**Step 1: Copy the provided csv files in a directory named AirlinesOntimeAnalysis**

hdfs dfs -copyFromLocal airports.csv /user/$USER/AirlinesOntimeAnalysis

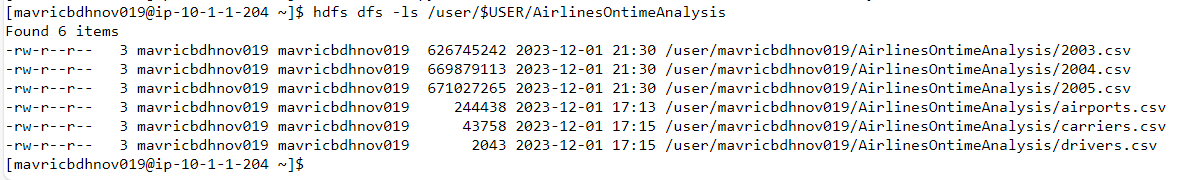
hdfs dfs -copyFromLocal carriers.csv /user/$USER/AirlinesOntimeAnalysis

hdfs dfs -copyFromLocal drivers.csv /user/$USER/AirlinesOntimeAnalysis

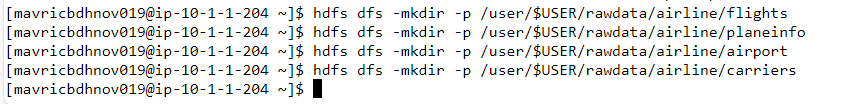
hdfs dfs -copyFromLocal 2003.csv /user/$USER/AirlinesOntimeAnalysis

hdfs dfs -copyFromLocal 2004.csv /user/$USER/AirlinesOntimeAnalysis

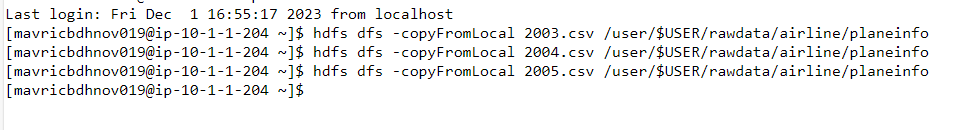
hdfs dfs -copyFromLocal 2004.csv /user/$USER/AirlinesOntimeAnalysis



**Step 3: Create the directories in hdfs where your data will reside**



**Step 4: copy the respective files to the above created hdfs structure**



**Create all the tables in a different database in Hive(airline\_db)**

A black text with black letters

Description automatically generated with medium confidence

**Create tables for flights , plane\_info, carriers and airports in the staging area.**

CREATE EXTERNAL TABLE Airports\_stg2 (

IATA STRING,

Name STRING,

City STRING,

State STRING,

Country STRING,

Lat DECIMAL(9, 6),

Long DECIMAL(9, 6)

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/user/$USER/rawdata/airline/airport'

tblproperties(

"skip.header.line.count"="1"

);

Getting permission issue while creating the table.



CREATE EXTERNAL TABLE Carrier\_stg2 (

Code STRING,

Description STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION 'user/$USER/rawdata/airline/carriers'

tblproperties(

"skip.header.line.count"="1"

);

CREATE EXTERNAL TABLE PlaneInfo\_stg2 (

tailNum STRING,

type STRING,

manufacturer STRING,

issue\_date STRING,

model STRING,

status STRING,

aircraft\_type STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/user/$USER/rawdata/airline/planeinfo'

tblproperties(

"skip.header.line.count"="1"

);

CREATE EXTERNAL TABLE Flights\_stg2 (

Year INT,

Month INT,

DayofMonth INT,

DayOfWeek INT,

DepTime INT,

CRSDepTime INT,

ArrTime INT,

CRSArrTime INT,

UniqueCarrier STRING,

FlightNum STRING,

TailNum STRING,

ActualElapsedTime INT,

CRSElapsedTime INT,

AirTime INT,

ArrDelay INT,

DepDelay INT,

Origin STRING,

Dest STRING,

Distance INT,

TaxiIn STRING,

TaxiOut STRING,

Cancelled STRING,

CancellationCode STRING,

Diverted STRING,

CarrierDelay INT,

WeatherDelay INT,

NASDelay INT,

SecurityDelay INT,

LateAircraftDelay INT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION '/user/$USER/rawdata/airline/flights'

tblproperties(

"skip.header.line.count"="1"

);

**Check if the tables have data.**

LOAD DATA LOCAL INPATH 'AirlinesOntimeAnalysis/airports.csv' OVERWRITE INTO TABLE Airports\_stg2;

LOAD DATA LOCAL INPATH 'AirlinesOntimeAnalysis/carriers.csv' OVERWRITE INTO TABLE Carrier\_stg2;

LOAD DATA LOCAL INPATH 'AirlinesOntimeAnalysis/plane-data.csv' OVERWRITE INTO TABLE PlaneInfo\_stg2;

LOAD DATA LOCAL INPATH 'AirlinesOntimeAnalysis/2003.csv' OVERWRITE INTO TABLE Flights\_stg2;

LOAD DATA LOCAL INPATH 'AirlinesOntimeAnalysis/2004.csv' OVERWRITE INTO TABLE Flights\_stg2;

LOAD DATA LOCAL INPATH 'AirlinesOntimeAnalysis/2005.csv' OVERWRITE INTO TABLE Flights\_stg2;