***********************Create External table in Hive*******************************
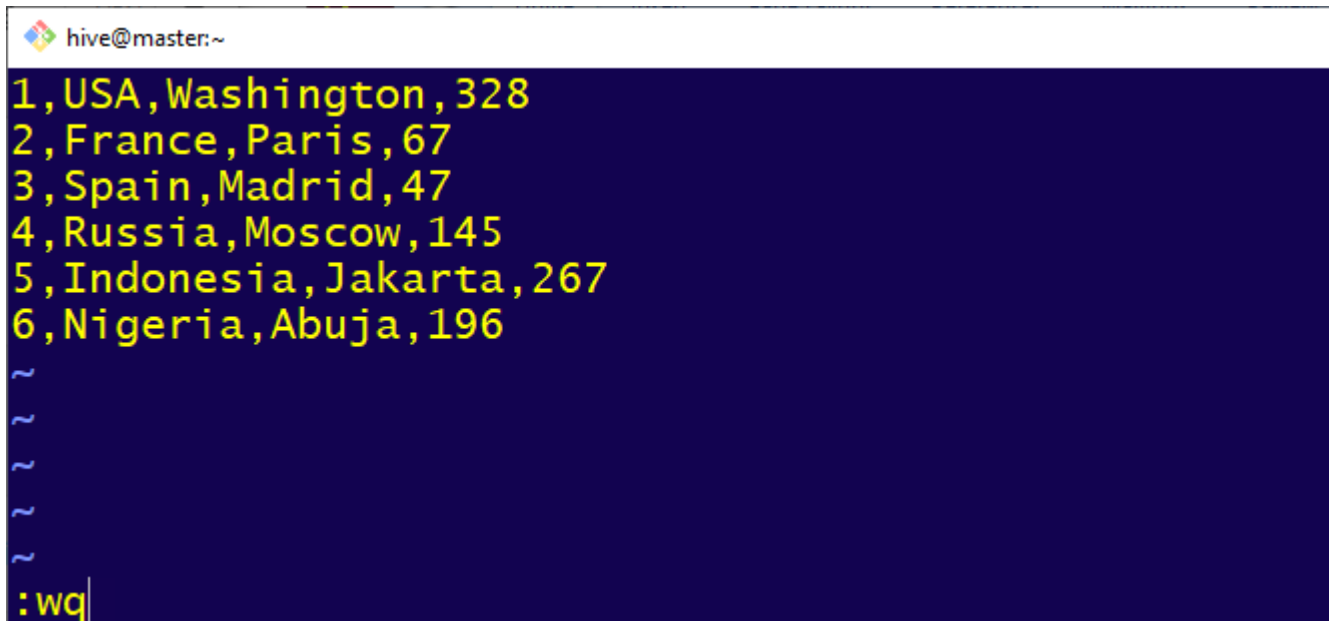
Step 1: Prepare the Data File
Create a CSV file titled 'countries.csv':
[hive@master ~]$ vim countries.csv

For each country in the list, write a row number, the country's name, its capital city, and its population in millions

1,USA,Washington,328
2,France,Paris,67
3,Spain,Madrid,47
4,Russia,Moscow,145
5,Indonesia,Jakarta,267
6,Nigeria,Abuja,196

```
hive@master:~
1,USA,Washington,328
2,France,Paris,67
3,Spain,Madrid,47
4,Russia,Moscow,145
5,Indonesia,Jakarta,267
6,Nigeria,Abuja,196
~
~
~
~
~
:wq
```

3. Save the file and make a note of its location.

Step 2: Import the File to HDFS

1. Create an HDFS directory. You will use this directory as an HDFS location of the file you created.

   hdfs dfs -mkdir /user/hive/Bharat

hive@master:~

```
[hive@master ~]$ hdfs dfs -mkdir /user/hive/Bharat
```

2.Import the CSV file into HDFS:

   hdfs dfs -put countries.csv "/user/hive/Bharat"

hive@master:~

```
[hive@master ~]$ hdfs dfs -put countries.csv "/user/hive/Bharat"
[hive@master ~]$
```

3.Use the -ls command to verify that the file is in the HDFS folder:

   hdfs dfs -ls /user/hive/Bharat

hive@master:~   —   [

```
[hive@master ~]$ hdfs dfs -put countries.csv "/user/hive/Bharat"
[hive@master ~]$ hdfs dfs -ls  /user/hive/Bharat
Found 1 items
-rw-r--r--   3 hive hive        121 2023-06-27 16:59 /user/hive/Bharat/countries.csv
[hive@master ~]$
```

The output displays all the files currently in the directory.

Step 3: Create an External Table

1. After you import the data file to HDFS, initiate Hive and use the syntax explained above to create an external table.

create external table if not exists countries_list (id int, name string, capital string, population string) comment 'List of Countries' row format delimited fields terminated by ',' stored as textfile location '/user/hive/Bharat';

```
root@master:~                                                    —    □    ×
0: jdbc:hive2://localhost:10000/default>  create external table if not exists
countries_list (id int, name string, capital string, population string ) comme
nt 'List of Countries' row format delimited fields terminated by ',' stored as
 textfile location '/user/hive/Bharat';
```
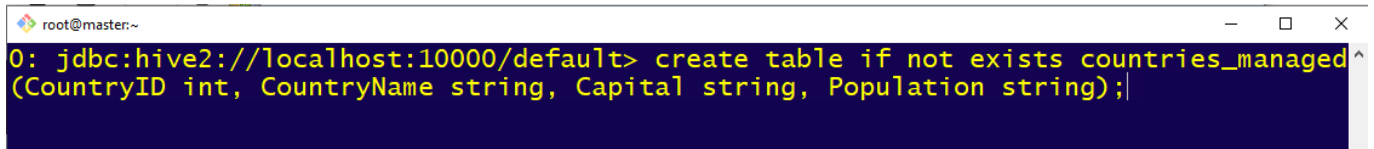
2. To verify that the external table creation was successful, type:

select * from countries_list;

```
root@master:~                                                                —
0: jdbc:hive2://localhost:10000/default> select * from countries_list;
INFO  : Compiling command(queryId=hive_20230627180652_500029cf-dd21-4536-851e-d6dbb6b8b6c6): select *
untries_list
INFO  : No Stats for default@countries_list, Columns: capital, name, id, population
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Created Hive schema: Schema(fieldSchemas:[FieldSchema(name:countries_list.id, type:int, comment
, FieldSchema(name:countries_list.name, type:string, comment:null), FieldSchema(name:countries_list.cap
type:string, comment:null), FieldSchema(name:countries_list.population, type:string, comment:null)], pr
es:null)
INFO  : Completed compiling command(queryId=hive_20230627180652_500029cf-dd21-4536-851e-d6dbb6b8b6c6);
aken: 0.034 seconds
INFO  : Executing command(queryId=hive_20230627180652_500029cf-dd21-4536-851e-d6dbb6b8b6c6): select *
untries_list
INFO  : Completed executing command(queryId=hive_20230627180652_500029cf-dd21-4536-851e-d6dbb6b8b6c6);
aken: 0.001 seconds
INFO  : OK
+------------------+--------------------+----------------------+-------------------------+
| countries_list.id | countries_list.name | countries_list.capital | countries_list.population |
+------------------+--------------------+----------------------+-------------------------+
| 1                | USA                | Washington           | 328                     |
| 2                | France             | Paris                | 67                      |
| 3                | Spain              | Madrid               | 47                      |
| 4                | Russia             | Moscow               | 145                     |
| 5                | Indonesia          | Jakarta              | 267                     |
| 6                | Nigeria            | Abuja                | 196                     |
+------------------+--------------------+----------------------+-------------------------+
6 rows selected (0.118 seconds)
0: jdbc:hive2://localhost:10000/default>
```

3. If you wish to create a managed table using the data from an external table, type:

Create table if not exists countries_managed(CountryID int, CountryName string, Capital string, Population string);

4. Next, import the data from the external table:

insert overwrite table countries_managed select * from countries_list;

5. Verify that the data is successfully inserted into the manged tables.

Select * from countries_managed;

How to Query a Hive External table

To display all the data stored in a table, you will use the select * from command by the table name. Hive offers an expensive list of query commands to let you narrow down your searches and sort the data according to your preferences.

For example, you can use the where command after select * from to specify a condition:

Select * from countries_managed where population<150;

Hive will output only the rows which satisfy the condition given in the query:

```
0: jdbc:hive2://master:2181/default> select * from countries_managed where population<150;
INFO  : Compiling command(queryId=hive_20230628121019_0570b7c1-94e1-4a49-af4d-6002d8a509a8): select * from countries_managed whe
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Created Hive schema: Schema(fieldSchemas:[FieldSchema(name:countries_managed.countryid, type:int, comment:null), FieldSch
s_managed.countryname, type:string, comment:null), FieldSchema(name:countries_managed.capital, type:string, comment:null), Field
ies_managed.population, type:string, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20230628121019_0570b7c1-94e1-4a49-af4d-6002d8a509a8); Time taken: 0.048 seconds
INFO  : Executing command(queryId=hive_20230628121019_0570b7c1-94e1-4a49-af4d-6002d8a509a8): select * from countries_managed whe
INFO  : Completed executing command(queryId=hive_20230628121019_0570b7c1-94e1-4a49-af4d-6002d8a509a8); Time taken: 0.003 seconds
INFO  : OK
+---------------------------+-----------------------------+--------------------------+-----------------------------+
| countries_managed.countryid | countries_managed.countryname | countries_managed.capital | countries_managed.population |
+---------------------------+-----------------------------+--------------------------+-----------------------------+
| 2                         | France                      | Paris                    | 67                          |
| 3                         | Spain                       | Madrid                   | 47                          |
| 4                         | Russia                      | Moscow                   | 145                         |
| 7                         | India                       | Delhi                    | 143                         |
+---------------------------+-----------------------------+--------------------------+-----------------------------+
4 rows selected (0.07 seconds)
```