**ICP-4**

**Problem Statement 1-**

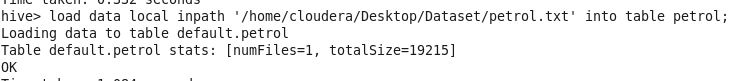
Create Hive Tables and Perform Queries for Use Case based on Petrol Data.

**Create table query -**

Create table petrol(distributer\_id string, distributer\_name string, amt\_in string, amt\_out string,vol\_in int, vol\_out int,year int) row format delimited fields terminated by ',' stored as textfile;

https://github.com/NikithaPateel/Big_Data_Programming/raw/master/ICP_4/Screenshots/petrol/create_table.JPG

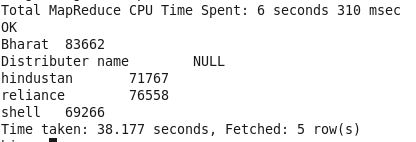
**Loading the data into the table**



**Query 1**- Total amount of petrol in volume sold by every distributor

Select distributer\_name ,sum(vol\_out) from petrol group by distributer\_name;

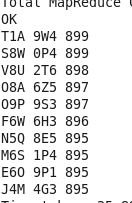
https://github.com/NikithaPateel/Big_Data_Programming/raw/master/ICP_4/Screenshots/petrol/query1.JPG



**Query 2** -Top 10 distributors ID’s for selling petrol and also display the amount of petrol sold in volume by them individually

Select distributer\_id ,vol\_out from petrol order by vol\_out desc limit 10;

https://github.com/NikithaPateel/Big_Data_Programming/raw/master/ICP_4/Screenshots/petrol/query_2.JPG



**Query 3 -** Top 10 distributor name who sold petrol in the least amount

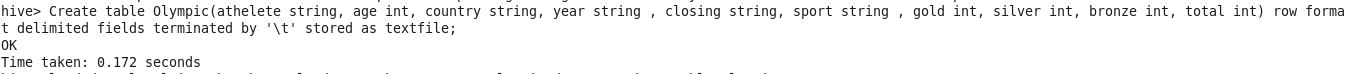
Select distributer\_name, vol\_out from petrol order by vol\_out limit 10;

**Problem Statement 2-**

Create Hive Tables and Perform Queries for Use Case based on Olympics Data.

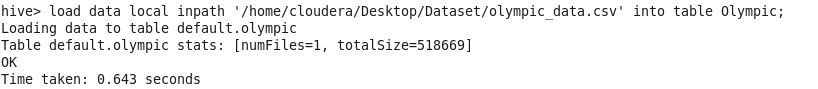
**Create table query:**

Create table Olympic(athelete string, age int, country string, year string , closing string, sport string , gold int, silver int, bronze int, total int) row format delimited fields terminated by '\t' stored as textfile;



**Loading Data into the table:**

load data local inpath '/home/cloudera/Desktop/Dataset/olympic\_data.csv' into table Olympic;



\*\*Query 1 \*\*- Total number of medals won by each country in swimming.

select country,sum(total) from Olympic where sport = "swimming" group by country;

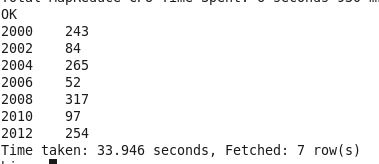
https://github.com/NikithaPateel/Big_Data_Programming/raw/master/ICP_4/Screenshots/olympics/query_1.JPG



\*\*Query 2 \*\*- Displaying number of medals United States won Year wise

Select year, sum(total) from Olympic where country="United States" group by year;

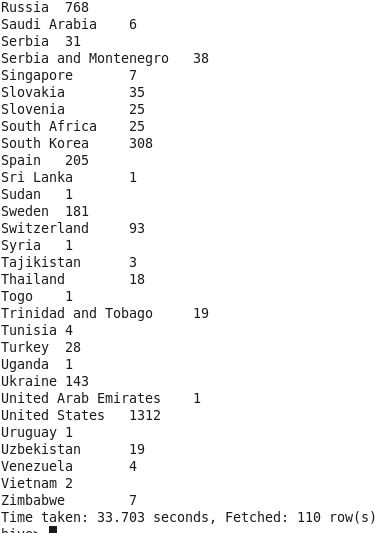
https://github.com/NikithaPateel/Big_Data_Programming/raw/master/ICP_4/Screenshots/olympics/query_2.JPG



**Query 3-** Total number of medals each country won display the name along with total medals.

Select country, sum(total) from Olympic group by country;

https://github.com/NikithaPateel/Big_Data_Programming/raw/master/ICP_4/Screenshots/olympics/query_3.JPG



**Query 4-** Total number of gold medals won by each country in Cycling

select country,sum(gold) from Olympic where sport = "Cycling" group by country;

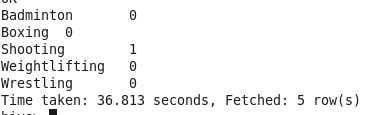
https://github.com/NikithaPateel/Big_Data_Programming/raw/master/ICP_4/Screenshots/olympics/query_4.JPG



**Query 5** - Number of gold medals won by India in different sports

select sport,sum(gold) from Olympic where country ="India" group by sport;

https://github.com/NikithaPateel/Big_Data_Programming/raw/master/ICP_4/Screenshots/olympics/query_5.JPG



**Limitation:**

* No easy way to append data
* Files in HDFS are immutable

**References**

<https://www.tutorialspoint.com/hive/>