

# Country Project

## Creating and loading data into tables:



### Create flag table and load the data into it

```
CREATE TABLE flag (name String, landmass int, zone int, area int, population int, language int, religion int, bars int, stripes int,
colours int, red int, green int, blue int, gold int,
white int, black int, orange int, mainhue String, circles int, crosses int, saltires int, quaters int, sunstars int, crescent int,
trangle int, icon int, animate int, text int,
topleft String, botright String)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
STORED AS textfile LOCATION '/user/praveenkum79edu/Country_Analysis';
```

## Hive Screenshot:

The screenshot shows the Hive interface with a query editor and a results pane. The query is:   
`CREATE TABLE flag (name String, landmass int, zone int, area int, population int, language int, religion int, bars int, stripes int, colours int, red int, green int, blue int, gold int, white int, black int, orange int, mainhue String, circles int, crosses int, saltires int, quaters int, sunstars int, crescent int, trangle int, icon int, animate int, text int, topleft String, botright String) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS textfile LOCATION '/user/praveenkum79edu/Country_Analysis';`   
The results pane shows the execution of the query, indicating it was successful and took 8.963 seconds.

## Loading Data in to Flag table:

LOAD DATA INPATH '/user/praveenkum79edu/CoutryDataSet\_Data/flag.data\_dataset.txt' OVERWRITE INTO TABLE flag;

## Hive Screenshot:

The screenshot shows the Hive interface with a query editor and a results pane. The query is:   
`LOAD DATA INPATH '/user/praveenkum79edu/CoutryDataSet_Data/flag.data_dataset.txt' OVERWRITE INTO TABLE flag;`   
The results pane shows the execution of the query, indicating it was successful and took 9.626 seconds.

## Creating Reference tables to support flag table

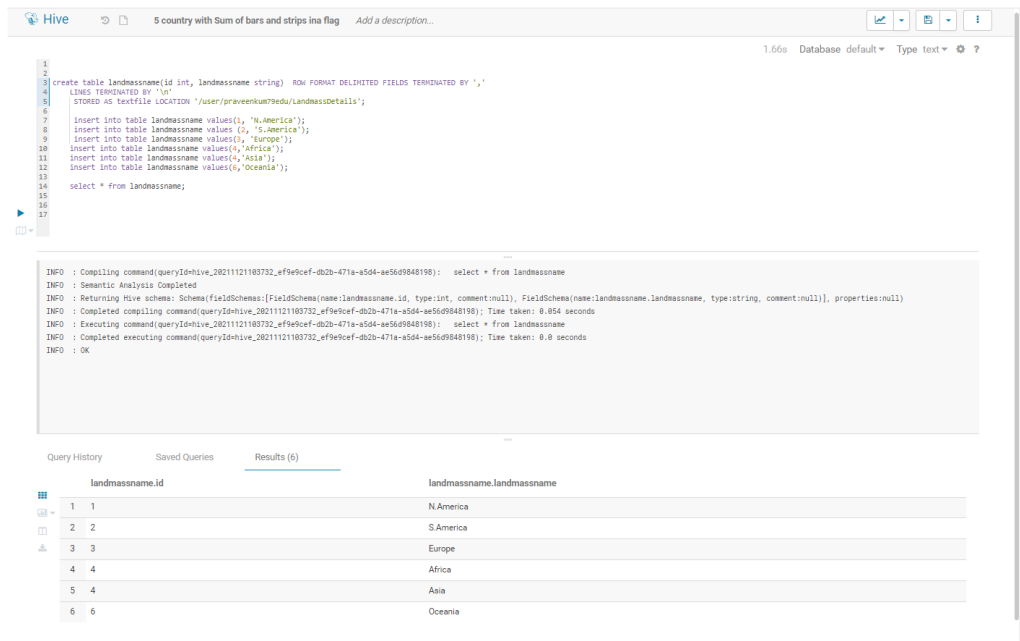
### Creating landmass table and loading data into it

```
create table landmassname(id int, landmassname string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
  LINES TERMINATED BY '\n'
  STORED AS textfile LOCATION '/user/praveenkum79edu/LandmassDetails';

insert into table landmassname values(1, 'N.America');
insert into table landmassname values (2, 'S.America');
insert into table landmassname values(3, 'Europe');
insert into table landmassname values(4,'Africa');
insert into table landmassname values(4,'Asia');
insert into table landmassname values(6,'Oceania');

select * from landmassname;
```

### Hive Screenshot:



The screenshot shows the Hive CLI interface. The top bar displays the Hive logo, a refresh icon, a description field, and a toolbar. The main area shows the execution of a Hive query. The query is as follows:

```
1 create table landmassname(id int, landmassname string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
2   LINES TERMINATED BY '\n'
3   STORED AS textfile LOCATION '/user/praveenkum79edu/LandmassDetails';
4
5 insert into table landmassname values(1, 'N.America');
6 insert into table landmassname values (2, 'S.America');
7 insert into table landmassname values(3, 'Europe');
8 insert into table landmassname values(4,'Africa');
9 insert into table landmassname values(4,'Asia');
10 insert into table landmassname values(6,'Oceania');
11
12 select * from landmassname;
```

The output shows the execution details and the resulting data:

```
INFO : Compiling command(queryId=hive_20211121183732_efe9ecef-d02b-471a-a5d4-ae56d9d48198): select * from landmassname
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(FieldSchema(name=landmassname.id, type=int, comment=null), FieldSchema(name=landmassname.landmassname, type=string, comment=null), properties=null)
INFO : Completed compiling command(queryId=hive_20211121183732_efe9ecef-d02b-471a-a5d4-ae56d9d48198): Time taken: 0.054 seconds
INFO : Executing command(queryId=hive_20211121183732_efe9ecef-d02b-471a-a5d4-ae56d9d48198): select * from landmassname
INFO : Completed executing command(queryId=hive_20211121183732_efe9ecef-d02b-471a-a5d4-ae56d9d48198): Time taken: 0.0 seconds
INFO : OK
```

The results are displayed in a table with 6 rows and 2 columns:

| landmassname.id | landmassname.landmassname |
|-----------------|---------------------------|
| 1               | N.America                 |
| 2               | S.America                 |
| 3               | Europe                    |
| 4               | Africa                    |
| 5               | Asia                      |
| 6               | Oceania                   |

### Creating ZoneDetails table and loading data into it

```
create table zoneDetails(id int, zonename string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
  LINES TERMINATED BY '\n'
  STORED AS textfile LOCATION '/user/praveenkum79edu/ZoneDetails';

insert into table zoneDetails values(1, 'NE');
insert into table zoneDetails values (2, 'SE');
insert into table zoneDetails values(3, 'SW');
insert into table zoneDetails values(4,'NW');

select * from zoneDetails;
```

## Hive Screenshot:

The screenshot shows the Hive interface with a query editor and a results pane. The query editor contains the following SQL:

```
1 create table zoneDetails(id int, zonename string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
2 LINES TERMINATED BY '\n'
3
4 STORED AS textfile LOCATION '/user/praveenkum79edu/zoneDetails';
5
6
7 insert into table zoneDetails values(1, 'NE');
8 insert into table zoneDetails values(2, 'SE');
9 insert into table zoneDetails values(3, 'SW');
10 insert into table zoneDetails values(4, 'NW');
11
12 select * from zoneDetails;
```

The results pane shows the output of the query:

| zoneDetails.id | zoneDetails.zonename |
|----------------|----------------------|
| 1              | NE                   |
| 2              | SE                   |
| 3              | SW                   |
| 4              | NW                   |

## Create language details table and load data into it

```
create table languageDetails(id int, zonename string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
STORED AS textfile LOCATION '/user/praveenkum79edu/languageDetails';
```

```
insert into table languageDetails values(1, 'English');
insert into table languageDetails values (2, 'Spanish');
insert into table languageDetails values(3, 'French');
insert into table languageDetails values(4,'German');
insert into table languageDetails values(5,'Slavic');
insert into table languageDetails values(6,'Other Indo-European');
insert into table languageDetails values(7,'Chinese');
insert into table languageDetails values(8,'Arabic');
insert into table languageDetails values(9,'Japanese/Turkish/Finnish/Magyar');
insert into table languageDetails values(10,'Others');
```

## Hive Screenshot

The screenshot shows the Hive interface with a query editor and a results pane. The query editor contains the following SQL:

```
1 create table languageDetails(id int, zonename string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
2 LINES TERMINATED BY '\n'
3
4 STORED AS textfile LOCATION '/user/praveenkum79edu/languageDetails';
5
6
7 insert into table languageDetails values(1, 'English');
8 insert into table languageDetails values (2, 'Spanish');
9 insert into table languageDetails values(3, 'French');
10 insert into table languageDetails values(4,'German');
11 insert into table languageDetails values(5,'Slavic');
12 insert into table languageDetails values(6,'Other Indo-European');
13 insert into table languageDetails values(7,'Chinese');
14 insert into table languageDetails values(8,'Arabic');
15 insert into table languageDetails values(9,'Japanese/Turkish/Finnish/Magyar');
16 insert into table languageDetails values(10,'Others');
```

The results pane shows the output of the query:

| languageDetails.id | languageDetails.zonename        |
|--------------------|---------------------------------|
| 1                  | English                         |
| 2                  | Spanish                         |
| 3                  | French                          |
| 4                  | German                          |
| 5                  | Slavic                          |
| 6                  | Other Indo-European             |
| 7                  | Chinese                         |
| 8                  | Arabic                          |
| 9                  | Japanese/Turkish/Finnish/Magyar |
| 10                 | Others                          |

## Problem Statement

Count number of countries based on landmass.

- ✓ select landmassname.landmassname as Countries , count(flag.name) from flag JOIN landmassname on landmassname.id=flag.landmass group by landmassname.landmassname;

The screenshot shows a Hive query execution interface. The query is: `select landmassname.landmassname as Countries , count(flag.name) from flag JOIN landmassname on landmassname.id=flag.landmass group by landmassname.landmassname;`. The results table has two columns: 'countries' and 'c1'. The data is as follows:

| countries   | c1 |
|-------------|----|
| 1 Africa    | 52 |
| 2 Asia      | 52 |
| 3 Europe    | 35 |
| 4 N.America | 31 |
| 5 Oceania   | 20 |
| 6 S.America | 17 |

Find out top 5 country with Sum of bars and strips in a flag.

- ✓ SELECT name, (bars+strips) as barstrip FROM flag ORDER BY barstrip DESC LIMIT 5;

The screenshot shows a Hive query execution interface. The query is: `SELECT name, (bars+strips) as barstrip FROM flag ORDER BY barstrip DESC LIMIT 5;`. The results table has two columns: 'name' and 'barstrip'. The data is as follows:

| name       | barstrip |
|------------|----------|
| 1 Malaysia | 14       |
| 2 USA      | 13       |
| 3 Liberia  | 11       |
| 4 Uruguay  | 9        |
| 5 Greece   | 9        |



Count of countries with icon.

✓ SELECT count(name) as countries\_with\_Icon FROM flag WHERE icon=1;

Hive 5 country with Sum of bars and strips ina flag Add a description...

28.33s Database default Type text ?

```
1 SELECT count(name) as countries_with_Icon FROM flag WHERE icon=1;
```

INFO : Starting Job = job\_1635139249191\_4523, Tracking URL = http://ip-10-1-1-284.ap-south-1.compute.internal:6066/proxy/application\_1635139249191\_4523/  
INFO : Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job\_1635139249191\_4523  
INFO : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
INFO : 2021-11-21 10:04:02,353 Stage-1 map = 0%, reduce = 0%  
INFO : 2021-11-21 10:04:10,573 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.35 sec  
INFO : 2021-11-21 10:04:10,743 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.38 sec  
INFO : MapReduce Total cumulative CPU time: 6 seconds 388 msec  
INFO : Ended Job = job\_1635139249191\_4523  
INFO : MapReduce Jobs Launched:  
INFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.38 sec HDFS Read: 28129 HDFS Write: 182 HDFS EC Read: 0 SUCCESS  
INFO : Total MapReduce CPU Time Spent: 6 seconds 388 msec  
INFO : Completed executing command(queryId=hive\_20211121100352\_f23c6f86-99ba-4d07-02de-129dbae332b3); Time taken: 28.276 seconds  
INFO : OK

Query History Saved Queries Results (1)

| countries_with_Icon |
|---------------------|
| 1 49                |



Count of countries which have same top left and top right color in flag.

✓ SELECT COUNT(name) FROM flag where topleft=botright;

Hive 5 country with Sum of bars and strips ina flag Add a description...

28.100s Database default Type text ?

```
1 SELECT COUNT(name) FROM flag where topleft=botright;
```

INFO : Starting Job = job\_1635139249191\_4526, Tracking URL = http://ip-10-1-1-284.ap-south-1.compute.internal:6066/proxy/application\_1635139249191\_4526/  
INFO : Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job\_1635139249191\_4526  
INFO : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
INFO : 2021-11-21 10:09:48,577 Stage-1 map = 0%, reduce = 0%  
INFO : 2021-11-21 10:09:56,748 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.01 sec  
INFO : 2021-11-21 10:10:04,920 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.26 sec  
INFO : MapReduce Total cumulative CPU time: 6 seconds 260 msec  
INFO : Ended Job = job\_1635139249191\_4526  
INFO : MapReduce Jobs Launched:  
INFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.26 sec HDFS Read: 27794 HDFS Write: 182 HDFS EC Read: 0 SUCCESS  
INFO : Total MapReduce CPU Time Spent: 6 seconds 260 msec  
INFO : Completed executing command(queryId=hive\_20211121100938\_6a7bf041-48d7-4c15-a955-91866cf1ac5a); Time taken: 28.041 seconds  
INFO : OK

Query History Saved Queries Results (1)

| _c0  |
|------|
| 1 76 |



Count number of countries based on zone.

✓ SELECT zoneDetails.zonename, COUNT(name) as Count\_Of\_Countries FROM flag JOIN zoneDetails on zonedetails.id=flag.zone group by zoneDetails.zonename;

**Hive** 5 country with Sum of bars and strips in flag Add a description...

```
1 SELECT zonename, COUNT(name) as Count_of_Countries FROM flag JOIN zonename ON zonename.id=flag.zone group by zonename
```

1m, 0s Database default Type text

INFO : Compiling command(queryId=hive\_2021121113324\_18a801d-745b-403b-85c2-21384ceaff11): SELECT zonename, COUNT(name) as Count\_of\_Countries FROM flag JOIN zonename ON zonename.id=flag.zone group by zonename

INFO : Semantic Analysis Completed

INFO : Returning Hive schema: Schema(FieldSchema(FieldSchema(name=zonename, type=string, comment=null), FieldSchema(name=count\_of\_countries, type=bigint, comment=null)), properties=null)

INFO : Completed compiling command(queryId=hive\_2021121113324\_18a801d-745b-403b-85c2-21384ceaff11): Time taken: 0.88 seconds

INFO : Executing command(queryId=hive\_2021121113324\_18a801d-745b-403b-85c2-21384ceaff11): SELECT zonename, COUNT(name) as Count\_of\_Countries FROM flag JOIN zonename ON zonename.id=flag.zone group by zonename

WARN :

INFO : Query ID = hive\_2021121113324\_18a801d-745b-403b-85c2-21384ceaff11

INFO : Total jobs = 3

INFO : Launching Job 1 out of 2

INFO : Starting task [Stage-1:MAPRED] in serial mode

INFO : Number of reduce tasks not specified. Estimated from input data size: 1

INFO : In order to change the average load for a reducer (in bytes):

INFO : ... set hive.exec.reducers.bytes.per.reducer=<number>

Query History Saved Queries Results (4)

| zonename | count_of_countries |
|----------|--------------------|
| NE       | 91                 |
| NW       | 58                 |
| SE       | 29                 |
| SW       | 16                 |



Find out largest county in terms of area in NE zone

- ✓ SELECT MAX(area) as max\_area, flag.name FROM flag WHERE flag.zone IN (SELECT ZoneDetails.id from zoneDetails WHERE zoneDetails.zonename='NE') GROUP BY flag.name order by max\_area desc limit 1;

**Hive** 5 country with Sum of bars and strips in flag Add a description...

```
1 SELECT MAX(area) as max_area, flag.name FROM flag WHERE flag.zone IN (SELECT ZoneDetails.id from zoneDetails WHERE zoneDetails.zonename='NE') GROUP BY flag.name order by max_area desc limit 1;
```

1m, 28s Database default Type text

INFO : Compiling command(queryId=hive\_2021121115416\_19e9f64-83d5-45c3-842f-c126e9f6eeae):

SELECT MAX(area) as max\_area, flag.name FROM flag WHERE flag.zone IN (SELECT ZoneDetails.id from zoneDetails WHERE zoneDetails.zonename='NE') GROUP BY flag.name order by max\_area desc limit 1

INFO : Semantic Analysis Completed

INFO : Returning Hive schema: Schema(FieldSchema(FieldSchema(name=max\_area, type=int, comment=null), FieldSchema(name=flag.name, type=string, comment=null)), properties=null)

INFO : Completed compiling command(queryId=hive\_2021121115416\_19e9f64-83d5-45c3-842f-c126e9f6eeae): Time taken: 0.118 seconds

INFO : Executing command(queryId=hive\_2021121115416\_19e9f64-83d5-45c3-842f-c126e9f6eeae):

SELECT MAX(area) as max\_area, flag.name FROM flag WHERE flag.zone IN (SELECT ZoneDetails.id from zoneDetails WHERE zoneDetails.zonename='NE') GROUP BY flag.name order by max\_area desc limit 1

WARN :

INFO : Query ID = hive\_2021121115416\_19e9f64-83d5-45c3-842f-c126e9f6eeae

INFO : Total jobs = 3

INFO : Launching Job 1 out of 3

INFO : Starting task [Stage-1:MAPRED] in serial mode

INFO : Number of reduce tasks not specified. Estimated from input data size: 1

Query History Saved Queries Results (1)

| max_area | flag.name |
|----------|-----------|
| 22402    | USSR      |



Find out least populated country in S.America landmass

- ✓ select flag.name, flag.population, flag.landmass, landmassname.landmassname from flag join landmassname on flag.landmass= landmassname.id where landmassname.landmassname='S.America' order by flag.population ASC limit 1;

**Hive** 5 country with Sum of bars and strips in flag Add a description...

```
1 select flag.name, flag.population, flag.landmass, landmassname.landmassname from flag join landmassname on flag.landmass= landmassname.id where landmassname.landmassname='S.America' order by flag.population ASC limit 1;
```

1m, 0s Database default Type text

INFO : Compiling command(queryId=hive\_2021121122084\_4378469-71ef-4d5b-9d2a-4a677b18444):

select flag.name, flag.population, flag.landmass, landmassname.landmassname from flag join landmassname on flag.landmass= landmassname.id where landmassname.landmassname='S.America' order by flag.population ASC limit 1

INFO : Semantic Analysis Completed

INFO : Returning Hive schema: Schema(FieldSchema(FieldSchema(name=flag.name, type=string, comment=null), FieldSchema(name=flag.population, type=int, comment=null), FieldSchema(name=flag.landmass, type=int, comment=null), FieldSchema(name=landmassname.landmassname, type=string, comment=null)), properties=null)

INFO : Completed compiling command(queryId=hive\_2021121122084\_4378469-71ef-4d5b-9d2a-4a677b18444): Time taken: 0.88 seconds

INFO : Executing command(queryId=hive\_2021121122084\_4378469-71ef-4d5b-9d2a-4a677b18444):

select flag.name, flag.population, flag.landmass, landmassname.landmassname from flag join landmassname on flag.landmass= landmassname.id where landmassname.landmassname='S.America' order by flag.population ASC limit 1

WARN :

INFO : Query ID = hive\_2021121122084\_4378469-71ef-4d5b-9d2a-4a677b18444

INFO : Total jobs = 2

INFO : Launching Job 1 out of 2

INFO : Starting task [Stage-1:MAPRED] in serial mode

Query History Saved Queries Results (1)

| flag.name | flag.population | flag.landmass | landmassname.landmassname |
|-----------|-----------------|---------------|---------------------------|
| Burkina   | 0               | 2             | S.America                 |



## Find out largest speaking language among all countries

- ✓ `SELECT zonename from languageDetails WHERE id IN (SELECT tmpflag.language FROM (SELECT language, COUNT(language) AS clang FROM flag GROUP BY language ORDER BY clang DESC LIMIT 1) tmpflag);`

Query History Saved Queries Results (1)

| zonename |
|----------|
| Others   |



## Find most common colour among flags from all countries

- ✓ `select colorName,count(name) as total  
from  
(select  
case  
when red == 1 then 'red'  
when black == 1 then 'black'  
when green == 1 then 'green'  
when blue == 1 then 'blue'  
when gold == 1 then 'gold'  
when white == 1 then 'white'  
when orange == 1 then 'orange'  
end as colorName, name from flag) a  
group by colorName order by total desc limit 1;`

Query History Saved Queries Results (1)

| colorname | total |
|-----------|-------|
| red       | 153   |



## Sum of all circles present in all country flags.

- ✓ `SELECT SUM(circles) AS SUM_OF_CIRCLE FROM flag;`

Hive 5 country with Sum of bars and stripes in a flag Add a description...

24.24s Database default Type text

```
1 SELECT SUM(circles) AS SUM_OF_CIRCLE FROM flag;
```

INFO : Compiling command(queryId=hive\_20211121133118\_3c83e8e4-3f22-4768-8cda-dba819163e5): SELECT SUM(circles) AS SUM\_OF\_CIRCLE FROM flag  
INFO : Semantic Analysis Completed  
INFO : Returning Hive schema: Schema(fieldSchemas([fieldSchema(name=sum\_of\_circle, type=bigint, comment=null)], properties=null))  
INFO : Completed compiling command(queryId=hive\_20211121133118\_3c83e8e4-3f22-4768-8cda-dba819163e5): Time taken: 0.00 seconds  
INFO : Executing command(queryId=hive\_20211121133118\_3c83e8e4-3f22-4768-8cda-dba819163e5): SELECT SUM(circles) AS SUM\_OF\_CIRCLE FROM flag  
WARN :  
INFO : Query ID = hive\_20211121133118\_3c83e8e4-3f22-4768-8cda-dba819163e5  
INFO : Total jobs = 1  
INFO : Launching Job 1 out of 1  
INFO : Starting task [Stage-1:MAPRED] in serial mode  
INFO : Number of reduce tasks determined at compile time: 1  
INFO : In order to change the average load for a reducer (in bytes):  
INFO : set hive.mapred.reducebytes.per.reducer=number>

Query History Saved Queries Results (1)

| sum_of_circle |
|---------------|
| 1 33          |

Count of countries which have both icon and text in flag  
✓ SELECT name FROM flag WHERE icon=1 AND text=1;

Hive 5 country with Sum of bars and stripes in a flag Add a description...

17.92s Database default Type text

```
1 SELECT name FROM flag WHERE icon=1 AND text=1;
```

INFO : Compiling command(queryId=hive\_20211121133239\_1b08723f-6e9e-496d-9ace-1a802c5982d3): SELECT name FROM flag WHERE icon=1 AND text=1  
INFO : Semantic Analysis Completed  
INFO : Returning Hive schema: Schema(fieldSchemas([fieldSchema(name=name, type=string, comment=null)], properties=null))  
INFO : Completed compiling command(queryId=hive\_20211121133239\_1b08723f-6e9e-496d-9ace-1a802c5982d3): Time taken: 0.00 seconds  
INFO : Executing command(queryId=hive\_20211121133239\_1b08723f-6e9e-496d-9ace-1a802c5982d3): SELECT name FROM flag WHERE icon=1 AND text=1  
WARN :  
INFO : Query ID = hive\_20211121133239\_1b08723f-6e9e-496d-9ace-1a802c5982d3  
INFO : Total jobs = 1  
INFO : Launching Job 1 out of 1  
INFO : Starting task [Stage-1:MAPRED] in serial mode  
INFO : Number of reduce tasks is set to 0 since there's no reduce operator  
INFO : number of splits:1  
INFO : Submitting tokens for job: job\_1635139249191\_4631

Query History Saved Queries Results (13)

| name                   |
|------------------------|
| 1 Belize               |
| 2 British-Virgin-Isles |
| 3 Brunei               |
| 4 Cayman-Islands       |
| 5 Falklands-Malvinas   |
| 6 Guam                 |
| 7 Hong-Kong            |
| 8 Iran                 |
| 9 Paraguay             |
| 10 Romania             |
| 11 Saudi-Arabia        |
| 12 St-Vincent          |
| 13 US-Virgin-Isles     |