CREATE A TABLE WITH THE MENTIONED FIELD VALUES

CREATE TABLE flights (`Year` int,

`Month` int,

`DayofMonth` int,

`DayOfWeek` int,

`Carrier` string,

`OriginAirportID` int,

`OriginAirportName` string,

`OriginCity` string,

`OriginState` string,

`DestAirportID` int,

`DestAirportName` string,

`DestCity` string,

`DestState` string,

`CRSDepTime` int,

`DepDelay` int,

`DepDel15` int,

`CRSArrTime` int,

`ArrDelay` int,

`ArrDel15` int,

`Cancelled` int)

row format delimited fields terminated by ',';

A computer screen shot of a black screen

Description automatically generated

LOAD FLIGHTS.csv data into hive.

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

hdfs dfs -put flights2.csv ./tmp

A computer screen with text and numbers

Description automatically generated

LOAD DATA LOCAL INPATH 'flights2.csv' OVERWRITE INTO TABLE flights;

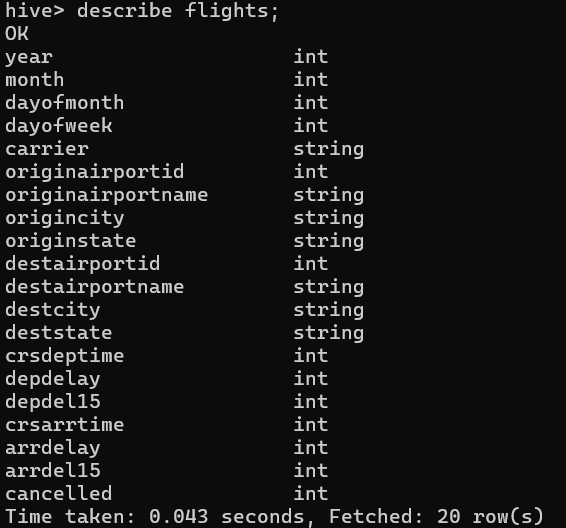


select \* from flights limit 10;

A black screen with white text

Description automatically generated

describe flights;



Write Hive Query to

Find the top 2 busiest airports based on the total number of flights departing from or arriving at each airport.

Report the airport code along with the total number of flights.

SELECT

airport\_id,

airport\_name,

SUM(total\_flights) AS total\_flights

FROM (

SELECT

originairportid AS airport\_id,

originairportname AS airport\_name,

COUNT(\*) AS total\_flights

FROM

flights

GROUP BY

originairportid, originairportname

UNION ALL

SELECT

destairportid AS airport\_id,

destairportname AS airport\_name,

COUNT(\*) AS total\_flights

FROM

flights

GROUP BY

destairportid, destairportname

) t

GROUP BY

airport\_id, airport\_name

ORDER BY

total\_flights DESC

LIMIT 2;

A screenshot of a computer program

Description automatically generated

Write Hive Query to

Find the longest flight duration (in minutes) for each unique carrier.

Report the carrier along with the longest flight duration.

SELECT

carrier,

MAX(crsarrtime - crsdeptime) AS longest\_flight\_duration\_minutes

FROM

flights

GROUP BY

carrier;

SELECT

carrier,

MAX(crsdeptime - crsarrtime) AS shortest\_flight\_duration\_minutes

FROM

flights

GROUP BY

carrier;

A screenshot of a computer

Description automatically generated

Write Hive Query to

Determine the day of the month with the highest number of flights.

Report the day along with the total number of flights.

SELECT

dayofmonth,

COUNT(\*) AS num\_flights

FROM

flights

GROUP BY

dayofmonth

ORDER BY

num\_flights DESC

LIMIT

1;

A screen shot of a computer

Description automatically generated

A screen shot of a computer

Description automatically generated