**Assignment 11.1**

**Question:** Write the Pig codes to analyze the petroleum data set.

*DATA SET DESCRIPTION*

Column 1: District ID I4N 1M1 varchar

Column 2: Distributor name shell varchar

Column 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

**Question:** What is the total amount of petrol in volume sold by every distributor?

**Code:**

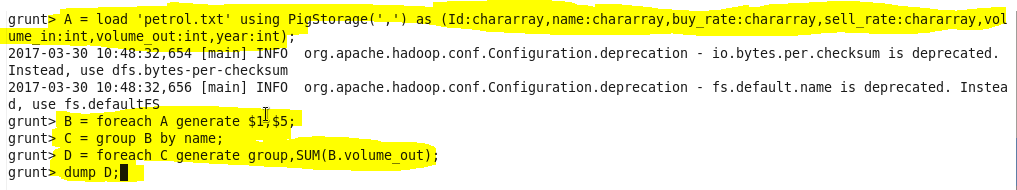
A = load ‘petrol.txt’ using PigStorage(‘,’)as (id:chararray,name:chararray,buy\_rate:chararray,sell\_rate:chararray,volume\_in:int,volume\_out:int,year:int);

B = foreach A generate $1,$5;

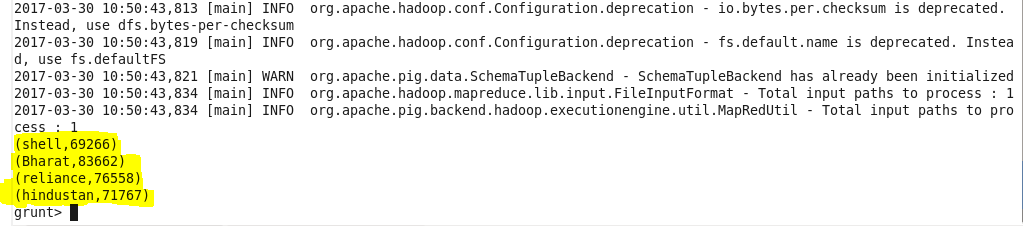
C = group B by name;

D = foreach C generate group,SUM(B.volume\_out);

Dump D;



**Output:**

****

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

**Code:**

A = load ‘petrol.txt’ using PigStorage(‘,’)as (id:chararray,name:chararray,buy\_rate:chararray,sell\_rate:chararray,volume\_in:int,volume\_out:int,year:int);

B= foreach A generate $0,$5;

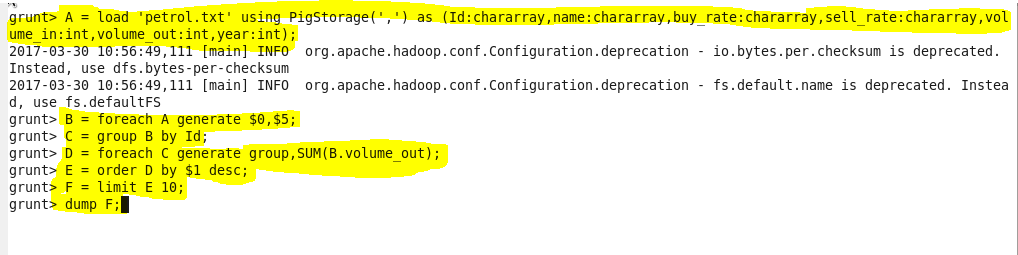
C = group B by Id;

D = foreach C generate group,SUM(B.volume\_out);

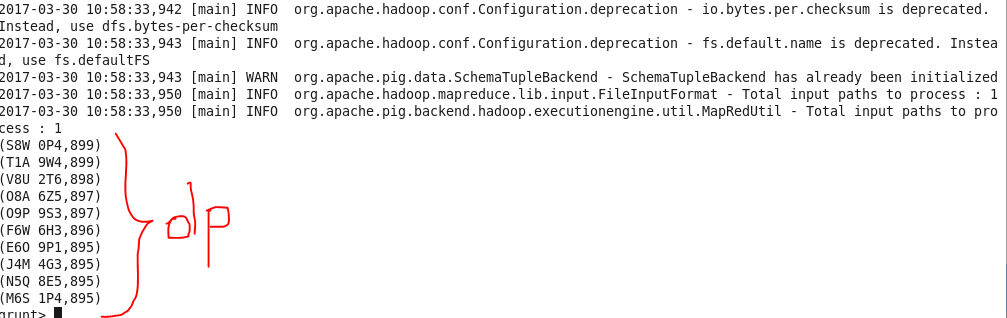
E = order D by $1 desc;

F = limit E 10;

Dump F;



**Output:**

****

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

**Code:**

A = load ‘petrol.txt’ using PigStorage(‘,’)as (id:chararray,name:chararray,buy\_rate:chararray,sell\_rate:chararray,volume\_in:int,volume\_out:int,year:int);

B = foreach A generate $0,$5,$6;

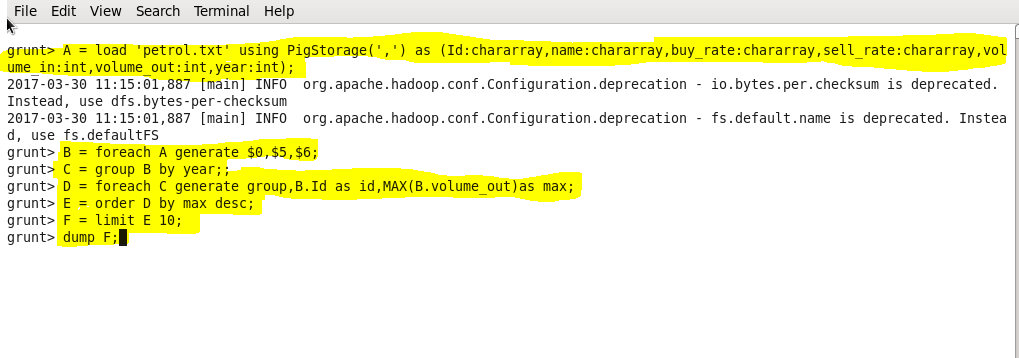
C = group B by year;

D = foreach C generate group,B.Id as id,MAX(B.volume\_out) as max;

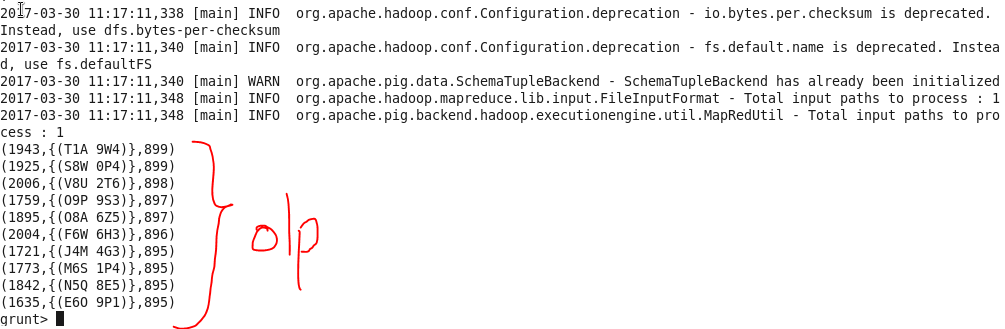
E = order D by max desc;

F = limit E 10;

Dump F;

****

**Output:**

****

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

**Code:**

A = load ‘petrol.txt’ using PigStorage(‘,’)as (id:chararray,name:chararray,buy\_rate:chararray,sell\_rate:chararray,volume\_in:int,volume\_out:int,year:int);

B= foreach A generate $1,$5;

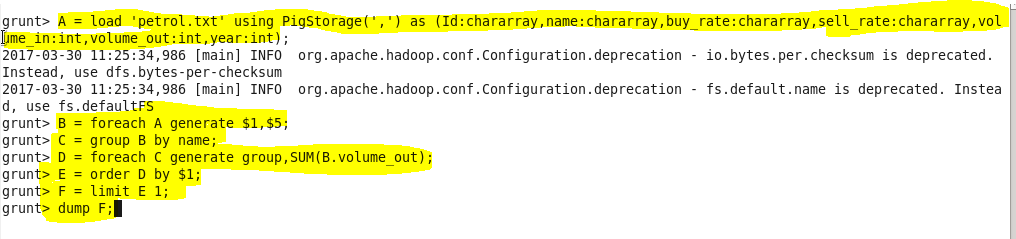
C = group B by name;

D = foreach C generate group,SUM(B.volume\_out);

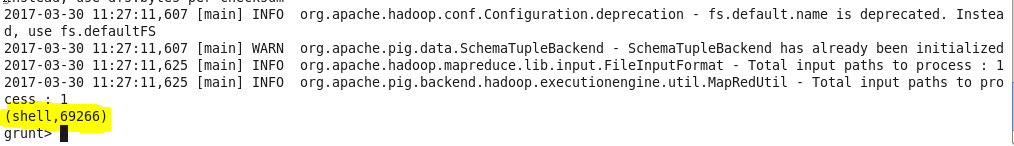
E = order D by $1 ;

F = limit E 1;

Dump F;



**Output:**

****