

Airlines Assignment

Create Airport Table

```
CREATE TABLE airport (airportid int,  
    name String,  
    city String,  
    Country String,  
    iatafaa String,  
    icao String,  
    latitude String,  
    longitude String,  
    altitude String,  
    dst String, tz String)  
    ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS textfile  
LOCATION '/user/praveenkum79edu/data';
```

Hive screen shot:

```
hive> CREATE TABLE airport (airportid int,  
>  
>     name String,  
>  
>     city String,  
>  
>     Country String,  
>  
>     iatafaa String,  
>  
>     icao String,  
>  
>     latitude String,  
>  
>     longitude String,  
>  
>     altitude String,  
>  
>     dst String, tz String)  
>  
>     ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS textfile LOCATION '/user/praveenkum79edu/data';  
OK  
Time taken: 1.93 seconds
```

Create Airlines Table

```
CREATE TABLE airlines  
(airline bigint,  
    name String,  
    alias String,  
    iata String,  
    icap String,  
    callsign String,  
    country String,  
    active String)  
    ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  
    LINES TERMINATED BY '\n'  
    STORED AS textfile LOCATION '/user/praveenkum79edu/data';
```

Hive Screenshot:

```
hive> CREATE TABLE airlines
>
> (airline bigint,
>
> name String,
>
> alias String,
>
> iata String,
>
> icao String,
>
> callsign String,
>
> country String,
>
> active String)
>
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> LINES TERMINATED BY '\n'
>
> STORED AS textfile LOCATION '/user/praveenkum79edu/data';
OK
Time taken: 1.763 seconds
hive> █
```

Create Routes Table

```
CREATE TABLE routes
(airline String,
airlineid int,
sourceairport String,
sourceairportid int,
destinationairport String,
destinationairportid int,
codeshare String,
stops int,
equipment String)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
STORED AS textfile LOCATION '/user/praveenkum79edu/data';
```

Hive Screen shot:

```
hive> CREATE TABLE routes
>
> (airline String,
>
> airlineid int,
>
> sourceairport String,
>
> sourceairportid int,
>
> destinationairport String,
>
> destinationairportid int,
>
> codeshare String,
>
> stops int,
>
> equipment String)
>
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
>
> LINES TERMINATED BY '\n'
>
> STORED AS textfile LOCATION '/user/praveenkum79edu/data';
OK
Time taken: 0.082 seconds
hive> █
```

Problem Statement

A. Find list of Airports operating in the Country India

Solution:

Select airportid, name, city
from airport
where country="India";

Result screenshot:

The screenshot displays the Hive web interface. At the top, the query is entered: `1 Select airportid, name, city`
`2 from airport`
`3 where country="India";`
`4`

Below the query, the execution logs are visible, showing the job ID `job_1635139249191_3968` and the command `hadoop jar /opt/cloudera/parcels/CDH-4.2.1-1.cdh4.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1635139249191_3968`. The logs indicate the job completed successfully.

The results table, titled "Results (140)", shows the following data:

airportid	name	city
1 2994	Ahmedabad	Ahmedabad
2 2995	Akola	Akola
3 2996	Aurangabad	Aurangabad
4 2997	Chhatrapati Shivaji Intl	Mumbai
5 2998	Bilaspur	Bilaspur

On the right side, the "Tables" section lists the columns for the `default.airport` table:

Column	Type
airportid	int
name	string
city	string
country	string
latitude	string
icao	string
longitude	string
altitude	string
dst	string
tz	string

B. Find the list of Airlines having zero stops.

Solution:

```
Select distinct(airlines.airline), airlines.name
From airlines join routes on
airlines.airline=routes.airlineid
where routes.stops=0 ;
```

C. List of Airlines operating with code share

Solution:

```
SELECT DISTINCT(airlines.name) FROM airlines JOIN routes ON
airlines.airline=routes.airlineid WHERE routes.codeshare='Y';
```

D. Which country (or) territory having highest Airports

Solution:

```
select      country, count(*) as AirportCount
from        airport
group      by      country
order      by      AirportCount desc limit 1;
```

Output Screen shot:

The screenshot shows a Hive query execution interface. At the top, the query is entered in a text area:

```
1 select country, count(*) as AirportCount
2   from airport
3   group by country
4   order by AirportCount desc limit 1;
5
```

Below the query, the execution log is displayed, showing various status messages and performance metrics. The log includes information about the kill command, Hadoop job information, and the results of the MapReduce jobs.

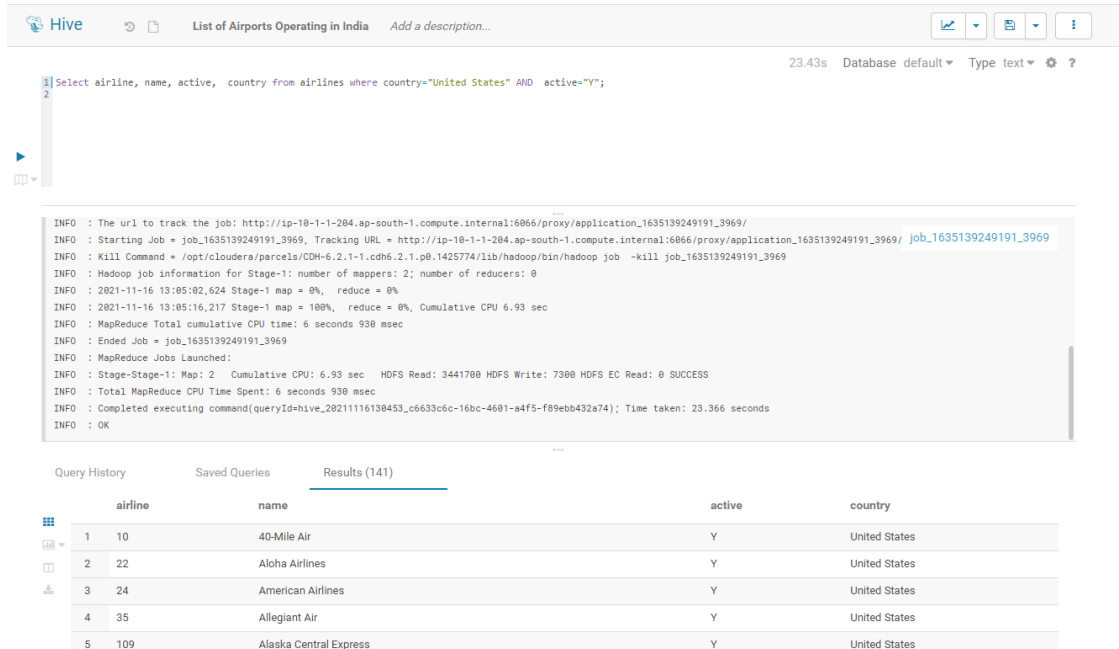
The results are shown in a table with two columns: **country** and **airportcount**. The table contains one row with the value 1 in the country column and 4585 in the airportcount column.

country	airportcount
1	4585

E. Find the list of Active Airlines in United state

Select airline, name, active, country from airlines where country="United States" AND active="Y";

Output Screen shot:



The screenshot displays the Hive web interface. At the top, the title bar shows 'List of Airports Operating in India' and 'Add a description...'. The query editor contains the following SQL query:

```
1 Select airline, name, active, country from airlines where country="United States" AND active="Y";
2
```

The execution status bar indicates '23.43s Database default Type text'. Below the query editor, the log output shows the following information:

```
INFO : The url to track the job: http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1635139249191_3969/
INFO : Starting Job = job_1635139249191_3969, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1635139249191_3969/ job_1635139249191_3969
INFO : Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1635139249191_3969
INFO : Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 0
INFO : 2021-11-16 13:05:02,624 Stage-1 map = 0%, reduce = 0%
INFO : 2021-11-16 13:05:16,217 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.93 sec
INFO : MapReduce Total cumulative CPU time: 6 seconds 930 msec
INFO : Ended Job = job_1635139249191_3969
INFO : MapReduce Jobs Launched:
INFO : Stage-Stage-1: Map: 2 Cumulative CPU: 6.93 sec HDFS Read: 3441700 HDFS Write: 7300 HDFS EC Read: 0 SUCCESS
INFO : Total MapReduce CPU Time Spent: 6 seconds 930 msec
INFO : Completed executing command(queryId=hive_20211116130453_c6633c6c-16bc-4601-a4f5-f89ebb432a74); Time taken: 23.366 seconds
INFO : OK
```

Below the log output, the 'Results (141)' tab is selected, showing a table with 5 rows of active airlines in the United States:

airline	name	active	country
1 10	40-Mile Air	Y	United States
2 22	Aloha Airlines	Y	United States
3 24	American Airlines	Y	United States
4 35	Allegiant Air	Y	United States
5 109	Alaska Central Express	Y	United States