2. HIVE 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.

ANS:

1.Go to terminal on cloudera then hive

2. Hive🡪create database folder2;

Ok

3. use folder2;

CREATING TABLE

4. create table flight (fno int, source varchar(10), year int, delay float);

ALTERING

1. Alter table flight rename to air\_flight;
2. Alter table air\_flight add columns(dest varchar(10));

DROP

* Drop table flight;

>create table flight (fno int, source varchar(10), year int, delay float)

>row format delimited

>fields terminated by ‘,’

>lines terminated by ‘\n’

>stored as textfile;

INSERT

* insert into flight values (215, “pune”, 2023, 15.00);

Then many commands appear then ok

* insert into flight values (216, “mumbai”, 2024, 10.00); ..keep this open on a terminal
* select \* from flight;

LOADING OF DATA

* gedit ipp.txt …this command on new terminal
* insert many such values like pune, Mumbai etc format then save go to ori terminal
* >load data load inpath “ipp.txt”
* >overwrite into table flight;
* OKKKKKK
* select \* from flight;

CREATE TABLE NEW AND THEN JOIN NOW ON SAME TERMINAL

* >create table nflight (fno int, dest varchar(10), year int)

>row format delimited

>fields terminated by ‘,’

>lines terminated by ‘\n’

>stored as textfile; then OKKKK

* Select \* from nflight
* Insert into nflight values(same data from ipp.txt SINGLE DATA VALUES NOT ALL)
* >select a.fno, a.source,a.year,a.delay,b.dest
* >from flight a join nflight b
* > on (a.fno = b.fno ) ………many commands then OKK will display that 1 value

CREATE EXTERNAL TABLE AND LOAD DATA FROM INT TO EXT

* Hive
* Create database mydb3;
* Use mydb3;
* >create table hive\_int(id int, name varchar(10), sal float)

>row format delimited

>fields terminated by ‘,’

>lines terminated by ‘\n’

>stored as textfile;

OPEN new terminal now :

* Gedit data.txt then enter assign record into it ref above steps:
* 1, amit,50,20
* 2,neha,60000
* 3,rahul,70000………like that 4 records

GO BACK TO ORI TEMINAL:

* load data local inpath “data.txt” into table hive int;
* select \* from hive\_int;

NOW CREATE EXT TABLE:

* >create external table hive\_ext(id int, name varchar(10), sal float)

>row format delimited

>fields terminated by ‘,’

>lines terminated by ‘\n’

>stored as textfile;

* insert into hive\_ext select \* from hive\_int
* select \* from hive\_ext

CREATE INDEX ON FLIGHT TABLE -QQQ

CREATE INDEX idx\_flight\_fno

ON TABLE air\_flight (fno)

AS 'COMPACT'

WITH DEFERRED REBUILD;

ALTER INDEX idx\_flight\_fno ON air\_flight REBUILD;

FIND AVERAGE DELAY PER DAY IN 2008

1. Add a date col: ALTER TABLE air\_flight ADD COLUMNS (flight\_date STRING);
2. Insert data with date values:

INSERT INTO TABLE air\_flight VALUES (219, 'delhi', 2008, 30.0, 'mumbai', '2008-06-15');

INSERT INTO TABLE air\_flight VALUES (220, 'bangalore', 2008, 20.0, 'pune', '2008-06-15');

INSERT INTO TABLE air\_flight VALUES (221, 'chennai', 2008, 25.0, 'delhi', '2008-06-16');

1. Run avg. delay query:

SELECT flight\_date, AVG(delay) AS avg\_delay

FROM air\_flight

WHERE year = 2008

GROUP BY flight\_date;