

Olympic Data analysis using Hive

DATA SET:

<https://drive.google.com/file/d/0B1QaXx7tpw3SaEE3bEFTQTMzNzg/view?usp=sharing>

Olympic Data Column Description:

ColumnNO.	Name	Example	DataType
Column1:	Athlete Name	Michael Phelps	STRING
Column2:	Age	23	INT
Column3:	Country	United States	STRING
Column4:	Year	2008	INT
Column5:	Closing Date	8/24/2008	STRING
Column6:	Sport	Swimming	STRING
Column7:	Gold Medals	8	INT
Column8:	Silver Medals	0	INT
Column9:	Bronze Medals	0	INT
Column10:	Total Medals	8	INT

```
Hive> CREATE DATABASE IF NOT EXISTS olympicdb;
```

```
Hive> USE olympicdb;
```

```
Hive> CREATE TABLE IF NOT EXISTS olympic( athletename string, age int, country  
string, year int, closing_date string,sport string, gold_medal int, silver_medal int,  
bronze_medal int, total_medals int)
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY '\t'
```

```
LINES TERMINATED BY '\n'
```

```
STORED AS textfile;
```

```
hive> load data local inpath '/home/cloudera/Downloads/olympic_data.csv' into table  
olympic;
```

Problem Statement 1:

Using the dataset list the total number of medals won by each country in swimming.

```
SELECT country, sum(total_medals) from olympic where sport = 'Swimming' group by  
country;
```

Problem Statement 2:

Display real life number of medals India won year wise.

```
SELECT year, sum(total_medals) from olympic where country == 'India' GROUP by  
year;
```

Problem Statement 3:

Find the total number of medals each country won display the name along with total medals.

```
SELECT country, sum(total_medals) from Olympic group by country;
```

Problem Statement 4:

Find the real life number of gold medals each country won.

```
SELECT country, sum(gold_medal) from olympic group by country;
```

Problem Statement 5:

Which country got medals for Shooting, year wise classification?

```
select country, year, sum(total_medals) from olympic where sport == 'Shooting' group  
by country, year;
```