Task 1-

1. **Create a database named as OLYMPICS-**

```
hive (demo)> create database OLYMPICS;
OK
Time taken: 0.494 seconds
hive (demo)> use OLYMPICS;
OK
Time taken: 0.069 seconds
```

2. **CREATE a TABLE named OLYMPIC inside the database OLYMPICS with below columns**
3. **LOAD DATA from LOCAL File system to the table OLYMPIC**

Contents are also shown below after loading the table-

```
hive (OLYMPICS)> CREATE TABLE OLYMPIC
               > (
               > athlete STRING,
               > age int,
               > country STRING,
               > year int,
               > closing_date STRING,
               > sport STRING,
               > gold_medal int,
               > silver_medal int,
               > bronze_medal int,
               > total_medal int
               > )
               > ROW FORMAT DELIMITED
               > FIELDS TERMINATED BY '\t';
OK
Time taken: 0.446 seconds
hive (OLYMPICS)> LOAD DATA LOCAL INPATH '/home/acadgild/hive/olympix_data.csv'
               > INTO TABLE OLYMPICS.OLYMPIC;
Loading data to table olympics.olympic
Table olympics.olympic stats: [numFiles=1, totalSize=518669]
OK
Time taken: 1.449 seconds
hive (OLYMPICS)> select * from OLYMPIC LIMIT 5;
OK
olympic.athlete olympic.age     olympic.country olympic.year    olympic.closing_date    olympic.sport   olympic.gold_medal      olympic.silver_medal    olympic.bronz
e_medal olympic.total_medal
Michael Phelps  23      United States   2008    08-24-08        Swimming        8       0       0       8
Michael Phelps  19      United States   2004    08-29-04        Swimming        6       0       2       8
Michael Phelps  27      United States   2012    08-12-12        Swimming        4       2       0       6
Natalie Coughlin        25      United States   2008    08-24-08        Swimming        1       2       3       6
Aleksey Nemov   24      Russia  2000    10-01-00        Gymnastics      2       1       3       6
Time taken: 0.255 seconds, Fetched: 5 row(s)
hive (OLYMPICS)>
```

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

Run below query to get total number of each type of medals for each country for Swimming-

```
SELECT country, SUM(gold_medal) as TOTAL_GOLD,
SUM(silver_medal) as TOTAL_SILVER,
SUM(bronze_medal) as TOTAL_BRONZE,
SUM(total_medal) as TOTAL_MEDALS
FROM OLYMPIC WHERE sport='Swimming' GROUP BY country;
```

```
hive (OLYMPICS)> SELECT country,SUM(gold_medal) as TOTAL_GOLD,SUM(silver_medal) as TOTAL_SILVER,SUM(bronze_medal) as TOTAL_BRONZE,
               > SUM(total_medal) as TOTAL_MEDALS FROM OLYMPIC WHERE sport='Swimming' GROUP BY country;
```

Below is the OUTPUT of above query

```
OK
country total_gold      total_silver    total_bronze    total_medals
Argentina       0       0       1       1
Australia       58      68      37      163
Austria 0       2       1       3
Belarus 0       2       0       2
Brazil  1       1       6       8
Canada  0       1       4       5
China   7       14      14      35
Costa Rica      0       0       2       2
Croatia 0       1       0       1
Denmark 0       0       1       1
France  11      16      12      39
Germany 2       6       24      32
Great Britain   2       3       6       11
Hungary 3       4       2       9
Italy   4       3       9       16
Japan   5       9       29      43
Lithuania       1       0       0       1
Netherlands     16      17      13      46
Norway  0       1       1       2
Poland  1       2       0       3
Romania 3       1       2       6
Russia  1       9       10      20
Serbia  0       1       0       1
Slovakia        0       2       0       2
Slovenia        0       1       0       1
South Africa    6       3       2       11
South Korea     1       3       0       4
Spain   0       2       1       3
Sweden  1       2       6       9
Trinidad and Tobago     0       0       1       1
Tunisia 2       0       1       3
Ukraine 4       2       1       7
United States   139     77      51      267
Zimbabwe        2       4       1       7
Time taken: 131.413 seconds, Fetched: 34 row(s)
hive (OLYMPICS)>
```

2. **Write a Hive program to find the number of medals that India won year wise.**

Below is the query to find the number of medals India won year wise-

```
>SELECT year,
 > SUM(gold_medal) as TOTAL_GOLD,
 > SUM(silver_medal) as TOTAL_SILVER,
 > SUM(bronze_medal) as TOTAL_BRONZE,
 > SUM(total_medal) as TOTAL_MEDALS
 > FROM OLYMPIC WHERE country='India'
 > GROUP BY year;
```

```
hive (OLYMPICS)> SELECT year,
               > SUM(gold_medal) as TOTAL_GOLD,
               > SUM(silver_medal) as TOTAL_SILVER,
               > SUM(bronze_medal) as TOTAL_BRONZE,
               > SUM(total_medal) as TOTAL_MEDALS
               > FROM OLYMPIC WHERE country='India'
               > GROUP BY year;
```

Below is the output of above query-

```
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 35.8 sec   HDFS Read: 518901 HDFS Write: 52 SUCCESS
Total MapReduce CPU Time Spent: 35 seconds 800 msec
OK
year    total_gold      total_silver    total_bronze    total_medals
2000    0       0       1       1
2004    0       1       0       1
2008    1       0       2       3
2012    0       2       4       6
Time taken: 122.924 seconds, Fetched: 4 row(s)
hive (OLYMPICS)>
```

3. **<u>Write a Hive Program to find the total number of medals each country won.</u>**

Below is the query used to find the total number of medals each country has won-

```
> SELECT country,
> SUM(gold_medal) as TOTAL_GOLD,
> SUM(silver_medal) as TOTAL_SILVER,
> SUM(bronze_medal) as TOTAL_BRONZE,
> SUM(total_medal) as TOTAL_MEDALS
> FROM OLYMPIC
> GROUP BY country;
```

```
hive (OLYMPICS)> SELECT country,
               > SUM(gold_medal) as TOTAL_GOLD,
               > SUM(silver_medal) as TOTAL_SILVER,
               > SUM(bronze_medal) as TOTAL_BRONZE,
               > SUM(total_medal) as TOTAL_MEDALS
               > FROM OLYMPIC
               > GROUP BY country;
```

Below are the screenshots for the result of output-

```
OK
country total_gold      total_silver    total_bronze    total_medals
Afghanistan     0       0       2       2
Algeria 2       2       4       8
Argentina       49      34      58      141
Armenia 0       1       9       10
Australia       163     226     220     609
Austria 36      26      29      91
Azerbaijan      6       4       15      25
Bahamas 11      6       7       24
Bahrain 0       0       1       1
Barbados        0       0       1       1
Belarus 17      33      47      97
Belgium 2       8       8       18
Botswana        0       1       0       1
Brazil  46      99      76      221
Bulgaria        8       11      22      41
Cameroon        20      0       0       20
Canada  168     98      104     370
Chile   3       1       18      22
China   234     156     140     530
Chinese Taipei  2       6       12      20
Colombia        2       4       7       13
Costa Rica      0       0       2       2
Croatia 35      14      32      81
Cuba    57      80      51      188
Cyprus  0       1       0       1
Czech Republic  14      21      46      81
Denmark 46      15      28      89
Dominican Republic      3       2       0       5
Ecuador 0       1       0       1
Egypt   1       3       4       8
Eritrea 0       0       1       1
Estonia 6       6       6       18
Ethiopia        13      6       10      29
Finland 11      46      61      118
France  108     107     103     318
Gabon   0       1       0       1
```

```
Georgia 6       5       12      23
Germany 223     183     223     629
Great Britain   124     101     97      322
Greece  12      27      20      59
Grenada 1       0       0       1
Guatemala       0       1       0       1
Hong Kong       0       2       1       3
Hungary 77      52      16      145
Iceland 0       14      1       15
India   1       3       7       11
Indonesia       5       8       9       22
Iran    10      7       7       24
Ireland 1       3       5       9
Israel  1       0       3       4
Italy   86      103     142     331
Jamaica 24      33      23      80
Japan   57      112     113     282
Kazakhstan      13      14      15      42
Kenya   11      15      13      39
Kuwait  0       0       2       2
Kyrgyzstan      0       1       2       3
Latvia  3       9       5       17
Lithuania       5       5       20      30
Macedonia       0       0       1       1
Malaysia        0       2       1       3
Mauritius       0       0       1       1
Mexico  19      10      9       38
Moldova 0       1       4       5
Mongolia        2       4       4       10
Montenegro      0       14      0       14
Morocco 2       3       6       11
Mozambique      1       0       0       1
Netherlands     101     135     82      318
New Zealand     18      7       27      52
Nigeria 6       18      15      39
North Korea     6       6       9       21
Norway  97      44      51      192
Panama  1       0       0       1
```

```
Paraguay        0       17      0       17
Poland  20      32      28      80
Portugal        1       5       3       9
Puerto Rico     0       1       1       2
Qatar   0       0       3       3
Romania 57      20      46      123
Russia  234     221     313     768
Saudi Arabia    0       1       5       6
Serbia  1       2       28      31
Serbia and Montenegro   11      14      13      38
Singapore       0       3       4       7
Slovakia        10      13      12      35
Slovenia        5       7       13      25
South Africa    10      8       7       25
South Korea     110     93      105     308
Spain   19      116     70      205
Sri Lanka       0       1       0       1
Sudan   0       1       0       1
Sweden  57      73      51      181
Switzerland     21      30      42      93
Syria   0       0       1       1
Tajikistan      0       1       2       3
Thailand        6       5       7       18
Togo    0       0       1       1
Trinidad and Tobago     1       7       11      19
Tunisia 2       1       1       4
Turkey  9       9       10      28
Uganda  1       0       0       1
Ukraine 31      38      74      143
United Arab Emirates    1       0       0       1
United States   552     440     320     1312
Uruguay 0       1       0       1
Uzbekistan      5       4       10      19
Venezuela       1       0       3       4
Vietnam 0       2       0       2
Zimbabwe        2       4       1       7
Time taken: 113.98 seconds, Fetched: 110 row(s)
hive (OLYMPICS)>
```

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

Below is the query used to find the total number of GOLD medals each country has won-

```
> SELECT country,
 > SUM(gold_medal) as TOTAL_GOLD
 > FROM OLYMPIC
 > GROUP BY country;
```

```
hive (OLYMPICS)> SELECT country,
               > SUM(gold_medal) as TOTAL_GOLD
               > FROM OLYMPIC
               > GROUP BY country;
```

Below is the output of above query which shows country name along with count of gold medals won by them -

```
OK
country total_gold
Afghanistan     0
Algeria 2
Argentina       49
Armenia 0
Australia       163
Austria 36
Azerbaijan      6
Bahamas 11
Bahrain 0
Barbados        0
Belarus 17
Belgium 2
Botswana        0
Brazil  46
Bulgaria        8
Cameroon        20
Canada  168
Chile   3
China   234
Chinese Taipei  2
Colombia        2
Costa Rica      0
Croatia 35
Cuba    57
Cyprus  0
Czech Republic  14
Denmark 46
Dominican Republic      3
Ecuador 0
Egypt   1
Eritrea 0
Estonia 6
Ethiopia        13
Finland 11
France  108
Gabon   0
```

```
Georgia 6
Germany 223
Great Britain   124
Greece  12
Grenada 1
Guatemala       0
Hong Kong       0
Hungary 77
Iceland 0
India   1
Indonesia       5
Iran    10
Ireland 1
Israel  1
Italy   86
Jamaica 24
Japan   57
Kazakhstan      13
Kenya   11
Kuwait  0
Kyrgyzstan      0
Latvia  3
Lithuania       5
Macedonia       0
Malaysia        0
Mauritius       0
Mexico  19
Moldova 0
Mongolia        2
Montenegro      0
Morocco 2
Mozambique      1
Netherlands     101
New Zealand     18
Nigeria 6
North Korea     6
Norway  97
Panama  1
```

```
Paraguay        0
Poland  20
Portugal        1
Puerto Rico     0
Qatar   0
Romania 57
Russia  234
Saudi Arabia    0
Serbia  1
Serbia and Montenegro   11
Singapore       0
Slovakia        10
Slovenia        5
South Africa    10
South Korea     110
Spain   19
Sri Lanka       0
Sudan   0
Sweden  57
Switzerland     21
Syria   0
Tajikistan      0
Thailand        6
Togo    0
Trinidad and Tobago     1
Tunisia 2
Turkey  9
Uganda  1
Ukraine 31
United Arab Emirates    1
United States   552
Uruguay 0
Uzbekistan      5
Venezuela       1
Vietnam 0
Zimbabwe        2
Time taken: 104.606 seconds, Fetched: 110 row(s)
hive (OLYMPICS)>
```

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.

Below screenshot shows the creation of database and table-

```
hive> CREATE database employees;
OK
Time taken: 0.708 seconds
hive>
```

We are using similar file used as that in Task-1 but we are adding an extra column as “greet” . Below is the DDL query-

```
create table emp3
(
emp_id int,
emp_name string,
sal int,
dept int,
greet string
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
```

Below is the screenshot-

```
hive (employees)> create table emp3
                > (
                > emp_id int,
                > emp_name string,
                > sal int,
                > dept int,
                > greet string
                > )
                > ROW FORMAT DELIMITED
                > FIELDS TERMINATED BY ',';
OK
Time taken: 0.61 seconds
```

After that we loaded data in the table emp3 using below command-

- **LOAD DATA LOCAL INPATH ‘/home/acadgild/hve/employee_details2.txt’**
- **INTO TABLE employees.emp3;**

```
hive (employees)> LOAD DATA LOCAL INPATH '/home/acadgild/hive/employee_details2.txt'
                > INTO TABLE employees.emp3;
Loading data to table employees.emp3
Table employees.emp3 stats: [numFiles=1, totalSize=529]
OK
Time taken: 1.504 seconds
hive (employees)> select * from emp3;
OK
emp3.emp_id     emp3.emp_name   emp3.sal        emp3.dept       emp3.greet
101     Amitabh 20000   1       Hi This is Amitabh
102     Shahrukh        10000   2       Hi This is Shahrukh
103     Akshay  11000   3       Hi This is Akshay
104     Anubhav 5000    4       Hi This is Anubhav
105     Pawan   2500    5       Hi This is Pawan
106     Aamir   25000   1       Hi This is Aamir
107     Salman  17500   2       Hi This is Salman
108     Ranbir  14000   3       Hi This is Ranbir
109     Katrina 1000    4       Hi This is Katrina
110     Priyanka        2000    5       Hi This is Priyanka
111     Tushar  500     1       Hi This is Tushar
112     Ajay    5000    2       Hi This is Ajay
113     Jubeen  1000    1       Hi This is Jubeen
114     Madhuri 2000    2       Hi This is Madhuri
```

After that we are adding JAR created from the JAVA class which is defining the UDF using below syntax-

- **ADD JAR /home/acadgild/hive/hive-task2.jar;**

After that we are creating a temporary function “conct” using below syntax-

- **CREATE TEMPORARY FUNCTION conct AS ‘udf.ConcatStr’;**

After that we run below query to take one column (NAME) input as String and another column(greet) as Array of Strings and concatenate them-

- **SELECT emp_id, sal, dept, conct(emp_name, greet) FROM emp3;**

Below is the screenshot for the same-

```
hive (employees)> ADD JAR /home/acadgild/hive/hive-task2.jar;
Added [/home/acadgild/hive/hive-task2.jar] to class path
Added resources: [/home/acadgild/hive/hive-task2.jar]
hive (employees)> CREATE TEMPORARY FUNCTION conct AS 'udf.ConcatStr';
OK
Time taken: 0.075 seconds
hive (employees)> SELECT emp_id, sal, dept, conct(emp_name, greet) FROM emp3;
OK
emp_id  sal     dept    _c3
101     20000   1       AmitabhHi This is Amitabh
102     10000   2       ShahrukhHi This is Shahrukh
103     11000   3       AkshayHi This is Akshay
104     5000    4       AnubhavHi This is Anubhav
105     2500    5       PawanHi This is Pawan
106     25000   1       AamirHi This is Aamir
107     17500   2       SalmanHi This is Salman
108     14000   3       RanbirHi This is Ranbir
109     1000    4       KatrinaHi This is Katrina
110     2000    5       PriyankaHi This is Priyanka
111     500     1       TusharHi This is Tushar
112     5000    2       AjayHi This is Ajay
113     1000    1       JubeenHi This is Jubeen
114     2000    2       MadhuriHi This is Madhuri
Time taken: 0.283 seconds, Fetched: 14 row(s)
hive (employees)>
```