- Hadoop must be installed and running (HDFS, YARN).

- Hive must be installed.
- Linux OS (or use Virtual Machine like Cloudera Sandbox).

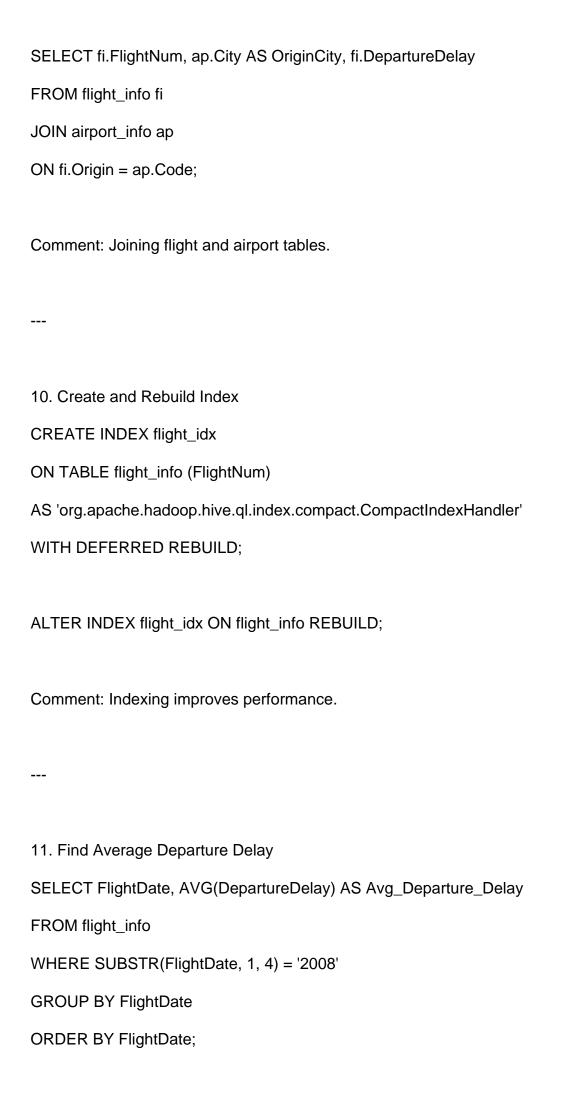
3. Step-by-Step Guide to Run Hive Queries
Step 1: Start Hadoop Services
start-dfs.sh
start-yarn.sh
Step 2: Start Hive
hive
4. Upload Data to HDFS
Create directory in HDFS:
hdfs dfs -mkdir /user/flightdata
Put your CSV file into HDFS:
hdfs dfs -put flightinfo.csv /user/flightdata/
Comment: HDFS is Hadoop's file system. Hive reads from HDFS.

```
Create a database:
CREATE DATABASE flights;
USE flights;
Comment: Database helps organize multiple tables.
Create External Table:
CREATE EXTERNAL TABLE flight_info (
  Year INT,
  Month INT,
  DayofMonth INT,
  FlightNum INT,
  DepartureDelay INT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
LOCATION '/user/flightdata/';
Comment: External Table: Data stays outside Hive in HDFS.
6. Load Data into Managed Table
LOAD DATA INPATH '/user/flightdata/flightinfo.csv' INTO TABLE flight_info;
```

5. Create Hive Database and Table

```
Comment: Managed tables store data inside Hive warehouse directory.
7. Insert New Values into Table
INSERT INTO TABLE flight_info VALUES ('AI-123', 'DEL', 'MUM', '2008-01-01', 15, 5);
Comment: Adding a new flight record manually.
8. Create and Insert into Another Table
CREATE TABLE airport_info (
  Code STRING,
  City STRING,
  State STRING
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE;
INSERT INTO TABLE airport_info VALUES ('DEL', 'New Delhi', 'DL');
INSERT INTO TABLE airport_info VALUES ('MUM', 'Mumbai', 'MH');
```

9. Perform Join Operation



12. Drop Table

DROP TABLE flight_info;

DROP TABLE airport_info;

Comment: Dropping a table.

Important Reminders:

- Hadoop + Hive must be installed properly.
- Hadoop must be running before starting Hive.
- Hive queries are slower than normal SQL because of MapReduce.
- Always end Hive queries with ';'.
- HDFS stores data across multiple nodes (distributed).
- External tables do not delete data on DROP.
- Managed tables delete data on DROP.