Hive Commands
Group Members:
Muhammad Mujtaba Afzal 14882
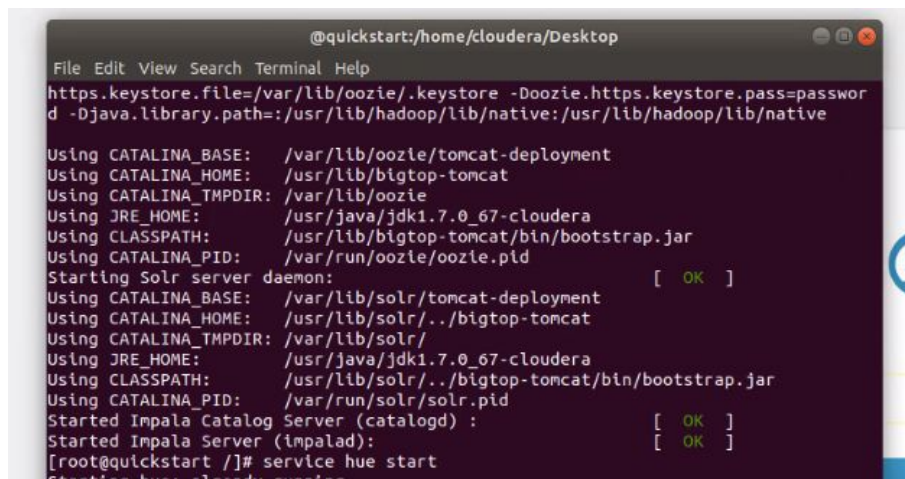Akash Memon 14884

Prerequisites:
Ubuntu 18
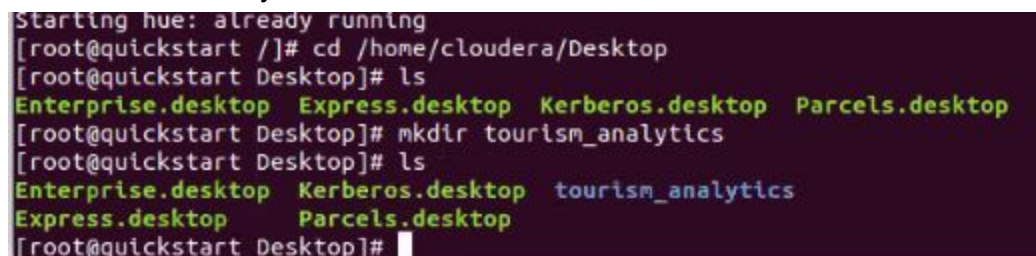Docker installed
Cloudera installed

Loading data to HDFS:
Run Cloudera container:
docker run --hostname=quickstart.cloudera --privileged=true -t -v
/home/akategra/Desktop/dataset:/user/cloudera/DWProject -i -p 8889:8888 -p
7180:7180 cloudera/quickstart /usr/bin/docker-quickstart

Create new directory in Container:
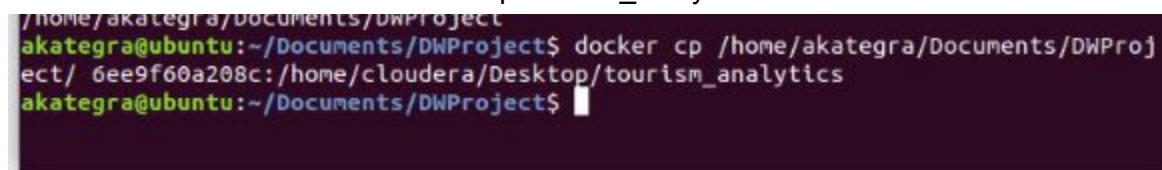
Copy data files from Host machine to Docker container:
In a new terminal run the following command:
docker cp ~/home/akategra/Documents/DWProject
6ee9f60a208c:/home/cloudera/Desktop/tourism_analyitics/

6ee9f60a208c is the container ID.

Create directory in HDFS and copy data from Docker:

```
[root@quickstart Desktop]# hadoop fs -ls
[root@quickstart Desktop]# hadoop fs -mkdir /user/data/
[root@quickstart Desktop]# hadoop fs -put -f /home/cloudera/Desktop/tourism_anal
ytics /user/data
[root@quickstart Desktop]#
```

Check the directory for uploaded files:

```
ytics /user/data
[root@quickstart Desktop]# hadoop fs -ls /user/data/tourism_analytics
Found 1 items
drwxr-xr-x   - root supergroup          0 2021-01-07 20:26 /user/data/tourism_an
alytics/DWProject
[root@quickstart Desktop]# hadoop fs -ls /user/data/tourism_analytics/DWProject
Found 8 items
-rw-r--r--   1 root supergroup         75 2021-01-07 20:25 /user/data/tourism_an
alytics/DWProject/TOI_Event_Type.csv
-rw-r--r--   1 root supergroup  272560325 2021-01-07 20:25 /user/data/tourism_an
alytics/DWProject/TOI_Events.csv
-rw-r--r--   1 root supergroup         75 2021-01-07 20:25 /user/data/tourism_an
alytics/DWProject/TOI_Type.csv
-rw-r--r--   1 root supergroup    1122179 2021-01-07 20:25 /user/data/tourism_an
alytics/DWProject/cities.csv
-rw-r--r--   1 root supergroup     167803 2021-01-07 20:25 /user/data/tourism_an
alytics/DWProject/states.csv
-rw-r--r--   1 root supergroup  185855372 2021-01-07 20:25 /user/data/tourism_an
alytics/DWProject/things_of_interest.csv
-rw-r--r--   1 root supergroup  143801718 2021-01-07 20:26 /user/data/tourism_an
alytics/DWProject/tourism_organisations.csv
-rw-r--r--   1 root supergroup         48 2021-01-07 20:26 /user/data/tourism_an
alytics/DWProject/travel_type.csv
[root@quickstart Desktop]#
```
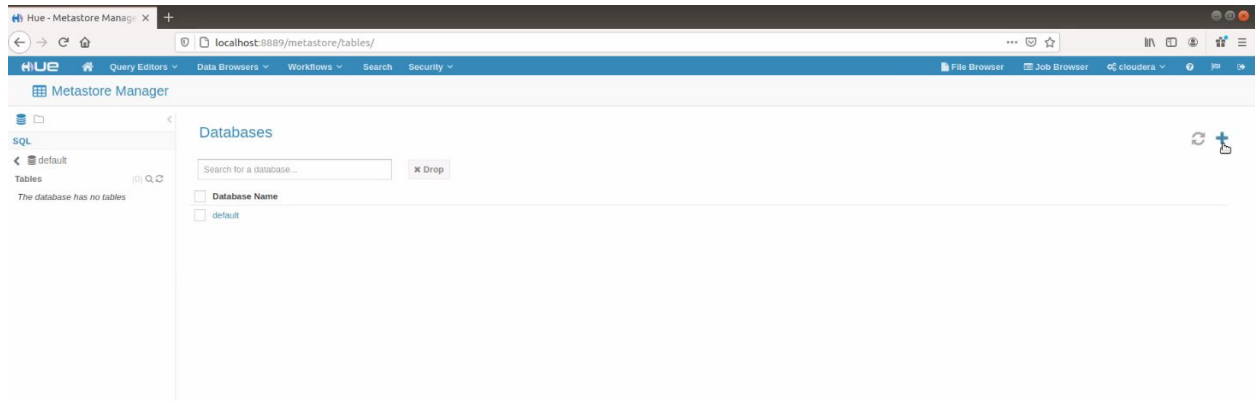
Open Browser and access localhost:8889
Goto databrowser> Metastore Manager
Click on plus sign in top right corner as shown below to create a new data base:

Create database by filing the required details:

Create new table using the csv:

Choose the file:

Click next to create the table:

A preview is shown:

Upload all the files in this way creating tables in your database.

Querying the data:

Goto Query editor and choose Hive.

```
1 select cities.name, floor(avg(things_of_interest.lowerpricers)) as Average_lower_price from cities, things_of_interest where cities.city_id = things_of_interest.city_fk
2 AND things_of_interest.lowerpricers < 5000 group by cities.name
```

Execute   Save as...   Explain   Format   or create a   New query

Recent queries   Query   Log   Columns   **Results**   Chart

cities.name	average_lower_price
City 10056	3083
City 10057	2900
City 10058	3500
City 10059	3500
City 1006	2571

```
1 select avg(things_of_interest.higherpricers) from things_of_interest, cities where cities.city_id = things_of_interest.city_fk
2 AND things_of_interest.toi_type_fk in (select toi_type.toi_type_id from toi_type where toi_type.toi_type = "Hotel") group by things_of_interest.toi_id
```

Execute   Save as...   Explain   Format   or create a   New query

Recent queries   Query   Log   Columns   **Results**   Chart

_c0
27000
15000
20000
25000

```
1 select toi_type.toi_type, count(distinct things_of_interest.toi_id) from toi_type, things_of_interest
2 where toi_type.toi_type_id = things_of_interest.toi_type_fk group by toi_type.toi_type
3
```

Execute   Save as...   Explain   Format   or create a   New query

Recent queries   Query   Log   Columns   **Results**   Chart

toi_type.toi_type	_c1
1	Cafe	6516
2	Hotel	6630
3	Monument	6511
4	Museum	6555
5	Restaurant	6555

```
1 select count(tourism_organisations.org_id) as Bus_Travel_Count from tourism_organisations where tourism_organisations.travelcode_fk = 2 AND tourism_organisations.state_fk = 5
```

Execute   Save as...   Explain   Format   or create a   New query

	bus_travel_count
1	3

```
1 SELECT * from tourism_organisations where tourism_organisations.state_fk == 1
```

Execute   Save as...   Explain   Format   or create a   New query

	tourism_organisations.org_id	tourism_organisations.state_fk	tourism_organisations.travelcode_fk	tourism_organisations.name	tourism_organisations.address	tourism_organisations.contact	tourism_organisation
1	8355	1	1	Organization 8355	"509 Building 1	Street 9"	(021) 92188279
2	14144	1	1	Organization 14144	"466 Building 18	Street 3"	(021) 92162306
3	28877	1	3	Organization 28877	"557 Building 11	Street 7"	(021) 92177630
4	NULL	1	1	Organization 36583	"1667 Building 11	Street 7"	(021) 92193848

```
1 SELECT * from tourism_organisations where tourism_organisations.state_fk == 2 and tourism_organisations.travelcode_fk == 3
```

Execute   Save as...   Explain   Format   or create a   New query

	tourism_organisations.org_id	tourism_organisations.state_fk	tourism_organisations.travelcode_fk	tourism_organisations.name	tourism_organisations.address	tourism_organisations.contact	tourism_organisation
1	NULL	2	3	Organization 47207	2406 Building 0 Street 1	(021) 92142753	https://www.organiza
2	NULL	2	3	Organization 61928	2897 Building 4 Street 1	(021) 92168818	https://www.organiza
3	NULL	2	3	Organization 81945	2951 Building 5 Street 3	(021) 92140295	https://www.organiza
4	NULL	2	3	Organization 83524	1434 Building 15 Street 6	(021) 92136106	https://www.organiza
5	NULL	2	3	Organization 123585	877 Building 6 Street 10	(021) 92120102	https://www.organiza

```
1 SELECT * from cities where cities.state_fk == 10
```

Execute   Save as...   Explain   Format   or create a   New query

	cities.city_id	cities.state_fk	cities.name
1	13776	10	City 13776
2	17809	10	City 17809
3	27446	10	City 27446
4	30648	10	City 30648

Time	Query	Result
01/07/2021 1:04:09 PM	select floor(avg(things_of_interest.higherpricers)) as Average_Hotel_Higher_Price from things_of_interest, cities where cities.city_id = things_of_interest.city_fk AND things_of_interest.toi_type_fk in (select toi_type.toi_type_id from toi_type where toi_type.toi_type = "Hotel") group by things_of_interest.toi_id	See results...
01/07/2021 1:03:26 PM	select floor(avg(things_of_interest.higherpricers)) from things_of_interest, cities where cities.city_id = things_of_interest.city_fk AND things_of_interest.toi_type_fk in (select toi_type.toi_type_id from toi_type where toi_type.toi_type = "Hotel") group by things_of_interest.toi_id	
01/07/2021 1:03:02 PM	select cities.name,floor(avg(things_of_interest.higherpricers)) from things_of_interest, cities where cities.city_id = things_of_interest.city_fk AND things_of_interest.toi_type_fk in (select toi_type.toi_type_id from toi_type where toi_type.toi_type = "Hotel") group by things_of_interest.toi_id	
01/07/2021 12:59:52 PM	select avg(things_of_interest.higherpricers) from things_of_interest, cities where cities.city_id = things_of_interest.city_fk AND things_of_interest.toi_type_fk in (select toi_type.toi_type_id from toi_type where toi_type.toi_type = "Hotel") group by things_of_interest.toi_id	See results...
01/07/2021 12:57:07 PM	select avg(things_of_interest.higherpricers) from things_of_interest, cities where cities.city_id = things_of_interest.city_fk AND things_of_interest.toi_type_fk = (select toi_type.toi_type_id from toi_type where toi_type.toi_type = "Hotel") group by things_of_interest.toi_id	
01/07/2021 12:56:37 PM	select avg(things_of_interest.higherpricers) from things_of_interest, cities where cities.city_id = things_of_interest.city_fk AND things_of_interest.toi_type_fk = (select toi_type.toi_type_id from toi_type where toi_type.toi_type = "Hotel") group by things_of_interest.toi_id	
01/07/2021 12:56:08 PM	select avg(things_of_interest.higherpricers) from things_of_interest, cities where cities.city_id = things_of_interest.city_fk AND toi_type_fk = (select toi_type.toi_type_id from toi_type where toi_type.toi_type = "Hotel") group by things_of_interest.toi_id	
01/07/2021 12:55:40 PM	select avg(things_of_interest.higherpricers) from things_of_interest, cities where cities.city_id = things_of_interest.city_fk AND toi_type_fk = (select toi_type_id from toi_type where toi_type.toi_type = "Hotel") group by things_of_interest.toi_id	
01/07/2021 12:47:43 PM	select cities.name, floor(avg(things_of_interest.lowerpricers)) as Average_lower_price from cities, things_of_interest where cities.city_id = things_of_interest.city_fk AND things_of_interest.lowerpricers < 5000 group by cities.name	See results...
01/07/2021 12:46:25 PM	select cities.name, floor(avg(things_of_interest.lowerpricers)) from cities, things_of_interest where cities.city_id = things_of_interest.city_fk AND floor(avg(things_of_interest.lowerpricers)) < 5000 group by cities.name	