**SESSION 9: PIG – DEMO II**

**ASSIGNMENT 1**

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

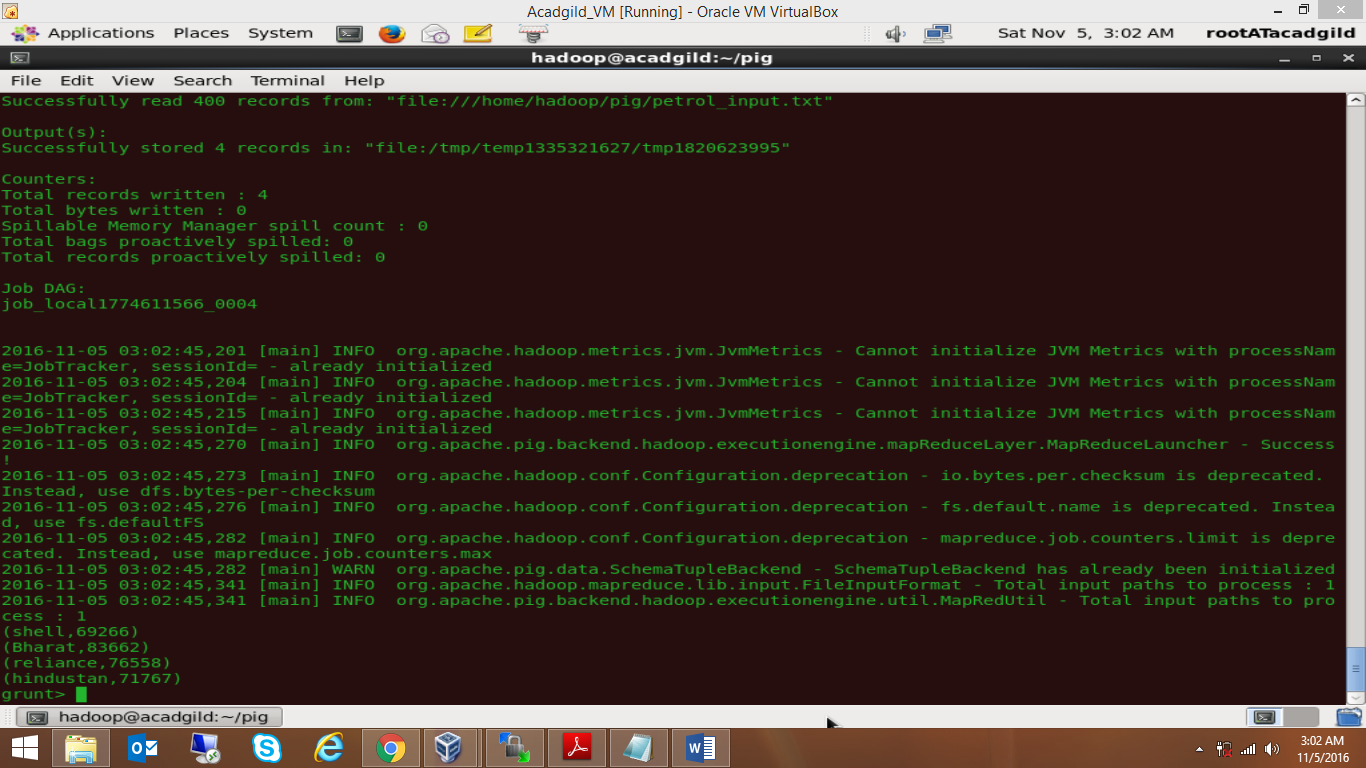
dataset = LOAD ‘petrol\_input.txt’ USING PigStorage(‘,’) AS (d\_id:chararray, d\_name:chararray,

buy\_rate:chararray, sell\_rate:chararray, volin:int, volout:int, year:int);

grouped\_dataset = GROUP dataset BY d\_name;

total\_vol\_sold = FOREACH grouped\_dataset GENERATE group, SUM(grouped\_dataset.volout);

dump total\_vol\_sold;



**ANS: (shell,69266)**

**(Bharat,83662)**

**(reliance,76558)**

**(hindustan,71767)**

1. **Which are the top 10 distributer’s IDs for selling petrol? Also, display the amount of petrol sold in volume.**

dataset = LOAD ‘petrol\_input.txt’ USING PigStorage(‘,’) AS (d\_id:chararray, d\_name:chararray,

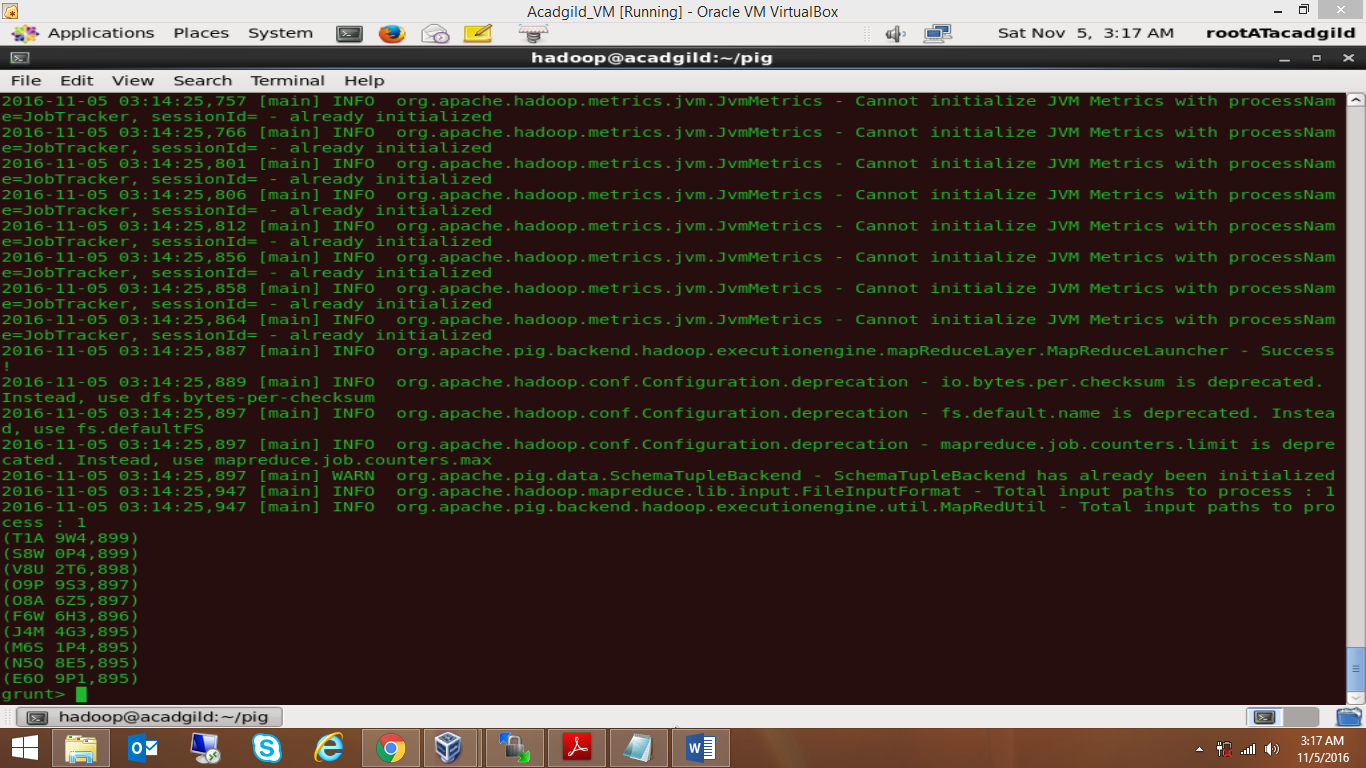
buy\_rate:chararray, sell\_rate:chararray, volin:int, volout:int, year:int);

did\_volout = FOREACH dataset GENERATE d\_id, volout;

sort\_did\_volout = ORDER did\_volout BY volout DESC;

top\_ten\_did = LIMIT sort\_did\_volout 10;

dump top\_ten\_did;



1. **List 10 years where consumption of petrol is more with the distributor ID who sold it.**

dataset = LOAD ‘petrol\_input.txt’ USING PigStorage(‘,’) AS (d\_id:chararray, d\_name:chararray,

buy\_rate:chararray, sell\_rate:chararray, volin:int, volout:int, year:int);

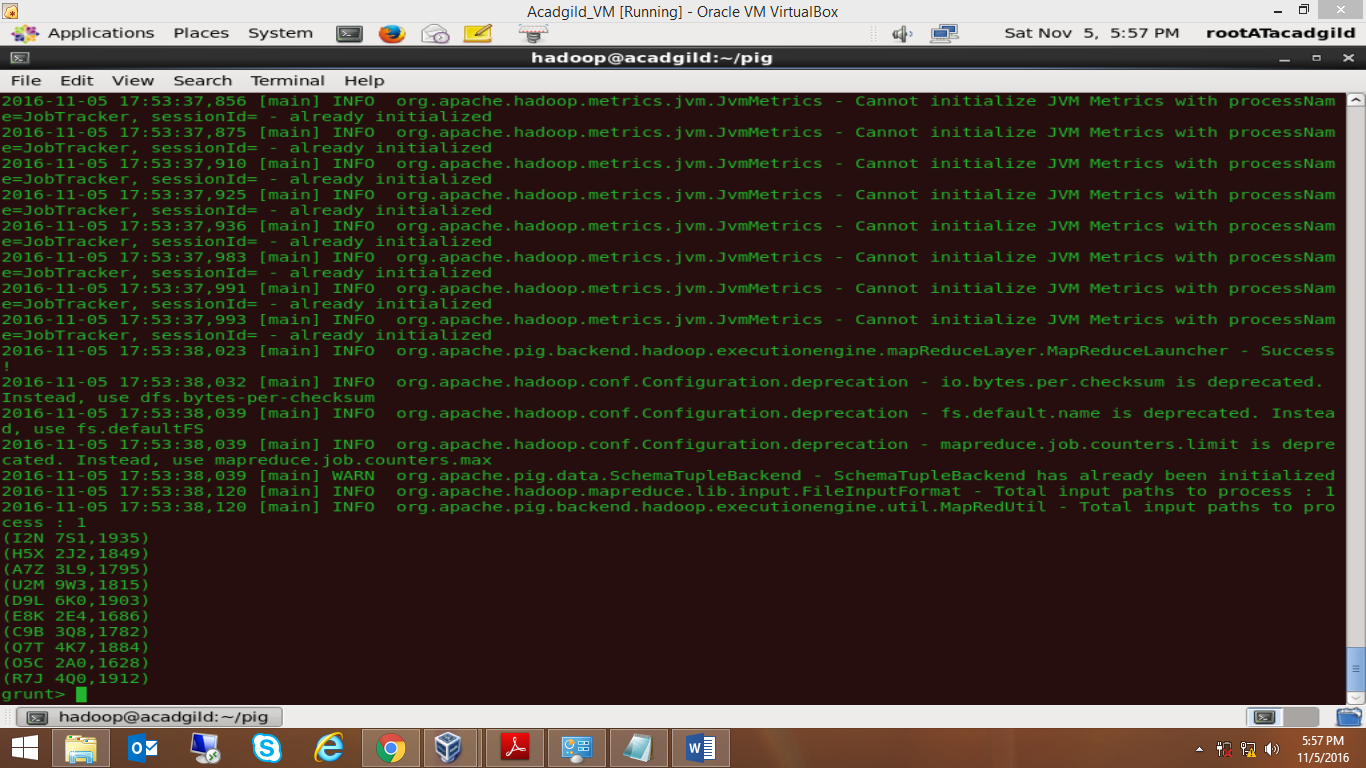
data = FOREACH dataset GENERATE d\_id, volin, year;

sorted\_data = ORDER data BY volin DESC;

limit\_sorted\_data = LIMIT sorted\_data 10;

highest\_consumer\_years = FOREACH limit\_sorted\_data GENERATE d\_id, year;

dump highest\_consumer\_years;



1. **Find the Distributor name who sold petrol in least amount.**

dataset = LOAD ‘petrol\_input.txt’ USING PigStorage(‘,’) AS (d\_id:chararray, d\_name:chararray,

buy\_rate:chararray, sell\_rate:chararray, volin:int, volout:int, year:int);

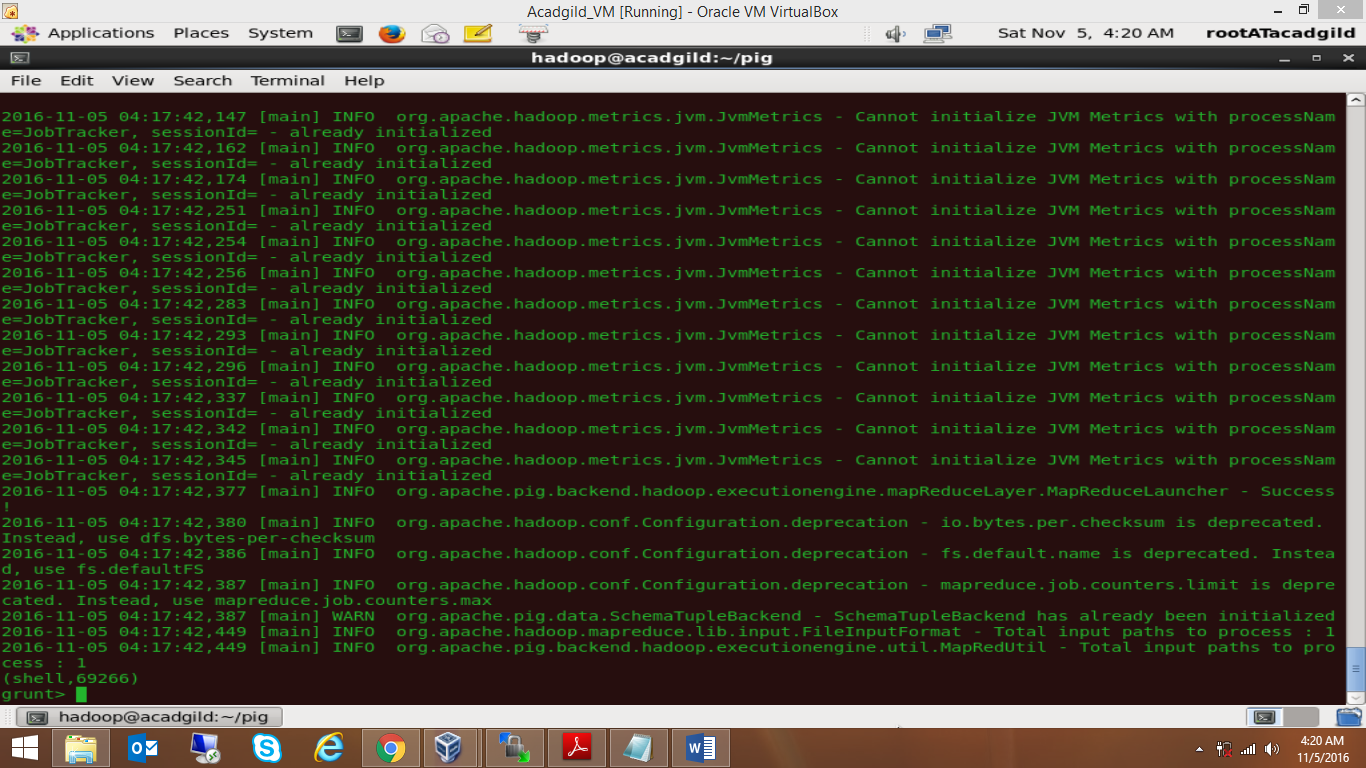
grouped\_dataset = GROUP dataset by d\_name;

dname\_volout = FOREACH grouped\_dataset GENERATE group, SUM(volout) as tvolout;

sorted\_dname\_volout = ORDER dname\_volout BY tvolout;

least\_seller = LIMIT sorted\_dname\_volout 1;

dump least\_seller;



**ANS: (shell, 69266)**