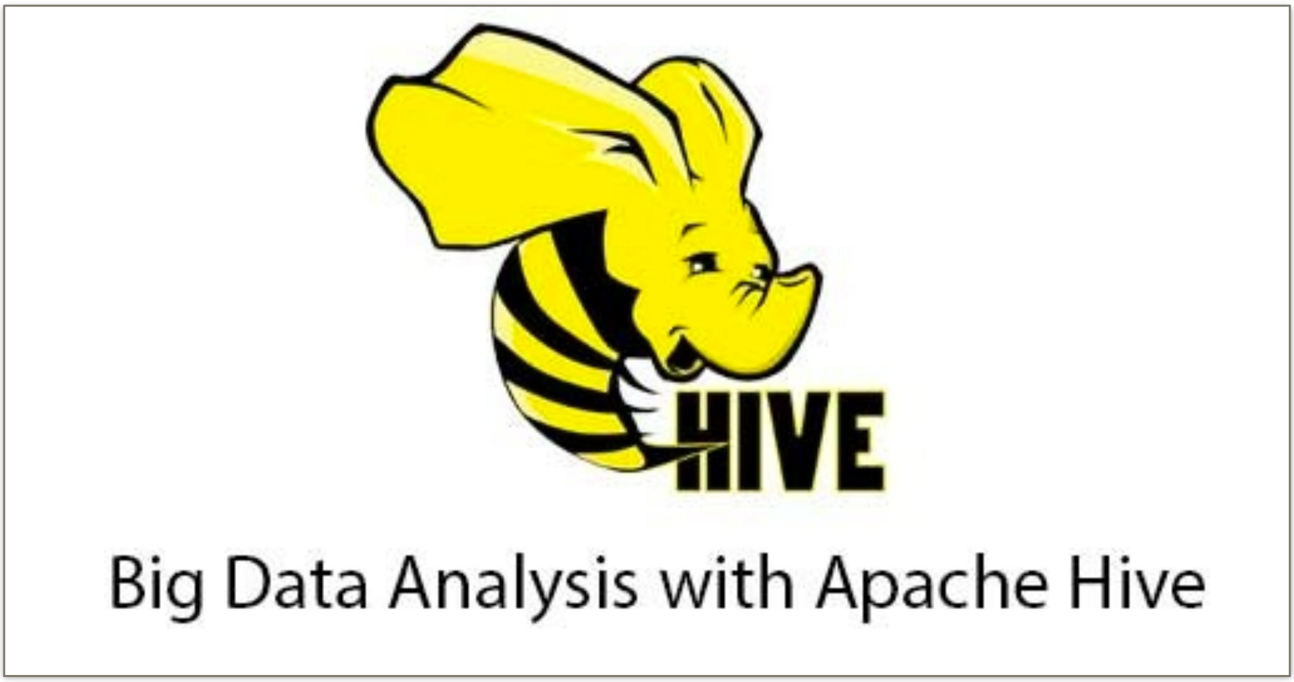


# HIVE ASSIGNMENT

# BIG DATA ANALYTICS

Performed on Red Hat (64 bit) CentOS 2019

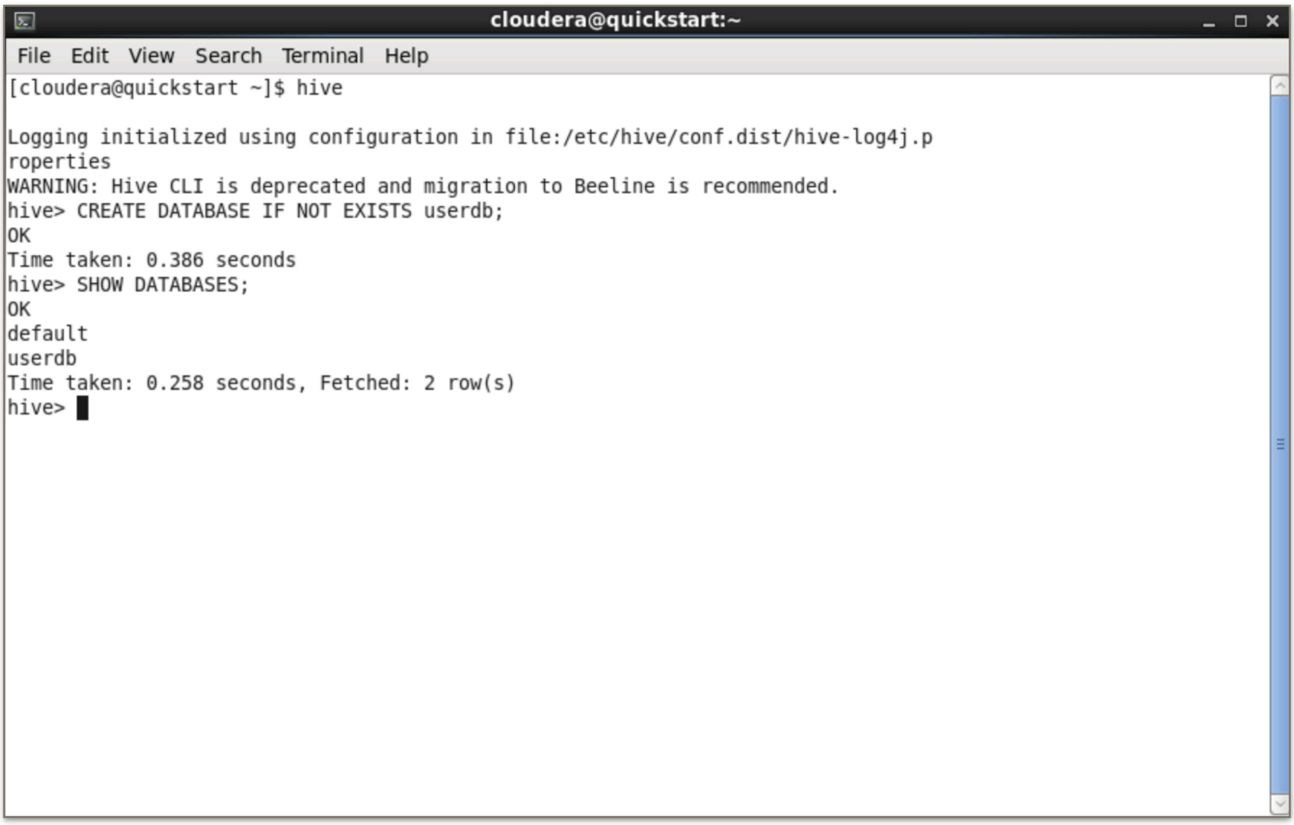
Submitted To: Prof. Aditya Bharadwaj
Big Data Analytics

Submitted By: SID - 16103104
Name - Lovedeep Singh
BTech- CSE 4th Year

Ensure that **HIVE** is setup in the linux system you are operating on.

**Note**: All these commands have been executed on **Red Hat (64 - bit) CentOS**.

Start the terminal and run ***hive*** command



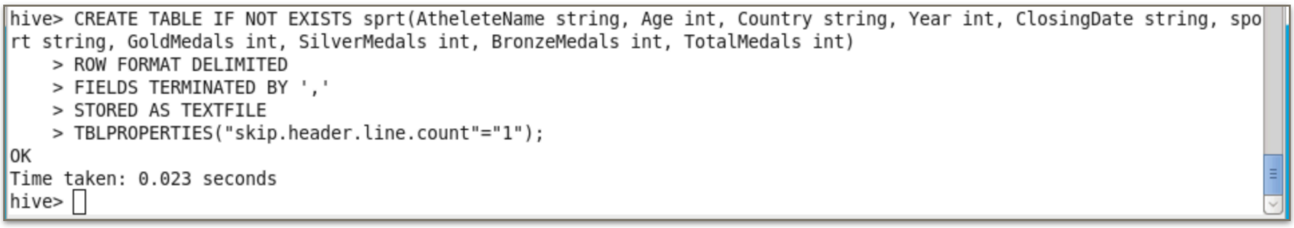
```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p
roperties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> CREATE DATABASE IF NOT EXISTS userdb;
OK
Time taken: 0.386 seconds
hive> SHOW DATABASES;
OK
default
userdb
Time taken: 0.258 seconds, Fetched: 2 row(s)
hive> 
```

## SPORTS DATA ANALYSIS

Before proceeding further, export the file sports.xslx in sports.csv format.

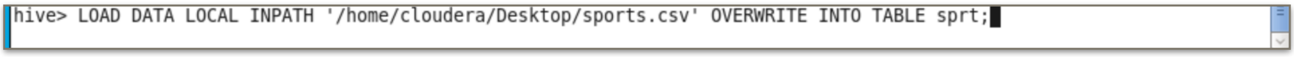
Q1. Creation of table in Hive and loading the data

*CREATE TABLE IF NOT EXISTS sprt(AtheleteName 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 ‘,’*
*STORED AS TEXTFILE*
*TBLPROPERTIES(“skip.header.line.count”=“1”);*



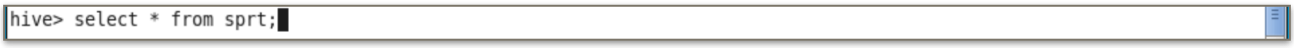
```
hive> CREATE TABLE IF NOT EXISTS sprt(AtheleteName string, Age int, Country string, Year int, ClosingDate string, spo
rt string, GoldMedals int, SilverMedals int, BronzeMedals int, TotalMedals int)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > STORED AS TEXTFILE
    > TBLPROPERTIES("skip.header.line.count"="1");
OK
Time taken: 0.023 seconds
hive> 
```

*LOAD DATA LOCAL INPATH ‘/home/cloudera/sports.csv’ OVERWRITE INTO TABLE sprt;*

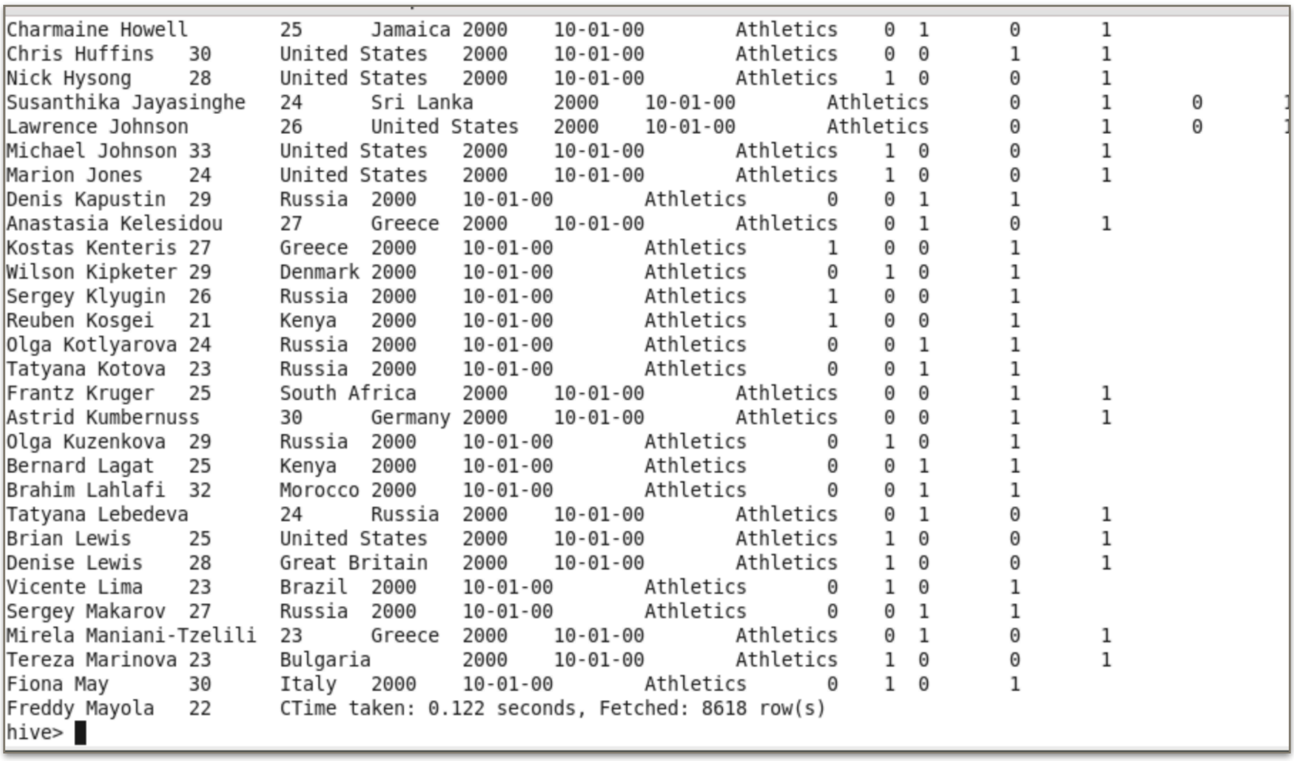


```
hive> LOAD DATA LOCAL INPATH '/home/cloudera/Desktop/sports.csv' OVERWRITE INTO TABLE sprt;
```

*SELECT * FROM sprt;*



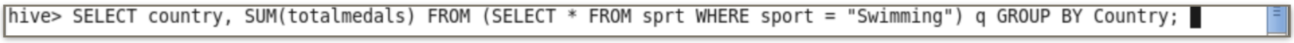
```
hive> select * from sprt;
```



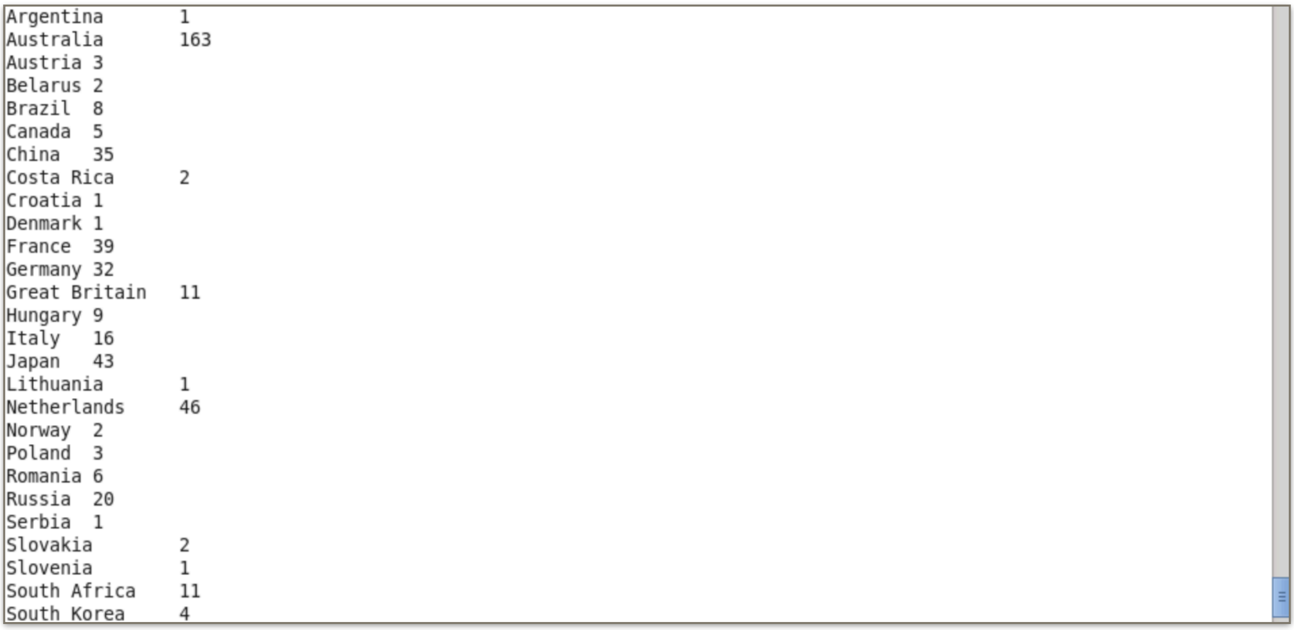
```
Charmaine Howell        25      Jamaica 2000    10-01-00        Athletics    0  1      0       1
Chris Huffins   30      United States   2000    10-01-00        Athletics    0  0      1       1
Nick Hysong     28      United States   2000    10-01-00        Athletics    1  0      0       1
Susanthika Jayasinghe   24      Sri Lanka       2000    10-01-00        Athletics      0       1       0       1
Lawrence Johnson        26      United States   2000    10-01-00        Athletics      0       1       0       1
Michael Johnson 33      United States   2000    10-01-00        Athletics    1  0      0       1
Marion Jones    24      United States   2000    10-01-00        Athletics    1  0      0       1
Denis Kapustin  29      Russia  2000    10-01-00        Athletics    0  0  1      1
Anastasia Kelesidou     27      Greece  2000    10-01-00        Athletics    0  1      0       1
Kostas Kenteris 27      Greece  2000    10-01-00        Athletics    1  0  0      1
Wilson Kipketer 29      Denmark 2000    10-01-00        Athletics    0  1  0      1
Sergey Klyugin  26      Russia  2000    10-01-00        Athletics    1  0  0      1
Reuben Kosgei   21      Kenya   2000    10-01-00        Athletics    1  0  0      1
Olga Kotlyarova 24      Russia  2000    10-01-00        Athletics    0  0  1      1
Tatyana Kotova  23      Russia  2000    10-01-00        Athletics    0  0  1      1
Frantz Kruger   25      South Africa    2000    10-01-00        Athletics    0  0      1       1
Astrid Kumbernuss       30      Germany 2000    10-01-00        Athletics    0  0      1       1
Olga Kuzenkova  29      Russia  2000    10-01-00        Athletics    0  1  0      1
Bernard Lagat   25      Kenya   2000    10-01-00        Athletics    0  0  1      1
Brahim Lahlafi  32      Morocco 2000    10-01-00        Athletics    0  0  1      1
Tatyana Lebedeva        24      Russia  2000    10-01-00        Athletics    0  1      0       1
Brian Lewis     25      United States   2000    10-01-00        Athletics    1  0      0       1
Denise Lewis    28      Great Britain   2000    10-01-00        Athletics    1  0      0       1
Vicente Lima    23      Brazil  2000    10-01-00        Athletics    0  1  0      1
Sergey Makarov  27      Russia  2000    10-01-00        Athletics    0  0  1      1
Mirela Maniani-Tzelili  23      Greece  2000    10-01-00        Athletics    0  1      0       1
Tereza Marinova 23      Bulgaria        2000    10-01-00        Athletics    1  0      0       1
Fiona May       30      Italy   2000    10-01-00        Athletics    0  1  0      1
Freddy Mayola   22      CTime taken: 0.122 seconds, Fetched: 8618 row(s)
hive> 
```

Q2. To list the total number of medals won by each country in swimming.

*SELECT country, SUM(totalmedals) FROM (SELECT * FROM sprt WHERE sport = "Swimming") q GROUP BY country;*



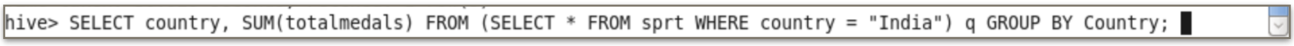
```
hive> SELECT country, SUM(totalmedals) FROM (SELECT * FROM sprt WHERE sport = "Swimming") q GROUP BY Country;
```



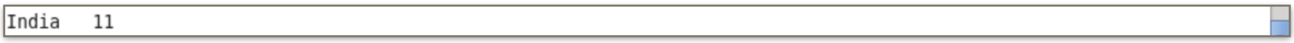
```
Argentina       1
Australia       163
Austria 3
Belarus 2
Brazil  8
Canada  5
China   35
Costa Rica      2
Croatia 1
Denmark 1
France  39
Germany 32
Great Britain   11
Hungary 9
Italy   16
Japan   43
Lithuania       1
Netherlands     46
Norway  2
Poland  3
Romania 6
Russia  20
Serbia  1
Slovakia        2
Slovenia        1
South Africa    11
South Korea     4
```

Q3. To list the total number of Gold medals won by India.

*SELECT country, SUM(totalmedals) FROM (SELECT * FROM sprt WHERE country = "India") q GROUP BY country;*



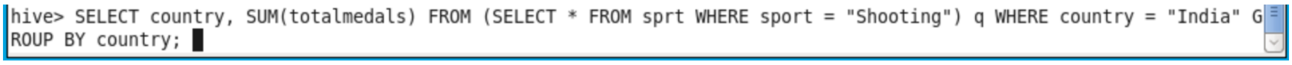
```
hive> SELECT country, SUM(totalmedals) FROM (SELECT * FROM sprt WHERE country = "India") q GROUP BY Country;
```



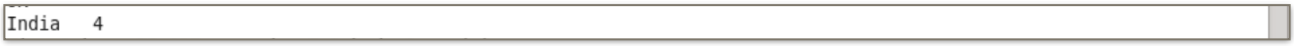
```
India   11
```

Q4 To list the number of medals won by India in Shooting.

*SELECT country, SUM(totalmedals) FROM (SELECT * FROM sprt WHERE sport = "Shooting") q WHERE country = "India" GROUP BY country;*



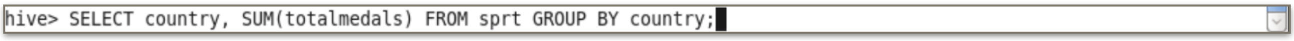
```
hive> SELECT country, SUM(totalmedals) FROM (SELECT * FROM sprt WHERE sport = "Shooting") q WHERE country = "India" G
ROUP BY country;
```



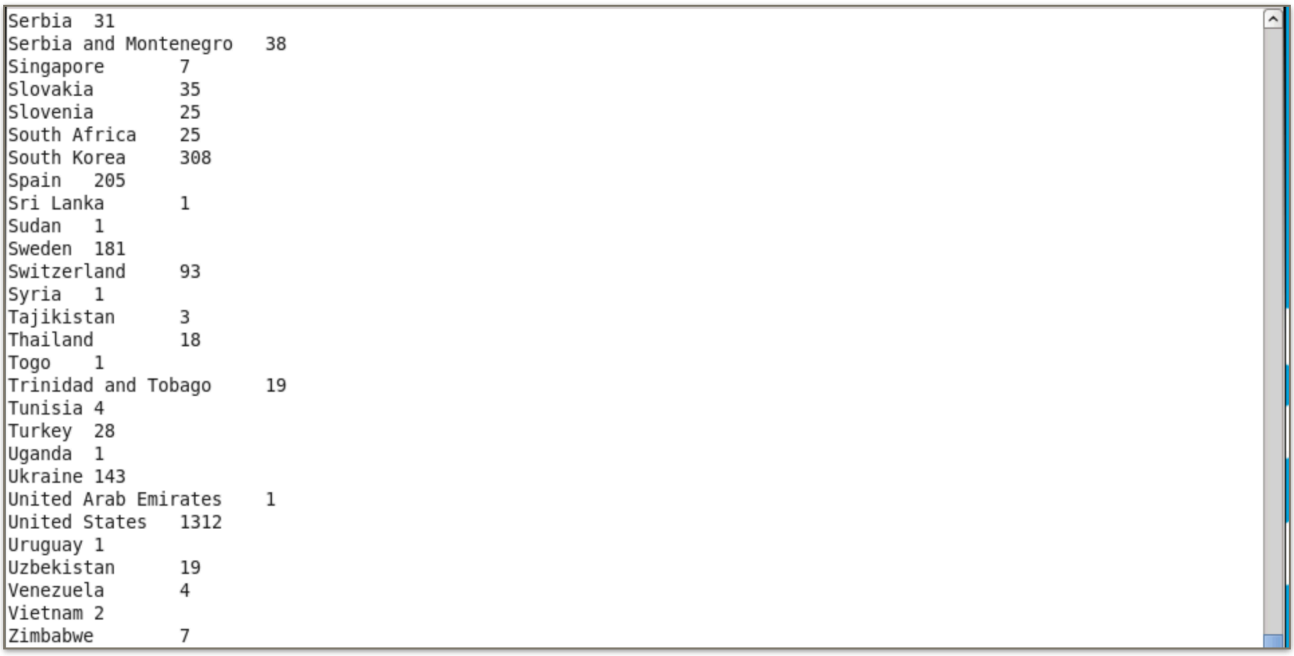
```
India   4
```

Q5. Find the total number of medals each country won display the name along with total medals.

*SELECT country, SUM(totalmedals) FROM sprt GROUP BY country;*



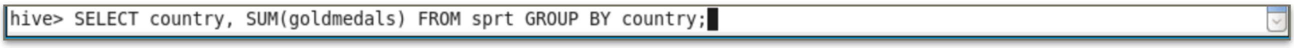
```
hive> SELECT country, SUM(totalmedals) FROM sprt GROUP BY country;
```



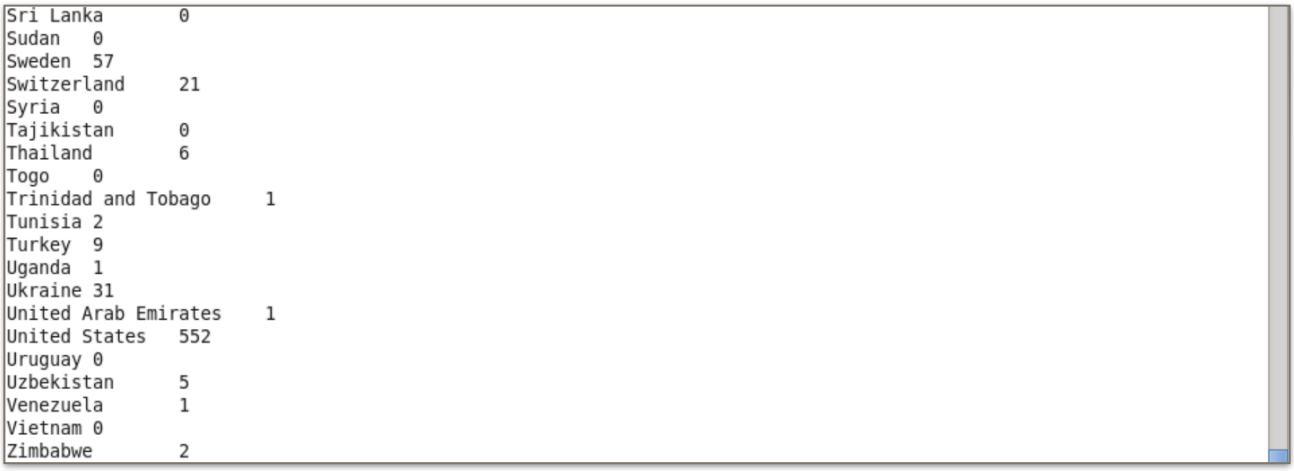
```
Serbia  31
Serbia and Montenegro   38
Singapore       7
Slovakia        35
Slovenia        25
South Africa    25
South Korea     308
Spain   205
Sri Lanka       1
Sudan   1
Sweden  181
Switzerland     93
Syria   1
Tajikistan      3
Thailand        18
Togo    1
Trinidad and Tobago     19
Tunisia 4
Turkey  28
Uganda  1
Ukraine 143
United Arab Emirates    1
United States   1312
Uruguay 1
Uzbekistan      19
Venezuela       4
Vietnam 2
Zimbabwe        7
```

Q6. List the data for number of gold medals each country won.

*SELECT country, SUM(goldmedals) FROM sprt GROUP BY country;*



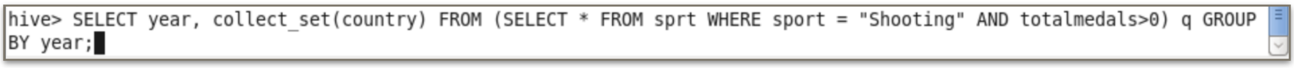
```
hive> SELECT country, SUM(goldmedals) FROM sprt GROUP BY country;
```



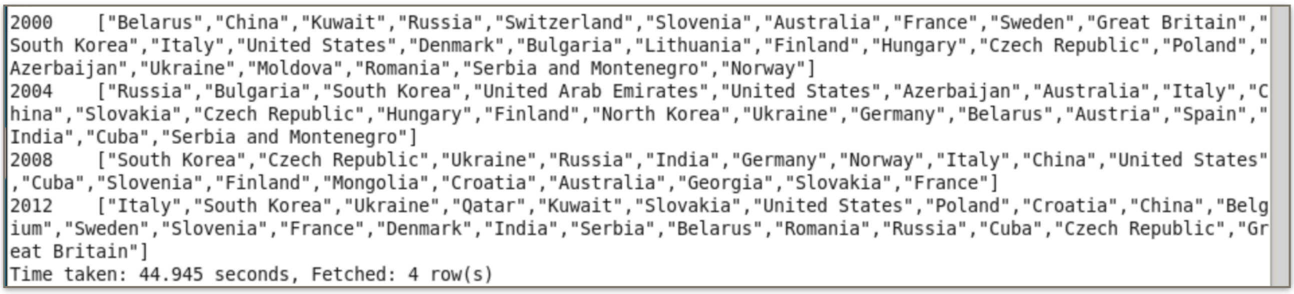
```
Sri Lanka       0
Sudan   0
Sweden  57
Switzerland     21
Syria   0
Tajikistan      0
Thailand        6
Togo    0
Trinidad and Tobago     1
Tunisia 2
Turkey  9
Uganda  1
Ukraine 31
United Arab Emirates    1
United States   552
Uruguay 0
Uzbekistan      5
Venezuela       1
Vietnam 0
Zimbabwe        2
```

Q7. Which country got medals for Shooting, year wise classification?

SELECT year, collect_set(country) FROM (SELECT * FROM sprt WHERE sport = “Shooting” AND totlamedals>0) q GROUP BY year;



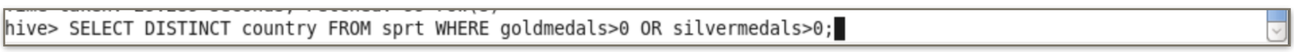
```
hive> SELECT year, collect_set(country) FROM (SELECT * FROM sprt WHERE sport = "Shooting" AND totalmedals>0) q GROUP
BY year;
```



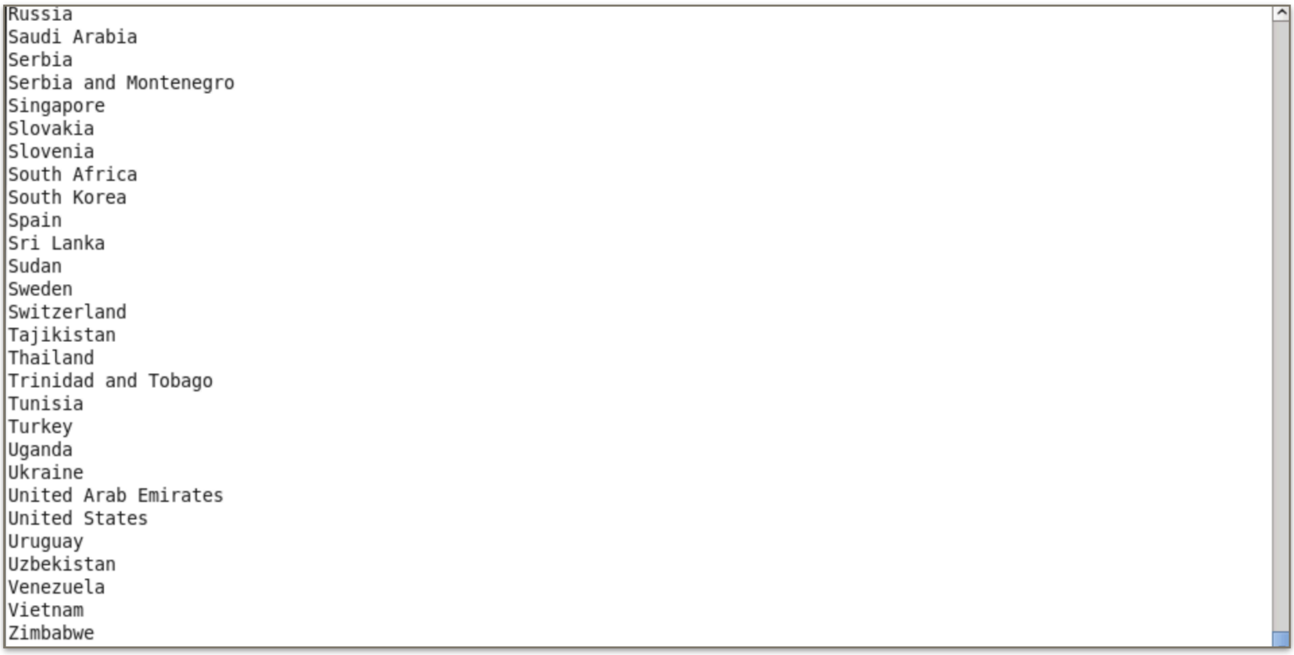
```
2000    ["Belarus","China","Kuwait","Russia","Switzerland","Slovenia","Australia","France","Sweden","Great Britain","
South Korea","Italy","United States","Denmark","Bulgaria","Lithuania","Finland","Hungary","Czech Republic","Poland","
Azerbaijan","Ukraine","Moldova","Romania","Serbia and Montenegro","Norway"]
2004    ["Russia","Bulgaria","South Korea","United Arab Emirates","United States","Azerbaijan","Australia","Italy","C
hina","Slovakia","Czech Republic","Hungary","Finland","North Korea","Ukraine","Germany","Belarus","Austria","Spain","
India","Cuba","Serbia and Montenegro"]
2008    ["South Korea","Czech Republic","Ukraine","Russia","India","Germany","Norway","Italy","China","United States"
,"Cuba","Slovenia","Finland","Mongolia","Croatia","Australia","Georgia","Slovakia","France"]
2012    ["Italy","South Korea","Ukraine","Qatar","Kuwait","Slovakia","United States","Poland","Croatia","China","Belg
ium","Sweden","Slovenia","France","Denmark","India","Serbia","Belarus","Romania","Russia","Cuba","Czech Republic","Gr
eat Britain"]
Time taken: 44.945 seconds, Fetched: 4 row(s)
```

Q8. To list the country that won gold and silver medals in Football.

SELECT DISTINCT country FROM sprt WHERE goldmedals > 0 OR silvermedals > 0;



```
hive> SELECT DISTINCT country FROM sprt WHERE goldmedals>0 OR silvermedals>0;
```

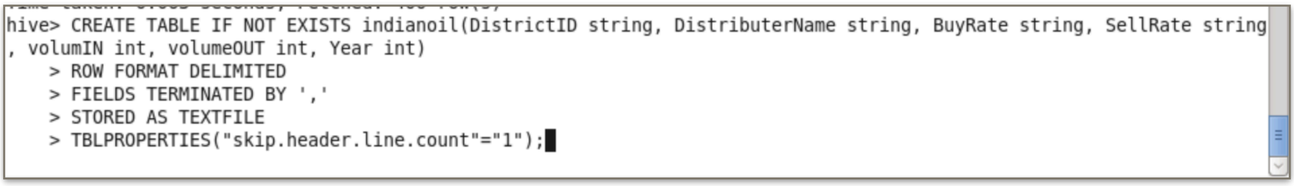


```
Russia
Saudi Arabia
Serbia
Serbia and Montenegro
Singapore
Slovakia
Slovenia
South Africa
South Korea
Spain
Sri Lanka
Sudan
Sweden
Switzerland
Tajikistan
Thailand
Trinidad and Tobago
Tunisia
Turkey
Uganda
Ukraine
United Arab Emirates
United States
Uruguay
Uzbekistan
Venezuela
Vietnam
Zimbabwe
```

# INDIAN OIL COMPANY ANALYSIS

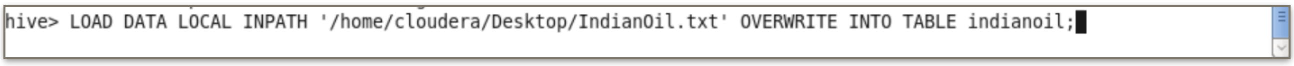
Q1. Creation of table in Hive and loading the data

*CREATE TABLE IF NOT EXISTS indianoil(DistrictID string, DistributerName string, BuyRate string, SellRate string, volumIN int, volumeOUT int, Year int)*
*ROW FORMAT DELIMITED*
*FIELDS TERMINATED BY ‘,’*
*STORED AS TEXTFILE*
*TBLPROPERTIES(“skip.header.line.count”=“1”);*



```
hive> CREATE TABLE IF NOT EXISTS indianoil(DistrictID string, DistributerName string, BuyRate string, SellRate string
, volumIN int, volumeOUT int, Year int)
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > STORED AS TEXTFILE
    > TBLPROPERTIES("skip.header.line.count"="1");
```

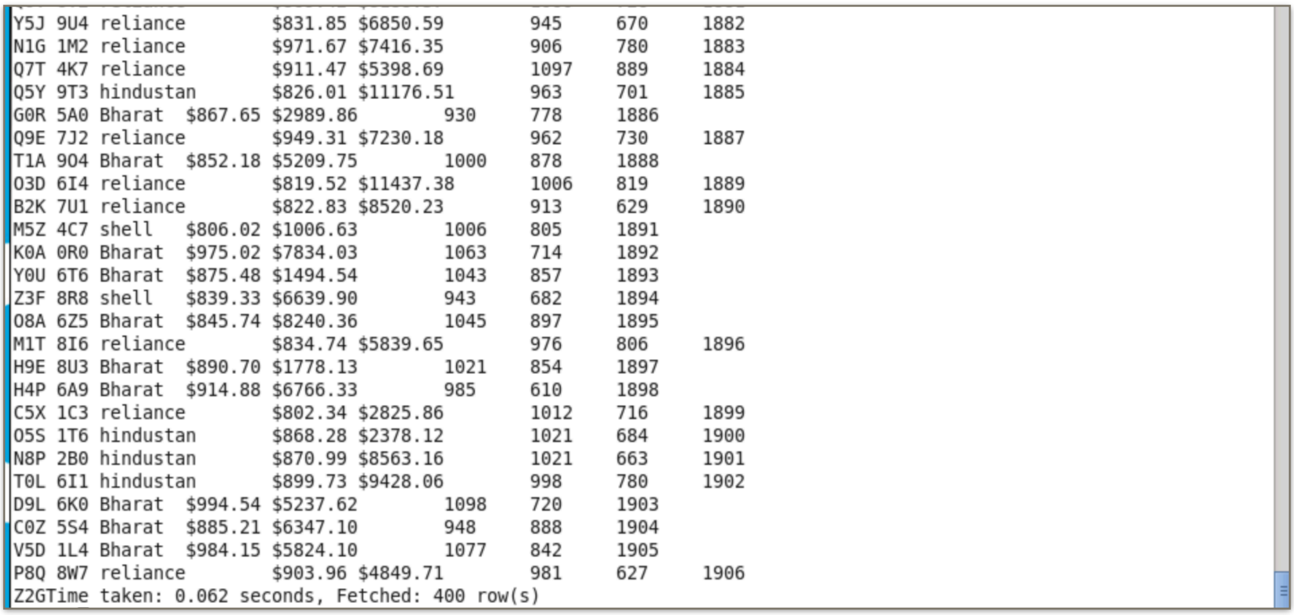
*LOAD DATA LOCAL INPATH ‘/home/cloudera/Desktop/IndinaOil.txt’ OVERWRITE INTO TABLE indianoil;*



```
hive> LOAD DATA LOCAL INPATH '/home/cloudera/Desktop/IndianOil.txt' OVERWRITE INTO TABLE indianoil;
```

*SELECT * FROM indianoil;*

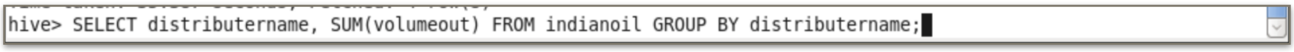
```
hive> SELECT * FROM indianoil;
```



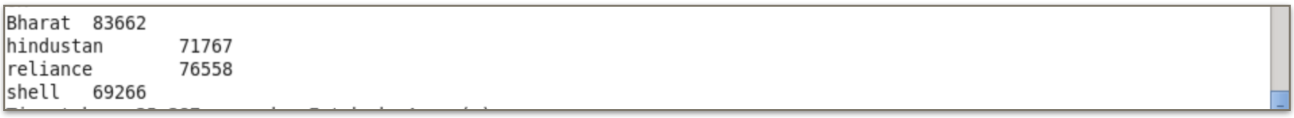
```
Y5J 9U4 reliance        $831.85 $6850.59        945     670     1882
N1G 1M2 reliance        $971.67 $7416.35        906     780     1883
Q7T 4K7 reliance        $911.47 $5398.69        1097    889     1884
Q5Y 9T3 hindustan       $826.01 $11176.51       963     701     1885
G0R 5A0 Bharat  $867.65 $2989.86        930     778     1886
Q9E 7J2 reliance        $949.31 $7230.18        962     730     1887
T1A 9O4 Bharat  $852.18 $5209.75        1000    878     1888
O3D 6I4 reliance        $819.52 $11437.38       1006    819     1889
B2K 7U1 reliance        $822.83 $8520.23        913     629     1890
M5Z 4C7 shell   $806.02 $1006.63        1006    805     1891
K0A 0R0 Bharat  $975.02 $7834.03        1063    714     1892
Y0U 6T6 Bharat  $875.48 $1494.54        1043    857     1893
Z3F 8R8 shell   $839.33 $6639.90        943     682     1894
O8A 6Z5 Bharat  $845.74 $8240.36        1045    897     1895
M1T 8I6 reliance        $834.74 $5839.65        976     806     1896
H9E 8U3 Bharat  $890.70 $1778.13        1021    854     1897
H4P 6A9 Bharat  $914.88 $6766.33        985     610     1898
C5X 1C3 reliance        $802.34 $2825.86        1012    716     1899
O5S 1T6 hindustan       $868.28 $2378.12        1021    684     1900
N8P 2B0 hindustan       $870.99 $8563.16        1021    663     1901
T0L 6I1 hindustan       $899.73 $9428.06        998     780     1902
D9L 6K0 Bharat  $994.54 $5237.62        1098    720     1903
C0Z 5S4 Bharat  $885.21 $6347.10        948     888     1904
V5D 1L4 Bharat  $984.15 $5824.10        1077    842     1905
P8Q 8W7 reliance        $903.96 $4849.71        981     627     1906
Z2GTime taken: 0.062 seconds, Fetched: 400 row(s)
```

Q2. To find what is the total amount of petrol in volume sold by every distributor?

*SELECT distributername, SUM(volumeout) FROM indianoil GROUP BY distributername;*



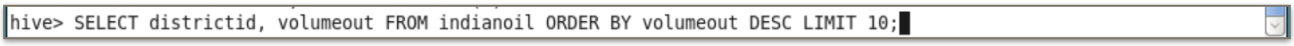
```
hive> SELECT distributername, SUM(volumeout) FROM indianoil GROUP BY distributername;
```



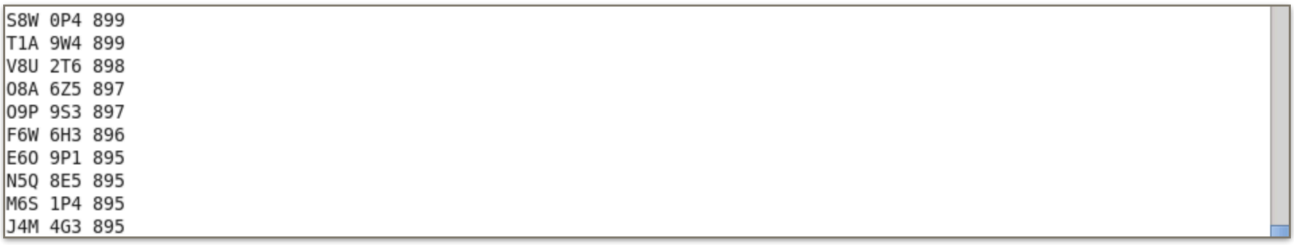
```
Bharat  83662
hindustan       71767
reliance        76558
shell   69266
```

Q3. Which are the top 10 distributors ID’s for selling petrol and also display the amount of petrol sold in volume by them individually?

*SELECT districtid, volumeout FROM indianoil ORDER BY volumeout DESC LIMIT 10;*



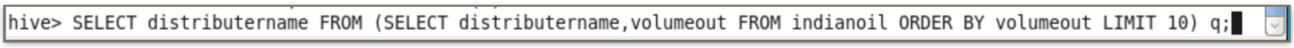
```
hive> SELECT districtid, volumeout FROM indianoil ORDER BY volumeout DESC LIMIT 10;
```



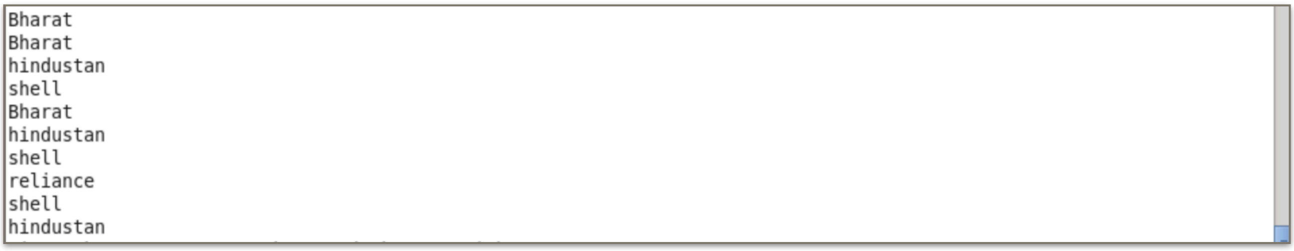
```
S8W 0P4 899
T1A 9W4 899
V8U 2T6 898
O8A 6Z5 897
O9P 9S3 897
F6W 6H3 896
E6O 9P1 895
N5Q 8E5 895
M6S 1P4 895
J4M 4G3 895
```

Q4. List 10 distributor name who sold petrol in the least amount.

*SELECT distributername FROM(SELECT distributername, volumeout FROM indianoil ORDER BY volumeout LIMIT 10) q;*



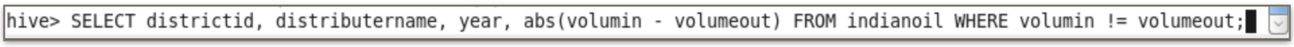
```
hive> SELECT distributername FROM (SELECT distributername,volumeout FROM indianoil ORDER BY volumeout LIMIT 10) q;
```



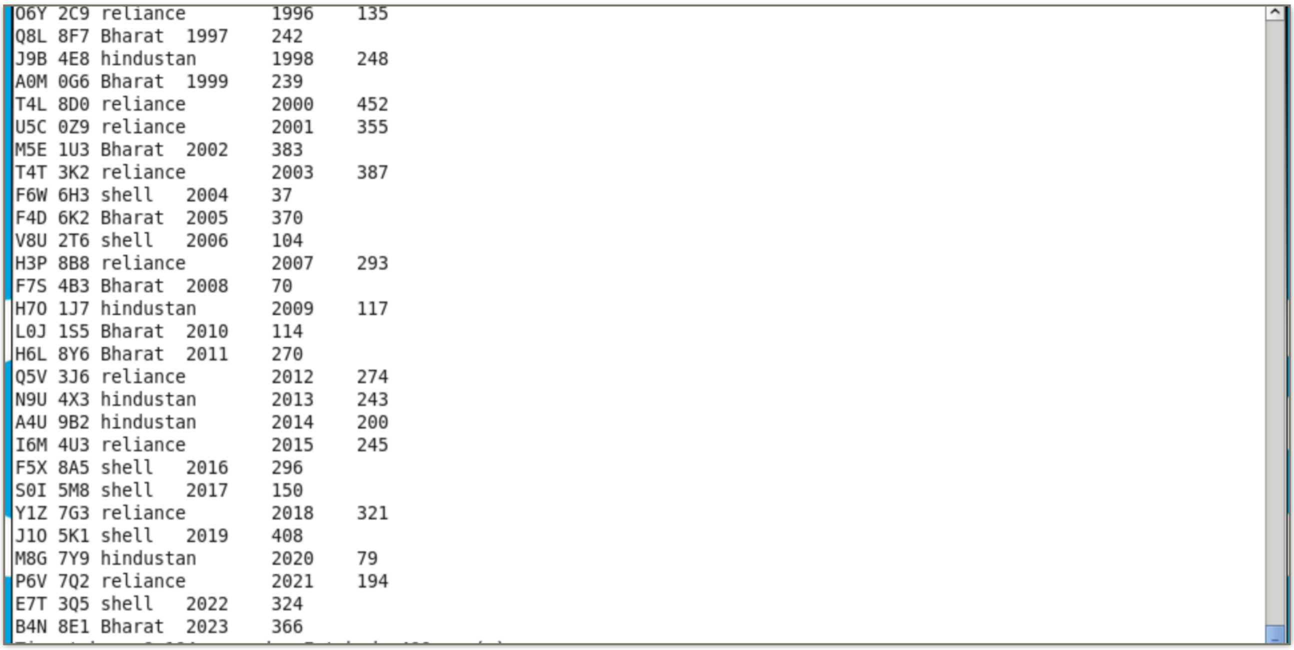
```
Bharat
Bharat
hindustan
shell
Bharat
hindustan
shell
reliance
shell
hindustan
```

Q5. List all distributors who have this difference, along with the year and the difference which they have in that year.

*SELECT districtid, distributername, year, abs(volumin-volumeout) FROM indianoil WHERE volumin != volumeout;*



```
hive> SELECT districtid, distributername, year, abs(volumin - volumeout) FROM indianoil WHERE volumin != volumeout;
```



```
O6Y 2C9 reliance        1996    135
Q8L 8F7 Bharat  1997    242
J9B 4E8 hindustan       1998    248
A0M 0G6 Bharat  1999    239
T4L 8D0 reliance        2000    452
U5C 0Z9 reliance        2001    355
M5E 1U3 Bharat  2002    383
T4T 3K2 reliance        2003    387
F6W 6H3 shell   2004    37
F4D 6K2 Bharat  2005    370
V8U 2T6 shell   2006    104
H3P 8B8 reliance        2007    293
F7S 4B3 Bharat  2008    70
H7O 1J7 hindustan       2009    117
L0J 1S5 Bharat  2010    114
H6L 8Y6 Bharat  2011    270
Q5V 3J6 reliance        2012    274
N9U 4X3 hindustan       2013    243
A4U 9B2 hindustan       2014    200
I6M 4U3 reliance        2015    245
F5X 8A5 shell   2016    296
S0I 5M8 shell   2017    150
Y1Z 7G3 reliance        2018    321
J1O 5K1 shell   2019    408
M8G 7Y9 hindustan       2020    79
P6V 7Q2 reliance        2021    194
E7T 3Q5 shell   2022    324
B4N 8E1 Bharat  2023    366
```