sudo usermod -aG vboxsf cloudera

sudo reboot

Hive Commnads:

```
cp /media/sf_file/flight_info.csv /home/cloudera
hdfs dfs -mkdir -p /user/hive/warehouse/flight_data
hdfs dfs -put /home/cloudera/flight_info.csv /user/hive/warehouse/flight_data
hdfs dfs -ls /user/hive/warehouse/flight_data
```

hive

1. show databases;
2. drop database (database_name) CASCADE;

3. create database flightdb;
4. use flightdb;

5. create table flight_info (
)
   row format delimited
   fields terminated by ','
   stored as textfile;

6. alter table flight_info add columns (aircraft_type string);

7. create external table flight_info_ext (
)
   row format delimited
   fields terminated by ','
   stored as textfile
   location '/user/hive/warehouse/flight_data/';

8. load data inpath '/user/hive/warehouse/flight_data/flight_info.csv' into table flight_info;

9. SELECT * FROM flight_info LIMIT 10;

10. insert into table flight_info values (….);

11. create table airports (
    airport_code string,
    airport_name string
    );

12. insert into airports values (…)

13.
```
SELECT
  f.flight_num,
  f.origin,
  a.airport_name
FROM
  flight_info f
JOIN
  airports a
ON
  (f.origin = a.airport_code)
LIMIT 10;
```

14.
```
create index flight_num_idx
  on table flight_info (flight_num)
as
 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'
with
 deferred rebuild;
```

15.
```
set hive.cli.print.header=true;
```

16.
```
SELECT year, month, day, AVG(dep_delay) AS avg_dep_delay
FROM flight_info
GROUP BY year, month, day
ORDER BY year, month, day;
```