4. Write an application using HBase and HiveQL for flight information

system which will include

a) Creating, Dropping, and altering Database tables

b) Creating an external Hive table to connect to the HBase for Customer

Information 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.

**a) Creating, Dropping, and altering Database tables**

**Using Hbase**

**#Create Table:**

hbase(main):002:0> create 'flight','finfo','fsch'

**#Table Created-list**

hbase(main):003:0> list

**#Insert records in created table**

hbase(main):004:0> put 'flight',1,'finfo:source','pune'

hbase(main):008:0> put 'flight',1,'finfo:dest','mumbai'

hbase(main):010:0> put 'flight',1,'fsch:at','10.25a.m.'

hbase(main):011:0> put 'flight',1,'fsch:dt','11.25 a.m.'

hbase(main):012:0> put 'flight',1,'fsch:delay','5'

hbase(main):015:0> put 'flight',2,'finfo:source','pune'

hbase(main):016:0> put 'flight',2,'finfo:dest','kolkata'

hbase(main):017:0> put 'flight',2,'fsch:at','7.00a.m.'

hbase(main):018:0> put 'flight',2,'fsch:dt','7.30a.m.'

hbase(main):019:0> put 'flight',2,'fsch:delay','2'

hbase(main):021:0> put 'flight',3,'finfo:source','mumbai'

hbase(main):022:0> put 'flight',3,'finfo:dest','pune'

hbase(main):023:0> put 'flight',3,'fsch:at','12.30p.m.'

hbase(main):024:0> put 'flight',3,'fsch:dt','12.45p.m.'

hbase(main):025:0> put 'flight',3,'fsch:delay','1'

hbase(main):026:0> put 'flight',4,'finfo:source','mumbai'

hbase(main):027:0> put 'flight',4,'finfo:dest','delhi'

hbase(main):028:0> put 'flight',4,'fsch:at','2.00p.m.'

hbase(main):029:0> put 'flight',4,'fsch:dt','2.45p.m.'

hbase(main):030:0> put 'flight',4,'fsch:delay','10'

**#Display Records from Table ‘flight’**

hbase(main):031:0> scan 'flight'

**#Alter Table (add one more column family)**

hbase(main):036:0> alter 'flight',NAME=>'revenue'

**#Insert records into added column family**

hbase(main):038:0> put 'flight',4,'revenue:rs','45000'

**#Check the updates**

hbase(main):039:0> scan 'flight'

**#Delete Column family**

hbase(main):040:0> alter 'flight',NAME=>'revenue',METHOD=>'delete'

**#changes Reflected in Table**

hbase(main):041:0> scan 'flight'

**#Drop Table**

**#Create Table for dropping**

hbase(main):046:0\* create 'tb1','cf'

**#Drop Table**

hbase(main):048:0> drop 'tb1'

**#Disable table**

hbase(main):049:0> disable 'tb1'

hbase(main):050:0> drop 'tb1'

hbase(main):051:0> list

**#Read data from table for row key 1:**

hbase(main):052:0> get 'flight',1

**Read data for particular column from HBase table:**

hbase(main):053:0> get 'flight','1',COLUMN=>'finfo:source'

**Read data for multiple columns in HBase Table:**

hbase(main):054:0> get 'flight','1',COLUMN=>['finfo:source','finfo:dest']

hbase(main):055:0> scan 'flight',COLUMNS=>'finfo:source'

**b) Creating an external Hive table to connect to the HBase for Customer Information Table**

**Covers===>**

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

**# Create the external table emp using hive**

hive>create external table empdata2 ( ename string, esal int)

row format delimited fields terminated by "," stored as textfile location

"/home/saurabh/Desktop/empdata2";

hive>load data local inpath '/home/saurabh/empdb.txt' into table empdata2;

**#Create External Table in hive referring to hbase table**

**# create hbase table emphive first**

hbase(main):003:0> create 'emphive', 'cf'

**#create hive external table**

CREATE external TABLE hive\_table\_emp(id int, name string, esal string)

STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'

WITH SERDEPROPERTIES ("hbase.columns.mapping" = ":key,cf:name,cf:esal")

TBLPROPERTIES ("hbase.table.name" = "emphive");

**# load data into hive\_table\_emp**

**(Hive doesn’t allow directly inserting data into external hive table)**

**#for that create one hive table(managed table in hive)**

**Managed table** and External **table in Hive**. There are two types of **tables in Hive** ,one is **Managed table** and second is external **table**. The difference is , when you drop a **table**, if it is **managed table hive** deletes both data and meta data,if it is external **table Hive** only deletes metadata.

hive>create table empdbnew(eno int, ename string, esal int) row format delimited fields terminated by ',' stored as textfile;

**#load data in managed table**

hive>load data local inpath '/home/saurabh/Desktop/empdbnew.txt' into table empdbnew;

**#Load data in external table from managed table.**

hive>INSERT INTO hive\_table\_emp select \* from empdbnew;

hive> select \* from hive\_table\_emp;

1 sachin 5000

2 yuvraj 3000

3 yogesh 2500

4 rahul 2000

5 sunil 4000

6 pranav 3000

**#display records where salary is greater than 4000**

hive> select \* from hive\_table\_emp where esal>4000;

**#Check hbase for updates(The records are available in associated Hbase table)**

hbase(main):008:0> scan 'emphive'

**# Creating external table in Hive referring to Hbase**

**#referring to flight table created in Hbase**

hive> CREATE external TABLE hbase\_flight\_new(fno int, fsource string,fdest string,fsh\_at string,fsh\_dt string,fsch\_delay string)

STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'

WITH SERDEPROPERTIES

("hbase.columns.mapping"=":key,finfo:source,finfo:dest,fsch:at,fsch:dt,fsch:delay")

TBLPROPERTIES ("hbase.table.name" = "flight1");

**#table created in hive**

hive> show tables;

**# Display records from that table**

hive> select \* from hbase\_flight\_new;

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

**#calculate average delay**

hive> select sum(delay) from hbase\_flight\_new;

**d) Create index on Flight information Table**

hive> CREATE INDEX hbasefltnew\_index1

> ON TABLE hbase\_flight\_new (fsch\_delay)

> AS 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'

> WITH DEFERRED REBUILD;

OK

**#show index on table hbase\_flight\_new**

hive> SHOW INDEX ON hbase\_flight\_new;

**#join two tables in Hive**

**#create table B for join**

hive> create table empinfo(empno int, empgrade string) row format delimited fields terminated by ',' stored as textfile;

**#Load Data into table**

hive> load data local inpath '/home/hduser/Desktop/empinfo.txt' into table empinfo;

**#insert data into the table**

hive> load data local inpath '/home/hduser/Desktop/empinfo.txt' into table empinfo;

**# Table A empdbnew**

hive> select \* from empdbnew;

1 sachin 5000

2 yuvraj 3000

3 yogesh 2500

4 rahul 2000

5 sunil 4000

6 pranav 3000

**# Table B empinfo**

hive> select \* from empinfo;

1 A

2 B

3 B

4 B

5 B

6 A

**#Join two tables(empdbnew with empinfo on empno)**

hive> SELECT eno, ename, empno, empgrade FROM empdbnew JOIN empinfo ON eno = empno;

**#Join==> Result**