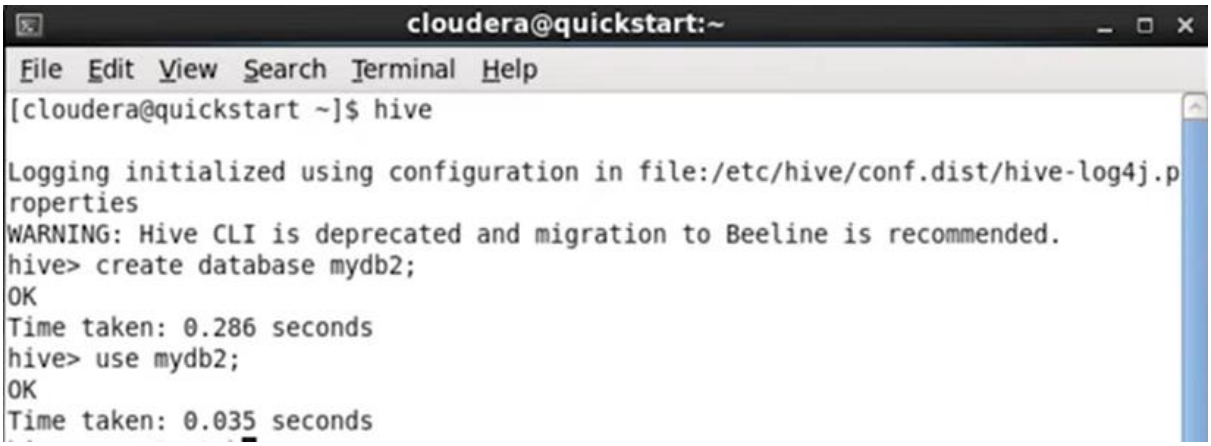
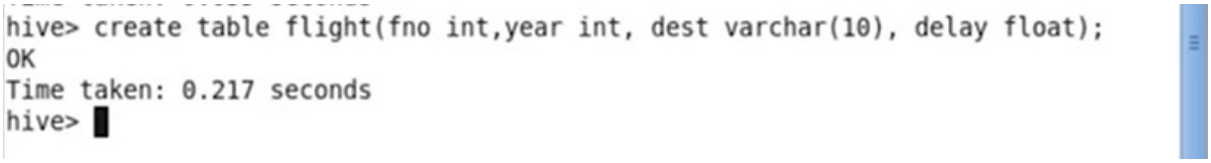
Name - Prathamesh Chaudhari
Roll no - TI10

Assignment No 3

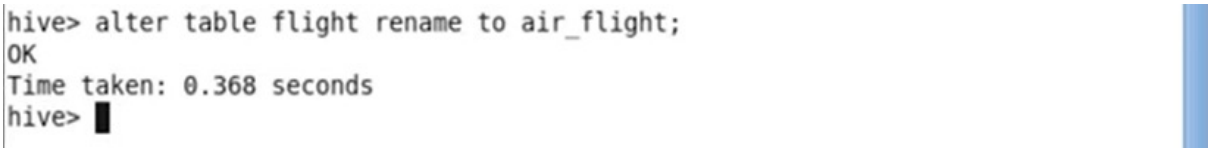
1. Creating Database



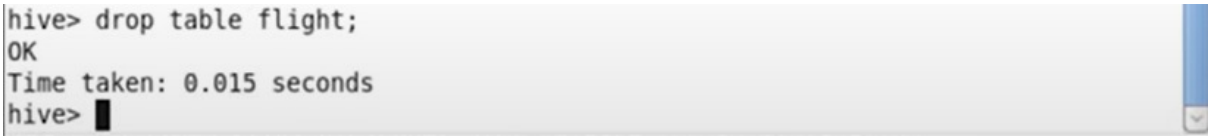
2. Creating Table



3. Alter Table



4. Drop Table



5. Creating an external Hive table

```
hive> create table flight(fno int, year int, dest varchar(10), delay float)
    > row format delimited
    > fields terminated by ','
    > lines terminated by '\n'
    > stored as textfile;
OK
Time taken: 0.072 seconds
hive>
```

6. Inserting records

```
hive> insert into flight values(123,2009,"mumbai",30.6);
Query ID = cloudera_20240212220707_24d53363-61a2-4bf2-8ec1-23ac57128f3e
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1707371758511_0001, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1707371758511_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1707371758511_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2024-02-12 22:07:28,715 Stage-1 map = 0%,  reduce = 0%
```

```
C
MapReduce Total cumulative CPU time: 1 seconds 30 msec
Ended Job = job_1707371758511_0001
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: hdfs://quickstart.cloudera:8020/user/hive/warehouse/mydb2.db/fli
ght/.hive-staging_hive_2024-02-12_22-07-17_828_8462008546613947698-1/-ext-10000
Loading data to table mydb2.flight
Table mydb2.flight stats: [numFiles=1, numRows=1, totalSize=21, rawDataSize=20]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 1.03 sec   HDFS Read: 4157 HDFS Write: 8
9 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 30 msec
OK
Time taken: 18.769 seconds
hive>
```

```
hive> select * from flight;
OK
123     2009    mumbai  30.6
124     2008    pune    50.6
Time taken: 0.076 seconds, Fetched: 2 row(s)
hive>
```

7. Load table with data

```
hive> load data local inpath "f.txt"
    > overwrite into table flight;
Loading data to table mydb2.flight
Table mydb2.flight stats: [numFiles=1, numRows=0, totalSize=112, rawDataSize=0]
OK
Time taken: 0.259 seconds
hive>
```

```
hive> select * from flight;
OK
923     2009    navi mumba      60.0
156     2009    kolhapur        78.0
112     2009    punr    0.0
322     2008    nagpur  90.0
132     2008    aurngabad       67.0
Time taken: 0.043 seconds, Fetched: 5 row(s)
hive>
```

8. Joining Tables

```
hive> create table nflight(fno int, year int, source varchar(10))
    > row format delimited
    > fields terminated by ','
    > lines terminated by '\n'
    > stored as textfile;
OK
Time taken: 0.053 seconds
hive>
```

```
hive> insert into nflight values(923,2009,"pune");
Query ID = cloudera_20240212222020_961c74f5-1a27-4a8e-bf71-44adc8602789
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1707371758511_0003, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1707371758511_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1707371758511_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2024-02-12 22:20:24,828 Stage-1 map = 0%,  reduce = 0%
2024-02-12 22:20:30,243 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.01 se
c
```

```
MapReduce Total cumulative CPU time: 1 seconds 10 msec
Ended Job = job_1707371758511_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: hdfs://quickstart.cloudera:8020/user/hive/warehouse/mydb2.db/nfl
ight/.hive-staging_hive_2024-02-12_22-20-17_980_7655139385168538320-1/-ext-10000
Loading data to table mydb2.nflight
Table mydb2.nflight stats: [numFiles=1, numRows=1, totalSize=14, rawDataSize=13]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 1.01 sec   HDFS Read: 3894 HDFS Write: 8
3 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 10 msec
OK
Time taken: 14.517 seconds
hive>
```

```
hive> select a.fno, a.year,a.dest,a.delay,b.source
    > from flight a join nflight b
    > on(a.fno=b.fno);
Query ID = cloudera_20240212222323_fdf12060-eef6-4b51-b17d-18a693772a26
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20240212222323_fdf12060-eef6-4b51-b17d-
18a693772a26.log
Total MapReduce CPU Time Spent: 1 seconds 60 msec
OK
923     2009    navi mumba      60.0    pune
Time taken: 19.787 seconds, Fetched: 1 row(s)
hive>
```

9. Creating index on Flight information Table

```
hive> create index flight_index on table flight(fno)
    > as 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'
    > WITH DEFERRED REBUILD;
OK
Time taken: 0.39 seconds
hive>
```

```
hive> show tables;
OK
air_flight
flight
mydb2__flight_flight_index__
nflight
values__tmp__table__1
values__tmp__table__2
values__tmp__table__4
Time taken: 0.059 seconds, Fetched: 7 row(s)
hive>
```