CREATE A TABLE WITH THE MENTIONED FIELD VALUES

LOAD FLIGHTS.csv data into ubuntu.

cp /mnt/f/assignment/BigData/flights2.csv ~/

hdfs dfs -put flights2.csv ./tmp

A computer screen with text and numbers

Description automatically generated

flights = LOAD ‘pig/flights2.csv' USING PigStorage(',') AS (

Year: int,

Month: int,

DayofMonth: int,

DayOfWeek: int,

Carrier: chararray,

OriginAirportID: int,

OriginAirportName: chararray,

OriginCity: chararray,

OriginState: chararray,

DestAirportID: int,

DestAirportName: chararray,

DestCity: chararray,

DestState: chararray,

CRSDepTime: chararray,

DepDelay: float,

DelDelay15: int,

CRSArrTime: chararray,

ArrDelay: float,

ArrDelay15: int,

Cancelled: int

);

dump flights;



A screen shot of a computer

Description automatically generated

illustrate flights;

A black and white screen with white text

Description automatically generated

A screen shot of a computer screen

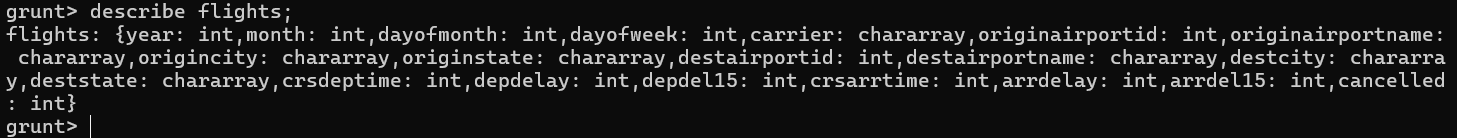
Description automatically generated

select \* from flights limit 10;

A black screen with white text

Description automatically generated

describe flights;



TASK 1

CODE:

filtered\_flights = FILTER flights BY Cancelled == 0;

grouped\_flights = GROUP filtered\_flights BY (OriginAirportID, OriginAirportName);

count\_flights = FOREACH grouped\_flights GENERATE group.OriginAirportID AS airport\_id, group.OriginAirportName AS airport\_name, COUNT(filtered\_flights) AS flight\_count;

sorted\_flights = ORDER count\_flights BY flight\_count DESC;

top\_airports = LIMIT sorted\_flights 2;

dump top\_airports;

PROCESS:

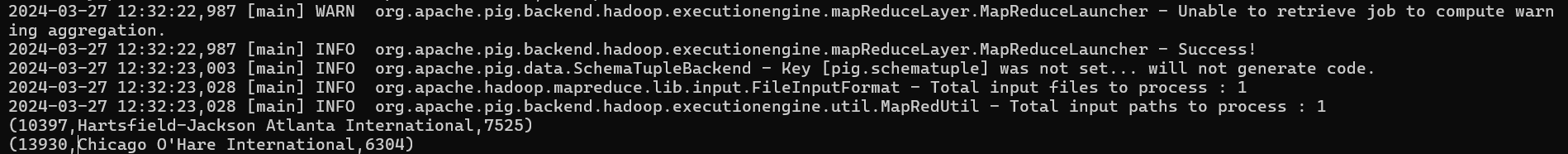
A screen shot of a computer

Description automatically generated

A black screen with white text

Description automatically generated

OUPTUT:



TASK 2

CODE:

flights\_not\_cancelled = FILTER flights BY Cancelled == 0;

flights\_with\_time = FOREACH flights\_not\_cancelled GENERATE Year, Month, DayofMonth, DayOfWeek, Carrier, OriginAirportID, OriginAirportName, OriginCity, OriginState,((int)SUBSTRING(CRSDepTime, 0, 2)) \* 60 + (int)SUBSTRING(CRSDepTime, 2, 2) AS CRSDepTime, DepDelay, ((int)SUBSTRING(CRSArrTime, 0, 2)) \* 60 + (int)SUBSTRING(CRSArrTime, 2, 2) AS CRSArrTime, ArrDelay, Cancelled;

flights\_with\_duration = FOREACH flights\_with\_time GENERATE Year, Month, DayofMonth, DayOfWeek, Carrier, OriginAirportID, OriginAirportName, OriginCity, OriginState, CRSDepTime, DepDelay, CRSArrTime, ArrDelay, CRSArrTime - CRSDepTime AS FlightDuration;

grouped\_by\_carrier = GROUP flights\_with\_duration BY Carrier;

max\_duration\_by\_carrier = FOREACH grouped\_by\_carrier {

max\_duration = ORDER flights\_with\_duration BY FlightDuration DESC;

max\_duration\_per\_group = LIMIT max\_duration 1;

GENERATE group AS Carrier, MAX(max\_duration\_per\_group.FlightDuration) AS LongestFlightDuration;

};

DUMP max\_duration\_by\_carrier;

PROCESS:

A computer screen with white text

Description automatically generated

OUTPUT:

A black screen with white text

Description automatically generated

TASK 3

CODE

flights\_per\_day = GROUP flights\_not\_cancelled BY DayofMonth;

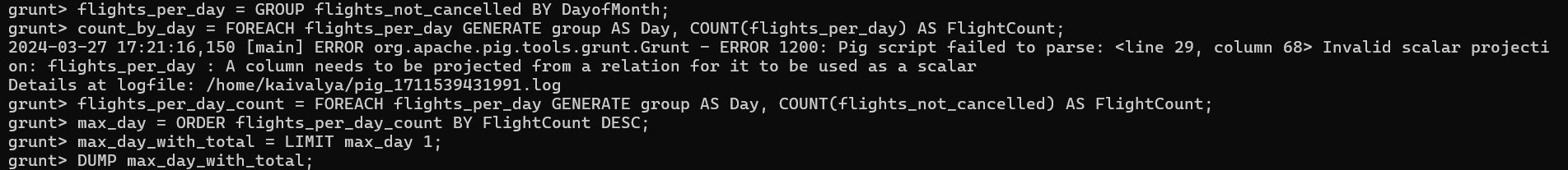
flights\_per\_day\_count = FOREACH flights\_per\_day GENERATE group AS Day, COUNT(flights\_not\_cancelled) AS FlightCount;

max\_day = ORDER flights\_per\_day\_count BY FlightCount DESC;

max\_day\_with\_total = LIMIT max\_day 1;

DUMP max\_day\_with\_total;

PROCESS



A black screen with white text

Description automatically generated

OUTPUT

A black screen with white text

Description automatically generated