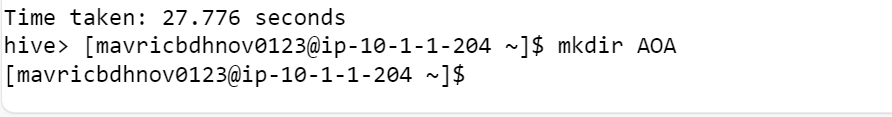
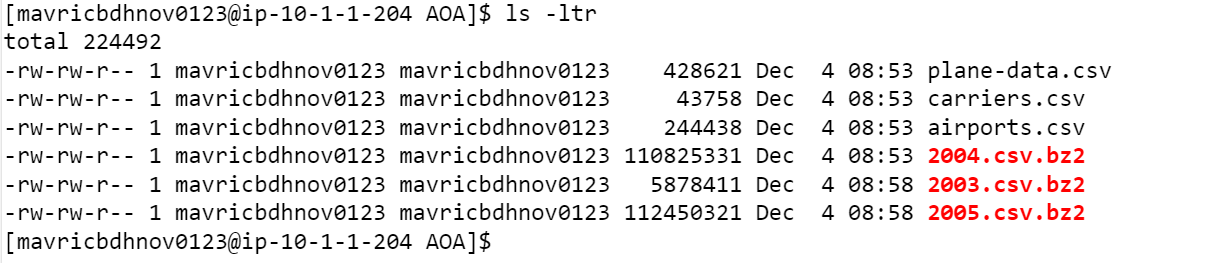
mkdir AOA(Airline online Analysis)





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

A close-up of a computer screen

Description automatically generated

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

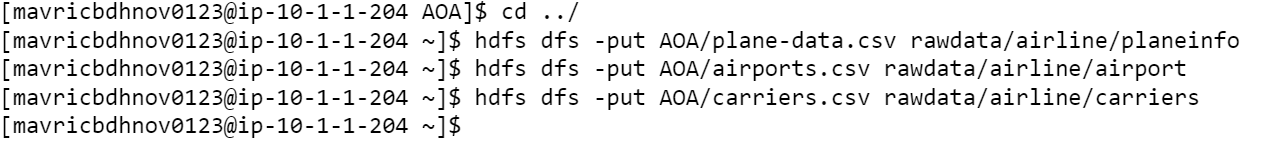
A close-up of a white background

Description automatically generated

A black text on a white background

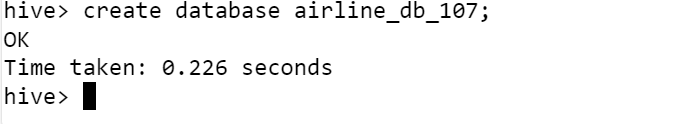
Description automatically generated

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



**2. Create tables using Hive**

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



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

**Hive QL:**

CREATE EXTERNAL TABLE carriers\_stg (Code string,Description string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' location '/user/mavricbdhnov0123/rawdata/airline/carriers';

CREATE EXTERNAL TABLE airport\_stg (iata string,airport string,city string,state string,country string,lat decimal,long decimal) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' location '/user/mavricbdhnov0123/rawdata/airline/airport';

CREATE EXTERNAL TABLE plane\_info\_stg (tailnum string,type string,manufacturer string,issue\_date string,model string,status string,aircraft\_type string,engine\_type string,year int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' location '/user/mavricbdhnov0123/rawdata/airline/planeinfo';

CREATE EXTERNAL TABLE flights\_stg (Year int,Month int,DayofMonth int,DayOfWeek int,DepTime int,CRSDepTime int,ArrTime int,CRSArrTime int,UniqueCarrier string,FlightNum strin

A computer screen shot of a program

Description automatically generated

A white background with black text

Description automatically generated

Hive>ALTER TABLE flights\_stg SET TBLPROPERTIES ("skip.header.line.count"="1");

hive> ALTER TABLE plane\_info\_stg SET TBLPROPERTIES ("skip.header.line.count"="1");

hive> ALTER TABLE airport\_stg SET TBLPROPERTIES ("skip.header.line.count"="1");

hive> ALTER TABLE carriers\_stg SET TBLPROPERTIES ("skip.header.line.count"="1");

A screenshot of a computer program

Description automatically generated

**Check if the tables have data.**

select \* from k\_airline\_db.flights\_stg limit 5;

select \* from k\_airline\_db.plane\_info\_stg limit 5;

select \* from k\_airline\_db.airport\_stg limit 5;

select \* from k\_airline\_db.carriers\_stg limit 5;

A group of black text

Description automatically generated

A white background with black text

Description automatically generated

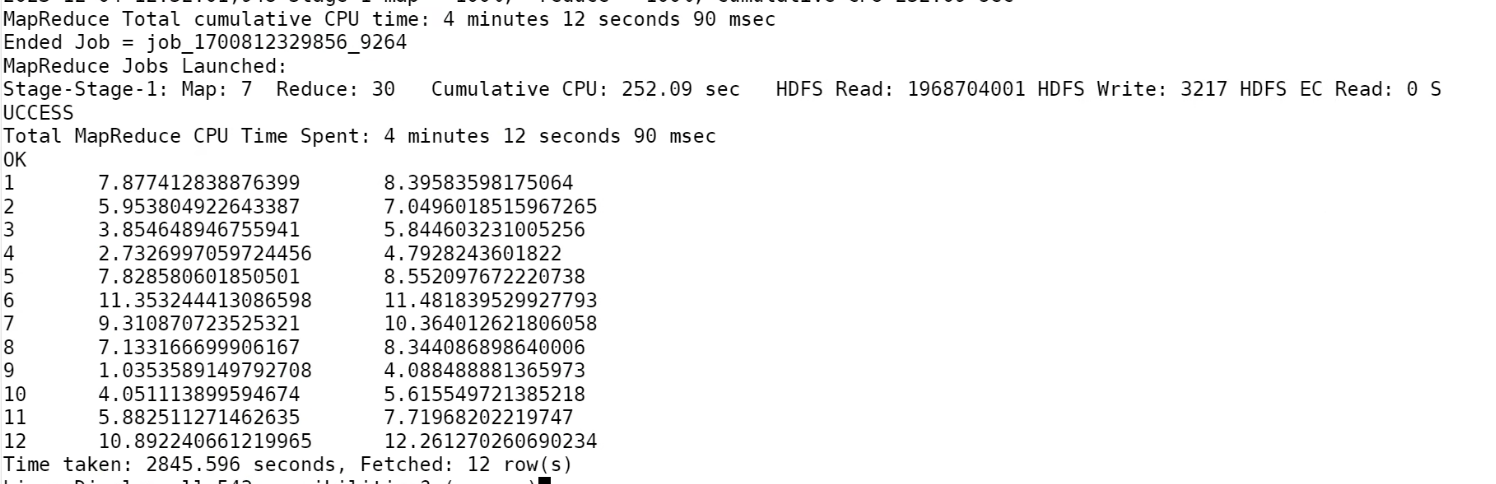
A white screen with black text

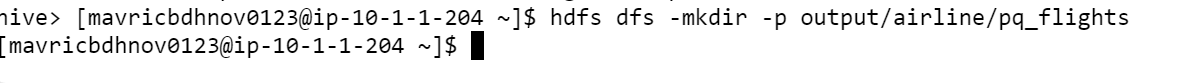
Description automatically generated

A screen shot of a computer

Description automatically generated

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





**3. Improving the performance of Hive tables - Parquet format**

CREATE EXTERNAL TABLE flight\_pq (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,Cance

llationCode string,Diverted string,CarrierDelay int,WeatherDelay int,NASDelay int,SecurityDelay int,LateAircraftDelay int

) STORED AS PARQUET location '/user/maverickavitha/output/airline/pq\_flights';

A screenshot of a computer program

Description automatically generated

insert overwrite table k\_airline\_db.flight\_pq select \* from k\_airline\_db.flights\_stg

A white paper with black text

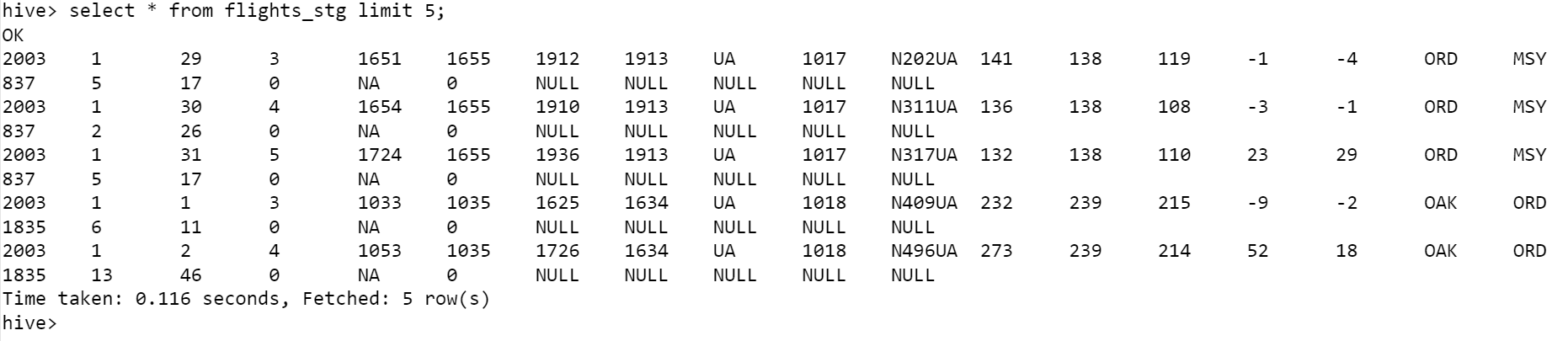
Description automatically generated

Select \* from flight\_pq limit 5;

A black and white text

Description automatically generated

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



**AVG Calculation in PARQUET format:**

A white background with black text

Description automatically generated

A white screen with black text

Description automatically generated

**AVG Calculation in normal STG table:**

A close-up of a computer code

Description automatically generated

A close-up of a computer screen

Description automatically generated