URL: https://acadgild.com/blog/hive-real-life-use-cases/

This is referring to the real life dataset of Petrol suppliers.

DATA SET:<https://drive.google.com/open?id=0B1QaXx7tpw3SMTBqLUQwX0lOWnM>

|  |  |  |  |
| --- | --- | --- | --- |
| ColumnNO. | Name | Example | DataType |
| Column1: | District.ID | I4N 1M1 | varchar |
| Column2: , | Distributor name | shell | varchar |
| Column3: | Buy rate (million) | $957.70 | varchar |
| Column4: | Sell rate(million) | $5779.92 | varchar |
| Column5: | volumeIN(millioncubic litter) | 933 | int |
| Column6: | volumeOUT(millioncubic litter) | 843, | int |
| Column7: | Year | 1624 | int |

**Create a table:**

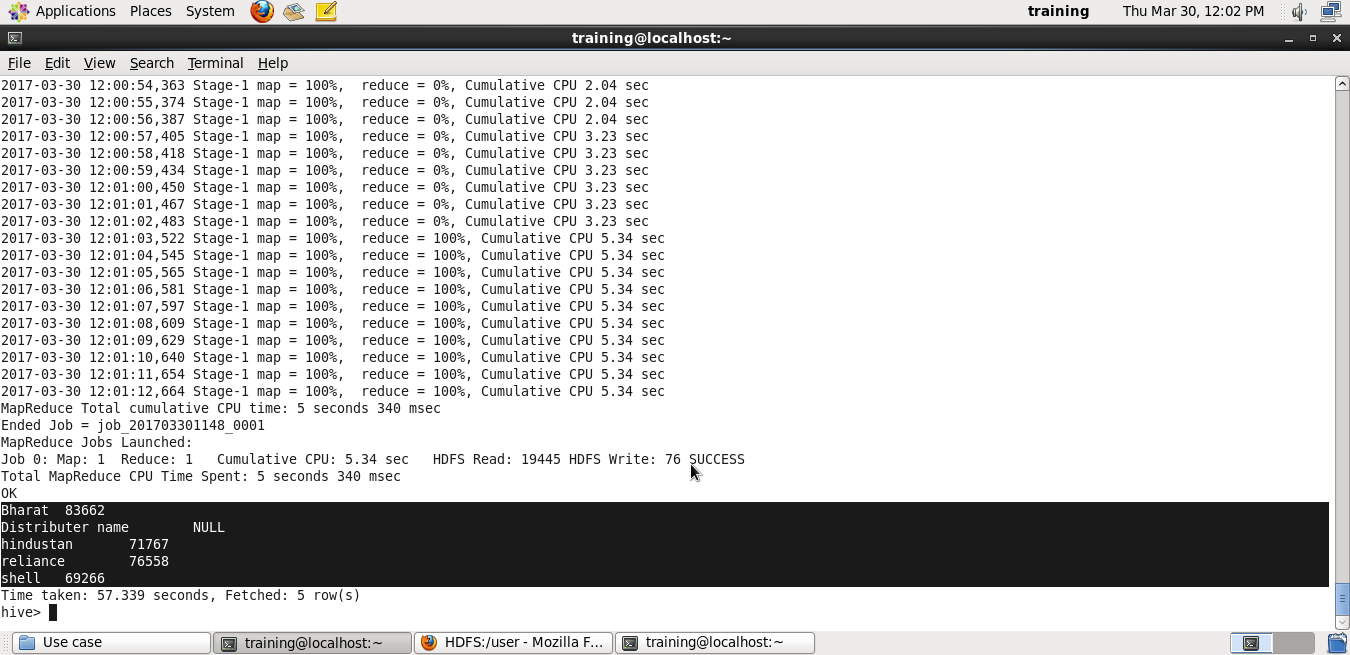
**create table petrol (distributer\_id STRING,distributer\_name STRING,amt\_IN STRING,amy\_OUT STRING,vol\_IN INT,vol\_OUT INT,year INT) row format delimited fields terminated by ‘,’ stored as textfile;**

**Load data:**

LOAD DATA LOCAL INPATH '/home/training/Desktop/Hive/Use case/hive use case - petrol.txt' INTO TABLE PETROL;

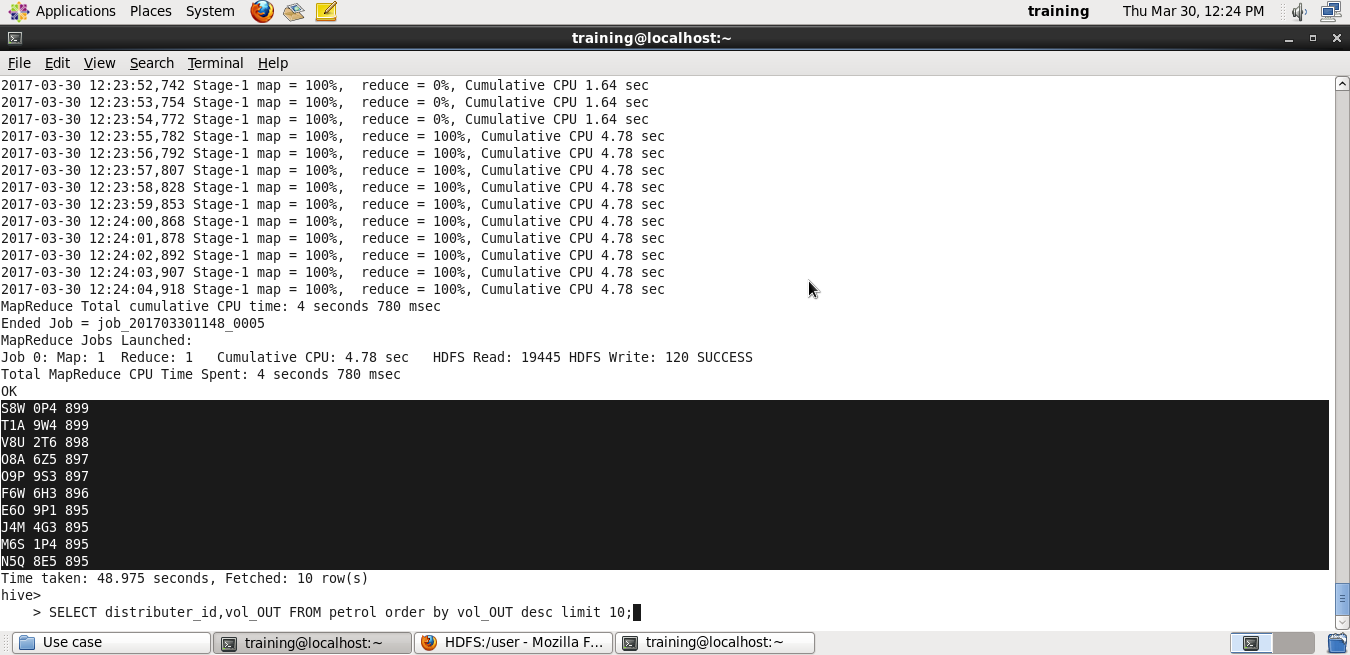
1. In real life what is the total amount of petrol in volume sold by every distributor?

SELECT distributer\_name, SUM(VOL\_OUT) FROM PETROL GROUP BY distributer\_name;



1. 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 distributer\_id,vol\_OUT FROM petrol order by vol\_OUT desc limit 10;



1. Find real life 10 distributor name who sold petrol in the least amount.

**SELECT distributer\_id,vol\_OUT FROM petrol order by vol\_OUT limit 10;**