SET A

Ques 2 HIVE

1. Which airports have the highest altitude ?

Select name from airport where altitude in (select altitude from airport order by altitude desc limit 1);

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

Select count(r.airline\_id) from routes r inner join airlines a on r.airline\_id=a.airline\_id where trim(upper(country))=”UNITED STATES” and trim(upper(active))=”Y”;

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

top 10 alphabetically?

select distinct(a.name) from airlines a inner join routes r on a.airline\_id=r.airline\_id where r.stops<3 order by a.name desc limit 10;

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

Select name from airlines where trim(upper(iata))=”W9”;

5. Find the airlines that operate routes with a specific equipment as ‘AN4’

and codeshare enabled.

select distinct(a.name) from airlines a inner join routes r on a.airline\_id=r.airline\_id where trim(upper(r.equipment))=”AN4” and trim(upper(r.codeshare))=”Y”;

QUES 3 PYSPARK

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

df\_with\_schema = spark.read.format("csv").option("header", "True").schema(schema).load("hdfs://nameservice1/user/bigdatalab456474/training/airlines.csv")

df\_with\_schema.registerTempTable("airlinespydata")

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

RevenueYearWise = spark.sql("select year, round(sum(arps\*booked\_seats)/100000,2) as total\_in\_lakhs from airlinespydata group by year order by total\_in\_lakhs desc")

RevenueYearWise.show(21)

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

HighAvgRevperSeat= spark.sql(“select year, max(avg(arps)) as aps from airlinespydata group by year”)

HighAvgRevperSeat.show()