[cloudera@quickstart ~]$ hive

hive> create database flight\_dbs;

hive> show databases;

a) Creating, Dropping, and altering Database tables.

a2)creating table

hive> use flight\_dbs;

hive> create table flight\_info(

> flight\_no INT,

> day\_of\_week INT,

> dep\_time INT,

> origin STRING,

> dest STRING

> )

> row format delimited

> fields terminated by ','

> stored as textfile;

hive> show table

hive> desc flight\_info;

a3)altering table by changing name or add column

hive> alter table flight\_info ADD columns(distance INT) ;

hive> desc flight\_info;

a4)dropping table

hive> drop table flight\_info;

hive> show tables;

[cloudera@quickstart ~]$ hdfs dfs -ls /

[cloudera@quickstart ~]$ hdfs dfs -mkdir -p /flight\_data

[cloudera@quickstart ~]$ hdfs dfs -ls /

[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/flight\_info.csv /flight\_data

[cloudera@quickstart ~]$ hdfs dfs -ls /flight\_data

b) Creating an external Hive table.(flight\_ext)

[cloudera@quickstart ~]$ hive

hive> create external table flight\_ext(

> year INT,

> month INT,

> day INT,

> day\_of\_week INT,

> dep\_time INT,

> crs\_dep\_time INT,

> arr\_time INT,

> crs\_arr\_time INT,

> unique\_carrier STRING,

> flight\_num INT,

> tail\_num STRING,

> actual\_elapsed\_time INT,

> crs\_elapsed\_time INT,

> air\_time INT,

> arr\_delay INT,

> dep\_delay INT,

> origin STRING,

> dest STRING,

> distance INT,

> taxi\_in INT,

> taxi\_out INT,

> cancelled INT,

> cancellation\_code STRING,

> diverted INT,

> carrier\_delay STRING,

> weather\_delay STRING,

> nas\_delay STRING,

> late\_aircraft\_delay STRING,

> security\_delay STRING

> )

> row format delimited

> fields terminated by ','

> stored as textfile

> location '/flight\_data';

hive> desc flight\_ext;

c) Load table with data, insert new values and field in the table, Join tables with Hive

c1)create internal table(flight\_int)

hive> create table flight\_int AS

> select

> year,month,day,flight\_num,dep\_delay,origin,dest

> FROM flight\_ext;

hive> select \* from flight\_int LIMIT 10;

c3)insert new values in table

hive> insert into flight\_int values(2010,2,21,505,6,'ENG','IND');

hive> select \* from flight\_int WHERE flight\_num = 505;

c4) join tables

hive> create table f\_price(flight\_num int,price float);

hive> insert into f\_price values (505,5000.0);

hive> select \* from f\_price;

hive> select a.flight\_num,a.day,a.month,a.year,b.price

> from flight\_int a JOIN f\_price b ON (a.flight\_num = b.flight\_num);

d)Create index on Flight Information Table

hive> create index flight\_index on table flight\_int(flight\_num)

> AS 'COMPACT'

> with DEFERRED REBUILD;

hive> show index on flight\_int;

e1)find avg departure delay

hive> select AVG(dep\_delay) FROM flight\_int;

e2)Find the average departure delay per day in 2008.

hive> select day,month,year,AVG(dep\_delay) AS avg\_dep\_del

> FROM flight\_int

> WHERE dep\_delay IS NOT NULL

> GROUP BY day,month,year;

e3) find monthly departure delay

hive> select month,AVG(dep\_delay)

> FROM flight\_int

> WHERE dep\_delay IS NOT NULL

> GROUP BY month;