

Airlines Analysis

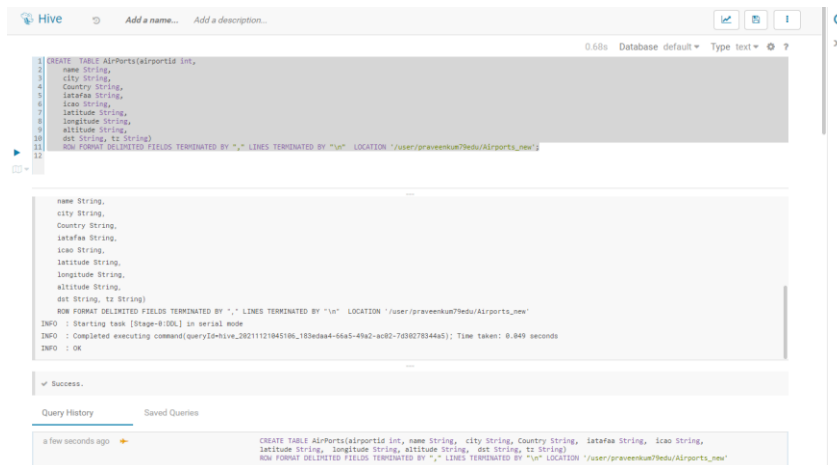
Creating and Loading Tables

- Create Airports Table and load airports_mod.data file into table

Query :

```
CREATE TABLE AirPorts(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" LOCATION '/user/praveenkum79edu/Airports_new';
```

Hive Screenshot:

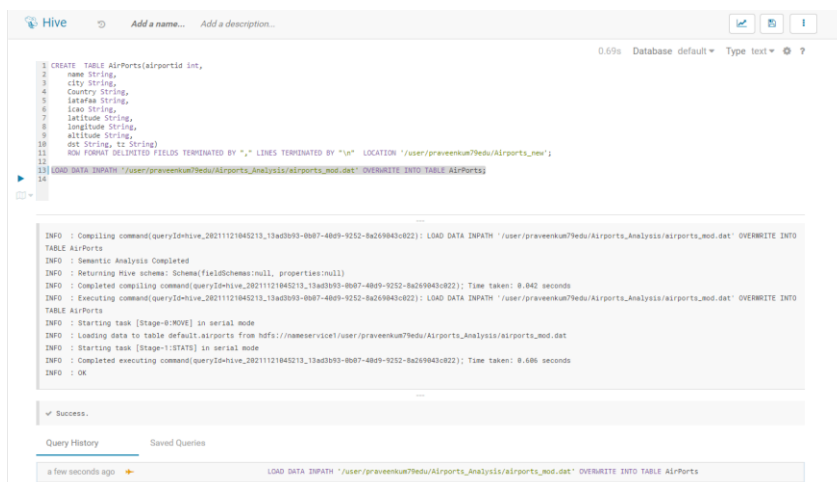


Load Data Into Airports Table:

Query:

LOAD DATA INPATH '/user/praveenkum79edu/Airports_Analysis/airports_mod.dat' OVERWRITE INTO TABLE AirPorts;

Hivwe Screenshot:



- **Create AirLines Table and load Final_airlines data file into table**

Query:

```
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/Airports_Analysis_new';
```

Hive Screenshot:

The screenshot shows the Hive web interface. The query editor contains the following SQL command:

```
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' LOCATION '/user/praveenkum79edu/AirLines_new';
```

The execution results show the table schema and the successful completion of the query:

```
(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' LOCATION '/user/praveenkum79edu/AirLines_new'
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryIdhive_2021121043117_35ef927d-5623-4d0c-918e-d5167cf27945); Time taken: 0.855 seconds
INFO : OK
```

The status bar indicates "Success." and the query history shows the command was executed a few seconds ago.

Load Data into Airliens Table :

LOAD DATA INPATH '/user/praveenkum79edu/Airports_Analysis/Final_airlines.txt' OVERWRITE INTO TABLE AirLines;

The screenshot shows the Hive web interface. The query editor contains the following SQL command:

```
LOAD DATA INPATH '/user/praveenkum79edu/Airports_Analysis/Final_airlines' OVERWRITE INTO TABLE AirLines;
```

The execution results show the successful completion of the query:

```
INFO : Compiling command(queryIdhive_2021121043117_35ef927d-5623-4d0c-918e-d5167cf27945): LOAD DATA INPATH '/user/praveenkum79edu/Airports_Analysis/Final_airlines' OVERWRITE INTO TABLE AirLines
INFO : Semantic Analysis Completed
INFO : Returning hive schema: Schema(fileIsSchema:null, properties:null)
INFO : Completed compiling command(queryIdhive_2021121043117_35ef927d-5623-4d0c-918e-d5167cf27945); Time taken: 0.84 seconds
INFO : Executing command(queryIdhive_2021121043117_35ef927d-5623-4d0c-918e-d5167cf27945): LOAD DATA INPATH '/user/praveenkum79edu/Airports_Analysis/Final_airlines' OVERWRITE INTO TABLE AirLines
INFO : Starting task [Stage-0:MOVE] in serial mode
INFO : Loading data to table default.airlines from hdfs://nameservice1/user/praveenkum79edu/Airports_Analysis/Final_airlines
INFO : Starting task [Stage-1:STATS] in serial mode
INFO : Completed executing command(queryIdhive_2021121043117_35ef927d-5623-4d0c-918e-d5167cf27945); Time taken: 0.651 seconds
INFO : OK
```

The status bar indicates "Success." and the query history shows the command was executed a few seconds ago.

- **Create Routes Table and load Routes.dat data file into table**

Query:

```
CREATE TABLE Routes_new
(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' LOCATION '/user/praveenkum79edu/Routes_new';
```

Hive Screenshot:

The screenshot shows the Hive query editor interface. The query to create the 'Routes_new' table is entered in the top text area. Below the query, the execution results are displayed, showing the table schema and the successful completion of the command. The bottom section shows the query history with the executed command.

```
CREATE TABLE Routes_new
(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' LOCATION '/user/praveenkum79edu/Routes_new';
```

INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command[queryId=hive_20211121040817_ea47eb9-f5ed-485b-8568-218cd5d54c93]; Time taken: 0.800 seconds
INFO : OK

Success.

Query History Saved Queries

a few seconds ago CREATE TABLE Routes_new (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' LOCATION '/user/praveenkum79edu/Routes_new'

Load data into Routes Table:

LOAD DATA INPATH '/user/praveenkum79edu/Airports_Analysis/routes.dat' OVERWRITE INTO TABLE Routes_new;

The screenshot shows the Hive query editor interface. The query to load data from the 'routes.dat' file into the 'Routes_new' table is entered in the top text area. Below the query, the execution results are displayed, showing the successful completion of the command. The bottom section shows the query history with the executed command.

```
LOAD DATA INPATH '/user/praveenkum79edu/Airports_Analysis/routes.dat' OVERWRITE INTO TABLE Routes_new;
```

INFO : Compiling command[queryId=hive_20211121050817_d8aee7e0-1cfd-45f0-ba7e-8579d691223d]; LOAD DATA INPATH '/user/praveenkum79edu/Airports_Analysis/routes.dat' OVERWRITE INTO TABLE Routes_new
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command[queryId=hive_20211121050817_d8aee7e0-1cfd-45f0-ba7e-8579d691223d]; Time taken: 0.041 seconds
INFO : Executing command[queryId=hive_20211121050817_d8aee7e0-1cfd-45f0-ba7e-8579d691223d]; LOAD DATA INPATH '/user/praveenkum79edu/Airports_Analysis/routes.dat' OVERWRITE INTO TABLE Routes_new
INFO : Starting task [Stage-0:MOVE] in serial mode
INFO : Loading data to table default.routes_new from hdfs://nameservice1/user/praveenkum79edu/Airports_Analysis/routes.dat
INFO : Starting task [Stage-1:STATS] in serial mode
INFO : Completed executing command[queryId=hive_20211121050817_d8aee7e0-1cfd-45f0-ba7e-8579d691223d]; Time taken: 0.596 seconds
INFO : OK

Success.

Query History Saved Queries

a few seconds ago LOAD DATA INPATH '/user/praveenkum79edu/Airports_Analysis/routes.dat' OVERWRITE INTO TABLE Routes_new

Problem statements:

Find list of Airports operating in the Country India

Solution:

Select airportid, name, city, country
from AirPorts
where country="India"

The screenshot shows a Hive query execution interface. The query is: `select airportid, name, city, country from AirPorts where country="India"`. The execution log shows the job starting, the command being executed, and the results being displayed. The results table has 5 rows and 4 columns: airportid, name, city, and country.

airportid	name	city	country
2994	Ahmedabad	Ahmedabad	India
2995	Akola	Akola	India
2996	Aurangabad	Aurangabad	India
2997	Chhatrapati Shivaji Intl	Mumbai	India
2998	Bilaspur	Bilaspur	India

Find the list of Airlines having zero stops.

Solution:

Select distinct(a.airline), a.name, Routes_new.stops
From AirLines a join Routes_new on
a.airline=Routes_new.airlineid
where stops=0;

The screenshot shows a Hive query execution interface. The query is: `select distinct(a.airline), a.name, Routes_new.stops from AirLines a join Routes_new on a.airline=Routes_new.airlineid where stops=0;`. The execution log shows the job starting, the command being executed, and the results being displayed. The results table has 6 rows and 3 columns: a.airline, a.name, and routes_new.stops.

a.airline	a.name	routes_new.stops
10	4D-Mile Air	0
21	Aigle Azur	0
24	American Airlines	0
28	Ariana Airlines	0
29	Ashari Aviation	0
32	Afghany Airways	0

List of Airlines operating with code share

Solution:

SELECT DISTINCT(AirLines.name),Routes_new.codeshare FROM AirLines JOIN Routes_new ON AirLines.airline=Routes_new.airlineid WHERE Routes_new.codeshare='Y';

Hive Airlines having zero stops Add a description...

1m, 2s Database default = Type text = ?

```
SELECT DISTINCT(airlines.name), routes_new.codeshare FROM Airlines JOIN routes_new ON Airlines.airline=routes_new.airlineid WHERE routes_new.codeshare="Y";
```

INFO : Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.jb.1425774/13b/hadoop/bin/hadoop job -kill job_1635139249191_4488

INFO : Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

INFO : 2021-11-21 05:14:43,938 Stage-2 map = 0%, reduce = 0%

INFO : 2021-11-21 05:14:58,522 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.8 sec

INFO : 2021-11-21 05:14:58,834 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.64 sec

INFO : MapReduce Total cumulative CPU time: 4 seconds 648 msec

INFO : Ended job = job_1635139249191_4488

INFO : MapReduce Jobs Launched:

INFO : Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 11.23 sec HDFS Read: 2718237 HDFS Write: 5115 HDFS EC Read: 0 SUCCESS

INFO : Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.64 sec HDFS Read: 11827 HDFS Write: 4244 HDFS EC Read: 0 SUCCESS

INFO : Total MapReduce CPU Time Spent: 15 seconds 879 msec

INFO : Completed executing command=queryIdhive_2021121051488_40863f9-ea83-4768-8465-416a16104877; Time taken: 57.21 seconds

INFO : OK

Query History Saved Queries Results (146)

airlines.name	routes_new.codeshare
1 Aloha Amer Flight	Y
2 Angkor Airlines	Y
3 Air Lingot	Y
4 Aeroflotus	Y
5 Aeroflot Russian Airlines	Y

Which country (or) territory having highest Airports

Solution:

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

Hive Airlines operating with code share Add a description...

55.99s Database default = Type text = ?

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

INFO : Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.jb.1425774/13b/hadoop/bin/hadoop job -kill job_1635139249191_4482

INFO : Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

INFO : 2021-11-21 05:18:01,885 Stage-2 map = 0%, reduce = 0%

INFO : 2021-11-21 05:18:08,945 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.84 sec

INFO : 2021-11-21 05:18:17,106 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.61 sec

INFO : MapReduce Total cumulative CPU time: 4 seconds 618 msec

INFO : Ended job = job_1635139249191_4482

INFO : MapReduce Jobs Launched:

INFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.17 sec HDFS Read: 748026 HDFS Write: 7822 HDFS EC Read: 0 SUCCESS

INFO : Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.61 sec HDFS Read: 12781 HDFS Write: 118 HDFS EC Read: 0 SUCCESS

INFO : Total MapReduce CPU Time Spent: 9 seconds 788 msec

INFO : Completed executing command=queryIdhive_2021121051724_7f0c76c3-4399-4817-9d99-7465835567c1; Time taken: 53.446 seconds

INFO : OK

Query History Saved Queries Results (1)

country	airportcount
1 United States	1697

Find the list of Active Airlines in United state

Solution:

```
select airline, name, active, country from AirLines where country="United States" AND
active="Y"
```

Hive AirportsOperating in India Add a description...

19.67s Database default = Type text = ?

```
select airline, name, active, country from Airlines where country="United States" AND active="Y"
```

INFO : Starting Job = job_1635139249191_4487, Tracking URL = http://10-10-1-1-284.ap-south-1.compute.internal:8066/proxy/application_1635139249191_4487/

INFO : Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.jb.1425774/13b/hadoop/bin/hadoop job -kill job_1635139249191_4487

INFO : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

INFO : 2021-11-21 05:25:04,685 Stage-1 map = 0%, reduce = 0%

INFO : 2021-11-21 05:25:12,910 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.52 sec

INFO : MapReduce Total cumulative CPU time: 3 seconds 520 msec

INFO : Ended job = job_1635139249191_4487

INFO : MapReduce Jobs Launched:

INFO : Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.39 sec HDFS Read: 12781 HDFS Write: 118 HDFS EC Read: 0 SUCCESS

INFO : Stage-Stage-1: Map: 1 Cumulative CPU: 3.52 sec HDFS Read: 321552 HDFS Write: 7213 HDFS EC Read: 0 SUCCESS

INFO : Total MapReduce CPU Time Spent: 7 seconds 918 msec

INFO : Completed executing command=queryIdhive_2021121052455_18141789-fbdc-47cb-ae62-47117244c201; Time taken: 19.62 seconds

INFO : OK

Query History Saved Queries Results (141)

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
6 149	Air Cargo Carriers	Y	United States