# Hive\_assignment\_pb3

## Data Setup in HDFS

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

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

A close-up of a computer code

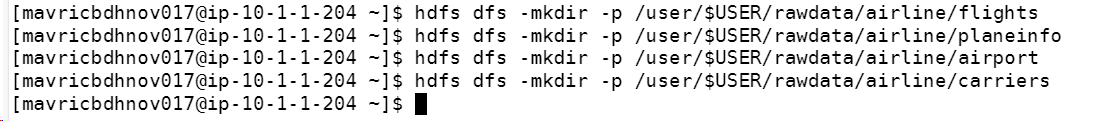
Description automatically generated

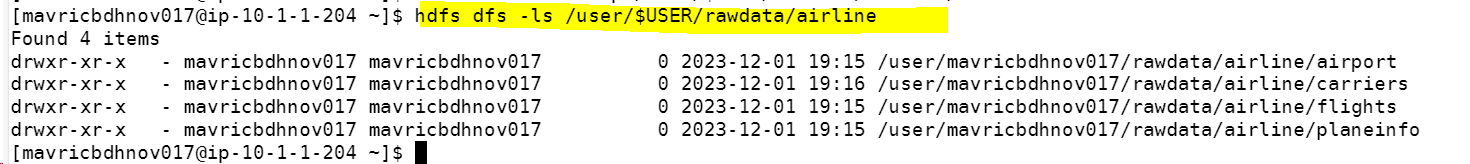
## **Step 2:** Unzip the contents for the bz2 files.

A computer screen with numbers and letters

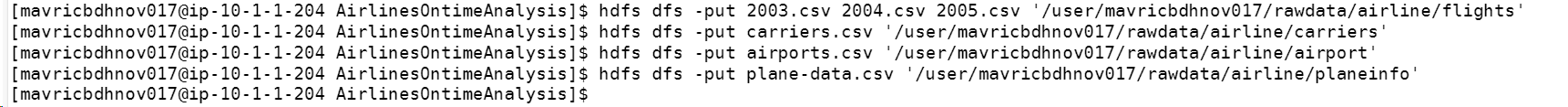
Description automatically generated

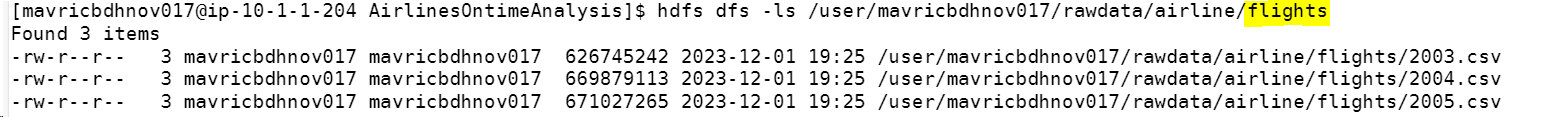
## Step 3: Create the directories in hdfs where your data will reside





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





A close up of a document

Description automatically generated

## Create tables using Hive

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

Create database dn\_airline\_db;

A close up of a white background

Description automatically generated

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

#### Airports

CREATE EXTERNAL TABLE airports\_stg

(IATA String COMMENT '3-letter IATA code. Null if not assigned/unknown',

Name String COMMENT 'Name of airport. May or may not contain the City name',

City String COMMENT 'Main city served by airport. May be spelled differently from Name',

State String COMMENT 'Main state served by airport. May be spelled differently from Name',

Country String COMMENT 'Country or territory where airport is located. See countries.dat to cross-reference ISO 3166-1 codes',

Lat Decimal COMMENT 'degreesCOMMENT usually to six significant digits. Negative is SouthCOMMENT positive is North',

Long Decimal COMMENT 'degreesCOMMENT usually to six significant digits. Negative is WestCOMMENT positive is East')

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

STORED As TextFile

LOCATION '/user/mavricbdhnov017/rawdata/airline/airport'

tblproperties(

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

);

A close-up of a letter

Description automatically generated

A screenshot of a computer

Description automatically generated

A screen shot of a computer

Description automatically generated

#### Carrier

CREATE EXTERNAL TABLE carrier\_stg

(Code String COMMENT 'Unique OpenFlights identifier for this airline',

Description String COMMENT 'Name of the airline')

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

STORED As TextFile

LOCATION '/user/mavricbdhnov017/rawdata/airline/carriers'

tblproperties(

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

);

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

#### Plane Info

CREATE EXTERNAL TABLE plane\_info\_stg

(tailNum String COMMENT 'Tail number painted on the tail of a plane',

type String COMMENT 'manufacturer String, Manufacturer of plane like Airbus BOEING',

issue\_date String,

model String COMMENT 'Model number of the plane',

status String ,

aircraft\_type String COMMENT 'Type of aircraft',

engine\_type String COMMENT 'Type of engine',

year Int COMMENT 'Year of manufacture')

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

STORED As TextFile

LOCATION '/user/mavricbdhnov017/rawdata/airline/planeinfo'

tblproperties(

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

);

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

#### Flights

CREATE EXTERNAL TABLE flights\_stg

(Year Int COMMENT '2003-2005',

Month Int COMMENT '1-12',

DayofMonth Int COMMENT '1-31',

DayOfWeek Int COMMENT '1 (Monday) - 7 (Sunday)',

DepTime Int COMMENT 'actual departure time (local, hhmm)',

CRSDepTime Int COMMENT 'scheduled departure time (local, hhmm)',

ArrTime Int COMMENT 'actual arrival time (local, hhmm)',

CRSArrTime Int COMMENT 'scheduled arrival time (local, hhmm)',

UniqueCarrier String COMMENT 'unique carrier code',

FlightNum String COMMENT 'flight number',

TailNum String COMMENT 'plane tail number',

ActualElapsedTime Int COMMENT 'in minutes',

CRSElapsedTime Int COMMENT 'in minutes',

AirTime Int COMMENT 'in minutes',

ArrDelay Int COMMENT 'arrival delay, in minutes',

DepDelay Int COMMENT 'departure delay, in minutes',

Origin String COMMENT 'origin IATA airport code',

Dest String COMMENT 'destination IATA airport code',

Distance Int COMMENT 'in miles',

TaxiIn String COMMENT 'taxi in time, in minutes',

TaxiOut String COMMENT 'taxi out time in minutes',

Cancelled String COMMENT 'was the flight cancelled?',

CancellationCode String COMMENT 'reason for cancellation (A = carrier,B = weather, C = NAS, D = security)',

Diverted String COMMENT '1 = yes, 0 = no',

CarrierDelay Int COMMENT 'in minutes',

WeatherDelay Int COMMENT 'in minutes',

NASDelay Int COMMENT 'in minutes',

SecurityDelay Int COMMENT 'in minutes',

LateAircraftDelay Int COMMENT 'in minutes')

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

STORED As TextFile

LOCATION '/user/mavricbdhnov017/rawdata/airline/flights'

tblproperties(

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

);

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

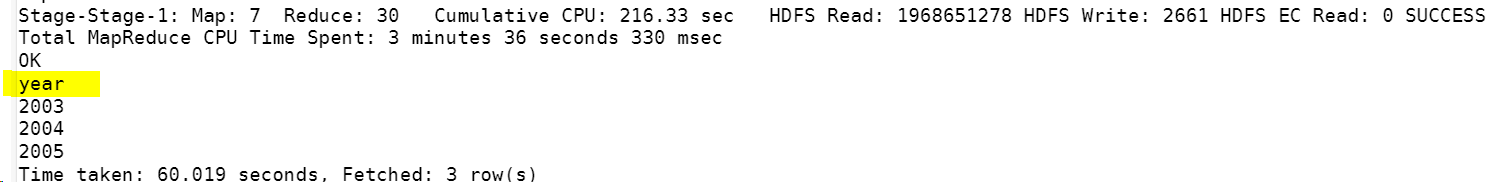
Description automatically generated

A black text on a white background

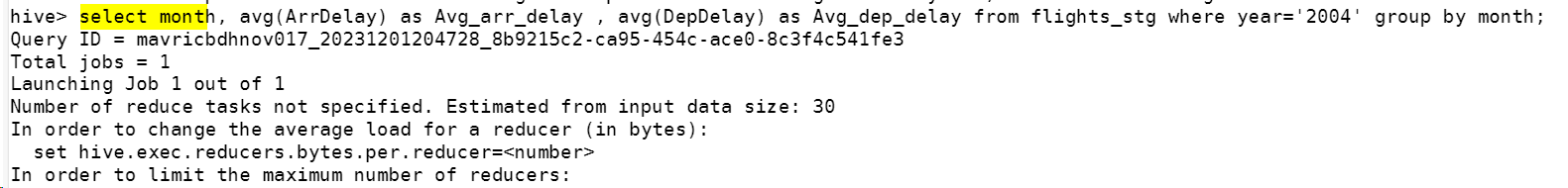
Description automatically generated

A screenshot of a computer

Description automatically generated



### What is the average arrival delay and average departure delay in each month of the year 2004?



A white screen with black text

Description automatically generated

## Improving the performance of Hive tables - Parquet format

Create the parquet table as external tables for flights from the above staging table.

Name the new table with \_pq as a suffix: Ex : flights\_pq

Point all the table location output/airline/pq\_flights

CREATE EXTERNAL TABLE flights\_pq

(Year Int COMMENT '2003-2005',

Month Int COMMENT '1-12',

DayofMonth Int COMMENT '1-31',

DayOfWeek Int COMMENT '1 (Monday) - 7 (Sunday)',

DepTime Int COMMENT 'actual departure time (local, hhmm)',

CRSDepTime Int COMMENT 'scheduled departure time (local, hhmm)',

ArrTime Int COMMENT 'actual arrival time (local, hhmm)',

CRSArrTime Int COMMENT 'scheduled arrival time (local, hhmm)',

UniqueCarrier String COMMENT 'unique carrier code',

FlightNum String COMMENT 'flight number',

TailNum String COMMENT 'plane tail number',

ActualElapsedTime Int COMMENT 'in minutes',

CRSElapsedTime Int COMMENT 'in minutes',

AirTime Int COMMENT 'in minutes',

ArrDelay Int COMMENT 'arrival delay, in minutes',

DepDelay Int COMMENT 'departure delay, in minutes',

Origin String COMMENT 'origin IATA airport code',

Dest String COMMENT 'destination IATA airport code',

Distance Int COMMENT 'in miles',

TaxiIn String COMMENT 'taxi in time, in minutes',

TaxiOut String COMMENT 'taxi out time in minutes',

Cancelled String COMMENT 'was the flight cancelled?',

CancellationCode String COMMENT 'reason for cancellation (A = carrier,B = weather, C = NAS, D = security)',

Diverted String COMMENT '1 = yes, 0 = no',

CarrierDelay Int COMMENT 'in minutes',

WeatherDelay Int COMMENT 'in minutes',

NASDelay Int COMMENT 'in minutes',

SecurityDelay Int COMMENT 'in minutes',

LateAircraftDelay Int COMMENT 'in minutes')

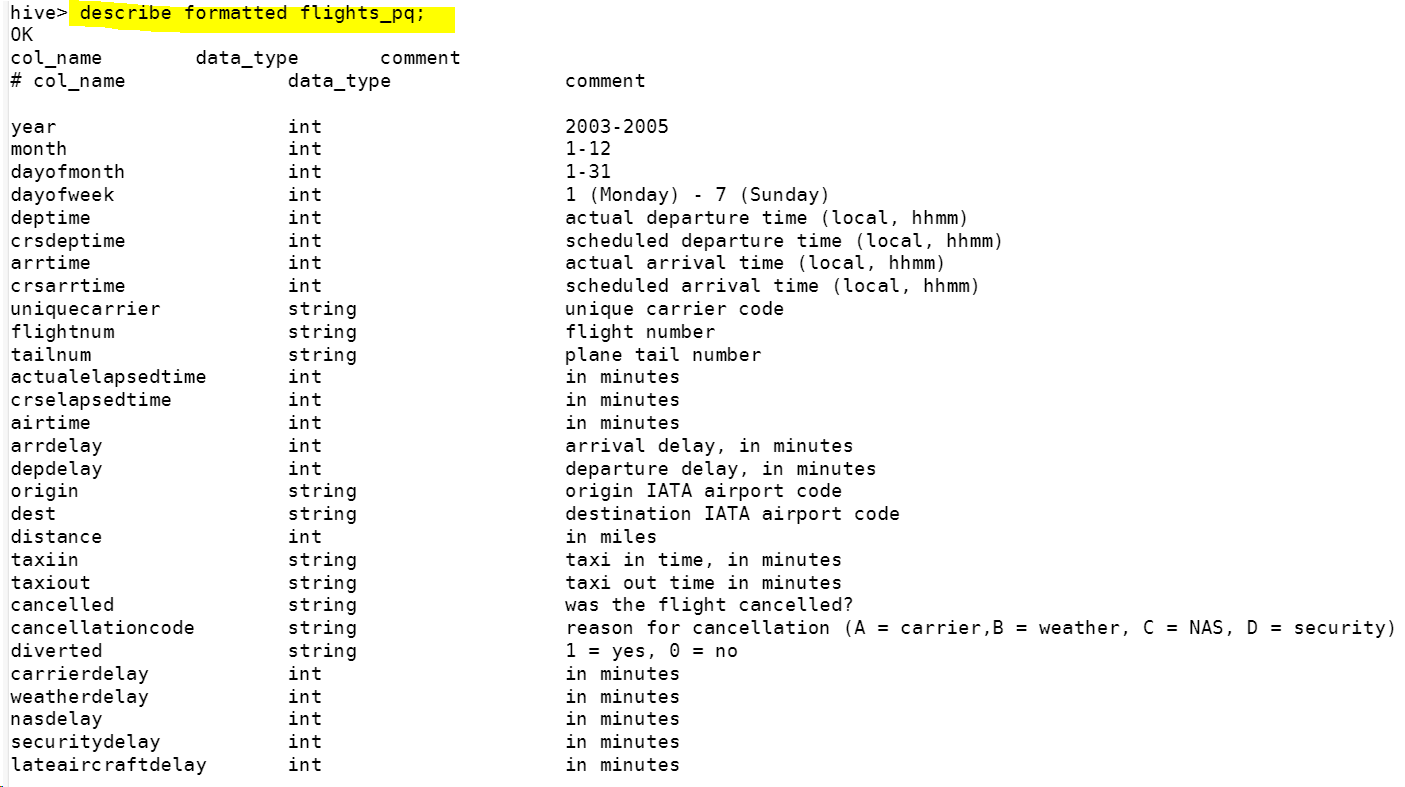
STORED AS PARQUET

LOCATION '/user/mavricbdhnov017/rawdata/airline/flights\_pq'

tblproperties(

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

);



A yellow line with black text

Description automatically generated

A close up of a document

Description automatically generated

### Run the same query as above to check if data is loaded properly

A white paper with black text

Description automatically generated

A yellow and black text

Description automatically generated

#### Is there a difference between the response times for the queries(2 and 3)

##### Query-2

A screenshot of a computer

Description automatically generated

A close-up of a computer screen

Description automatically generated

A white paper with black text

Description automatically generated

A yellow and black text

Description automatically generated

##### Query-3

###### Avg calculation

Staging table

A screenshot of a computer

Description automatically generated

parquet table

A screenshot of a computer

Description automatically generated