

This Data set is about Olympics. You can download the data set from the below link:

<https://drive.google.com/open?id=0ByJLBTmJojjzV1czX3Nha0R3bTQ>

DATE SET DESCRIPTION

The data set consists of the following fields.

Athlete: This field consists of the athlete name

Age: This field consists of athlete ages

Country: This field consists of the country names which participated in Olympics

Year: This field consists of the year

Closing Date: This field consists of the closing date of ceremony

Sport: Consists of the sports name

Gold Medals: No. of Gold medals

Silver Medals: No. of Silver medals

Bronze Medals: No. of Bronze medals

Total Medals: Consists of total no. of medals

Task 1

1. Write a Hive program to find the number of medals won by each country in swimming.
2. Write a Hive program to find the number of medals that India won year wise.
3. Write a Hive Program to find the total number of medals each country won.
4. Write a Hive program to find the number of gold medals each country won.

create table Olympics(Athlete string, Age int, Country string, Year int, ClosingDate string, Sport string, GoldMedals int, SilverMedals int, BronzeMedals int, TotalMedals int) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;

```
hive> create table Olympics(Athlete string, Age int, Country string, Year int, ClosingDate string, Sport string, GoldMedals int, SilverMedals int, BronzeMedals int, TotalMedals int) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' STORED AS TEXTFILE;
OK
Time taken: 1.042 seconds
```

LOAD DATA LOCAL INPATH '/home/acadgild/olympix_data.csv' OVERWRITE INTO TABLE olympics;

```
hive> LOAD DATA LOCAL INPATH '/home/acadgild/olympix_data.csv' OVERWRITE INTO TABLE olympics;
Loading data to table default.olympics
OK
Time taken: 2.76 seconds
```

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

select country, sport, sum(TotalMedals) from olympics where sport="Swimming" group by country, sport;

```
hive> select country, sport, sum(TotalMedals) from olympics where sport="Swimming" group by country, sport;
```

OK		
Argentina	Swimming	1
Australia	Swimming	163
Austria	Swimming	3
Belarus	Swimming	2
Brazil	Swimming	8
Canada	Swimming	5
China	Swimming	35
Costa Rica	Swimming	2
Croatia	Swimming	1
Denmark	Swimming	1
France	Swimming	39
Germany	Swimming	32
Great Britain	Swimming	11
Hungary	Swimming	9
Italy	Swimming	16
Japan	Swimming	43
Lithuania	Swimming	1
Netherlands	Swimming	46
Norway	Swimming	2
Poland	Swimming	3
Romania	Swimming	6
Russia	Swimming	20
Serbia	Swimming	1
Slovakia	Swimming	2
Slovenia	Swimming	1
South Africa	Swimming	11
South Korea	Swimming	4
Spain	Swimming	3
Sweden	Swimming	9
Trinidad and Tobago	Swimming	1
Tunisia	Swimming	3
Ukraine	Swimming	7
United States	Swimming	267
Zimbabwe	Swimming	7

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

```
select year,sum(TotalMedals) from olympics where country="India" group by year order by year desc;
```

```
hive> select year,sum(TotalMedals) from olympics where country="India" group by year order by year desc;
```

```
2012    6
2008    3
2004    1
2000    1
```

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

```
select country,sum(TotalMedals) from olympics group by country;
```

```
Time taken: 143.733 seconds, Fetched: 4 row(s)  
hive> select country,sum(TotalMedals) from olympics group by country;
```

Afghanistan	2
Algeria	8
Argentina	141
Armenia	10
Australia	609
Austria	91
Azerbaijan	25
Bahamas	24
Bahrain	1
Barbados	1
Belarus	97
Belgium	18
Botswana	1
Brazil	221
Bulgaria	41
Cameroon	20
Canada	370
Chile	22
China	530
Chinese Taipei	20
Colombia	13
Costa Rica	2
Croatia	81
Cuba	188
Cyprus	1
Czech Republic	81
Denmark	89

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

```
select country,sum(GoldMedals) from olympics group by country;
```

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	