Assignment 7-2

Creating table and loading data into the table:-

hive> create table if not exists athlete_details
(athlete_name string, age int, country string, year int, closing_date string, sport string, gold_medals int, silver_medals int, bronze_medals int, total_medals int)
row format delimited fields terminated by '\t';

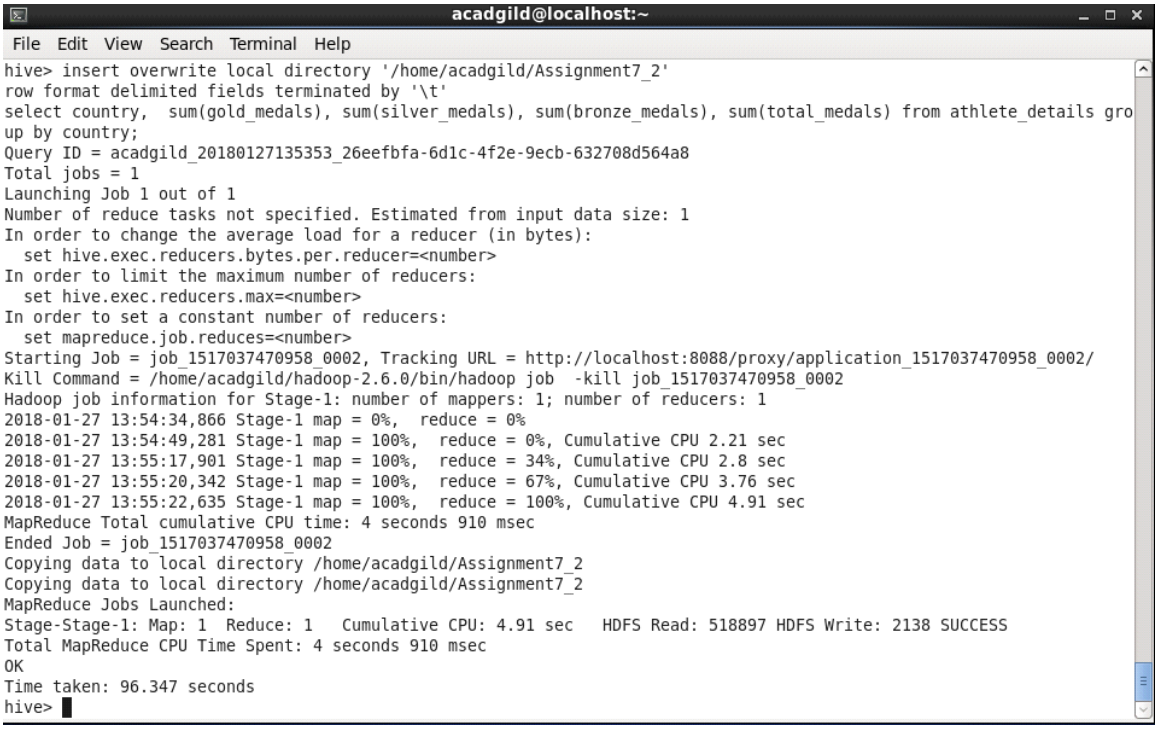
hive> load data local inpath '/home/acadgild/Assignment7-2/olympix_data.csv' into table athlete_details;

```
hive>  load data local inpath '/home/acadgild/Assignment7-2/olympix_data.csv' into table athlete_details;
Loading data to table default.athlete_details
Table default.athlete_details stats: [numFiles=1, totalSize=518669]
OK
Time taken: 2.815 seconds
hive>
```

```
hive> create table if not exists athlete_details
( athlete_name string, age int, country string, year int, closing_date string, s
port string, gold_medals int,
silver_medals int, bronze_medals int, total_medals int)
row format delimited fields terminated by '\t';
OK
Time taken: 1.638 seconds
hive>
```

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

hive> insert overwrite local directory '/home/acadgild/Assignment7_2'
row format delimited fields terminated by '\t'
select country, sum(gold_medals), sum(silver_medals), sum(bronze_medals), sum(total_medals) from athlete_details group by country;



Output file
https://drive.google.com/file/d/1-u2dlNtAnbXDf2CLZ_5CYho9MhEriREs/view?usp=sharing

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

Hive> insert overwrite local directory '/home/acadgild/Acadgild/Assignment7_2'
row format delimited fields terminated by '\t'
select year, country, sum(gold_medals), sum(silver_medals), sum(bronze_medals), sum(total_medals) from athlete_details where country = "India" group by year, country;

```
hive> insert overwrite local directory '/home/acadgild/Assignment7_2'
row format delimited fields terminated by '\t'
select year, country,  sum(gold_medals), sum(silver_medals), sum(bronze_medals), sum(total_medals) from athlete_detai
ls  where country = "India" group by year, country;
Query ID = acadgild_20180127135858_cae62471-7ac9-4f83-9c9f-26a0ecfee29c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1517037470958_0003, Tracking URL = http://localhost:8088/proxy/application_1517037470958_0003/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1517037470958_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-01-27 13:58:55,573 Stage-1 map = 0%,  reduce = 0%
2018-01-27 13:59:15,344 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.21 sec
2018-01-27 13:59:38,695 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.08 sec
MapReduce Total cumulative CPU time: 6 seconds 80 msec
Ended Job = job_1517037470958_0003
Copying data to local directory /home/acadgild/Assignment7_2
Copying data to local directory /home/acadgild/Assignment7_2
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.08 sec   HDFS Read: 518897 HDFS Write: 76 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 80 msec
OK
Time taken: 66.884 seconds
hive>
```

Output file -
https://drive.google.com/file/d/1IgAe9-BKIDKsYq9TsJKkkOlq5FLVP0nH/view?usp=sharing

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

Hive> insert overwrite local directory '/home/acadgild/Assignment7-2'
row format delimited fields terminated by '\t'
select country, sum(total_medals) from athlete_details group by country;

```
hive> insert overwrite local directory '/home/acadgild/Assignment7_2'
row format delimited fields terminated by '\t'
select country,  sum(total_medals) from athlete_details group by country;
Query ID = acadgild_20180127140202_2332c606-c1d2-4556-9521-f84e1958a473
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1517037470958_0004, Tracking URL = http://localhost:8088/proxy/application_1517037470958_0004/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1517037470958_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-01-27 14:02:21,459 Stage-1 map = 0%,  reduce = 0%
2018-01-27 14:02:36,823 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.19 sec
2018-01-27 14:02:57,533 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.99 sec
MapReduce Total cumulative CPU time: 4 seconds 990 msec
Ended Job = job_1517037470958_0004
Copying data to local directory /home/acadgild/Assignment7_2
Copying data to local directory /home/acadgild/Assignment7_2
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.99 sec   HDFS Read: 518897 HDFS Write: 1315 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 990 msec
OK
Time taken: 57.861 seconds
hive>
```

Output file -
https://drive.google.com/file/d/1ZfhDkv95slNGH3veufuXX0qujb8GXn0R/view?usp=sharing

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

Hive> insert overwrite local directory '/home/acadgild/Assignment7-2'
row format delimited fields terminated by '\t'
select country, sum(gold_medals) from athlete_details group by country;

```
hive> insert overwrite local directory '/home/acadgild/Assignment7_2'
row format delimited fields terminated by '\t'
select country,  sum(gold_medals) from athlete_details group by country;
Query ID = acadgild_20180127140505_eeed5df4-b533-4cf7-980d-5179bcc46e88
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1517037470958_0005, Tracking URL = http://localhost:8088/proxy/application_1517037470958_0005/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1517037470958_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-01-27 14:05:22,866 Stage-1 map = 0%,  reduce = 0%
2018-01-27 14:05:36,480 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.25 sec
2018-01-27 14:05:50,014 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.68 sec
MapReduce Total cumulative CPU time: 4 seconds 680 msec
Ended Job = job_1517037470958_0005
Copying data to local directory /home/acadgild/Assignment7_2
Copying data to local directory /home/acadgild/Assignment7_2
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.68 sec   HDFS Read: 518897 HDFS Write: 1276 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 680 msec
OK
Time taken: 46.426 seconds
hive>
```

Output file –

https://drive.google.com/file/d/1gzm3S2LEf_lRA4QLCsLVFlePIFddCVtL/view?usp=sharing