ASSIGNMENT 9.1 Advance Hive

Problem Statement:

Given Olympic data set, write a hive program for the below tasks.

Created table and loaded the data using below commands:

1.create table Olympic_data(name String, age Int, country String, year String, Cdate String, sport String, GMedal int, SMedal int, BMedal int, Totmedals int) row format delimited fields terminated by '\t';

2. LOAD DATA LOCAL INPATH '/home/acadgild/olympix_data.csv' into table Olympic_data;

Task 1:

---

1. Write a Hive program to find the number of medals won by each country in swimming.

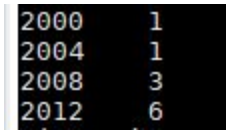
Query: select country , sum(totmedals) from olympic_data where sport = "Swimming" group by country;

output:

2. Write a Hive program to find the number of medals that India won year wise.
Query for the above task:

Query: select year , sum(totmedals) from olympic_data where country = "India" group by year;

output:



3. Write a Hive Program to find the total number of medals each country won.

Query: select country , sum(totmedals) from olympic_data group by country;

Output:

```
Poland  80
Portugal        9
Puerto Rico     2
Qatar   3
Romania 123
Russia  768
Saudi Arabia    6
Serbia  31
Serbia and Montenegro   38
Singapore       7
Slovakia        35
Slovenia        25
South Africa    25
South Korea     308
Spain   205
Sri Lanka       1
Sudan   1
Sweden  181
Switzerland     93
Syria   1
Tajikistan      3
Thailand        18
Togo    1
Trinidad and Tobago     19
Tunisia 4
Turkey  28
Uganda  1
Ukraine 143
United Arab Emirates    1
United States   1312
Uruguay 1
Uzbekistan      19
Venezuela       4
Vietnam 2
Zimbabwe        7
Time taken: 27.986 seconds, Fetched: 110 row(s)
```

4. Write a Hive program to find the number of gold medals each country won.

select country , sum(GMedals) from olympic_data group by country;

Output:

Task 2:

Write a hive UDF that implements functionality of string concat_ws(string SEP, array<string>).
This UDF will accept two arguments, one string and one array of string.
It will return a single string where all the elements of the array are separated by the SEP.
Steps to create a udf in Hive:
2. For Adding the import org.apache.hadoop.hive.ql.exec.UDF without error add the below external jar files as below:
install/hive/bin/hive-exec-2.3.2
3. Once after fixing all the code issues create a jar file(concat_ws1.jar) using eclipse IDE. Source code is attached as separate file explaining the code.
4. Move the concat_ws1.jar file to the VM local file system.
5. Add the jar file and create temporary function as below:
ADD JAR /home/acadgild/concat_ws1.jar;

CREATE TEMPORARY FUNCTION concat1 AS 'concat_ws.concat_ws';

```
hive> ADD JAR /home/acadgild/concat_ws1.jar;
Added [/home/acadgild/concat_ws1.jar] to class path
Added resources: [/home/acadgild/concat_ws1.jar]
hive> CREATE TEMPORARY FUNCTION concat1 AS 'concat_ws.concat_ws';
OK
Time taken: 0.152 seconds
hive>
```

7. Now use the created temporary file to concat1 to create the below output.
select concat1('-', zip,temp) from temperature_data;

Output:

```
hive> select concat1('-', zip,temp) from temperature_data;
OK
123112-10
283901-11
381920-15
302918-22
384902-9
123112-11
283901-12
381920-16
302918-23
384902-10
123112-11
283901-12
381920-16
302918-23
384902-10
123112-11
283901-12
381920-16
302918-23
384902-10
Time taken: 3.1 seconds, Fetched: 20 row(s)
```