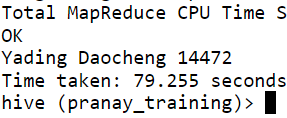
Q2) HIVE:

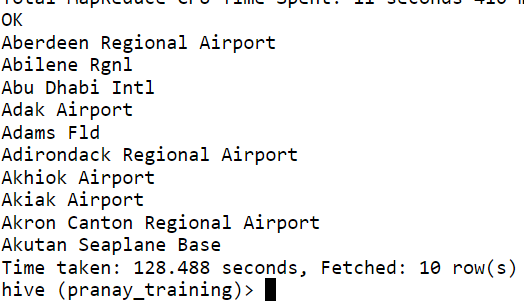
1. Which airports have highest altitude?

**select name, altitude from airport order by altitude desc limit 1;**



1. How many routes are operated by active airlines from the United States ?

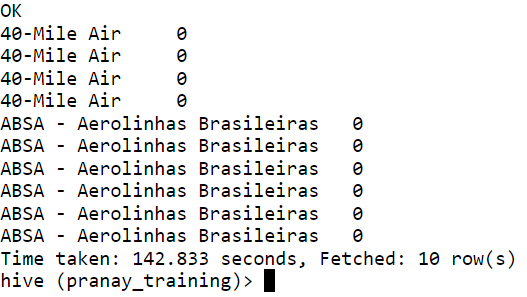
**select distinct a.name from airlines a join routes r on a.airline\_id=r.airline\_id join airport ar1 on ar1.airport\_id=src\_airport\_id join airport ar2 on ar2.airport\_id=dst\_airport\_id where trim(upper(ar1.country))=’UNITED STATES’ and trim(upper(ar2.country))=’UNITED STATES’ limit 10;**



1. Which airlines operate routes that have less than 3 stops number of stops

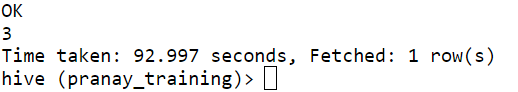
top 10 alphabetically?

**select a.name, r.stops from airlines a join routes r on a.airline\_id = r.airline\_id where r.stops<3 order by a.name limit 10;**



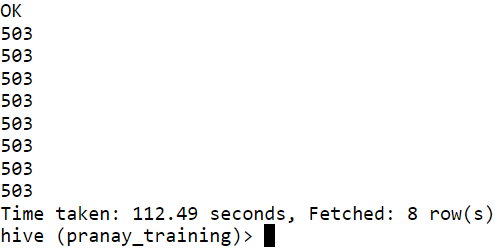
1. How many airlines have a specific IATA code ‘W9’?

select count(airlines\_id) from airlines where iata = ‘W9’;



1. Find the airlines that operate routes with a specific equipment as ‘AN4’ and codeshare enabled.

**Select airline\_id from routes where equipment = ‘AN4’ and codeshare = ‘Y’;**



Q3) PYSPARK

**from pyspark.sql.types import StructType, StringType, IntegerType, DoubleType, LongType**

**schema = StructType().add("Year",StringType(),True).add("Quarter",StringType(),True).add("ARPS",DoubleType(),True).add("Booked\_seats",IntegerType(),True)**

**data = spark.read.format("csv").option("header","True").schema(schema).load("hdfs://nameservice1/user/bigdatalab456455/training/airlines.csv")**

**data.registerTempTable("airlines")**

1. What is the total revenue generated in each year?

**TotRevYear = spark.sql(“select year,(arps\*booked\_seats) as total\_revenue from airlines group by year”)**

1. Which year had the highest average revenue per seat?

**HighAvgSeats = spark.sql("SELECT year, quarter, AVG(rev) AS Avg\_rev\_per\_seat FROM airlines GROUP BY Year, Quarter ORDER BY Avg\_rev\_per\_seat DESC LIMIT 1")**

1. What is the total number of booked seats for each quarter in a given year?

**TotalBookedSeats = spark.sql("SELECT quarter, SUM(seats) AS Total\_tickets FROM airlines GROUP BY quarter")**