A = LOAD '/home/acadgild/leena/ass/ass9/petrol-dataset.txt' USING PigStorage(',') AS (DistrictID:chararray, Dist\_Name:chararray, Buy\_rateD:chararray, Sell\_rateD:chararray, Vol\_IN:int, Vol\_OUT:int, Year:int); B = FOREACH A GENERATE DistrictID, Dist\_Name, (long)SUBSTRING(Buy\_rateD, 1, (int)(SIZE(Buy\_rateD))) AS Buy\_rate, (long)SUBSTRING(Sell\_rateD, 1, (int)(SIZE(Sell\_rateD))) AS Sell\_rate, Vol\_IN, Vol\_OUT, Year;

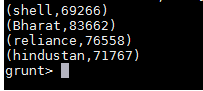
B = FOREACH A GENERATE DistrictID, Dist\_Name, (float)SUBSTRING(Buy\_rateD, 1, (int)(SIZE(Buy\_rateD))) AS Buy\_rate, (float)SUBSTRING(Sell\_rateD, 1, (int)(SIZE(Sell\_rateD))) AS Sell\_rate, Vol\_IN, Vol\_OUT, Year;

B: {DistrictID: chararray,Dist\_Name: chararray,Buy\_rate: float,Sell\_rate: float,Vol\_IN: int,Vol\_OUT: int,Year: int}

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

C = GROUP B BY Dist\_Name;

D = FOREACH C GENERATE Dist\_Name, SUM(B.Vol\_OUT);

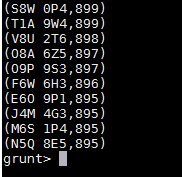


2. Which are the top 10 distributers ID's for selling petrol? Also display the amount of petrol sold in volume.

E = ORDER B BY Vol\_OUT DESC, DistrictID ASC

topTen = LIMIT E 10;

resultTopTen = FOREACH topTen GENERATE DistrictID, Vol\_OUT;

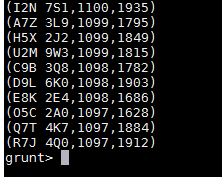


3. List 10 years where consumption of petrol is more with the distributer id who sold it.

F = ORDER B BY Vol\_IN DESC, DistrictID ASC;

topTenConsume = LIMIT F 10;

resultTopTenConsume = FOREACH topTenConsume GENERATE DistrictID, Vol\_OUT, Year;



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

G = ORDER B BY Vol\_OUT ASC, DistrictID DESC;

leastSold = LIMIT G 1;

resultLeastSold = FOREACH leastSold GENERATE DistrictID;

