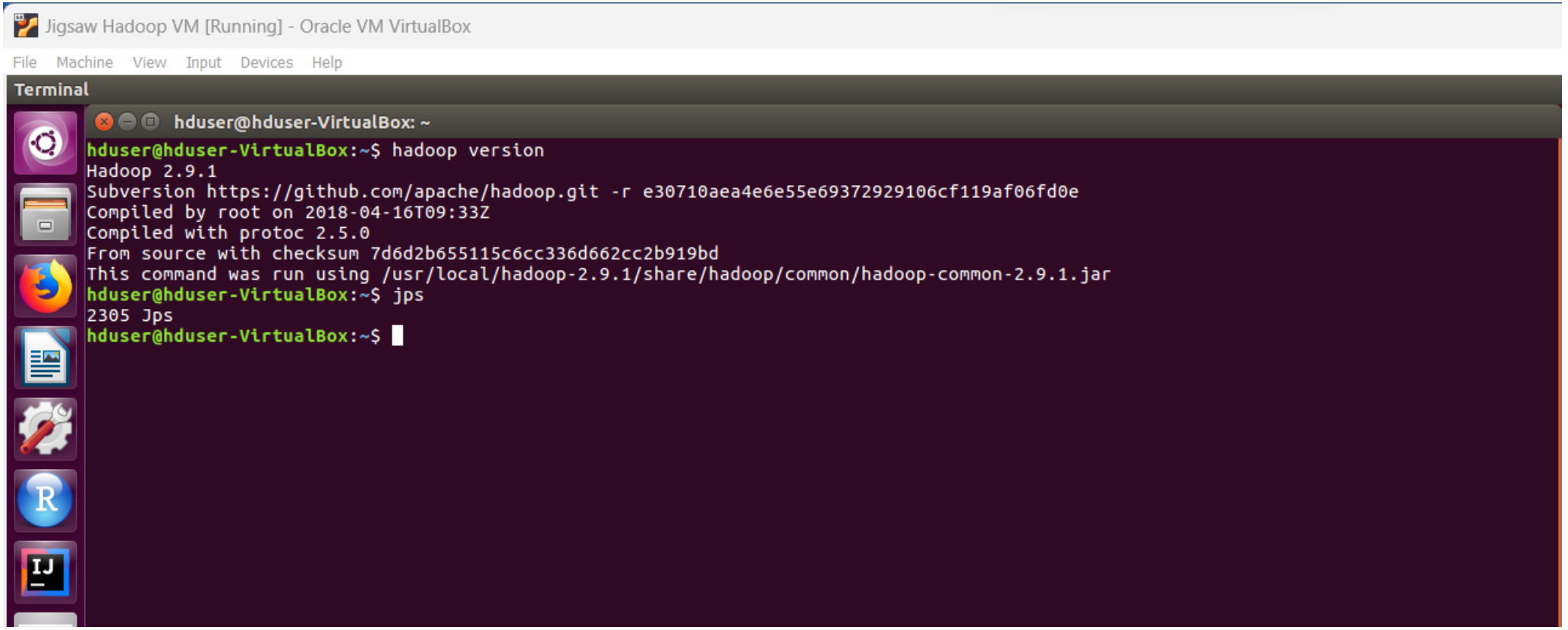
---

Solutions/Answer:

1st) Checking the Hadoop Version



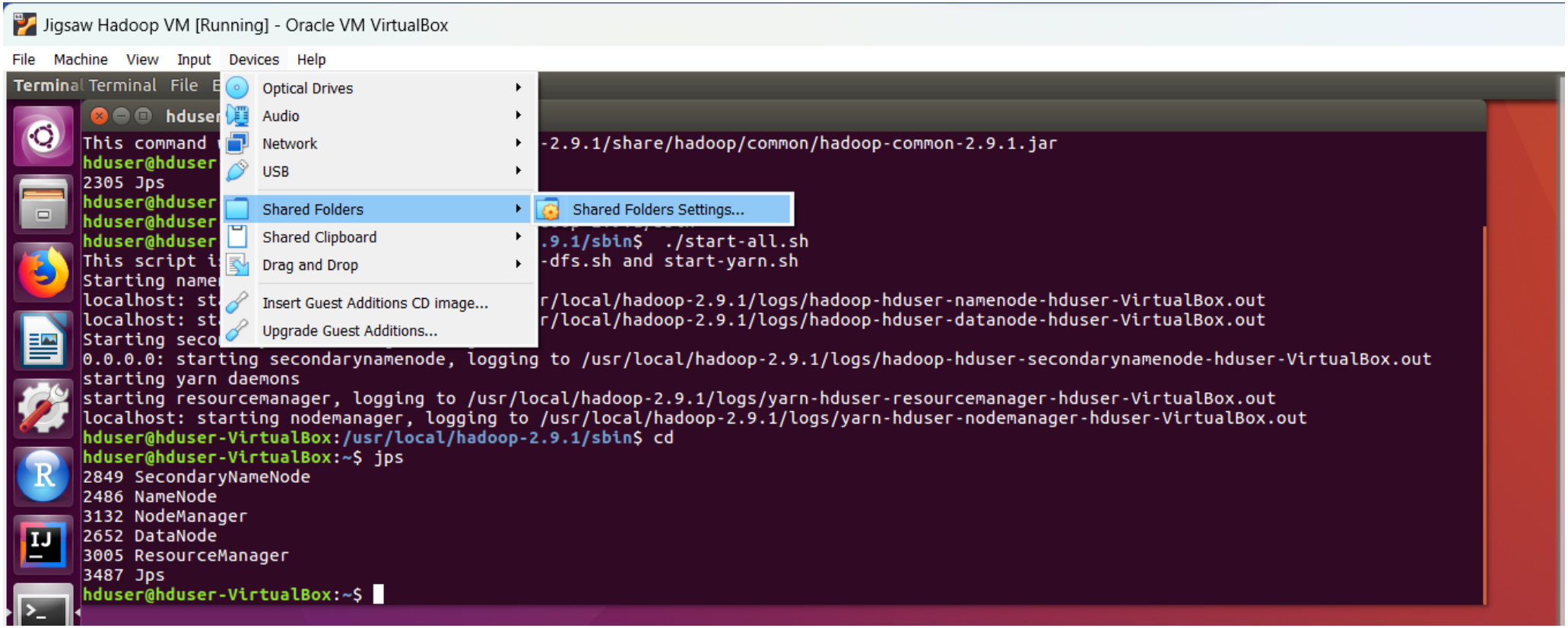
2nd) Checking Jps (Java Process Status) running or not as not

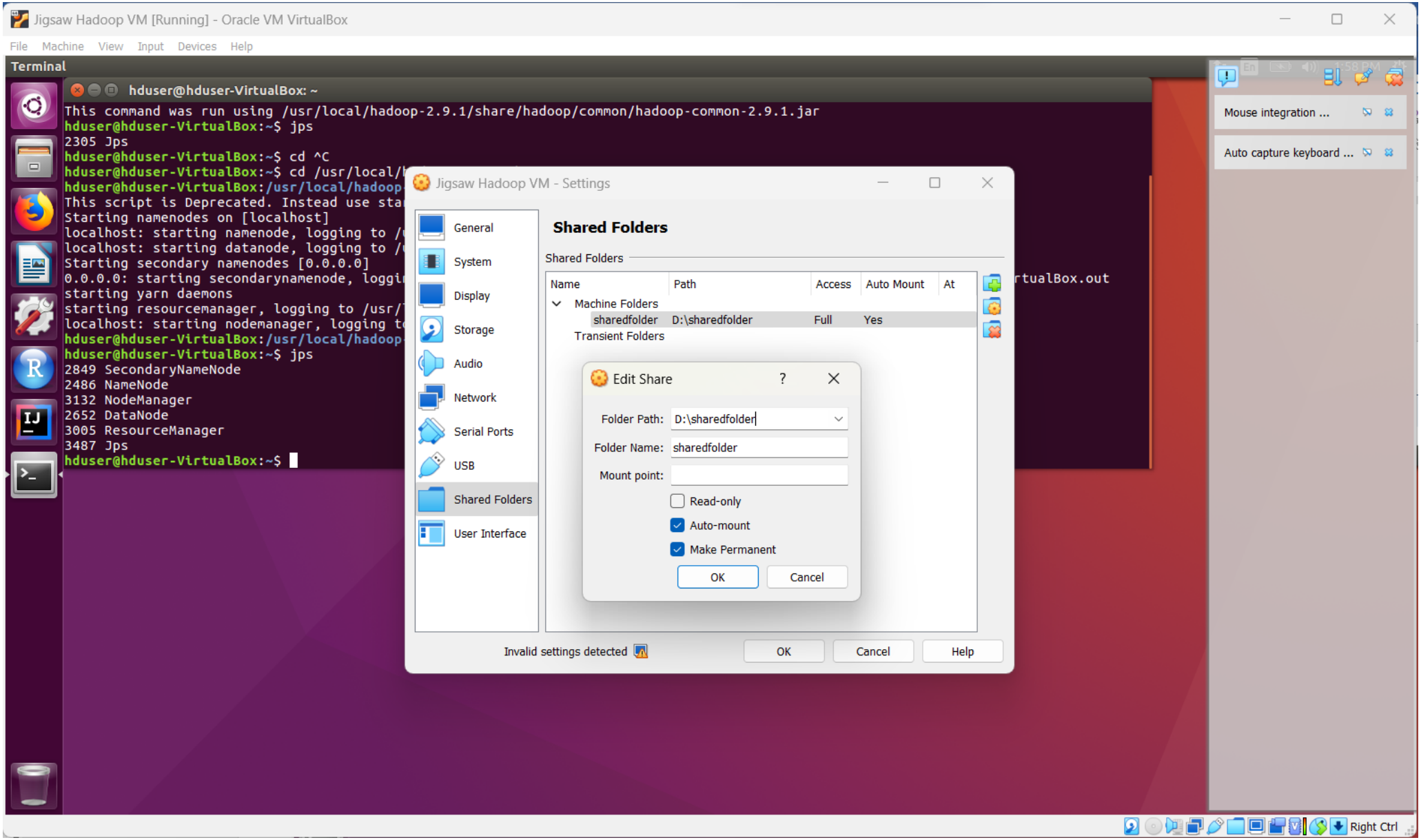


3rd) Starting jps

```
hduser@hduser-VirtualBox:~$ cd /usr/local/hadoop-2.9.1/sbin
hduser@hduser-VirtualBox:/usr/local/hadoop-2.9.1/sbin$  ./start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop-2.9.1/logs/hadoop-hduser-namenode-hduser-VirtualBox.out
localhost: starting datanode, logging to /usr/local/hadoop-2.9.1/logs/hadoop-hduser-datanode-hduser-VirtualBox.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop-2.9.1/logs/hadoop-hduser-secondarynamenode-hduser-VirtualBox.out
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop-2.9.1/logs/yarn-hduser-resourcemanager-hduser-VirtualBox.out
localhost: starting nodemanager, logging to /usr/local/hadoop-2.9.1/logs/yarn-hduser-nodemanager-hduser-VirtualBox.out
hduser@hduser-VirtualBox:/usr/local/hadoop-2.9.1/sbin$ cd
hduser@hduser-VirtualBox:~$ jps
2849 SecondaryNameNode
2486 NameNode
3132 NodeManager
2652 DataNode
3005 ResourceManager
3487 Jps
hduser@hduser-VirtualBox:~$
```

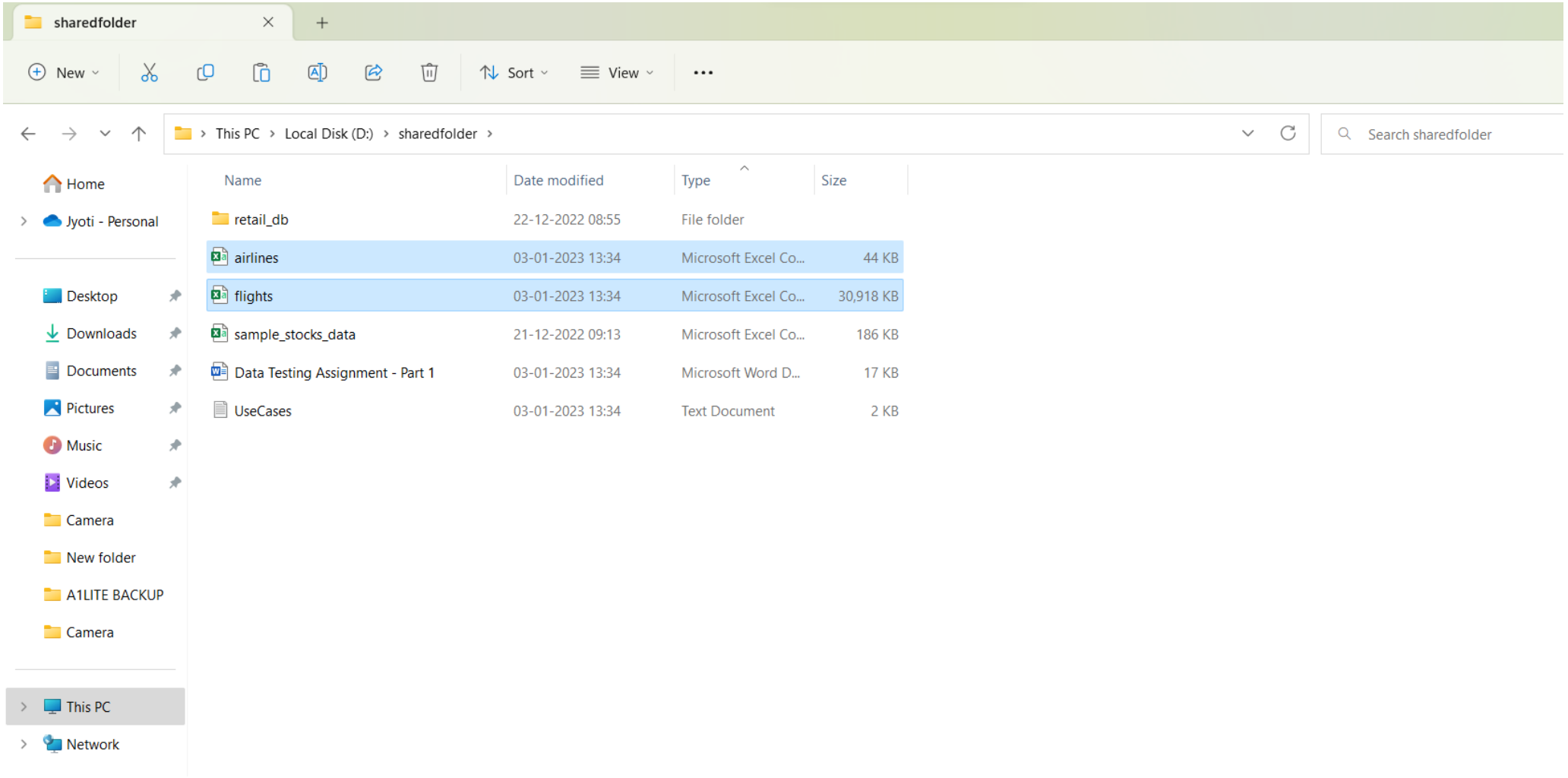
4th) Now Create a folder in the D: drive of the host OS (and copy the path)


Jigsaw Hadoop VM [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Optical Drives
Audio
Network
USB
Shared Folders
Shared Clipboard
Drag and Drop
Insert Guest Additions CD image...
Upgrade Guest Additions...
Shared Folders Settings...
-2.9.1/share/hadoop/common/hadoop-common-2.9.1.jar
2305 Jps
.9.1/sbin$ ./start-all.sh
-dfs.sh and start-yarn.sh
r/local/hadoop-2.9.1/logs/hadoop-hduser-namenode-hduser-VirtualBox.out
r/local/hadoop-2.9.1/logs/hadoop-hduser-datanode-hduser-VirtualBox.out
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop-2.9.1/logs/hadoop-hduser-secondarynamenode-hduser-VirtualBox.out
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop-2.9.1/logs/yarn-hduser-resourcemanager-hduser-VirtualBox.out
localhost: starting nodemanager, logging to /usr/local/hadoop-2.9.1/logs/yarn-hduser-nodemanager-hduser-VirtualBox.out
hduser@hduser-VirtualBox:/usr/local/hadoop-2.9.1/sbin$ cd
hduser@hduser-VirtualBox:~$ jps
2849 SecondaryNameNode
2486 NameNode
3132 NodeManager
2652 DataNode
3005 ResourceManager
3487 Jps
hduser@hduser-VirtualBox:~$

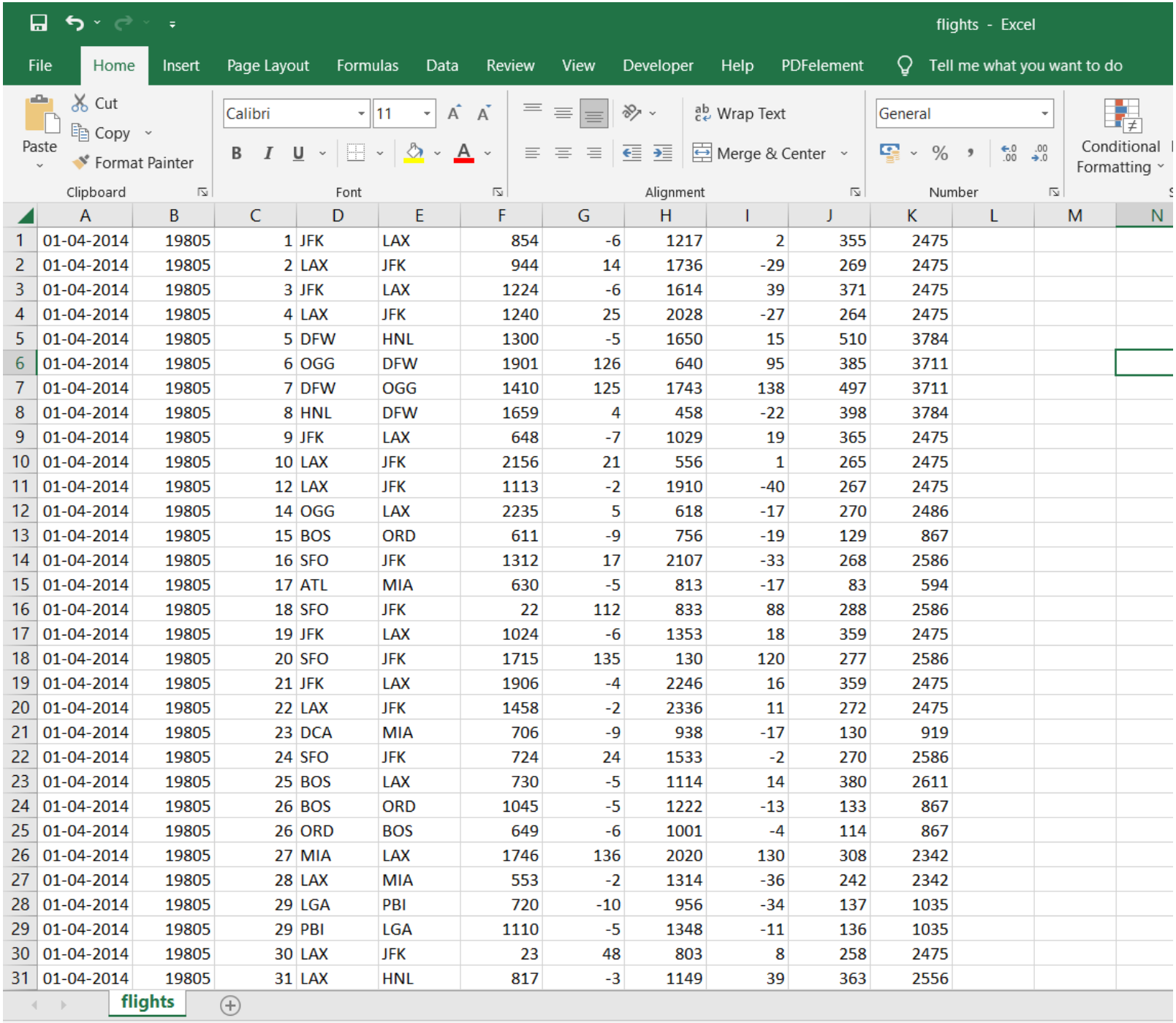


Path: D:\sharedfolder

Two .csv files airlines & flights are available on D drive (Host File system)

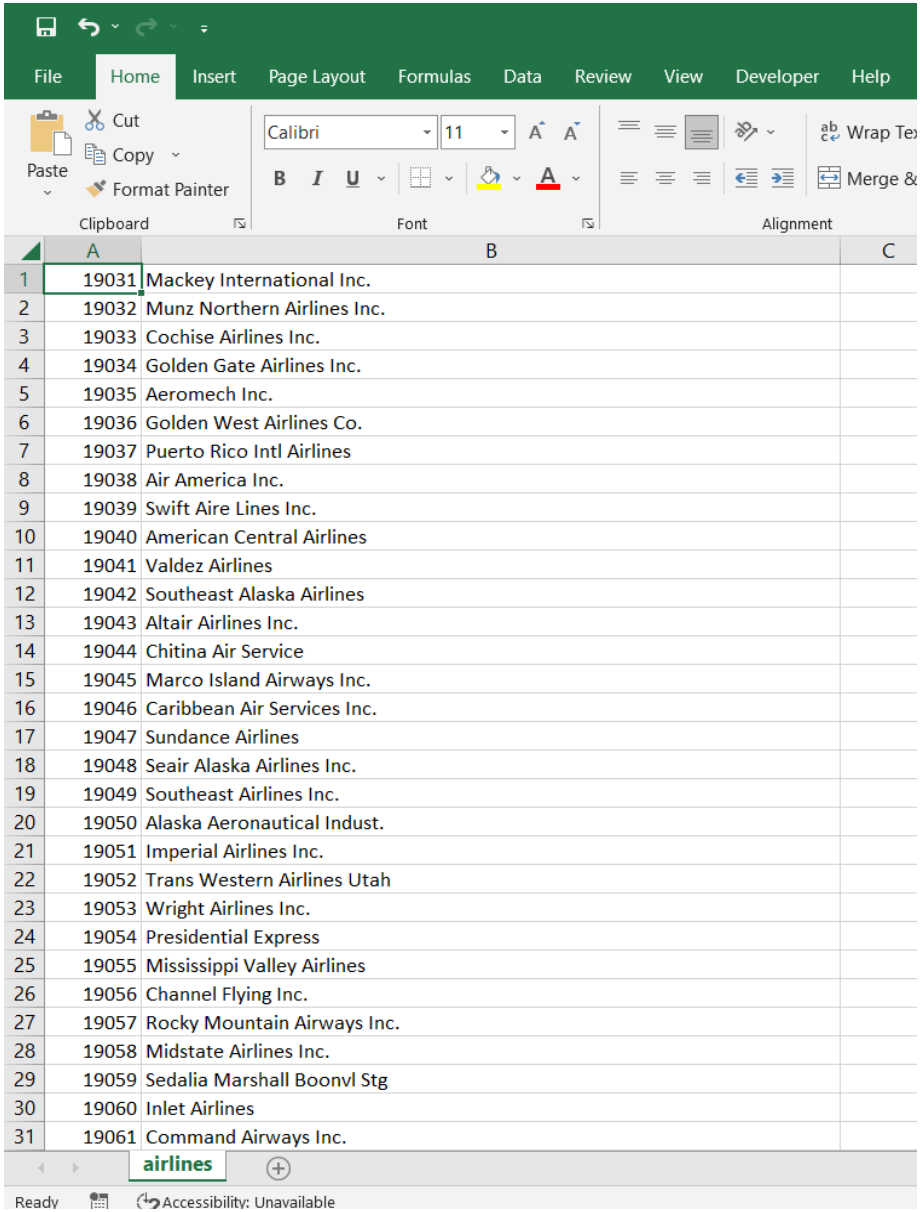


Flights.csv



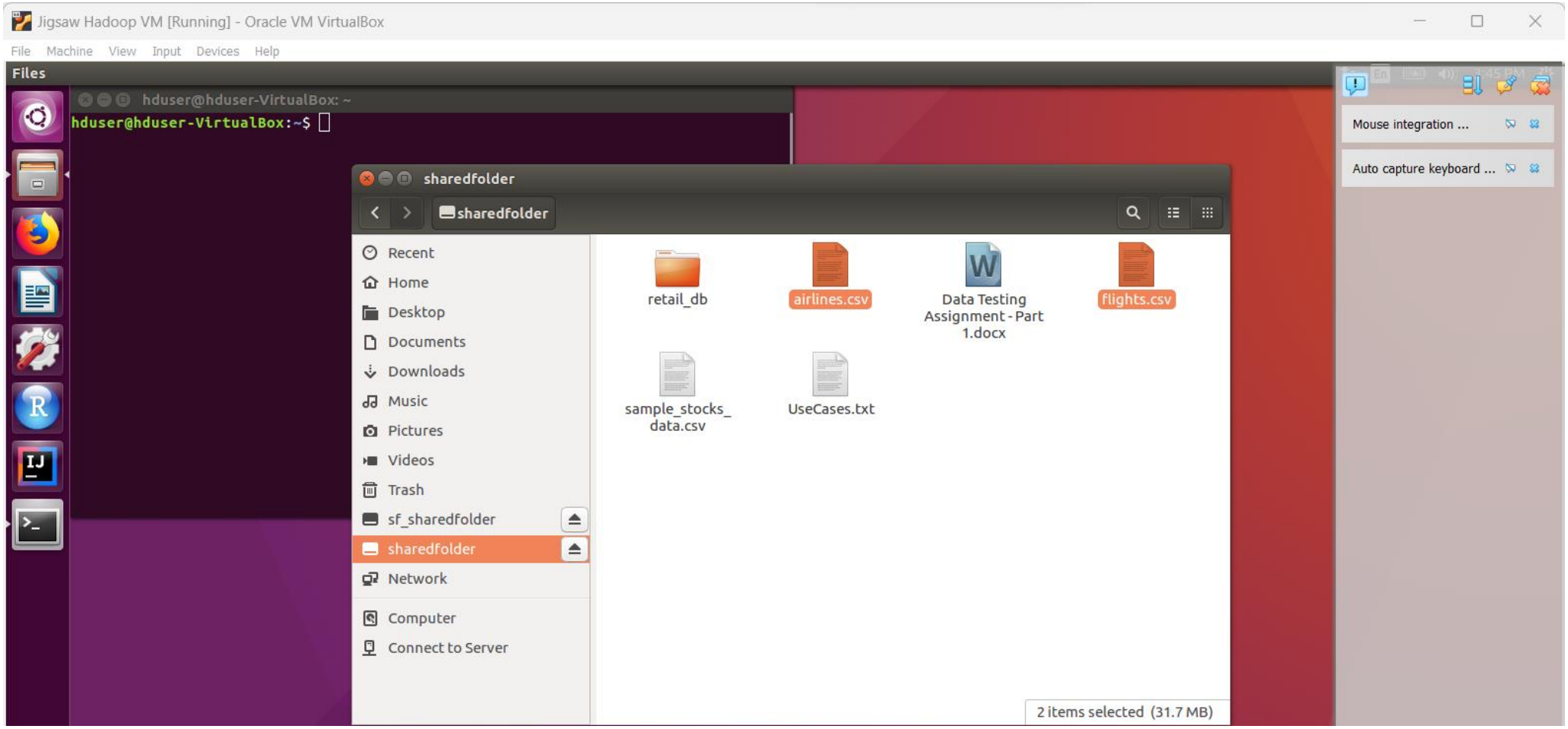
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	01-04-2014	19805	1	JFK	LAX	854	-6	1217	2	355	2475			
2	01-04-2014	19805	2	LAX	JFK	944	14	1736	-29	269	2475			
3	01-04-2014	19805	3	JFK	LAX	1224	-6	1614	39	371	2475			
4	01-04-2014	19805	4	LAX	JFK	1240	25	2028	-27	264	2475			
5	01-04-2014	19805	5	DFW	HNL	1300	-5	1650	15	510	3784			
6	01-04-2014	19805	6	OGG	DFW	1901	126	640	95	385	3711			
7	01-04-2014	19805	7	DFW	OGG	1410	125	1743	138	497	3711			
8	01-04-2014	19805	8	HNL	DFW	1659	4	458	-22	398	3784			
9	01-04-2014	19805	9	JFK	LAX	648	-7	1029	19	365	2475			
10	01-04-2014	19805	10	LAX	JFK	2156	21	556	1	265	2475			
11	01-04-2014	19805	12	LAX	JFK	1113	-2	1910	-40	267	2475			
12	01-04-2014	19805	14	OGG	LAX	2235	5	618	-17	270	2486			
13	01-04-2014	19805	15	BOS	ORD	611	-9	756	-19	129	867			
14	01-04-2014	19805	16	SFO	JFK	1312	17	2107	-33	268	2586			
15	01-04-2014	19805	17	ATL	MIA	630	-5	813	-17	83	594			
16	01-04-2014	19805	18	SFO	JFK	22	112	833	88	288	2586			
17	01-04-2014	19805	19	JFK	LAX	1024	-6	1353	18	359	2475			
18	01-04-2014	19805	20	SFO	JFK	1715	135	130	120	277	2586			
19	01-04-2014	19805	21	JFK	LAX	1906	-4	2246	16	359	2475			
20	01-04-2014	19805	22	LAX	JFK	1458	-2	2336	11	272	2475			
21	01-04-2014	19805	23	DCA	MIA	706	-9	938	-17	130	919			
22	01-04-2014	19805	24	SFO	JFK	724	24	1533	-2	270	2586			
23	01-04-2014	19805	25	BOS	LAX	730	-5	1114	14	380	2611			
24	01-04-2014	19805	26	BOS	ORD	1045	-5	1222	-13	133	867			
25	01-04-2014	19805	26	ORD	BOS	649	-6	1001	-4	114	867			
26	01-04-2014	19805	27	MIA	LAX	1746	136	2020	130	308	2342			
27	01-04-2014	19805	28	LAX	MIA	553	-2	1314	-36	242	2342			
28	01-04-2014	19805	29	LGA	PBI	720	-10	956	-34	137	1035			
29	01-04-2014	19805	29	PBI	LGA	1110	-5	1348	-11	136	1035			
30	01-04-2014	19805	30	LAX	JFK	23	48	803	8	258	2475			
31	01-04-2014	19805	31	LAX	HNL	817	-3	1149	39	363	2556			

Airlines.csv



	A	B	C
1	19031	Mackey International Inc.	
2	19032	Munz Northern Airlines Inc.	
3	19033	Cochise Airlines Inc.	
4	19034	Golden Gate Airlines Inc.	
5	19035	Aeromech Inc.	
6	19036	Golden West Airlines Co.	
7	19037	Puerto Rico Intl Airlines	
8	19038	Air America Inc.	
9	19039	Swift Aire Lines Inc.	
10	19040	American Central Airlines	
11	19041	Valdez Airlines	
12	19042	Southeast Alaska Airlines	
13	19043	Altair Airlines Inc.	
14	19044	Chitina Air Service	
15	19045	Marco Island Airways Inc.	
16	19046	Caribbean Air Services Inc.	
17	19047	Sundance Airlines	
18	19048	Seair Alaska Airlines Inc.	
19	19049	Southeast Airlines Inc.	
20	19050	Alaska Aeronautical Indust.	
21	19051	Imperial Airlines Inc.	
22	19052	Trans Western Airlines Utah	
23	19053	Wright Airlines Inc.	
24	19054	Presidential Express	
25	19055	Mississippi Valley Airlines	
26	19056	Channel Flying Inc.	
27	19057	Rocky Mountain Airways Inc.	
28	19058	Midstate Airlines Inc.	
29	19059	Sedalia Marshall Boonvl Stg	
30	19060	Inlet Airlines	
31	19061	Command Airways Inc.	

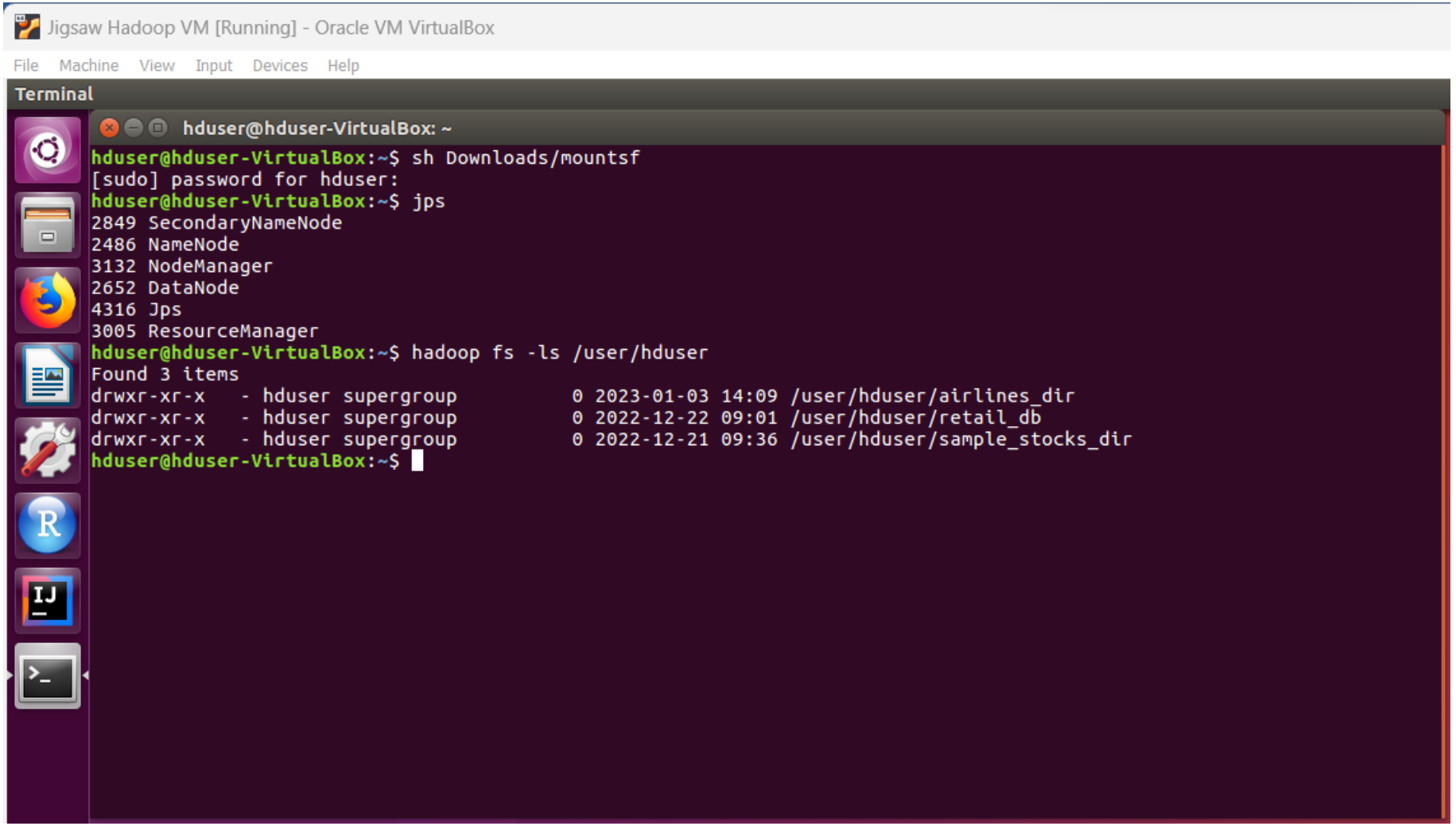
5TH) Now files are available on Ubuntu shared folder



6TH) Now Putting Data into HDFS

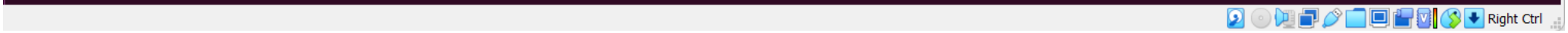
create an **airlines_dir** and put **airlines.csv** and **flights.csv** data into it

- /user/hduser is the working directory in the VM

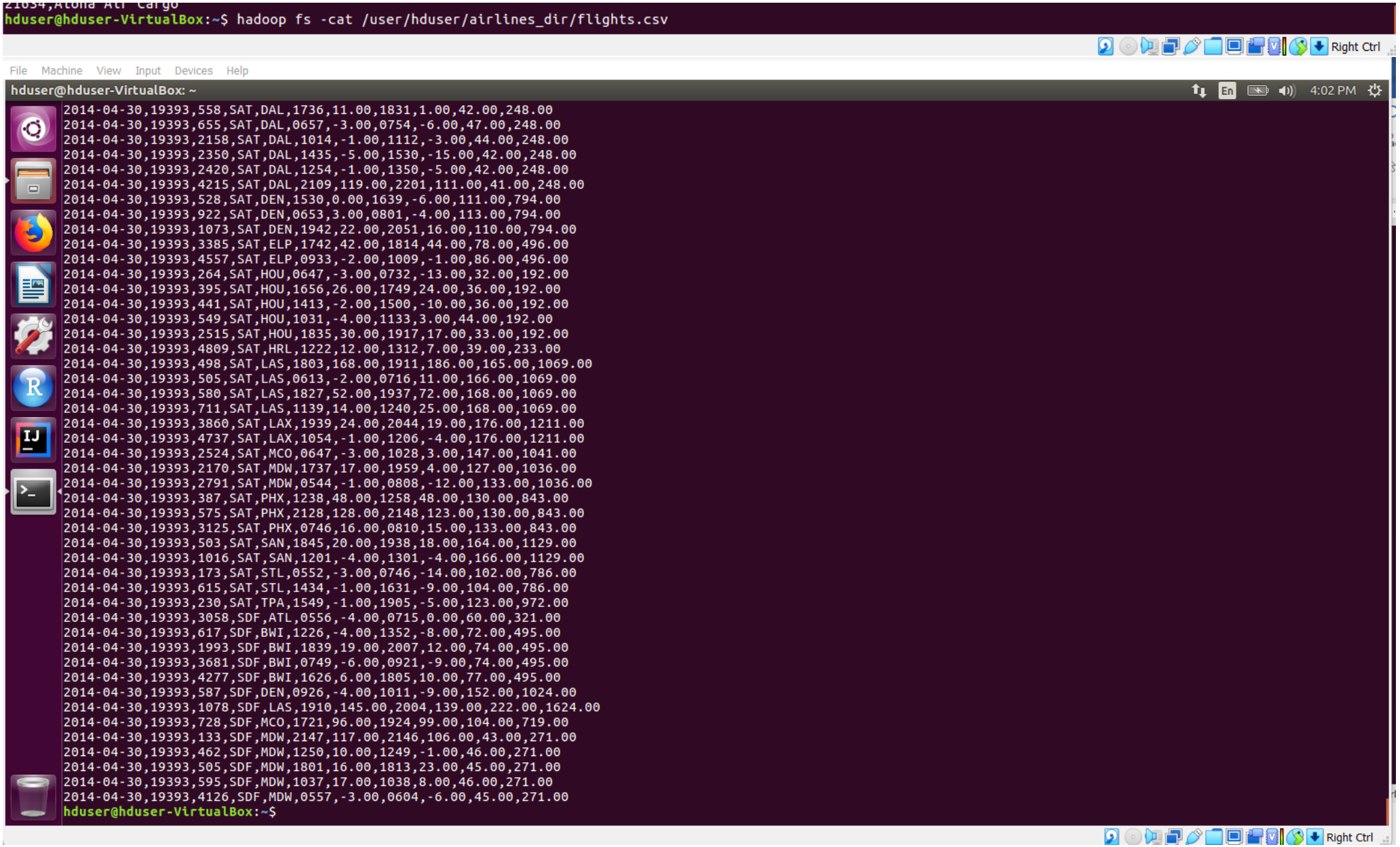


Now reading the airlines.csv file

```
hduser@hduser-VirtualBox:~$ hadoop fs -cat /user/hduser/airlines_dir/airlines.csv
19031,Mackey International Inc.
19032,Munz Northern Airlines Inc.
19033,Cochise Airlines Inc.
19034,Golden Gate Airlines Inc.
19035,Aeromech Inc.
19036,Golden West Airlines Co.
19037,Puerto Rico Intl Airlines
19038,Air America Inc.
19039,Swift Aire Lines Inc.
19040,American Central Airlines
19041,Valdez Airlines
19042,Southeast Alaska Airlines
19043,Altair Airlines Inc.
19044,Chitina Air Service
19045,Marco Island Airways Inc.
19046,Caribbean Air Services Inc.
19047,Sundance Airlines
19048,Seair Alaska Airlines Inc.
19049,Southeast Airlines Inc.
19050,Alaska Aeronautical Indust.
19051,Imperial Airlines Inc.
19052,Trans Western Airlines Utah
19053,Wright Airlines Inc.
19054,Presidential Express
19055,Mississippi Valley Airlines
19056,Channel Flying Inc.
19057,Rocky Mountain Airways Inc.
19058,Midstate Airlines Inc.
19059,Sedalia Marshall Boonvl Stg
19060,Inlet Airlines
19061,Command Airways Inc.
19062,Emerald Airlines
19063,Orion Air Inc.
19064,Pan Am Express
19065,Air Micronesia Inc.
19066,Sunbird Inc.
19067,Wings West Airlines
19068,Southcentral Air Inc.
19069,Air Cargo Express Inc.
19070,Associated Aviation Act.
19071,Antilles Air Boats Inc.
19072,AAA Airlines
19073,Argosy Air Lines Inc.
19074,Air Bahia
```

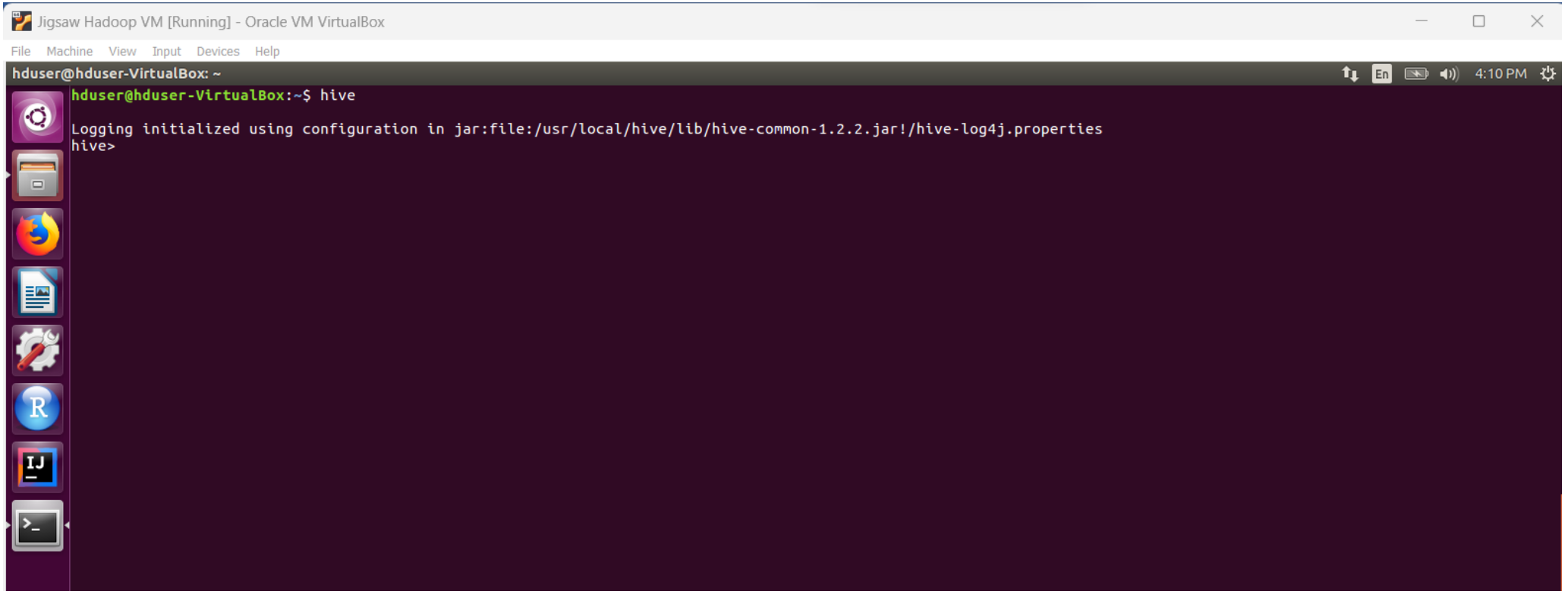


Now reading the flights.csv file

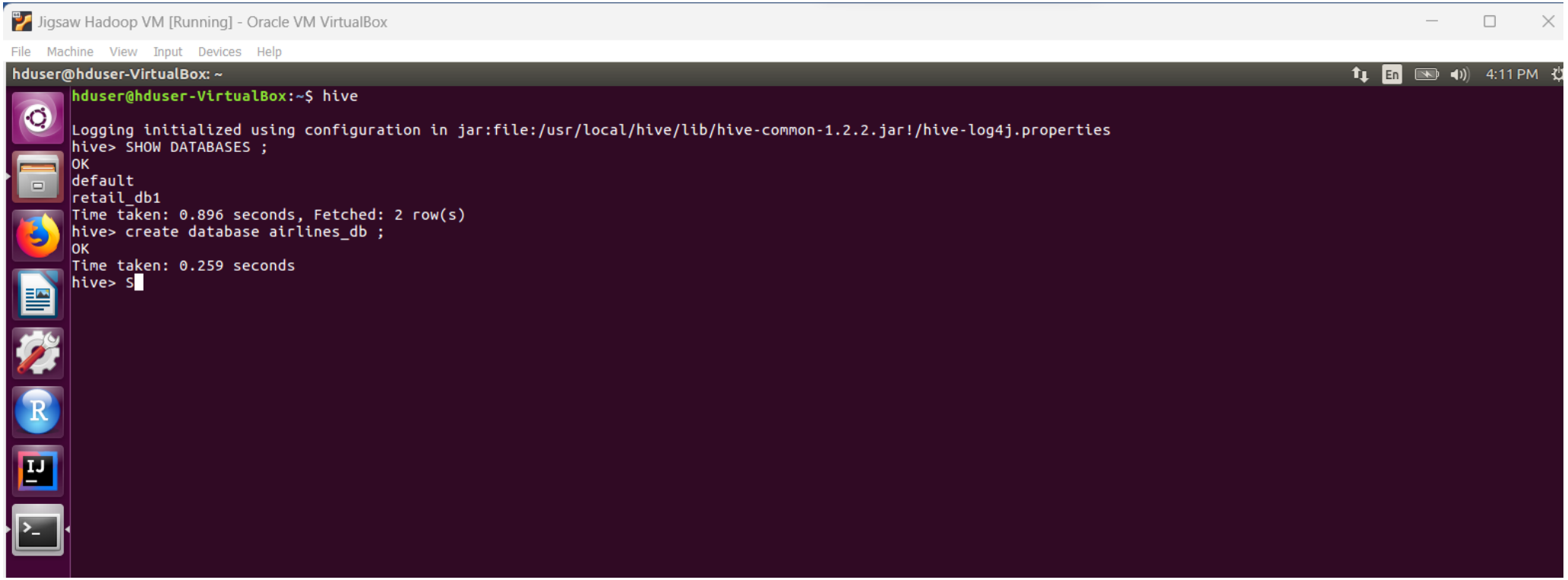

21654,Aloha Air Cargo
hduser@hduser-VirtualBox:~$ hadoop fs -cat /user/hduser/airlines_dir/flights.csv
Right Ctrl
File Machine View Input Devices Help
hduser@hduser-VirtualBox: ~
En
4:02 PM
2014-04-30,19393,558,SAT,DAL,1736,11.00,1831,1.00,42.00,248.00
2014-04-30,19393,655,SAT,DAL,0657,-3.00,0754,-6.00,47.00,248.00
2014-04-30,19393,2158,SAT,DAL,1014,-1.00,1112,-3.00,44.00,248.00
2014-04-30,19393,2350,SAT,DAL,1435,-5.00,1530,-15.00,42.00,248.00
2014-04-30,19393,2420,SAT,DAL,1254,-1.00,1350,-5.00,42.00,248.00
2014-04-30,19393,4215,SAT,DAL,2109,119.00,2201,111.00,41.00,248.00
2014-04-30,19393,528,SAT,DEN,1530,0.00,1639,-6.00,111.00,794.00
2014-04-30,19393,922,SAT,DEN,0653,3.00,0801,-4.00,113.00,794.00
2014-04-30,19393,1073,SAT,DEN,1942,22.00,2051,16.00,110.00,794.00
2014-04-30,19393,3385,SAT,ELP,1742,42.00,1814,44.00,78.00,496.00
2014-04-30,19393,4557,SAT,ELP,0933,-2.00,1009,-1.00,86.00,496.00
2014-04-30,19393,264,SAT,HOU,0647,-3.00,0732,-13.00,32.00,192.00
2014-04-30,19393,395,SAT,HOU,1656,26.00,1749,24.00,36.00,192.00
2014-04-30,19393,441,SAT,HOU,1413,-2.00,1500,-10.00,36.00,192.00
2014-04-30,19393,549,SAT,HOU,1031,-4.00,1133,3.00,44.00,192.00
2014-04-30,19393,2515,SAT,HOU,1835,30.00,1917,17.00,33.00,192.00
2014-04-30,19393,4809,SAT,HRL,1222,12.00,1312,7.00,39.00,233.00
2014-04-30,19393,498,SAT,LAS,1803,168.00,1911,186.00,165.00,1069.00
2014-04-30,19393,505,SAT,LAS,0613,-2.00,0716,11.00,166.00,1069.00
2014-04-30,19393,580,SAT,LAS,1827,52.00,1937,72.00,168.00,1069.00
2014-04-30,19393,711,SAT,LAS,1139,14.00,1240,25.00,168.00,1069.00
2014-04-30,19393,3860,SAT,LAX,1939,24.00,2044,19.00,176.00,1211.00
2014-04-30,19393,4737,SAT,LAX,1054,-1.00,1206,-4.00,176.00,1211.00
2014-04-30,19393,2524,SAT,MCO,0647,-3.00,1028,3.00,147.00,1041.00
2014-04-30,19393,2170,SAT,MDW,1737,17.00,1959,4.00,127.00,1036.00
2014-04-30,19393,2791,SAT,MDW,0544,-1.00,0808,-12.00,133.00,1036.00
2014-04-30,19393,387,SAT,PHX,1238,48.00,1258,48.00,130.00,843.00
2014-04-30,19393,575,SAT,PHX,2128,128.00,2148,123.00,130.00,843.00
2014-04-30,19393,3125,SAT,PHX,0746,16.00,0810,15.00,133.00,843.00
2014-04-30,19393,503,SAT,SAN,1845,20.00,1938,18.00,164.00,1129.00
2014-04-30,19393,1016,SAT,SAN,1201,-4.00,1301,-4.00,166.00,1129.00
2014-04-30,19393,173,SAT,STL,0552,-3.00,0746,-14.00,102.00,786.00
2014-04-30,19393,615,SAT,STL,1434,-1.00,1631,-9.00,104.00,786.00
2014-04-30,19393,230,SAT,TPA,1549,-1.00,1905,-5.00,123.00,972.00
2014-04-30,19393,3058,SDF,ATL,0556,-4.00,0715,0.00,60.00,321.00
2014-04-30,19393,617,SDF,BWI,1226,-4.00,1352,-8.00,72.00,495.00
2014-04-30,19393,1993,SDF,BWI,1839,19.00,2007,12.00,74.00,495.00
2014-04-30,19393,3681,SDF,BWI,0749,-6.00,0921,-9.00,74.00,495.00
2014-04-30,19393,4277,SDF,BWI,1626,6.00,1805,10.00,77.00,495.00
2014-04-30,19393,587,SDF,DEN,0926,-4.00,1011,-9.00,152.00,1024.00
2014-04-30,19393,1078,SDF,LAS,1910,145.00,2004,139.00,222.00,1624.00
2014-04-30,19393,728,SDF,MCO,1721,96.00,1924,99.00,104.00,719.00
2014-04-30,19393,133,SDF,MDW,2147,117.00,2146,106.00,43.00,271.00
2014-04-30,19393,462,SDF,MDW,1250,10.00,1249,-1.00,46.00,271.00
2014-04-30,19393,505,SDF,MDW,1801,16.00,1813,23.00,45.00,271.00
2014-04-30,19393,595,SDF,MDW,1037,17.00,1038,8.00,46.00,271.00
2014-04-30,19393,4126,SDF,MDW,0557,-3.00,0604,-6.00,45.00,271.00
hduser@hduser-VirtualBox:~$
Right Ctrl

7TH) Use Cases: 1. Create Hive tables for the above datasets by identifying the right datatypes

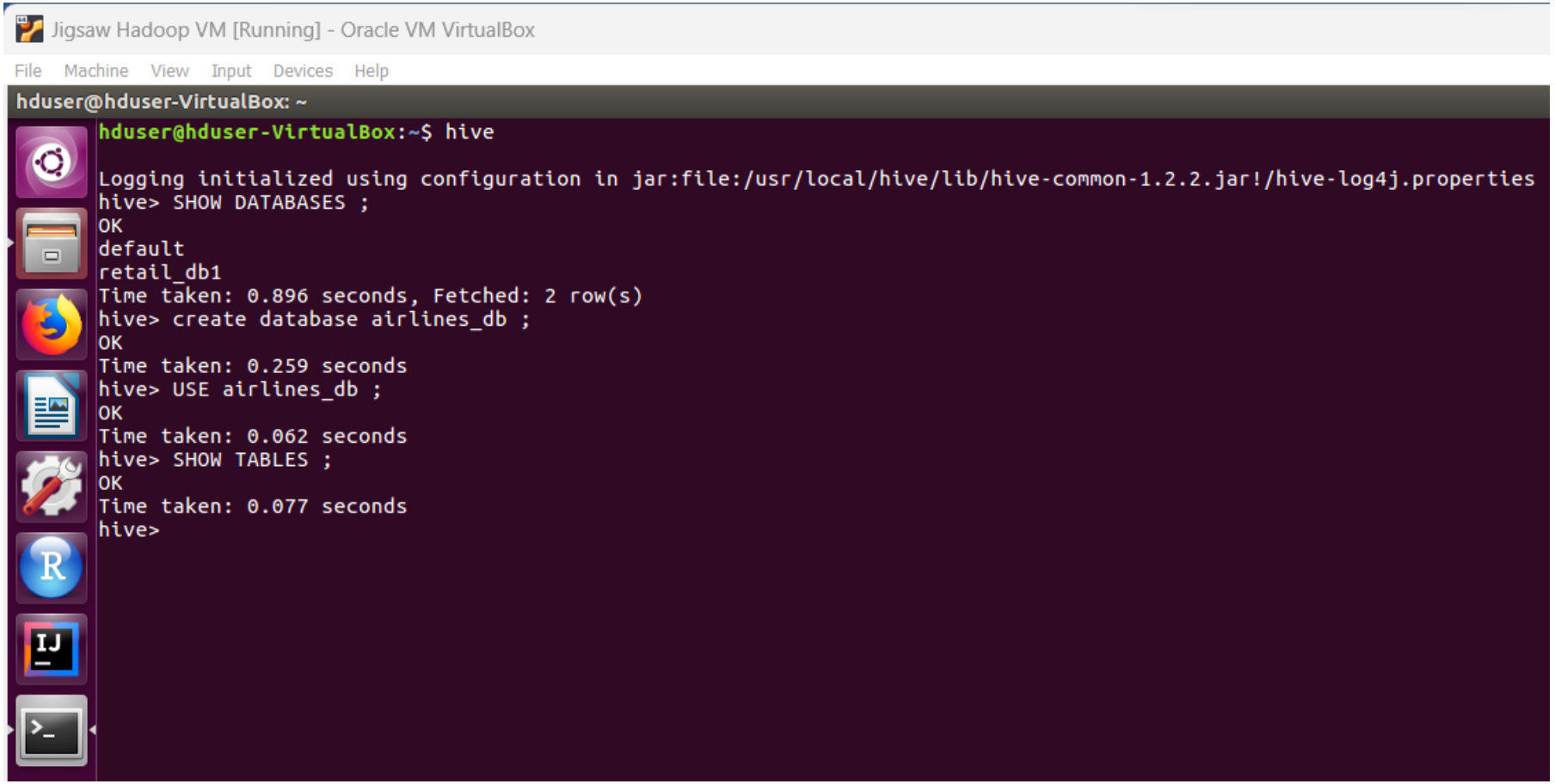
Starting hive



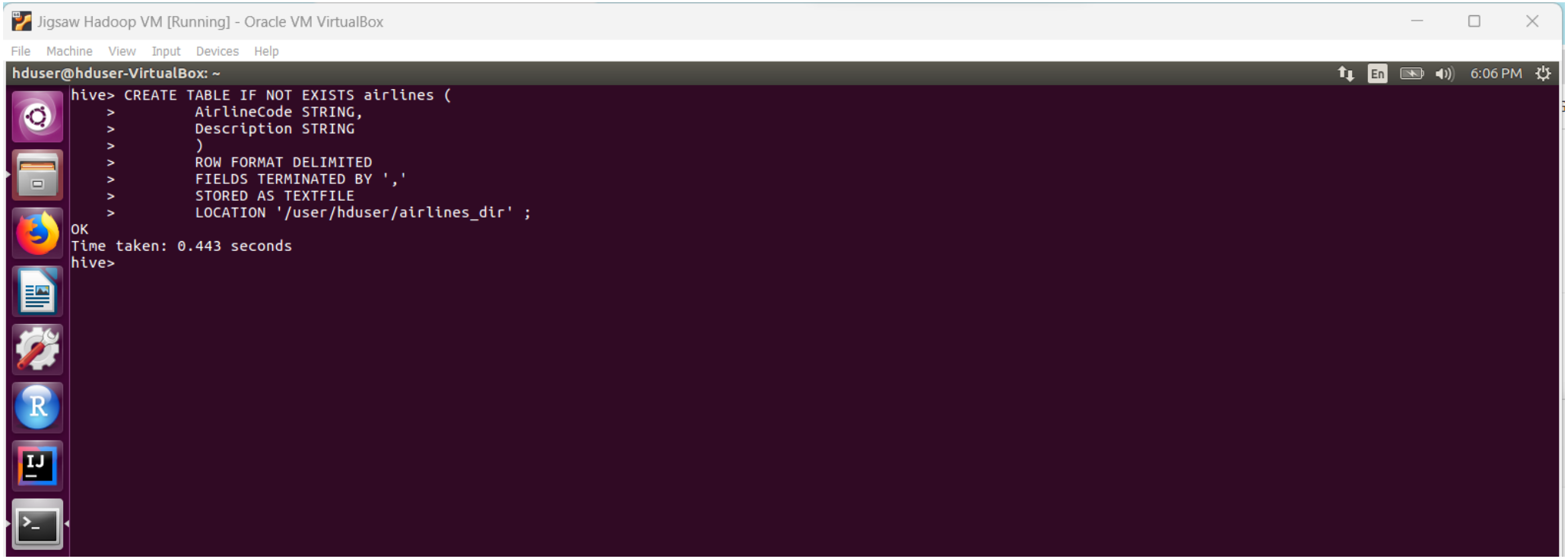
Creating airlines_db

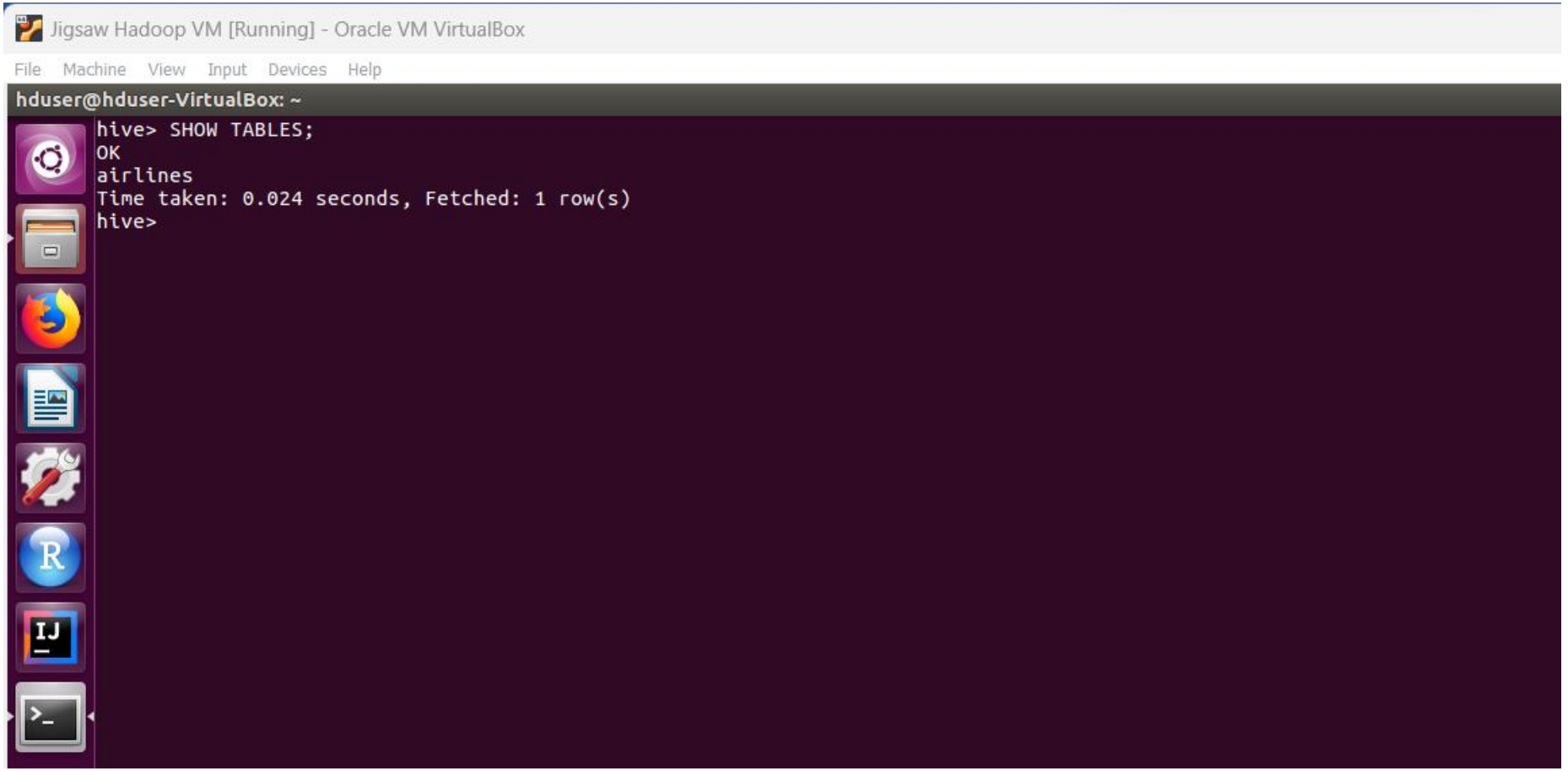


Now checking existing db & Tables

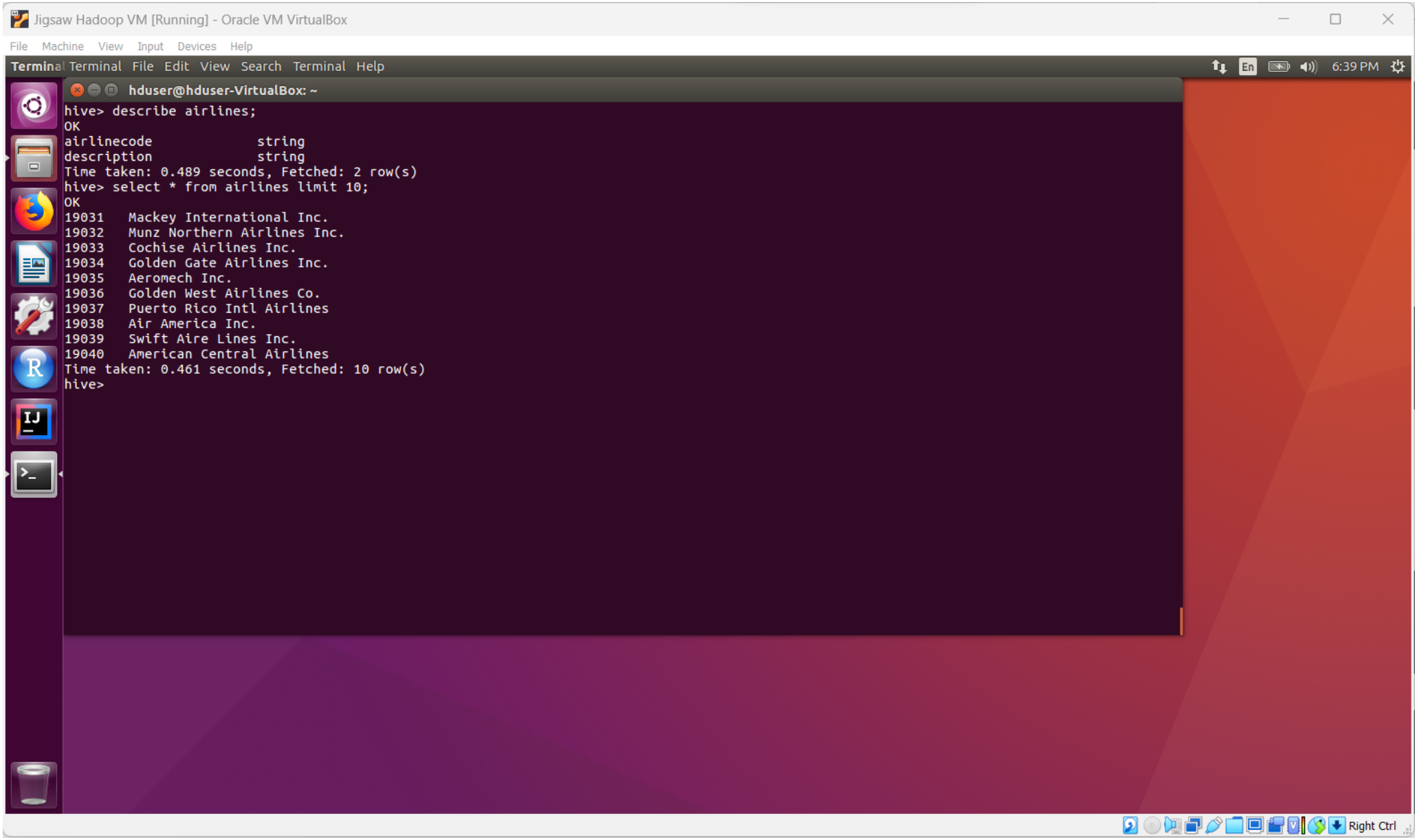


Now creating airlines table in hive


Jigsaw Hadoop VM [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
hduser@hduser-VirtualBox: ~
En
6:06 PM
hive> CREATE TABLE IF NOT EXISTS airlines (
> AirlineCode STRING,
> Description STRING
>)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> STORED AS TEXTFILE
> LOCATION '/user/hduser/airlines_dir' ;
OK
Time taken: 0.443 seconds
hive>



Airlines data has been loaded.



Now Creating another directory for flights

```
hduser@hduser-VirtualBox:~$ hadoop fs -ls /user/hduser
Found 4 items
drwxr-xr-x   - hduser supergroup          0 2023-01-03 15:57 /user/hduser/airlines_dir
drwxr-xr-x   - hduser supergroup          0 2023-01-03 18:18 /user/hduser/flights_dir
drwxr-xr-x   - hduser supergroup          0 2022-12-22 09:01 /user/hduser/retail_db
drwxr-xr-x   - hduser supergroup          0 2022-12-21 09:36 /user/hduser/sample_stocks_dir
hduser@hduser-VirtualBox:~$
```

Now loading flights data from .csv file

```
hduser@hduser-VirtualBox: ~
2014-04-30,19393,580,SAT,LAS,1827,52.00,1937,72.00,168.00,1069.00
2014-04-30,19393,711,SAT,LAS,1139,14.00,1240,25.00,168.00,1069.00
2014-04-30,19393,3860,SAT,LAX,1939,24.00,2044,19.00,176.00,1211.00
2014-04-30,19393,4737,SAT,LAX,1054,-1.00,1206,-4.00,176.00,1211.00
2014-04-30,19393,2524,SAT,MCO,0647,-3.00,1028,3.00,147.00,1041.00
2014-04-30,19393,2170,SAT,MDW,1737,17.00,1959,4.00,127.00,1036.00
2014-04-30,19393,2791,SAT,MDW,0544,-1.00,0808,-12.00,133.00,1036.00
2014-04-30,19393,387,SAT,PHX,1238,48.00,1258,48.00,130.00,843.00
2014-04-30,19393,575,SAT,PHX,2128,128.00,2148,123.00,130.00,843.00
2014-04-30,19393,3125,SAT,PHX,0746,16.00,0810,15.00,133.00,843.00
2014-04-30,19393,503,SAT,SAN,1845,20.00,1938,18.00,164.00,1129.00
2014-04-30,19393,1016,SAT,SAN,1201,-4.00,1301,-4.00,166.00,1129.00
2014-04-30,19393,173,SAT,STL,0552,-3.00,0746,-14.00,102.00,786.00
2014-04-30,19393,615,SAT,STL,1434,-1.00,1631,-9.00,104.00,786.00
2014-04-30,19393,230,SAT,TPA,1549,-1.00,1905,-5.00,123.00,972.00
2014-04-30,19393,3058,SDF,ATL,0556,-4.00,0715,0.00,60.00,321.00
2014-04-30,19393,617,SDF,BWI,1226,-4.00,1352,-8.00,72.00,495.00
2014-04-30,19393,1993,SDF,BWI,1839,19.00,2007,12.00,74.00,495.00
2014-04-30,19393,3681,SDF,BWI,0749,-6.00,0921,-9.00,74.00,495.00
2014-04-30,19393,4277,SDF,BWI,1626,6.00,1805,10.00,77.00,495.00
2014-04-30,19393,587,SDF,DEN,0926,-4.00,1011,-9.00,152.00,1024.00
2014-04-30,19393,1078,SDF,LAS,1910,145.00,2004,139.00,222.00,1624.00
2014-04-30,19393,728,SDF,MCO,1721,96.00,1924,99.00,104.00,719.00
2014-04-30,19393,133,SDF,MDW,2147,117.00,2146,106.00,43.00,271.00
2014-04-30,19393,462,SDF,MDW,1250,10.00,1249,-1.00,46.00,271.00
2014-04-30,19393,505,SDF,MDW,1801,16.00,1813,23.00,45.00,271.00
2014-04-30,19393,595,SDF,MDW,1037,17.00,1038,8.00,46.00,271.00
2014-04-30,19393,4126,SDF,MDW,0557,-3.00,0604,-6.00,45.00,271.00
```

Used below query to create flights table:

CREATE TABLE IF NOT EXISTS flights (

    Flight_Date DATE,

    AirlineCode STRING,

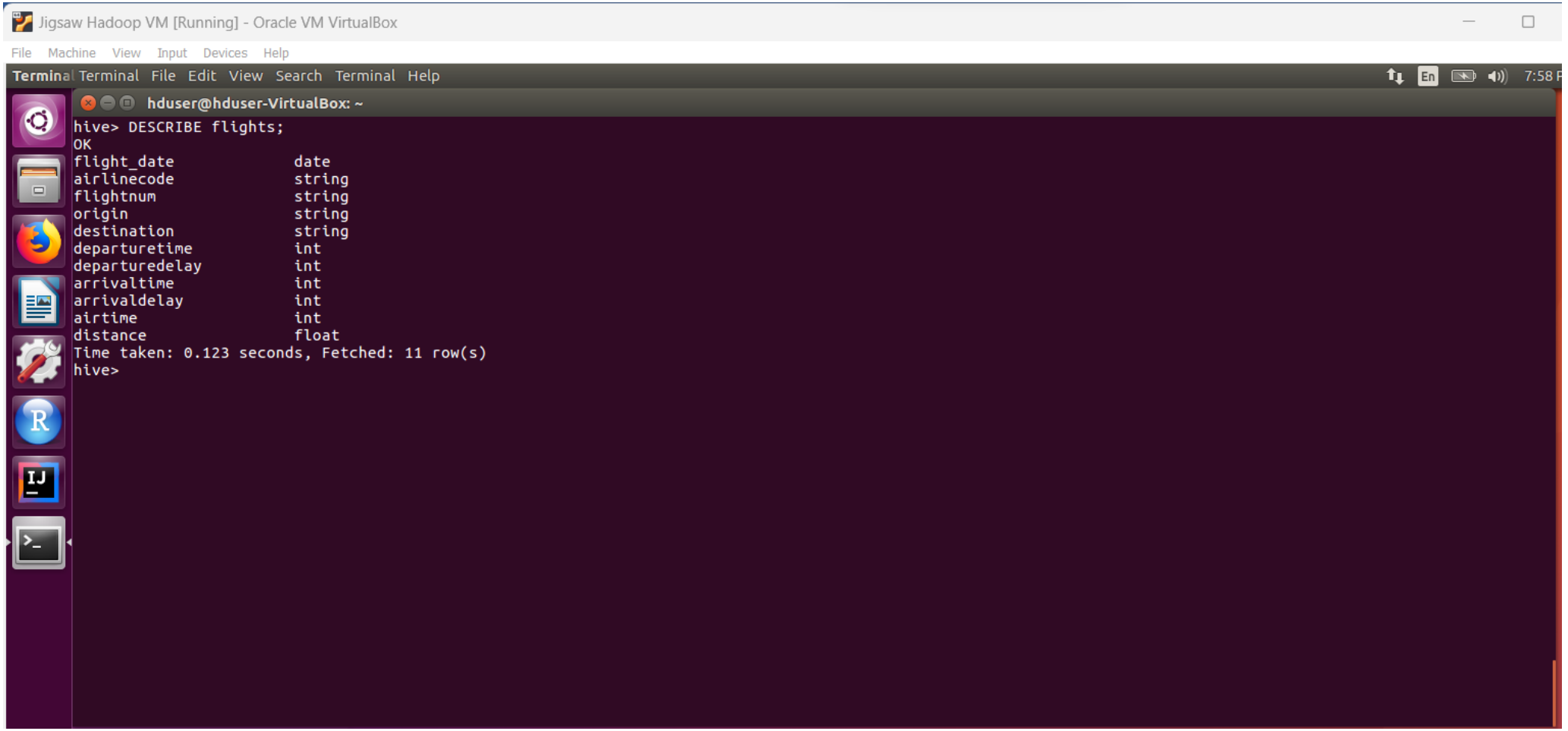
    FlightNum STRING,

```
Origin STRING,
Destination STRING,
DepartureTime INT,
DepartureDelay TIMESTAMP,
ArrivalTime TIMESTAMP,
ArrivalDelay TIMESTAMP,
Airtime TIMESTAMP,
Distance FLOAT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/hduser/flights_dir' ;
```

```
hive> CREATE TABLE IF NOT EXISTS flights (
    >         Flight_Date DATE,
    >         AirlineCode STRING,
    >         FlightNum STRING,
    >         Origin STRING,
    >         Destination STRING,
    >         DepartureTime INT,
    >         DepartureDelay TIMESTAMP,
    >         ArrivalTime TIMESTAMP,
    >         ArrivalDelay TIMESTAMP,
    >         Airtime TIMESTAMP,
    >         Distance FLOAT
    >         )
    >         ROW FORMAT DELIMITED
    >         FIELDS TERMINATED BY ','
    >         STORED AS TEXTFILE
    >         LOCATION '/user/hduser/flights_dir' ;
OK
Time taken: 0.119 seconds
hive>
```

```
hive> DESCRIBE flights;
OK
flight_date             date
airlinecode             string
flightnum               string
origin                  string
destination             string
departuretime           int
departuredelay          timestamp
arrivaltime             timestamp
arrivaldelay            timestamp
airtime                 timestamp
distance                float
Time taken: 0.404 seconds, Fetched: 11 row(s)
hive>
```

Now Altered table



Now again loading data from

Load data local INPATH '/home/hduser/Downloads/sharedfolder/flights.csv'

OVERWRITE into TABLE flights;

```
hive> Load data local INPATH '/home/hduser/Downloads/sharedfolder/flights.csv'
    > OVERWRITE into TABLE flights;
Loading data to table airlines_db.flights
Table airlines_db.flights stats: [numFiles=0, numRows=0, totalSize=0, rawDataSize=0]
OK
Time taken: 1.219 seconds
hive> SELECT * FROM flights LIMIT 10 ;
OK
2014-04-01      19805   1       JFK     LAX     854     -6      1217    2       355     2475.0
2014-04-01      19805   2       LAX     JFK     944     14      1736    -29     269     2475.0
2014-04-01      19805   3       JFK     LAX     1224    -6      1614    39      371     2475.0
2014-04-01      19805   4       LAX     JFK     1240    25      2028    -27     264     2475.0
2014-04-01      19805   5       DFW     HNL     1300    -5      1650    15      510     3784.0
2014-04-01      19805   6       OGG     DFW     1901    126     640     95      385     3711.0
2014-04-01      19805   7       DFW     OGG     1410    125     1743    138     497     3711.0
2014-04-01      19805   8       HNL     DFW     1659    4       458     -22     398     3784.0
2014-04-01      19805   9       JFK     LAX     648     -7      1029    19      365     2475.0
2014-04-01      19805   10      LAX     JFK     2156    21      556     1       265     2475.0
Time taken: 0.123 seconds, Fetched: 10 row(s)
hive>
```

Enabling column names

set hive.cli.print.header=true;

set hive.resultset.use.unique.column.names=false;

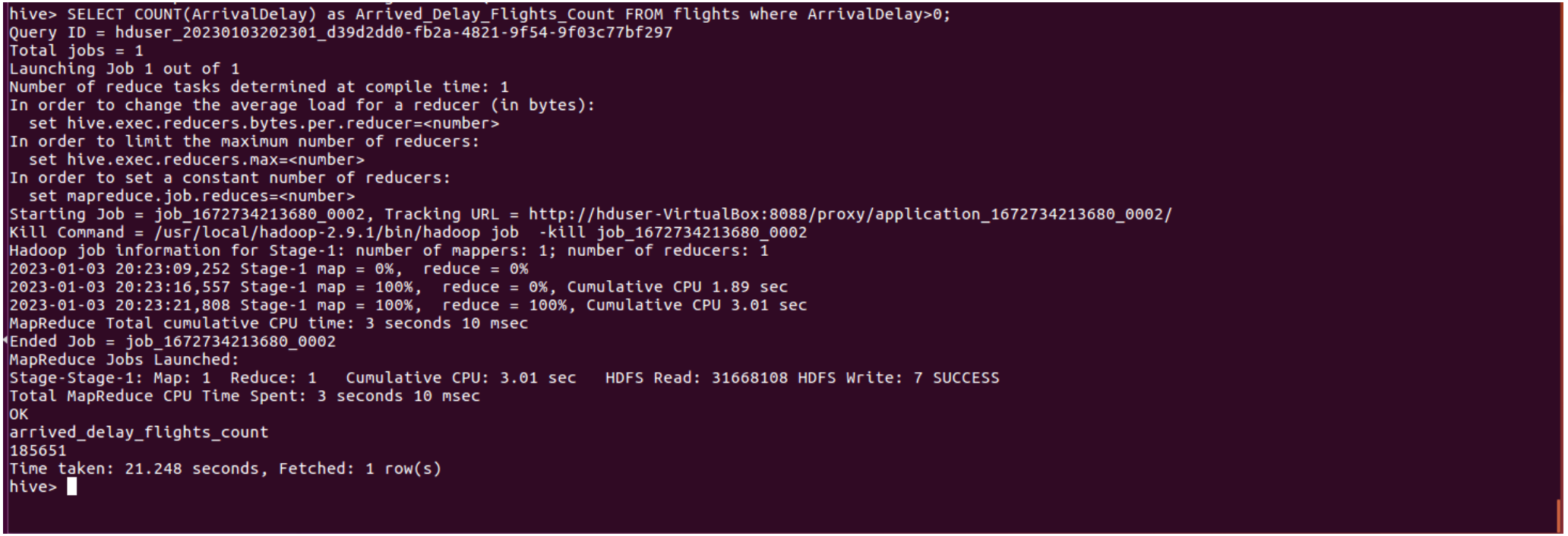
```
hive> SELECT * FROM flights LIMIT 10 ;
OK
flight_date     airlinecode     flightnum       origin  destination     departuretime   departuredelay  arrivaltime     arrivaldelay    airtime distance
2014-04-01      19805   1       JFK     LAX     854     -6      1217    2       355     2475.0
2014-04-01      19805   2       LAX     JFK     944     14      1736    -29     269     2475.0
2014-04-01      19805   3       JFK     LAX     1224    -6      1614    39      371     2475.0
2014-04-01      19805   4       LAX     JFK     1240    25      2028    -27     264     2475.0
2014-04-01      19805   5       DFW     HNL     1300    -5      1650    15      510     3784.0
2014-04-01      19805   6       OGG     DFW     1901    126     640     95      385     3711.0
2014-04-01      19805   7       DFW     OGG     1410    125     1743    138     497     3711.0
2014-04-01      19805   8       HNL     DFW     1659    4       458     -22     398     3784.0
2014-04-01      19805   9       JFK     LAX     648     -7      1029    19      365     2475.0
2014-04-01      19805   10      LAX     JFK     2156    21      556     1       265     2475.0
Time taken: 0.082 seconds, Fetched: 10 row(s)
hive>
```

Now,

2. Solve the following use cases

Find count of flights that had arrival delay

SELECT COUNT(ArrivalDelay) as Arrived_Delay_Flights_Count FROM flights where ArrivalDelay>0;

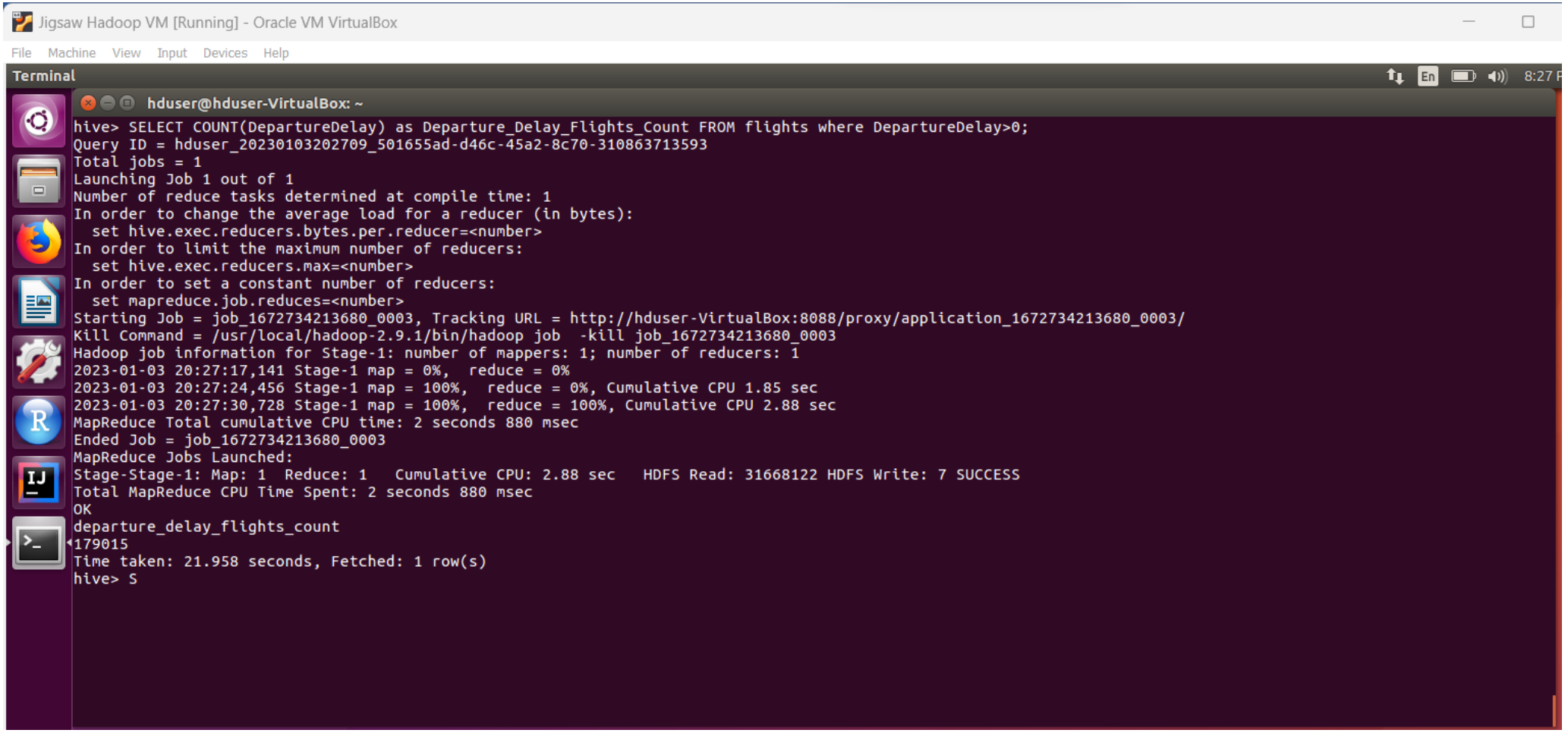


arrived_delay_flights_count

185651

Find count of flights that had departure delay

SELECT COUNT(DepartureDelay) as Departure_Delay_Flights_Count FROM flights where DepartureDelay>0;

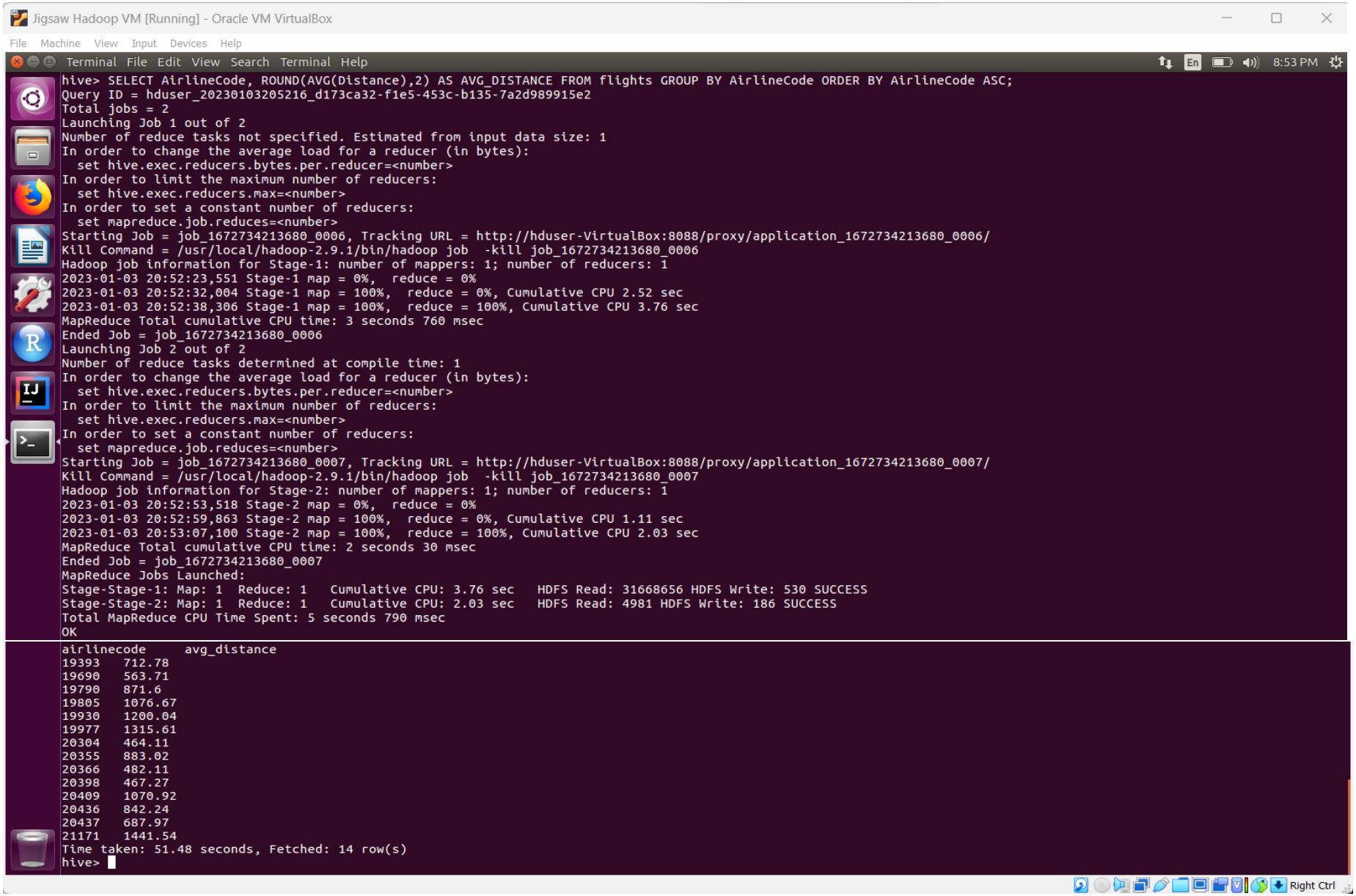


departure_delay_flights_count

179015

find the average distance travelled by a flight

SELECT AirlineCode, AVG(Distance) FROM flights GROUP BY AirlineCode ORDER BY ASC;



---

airlinecode avg_distance

19393 712.78

19690 563.71

19790 871.6

19805 1076.67

19930 1200.04

19977 1315.61

20304 464.11

20355 883.02

20366 482.11

20398 467.27

20409 1070.92

20436 842.24

20437 687.97

21171 1441.54

List the data that belong to the airline - American Airlines Inc

SELECT a.AirlineCode,a.Description,

f.flight_date,

f.flightnum,

f.origin,

f.destination,

f.departuretime,

f.departuredelay,

f.arrivaltime,

f.arrivaldelay,

f.airtime,

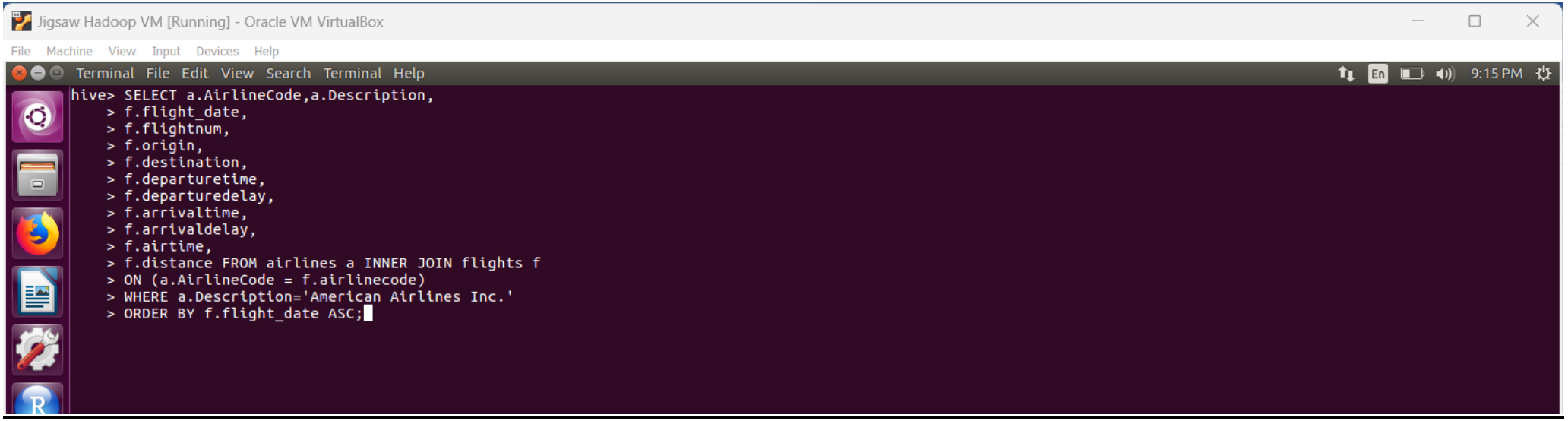
f.distance FROM airlines a INNER JOIN flights f

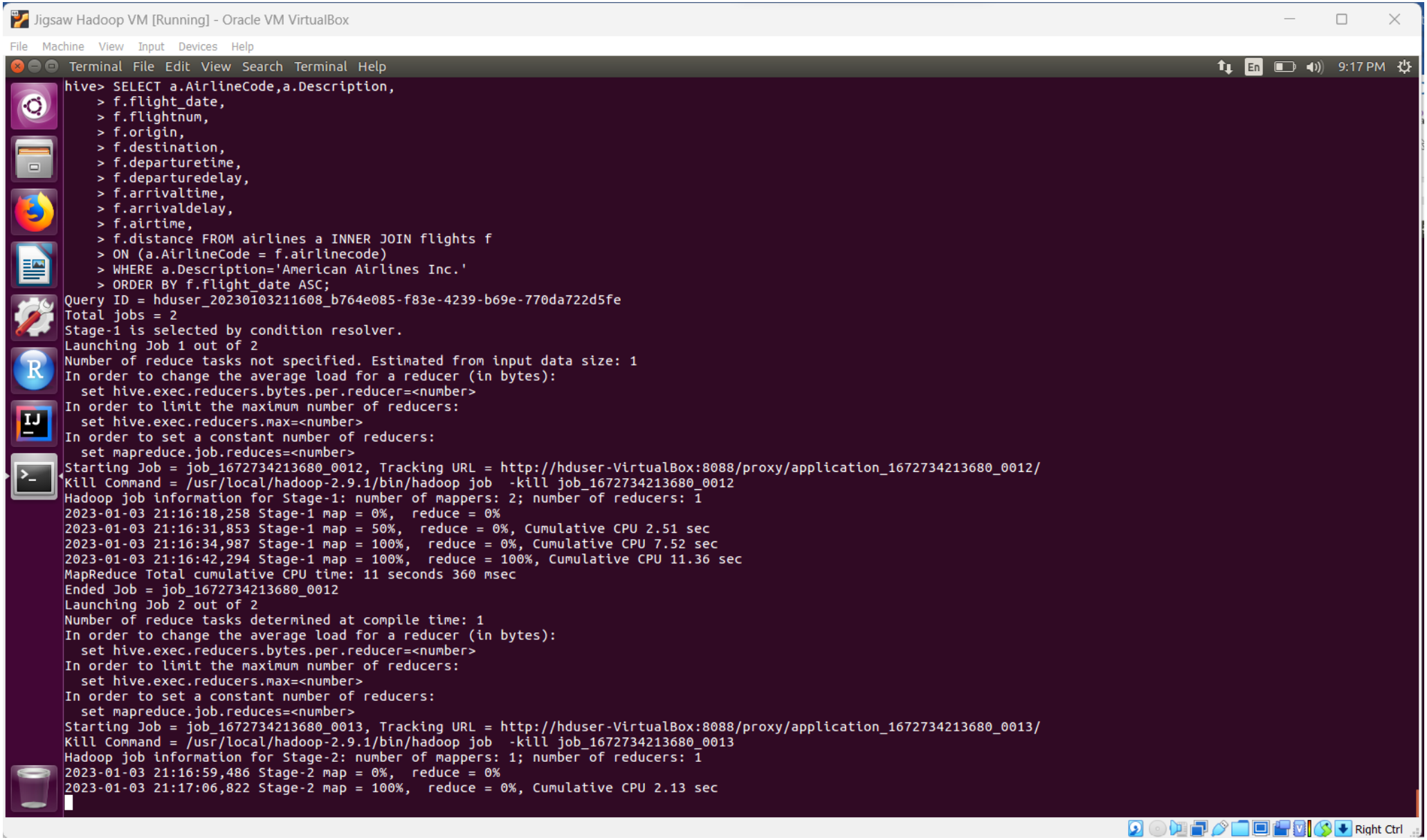
ON (a.AirlineCode = f.airlinecode)

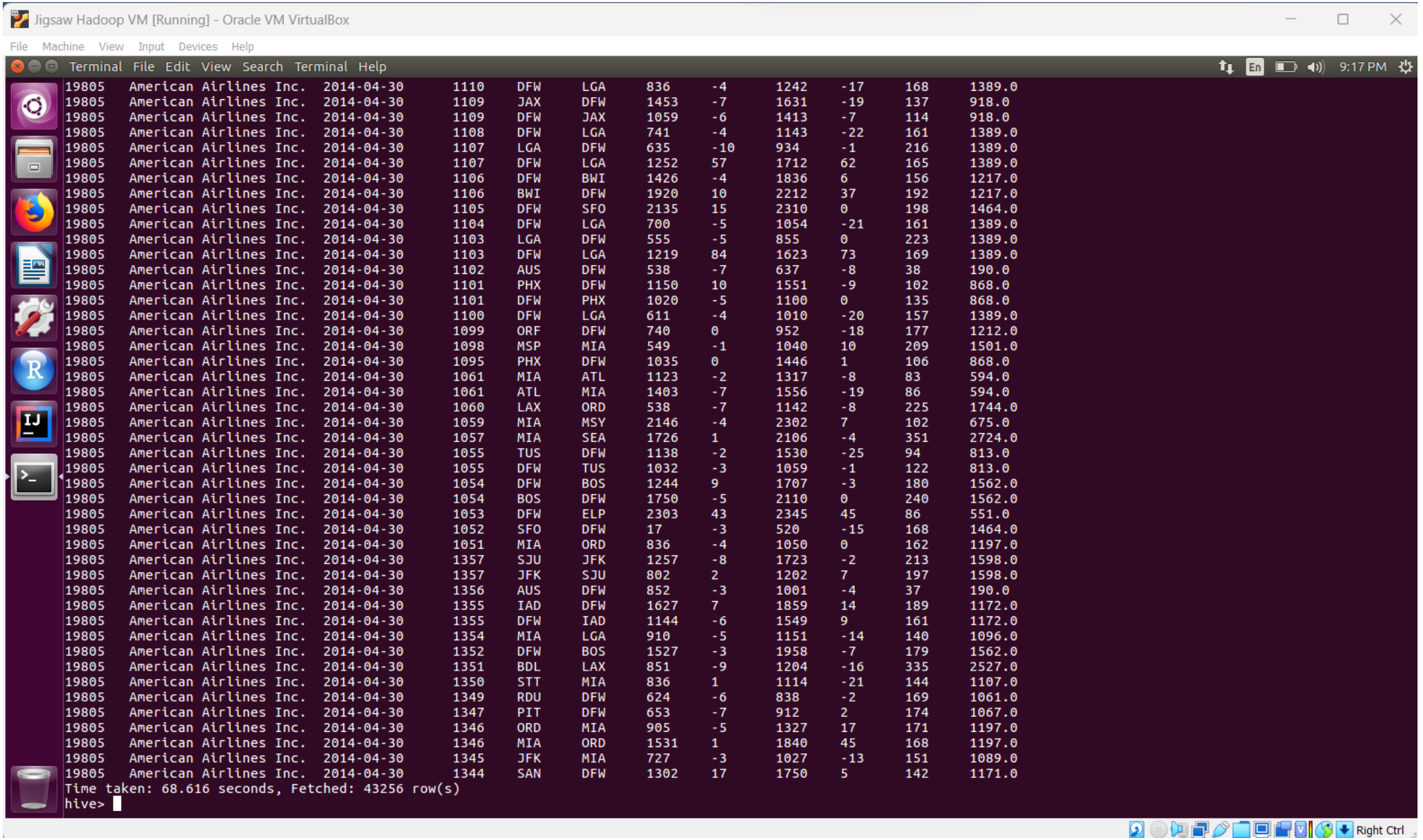
WHERE a.Description='American Airlines Inc.'

ORDER BY f.flight_date ASC;

RUNNING QUERY




Jigsaw Hadoop VM [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Terminal File Edit View Search Terminal Help
9:17 PM
hive> SELECT a.AirlineCode,a.Description,
> f.flight_date,
> f.flightnum,
> f.origin,
> f.destination,
> f.departuretime,
> f.departuredelay,
> f.arrivaltime,
> f.arrivaldelay,
> f.airtime,
> f.distance FROM airlines a INNER JOIN flights f
> ON (a.AirlineCode = f.airlinecode)
> WHERE a.Description='American Airlines Inc.'
> ORDER BY f.flight_date ASC;
Query ID = hduser_20230103211608_b764e085-f83e-4239-b69e-770da722d5fe
Total jobs = 2
Stage-1 is selected by condition resolver.
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 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_1672734213680_0012, Tracking URL = http://hduser-VirtualBox:8088/proxy/application_1672734213680_0012/
Kill Command = /usr/local/hadoop-2.9.1/bin/hadoop job -kill job_1672734213680_0012
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2023-01-03 21:16:18,258 Stage-1 map = 0%, reduce = 0%
2023-01-03 21:16:31,853 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 2.51 sec
2023-01-03 21:16:34,987 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.52 sec
2023-01-03 21:16:42,294 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 11.36 sec
MapReduce Total cumulative CPU time: 11 seconds 360 msec
Ended Job = job_1672734213680_0012
Launching Job 2 out of 2
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_1672734213680_0013, Tracking URL = http://hduser-VirtualBox:8088/proxy/application_1672734213680_0013/
Kill Command = /usr/local/hadoop-2.9.1/bin/hadoop job -kill job_1672734213680_0013
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-01-03 21:16:59,486 Stage-2 map = 0%, reduce = 0%
2023-01-03 21:17:06,822 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.13 sec
Right Ctrl



Thus, the above query displays the 43,256 records details for 'American Airlines Inc.' from airlines and flights both tables.