**Steps for Hive:**

1. (base) student@student-OptiPlex-3020:~$ su hduser

(base) student@student-OptiPlex-3020:~$ su - hduser

1. hduser@student-OptiPlex-3020:/home/student$ start-all.sh
2. hduser@student-OptiPlex-3020:/home/student$ start-dfs.sh
3. hduser@student-OptiPlex-3020:/home/student$ start-yarn.sh
4. hduser@student-OptiPlex-3020:/home/student$ cd $HIVE\_HOME
5. hduser@student-OptiPlex-3020:/usr/local/hive$ bin/hive
6. hive> CREATE DATABASE ourfirstdatabase;
7. USE ourfirstdatabase;
8. hive> CREATE TABLE our\_first\_table

> (FirstName STRING,LastName STRING,EmployeeId INT);

1. hive> CREATE TABLE IF NOT EXISTS FlightInfo2007(Year SMALLINT,Month TINYINT,DayofMonth TINYINT,DayofWeek TINYINT,DepTime SMALLINT,CRSDepTime SMALLINT,CRSArrTime SMALLINT,UniqueCarrier STRING,FlightNum STRING,TailNum STRING,ActualElapsedTime SMALLINT,CRSElapsedTime SMALLINT,AirTime SMALLINT,ArrDelay SMALLINT,DepDelay SMALLINT,Origin STRING,Dest STRING,Distance INT,TaxiIn SMALLINT,TaxiOut SMALLINT,Cancelled SMALLINT,CancellationCode STRING,Diverted SMALLINT,CarrierDelay SMALLINT,WeatherDelay SMALLINT,NASDelay SMALLINT,SecurityDelay SMALLINT,LateAircraftDelay SMALLINT);
2. hive> load data local inpath'/home/student/2007.csv'into table FlightInfo2007;
3. hive> CREATE TABLE IF NOT EXISTS FlightInfo2008 LIKE FlightInfo2007;
4. hive> load data local inpath'/home/student/2008.csv'into table FlightInfo2008;
5. hive> CREATE TABLE IF NOT EXISTS myFlightInfo(Year SMALLINT,DontQueryMonth TINYINT,DayofMonth TINYINT,DayOfweek TINYINT,DepTime SMALLINT, ArrTime SMALLINT,UniqueCarrier STRING,FlightNum STRING,ArrDelay SMALLINT,DepDelay SMALLINT,Origin STRING,Dest STRING,Cancelled SMALLINT,CancellationCode STRING);
6. hive> CREATE TABLE myflightinfo2007 AS SELECT Year,Month,DepTime,AirTime,FlightNum,Origin,Dest FROM FlightInfo2007 WHERE (Month=7 AND DayofMonth=3)AND(Origin='JFK'AND Dest='ORD');
7. hive> SELECT\*FROM myFlightInfo2007;
8. hive> CREATE TABLE myflightinfo2008 AS SELECT Year,Month,DepTime,AirTime,FlightNum,Origin,Dest FROM FlightInfo2007 WHERE (Month=7 AND DayofMonth=3)AND(Origin='JFK'AND Dest='ORD');
9. hive> SELECT\*FROM myFlightInfo2008;
10. hive> SELECT m8.Year,m8.Month,m8.FlightNum,m8.Origin,m8.Dest,m7.Year,m7.Month,m7.FlightNum,m7.Origin,m7.Dest FROM myFlightInfo2008 m8 JOIN myFlightInfo2007 m7 ON m8.FlightNum=m7.FlightNum;
11. hive> SELECT m8.FlightNum,m8.Origin,m8.Dest,m7.FlightNum,m7.Origin,m7.Dest FROM myFlightInfo2008 m8 FULL OUTER JOIN myFlightInfo2007 m7 ON m8.FlightNum=m7.FlightNum;
12. hive> SELECT m8.Year,m8.Month,m8.FlightNum,m8.Origin,m8.Dest,m7.Year,m7.Month,m7.FlightNum,m7.Origin,m7.Dest FROM myFlightInfo2008 m8 LEFT OUTER JOIN myFlightInfo2007 m7 ON m8.FlightNum=m7.FlightNum;
13. CREATE INDEX f08\_index ON TABLE flightinfo2008(Origin)AS 'COMPACT' WITH DEFERRED REBUILD;
14. SHOW INDEXES ON FlightInfo2008;
15. SELECT Origin,COUNT(1)FROM flightinfo2008 WHERE Origin='SYR'GROUP BY Origin;
16. hive> DESCRIBE flightinfo2008;
17. hive> CREATE VIEW avgdepdelay AS SELECT DayofWeek,AVG(DepDelay)FROM FlightInfo2008 GROUP By DayofWeek;
18. hive> SELECT\*FROM avgdepdelay;