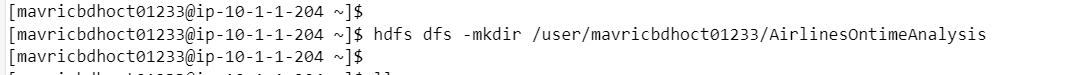
**Ques-1. Data Setup in HDFS**

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



A close-up of a document

Description automatically generated

Step 2: Unzip the contents for the bz2 files.

A screenshot of a computer

Description automatically generated

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

rawdata/airline/flights

rawdata/airline/planeinfo

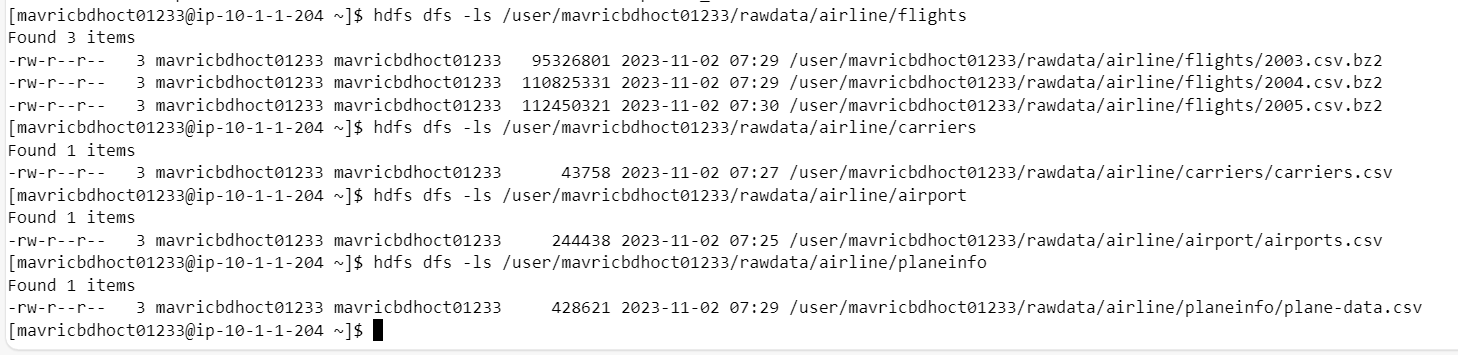
rawdata/airline/airport

rawdata/airline/carriers

A close-up of a computer code

Description automatically generated

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



**Ques-2. Create tables using Hive**

A screenshot of a computer

Description automatically generated

A close up of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A close up of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

A computer screen shot of a computer code

Description automatically generated

A screenshot of a computer

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 screenshot of a computer

Description automatically generated

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

**CREATE EXTERNAL TABLE flights\_pq**

**STORED AS PARQUET**

**LOCATION '/user/mavricbdhoct01233/output/airline/pq\_flights'**

**AS**

**SELECT \* FROM flights\_stg;**