Group A: Assignments based on the Hadoop

3. Write an application using HiveQL for flight information system which will include

a. Creating, Dropping, and altering Database tables.

b. Creating an external Hive table.

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

d. Create index on Flight Information Table

e. Find the average departure delay per day in 2008.

Hive Commands as Follows:

Open Terminal

-> Hive

Hive -> **Create database** mydb;

-> use mydb;

-> **create table** Flight (Fno int, year int, dest varchar(20), delay float);

-> **alter table** Flight rename to air-flight;

-> alter table air-flight **add columns** (source varchar(20));

-> **drop table** flight;

-> **desc** air-flight;

-> Create table flight (Fno int, year int, dest varchar(20), delay float)

-> row format delimited

-> fields terminated by ','

-> lines terminated by '\n'

-> stored as textfile;

-> **insert** into flight values (123, 2009, Mumbai, 30.6);

-> insert into flight values (124, 2008, Pune, 80.6);

-> **select** * from flight;

Open New Terminal

-> gedit f.txt

923, 2009, Navi Mumbai, 60.0

156, 2008, Kolp, 78.0 1

12, 2007, Pune, 0.0

322, 2008, Nagpur, 90.0

132, 2008, Aurangabad, 67.0

-> pwd

-> ls

Hive> **load** data local inpath 'f.txt'

-> **overwrite** into table flight;

-> select * from flight;

-> create table nflight (fno int, year int, source VARCHAR(50))

> row format delimited

```
        > fields terminated by ','
        > lines terminated by '\n'
        > stored as textfile;
-> select * from nflight;
-> insert into nflight values (205, 2001, 'Pune');
-> select a.fno, a.year, a.dest, a.delay, b.source
         > from flights a join nFlight b
        > on (a.fno = b.fno);
-> select * from flight;
-> select * from nflight;
-> create index flight_index  on table flight (fno)
        > as 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'
        > WITH DEFERRED REBUILD;
-> show tables;
```

Loading Data from Internal to External Hive Table

1. **Create an Internal Hive Table (hive_int).**
2. **Create a data.txt file containing the data to be loaded into the internal table.**
3. **Create an External Hive Table (hive_out).**
4. **Load data from the internal Hive table (hive_int) to the external Hive table (hive_out).**
5. **Check the contents of the external Hive table.**

Creating a Database and Tables

```
hive> create database mydb;
hive> use mydb;
hive> create table hive_int (id int, name varchar(10), sal float)
        >row format delimited
        >fields terminated by ','
        >lines terminated by '\n'
        >stored as textfile;
```

Creating a Data File

- Open a Nano or Gedit terminal and create **data.txt:**

1, Ram, 90.20
2, Shyam, 96.58
3, Sita, 80.86
4, Bajal, 67.89

Creating and Managing Hive Tables

Loading Data into Hive Table

```
hive> load data local inpath 'data.txt' into table hive_int;
hive> select * from hive_int;
```

Creating an External Hive Table

```
hive> create external table hive_table (id int, name varchar(10), sal float)
        >row format delimited
        >fields terminated by ','
        >lines terminated by '\n'
        >stored as textfile;
```

Inserting Data into a Table

```
hive> insert into hive_ext select * from hive_int;
hive> select * from hive_ext;
```
