**Problem Statement-11.1**

Write the Pig codes to analyze the petroleum data set.

**Data Definition:**

Column 1: District ID I4N 1M1 varchar

Column 2: Distributor name shell varchar C

olumn 3: Buy rate (million) $957.70 varchar

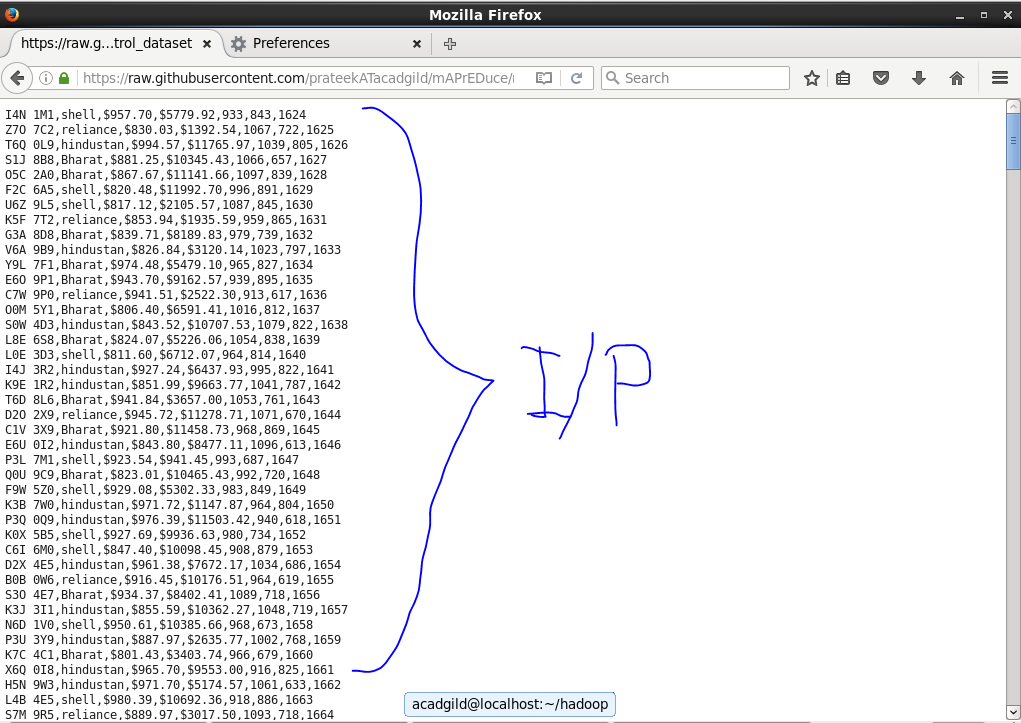
Column 4: Sell rate(million) $5779.92 varchar

Column 5: volume IN(million cubic litre) 933, int

Column 6: volume OUT(million cubic litre) 843, int

Column 7: Year 1624 int

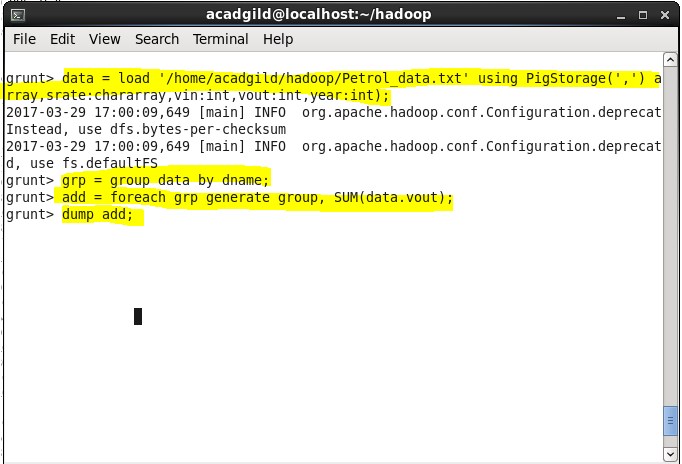
**Input File-**

****

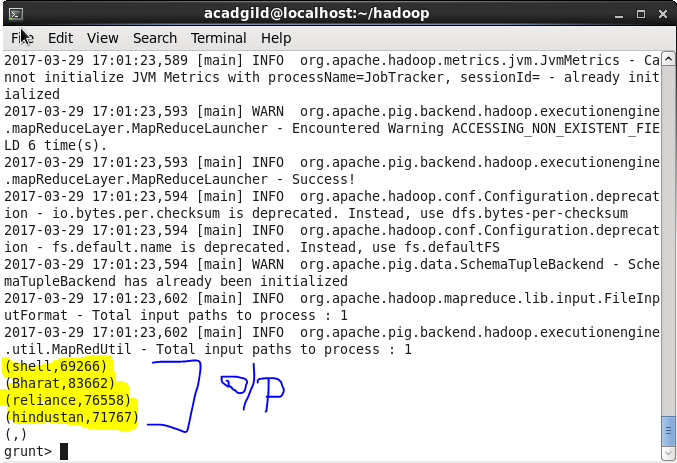
**1. What is the total amount of petrol in volume sold by every distributor?**

**Solution-**

1. Load data from input file.
2. Group data by Distributer Name.
3. Find the sum of volume sold.
4. Display total sold petrol from all the distributors.

****

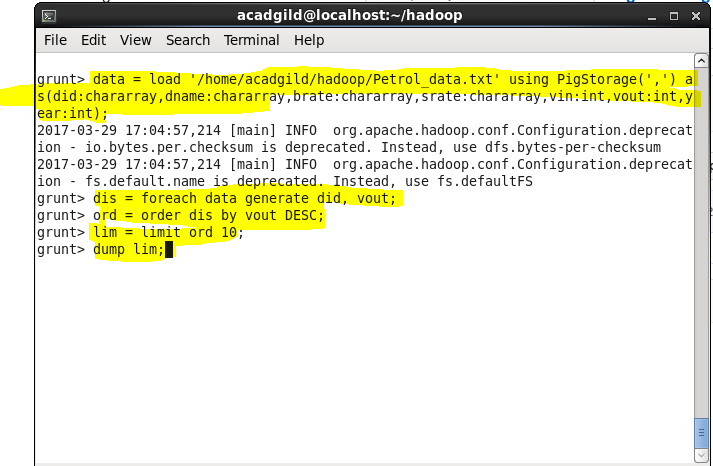
**Output-**

****

**2. Which are the top 10 distributors ID's for selling petrol? Also display the amount of petrol sold in volume.**

**Solution-**

1. Load data from input file.
2. Display Distributor Id and Volume out.
3. Order data by Volume out in descending order.
4. Limit them to display top 10.
5. Display the data.

****

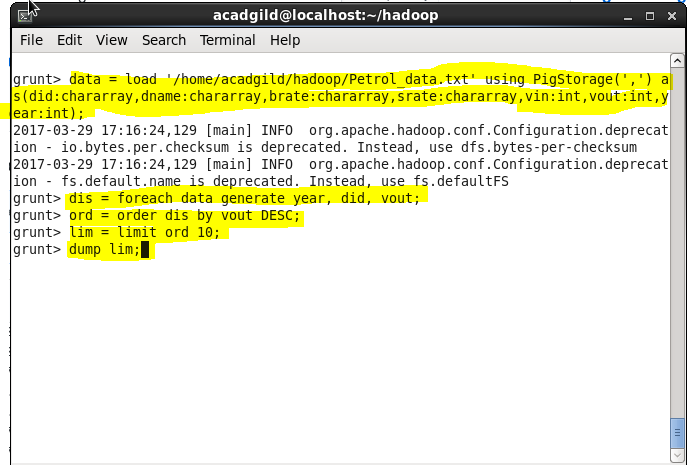
**Output-**

****

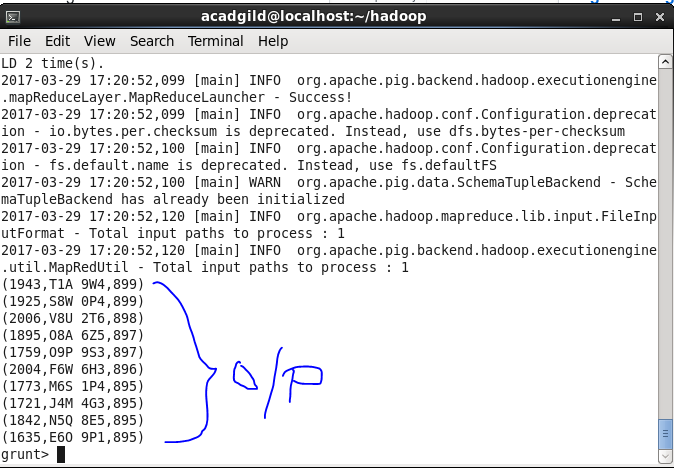
**3. List 10 years where consumption of petrol is more with the distributer id who sold it.**

**Solution-**

1. Load data from input file.
2. Display Distributor Id and Volume out and year.
3. Order data by Volume out in descending order.
4. Limit them to display top 10.
5. Display the data.

****

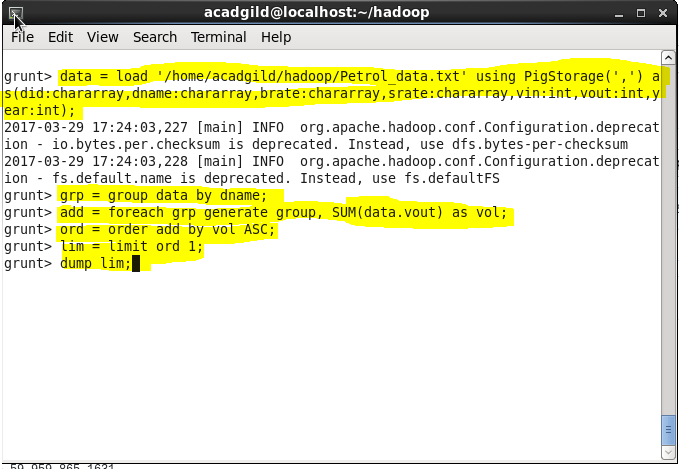
**Output-**

****

**4. Find the distributer name who sold petrol in least amount.**

**Solution-**

1. Load data from input file.
2. Group data on Distributor Name.
3. Find out the sum of amount of Petrol sold.
4. Order the sum in Ascending order.
5. Limit by 1 to find the least seller and display output.



**Output-**

