Task1: What is the total amount of petrol in volume sold by every distributer?

Pig Script:

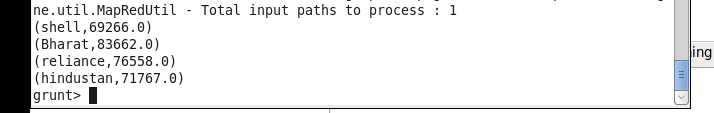
Data = LOAD '/demo/test1.txt' USING PigStorage(',') ;

B = group Data by $1 ;

C = foreach B generate group , SUM(Data.$5);

Dump C;

Output:



Task2 : Which are the top 10 distributers ID's for selling petrol? Also display the amount of petrol sold in volume.

Pig script:

Data = LOAD '/myfiles/pig/test1.txt' USING PigStorage(',') ;

B = order Data by $5 DESC ;

C = limit B 10;

D = foreach C generate $0, $5 ;

dump D;

Output:



Task3 : List 10 years where consumption of petrol is more with the distributer id who sold it.

Pig Script:

Data = LOAD '/myfiles/pig/test1.txt' USING PigStorage(',') ;

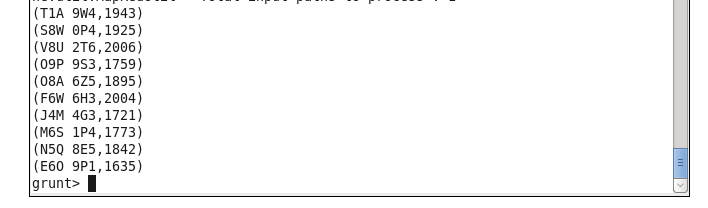
B = order Data by $5 desc;

C = limit B 10;

D = foreach C generate $0 , $6 ;

dump D ;

Output:



Task4 : Find the distributer name who sold petrol in least amount.

Data = LOAD '/myfiles/pig/test1.txt' USING PigStorage(',') ;

sub\_a = foreach Data generate $1 , SUBSTRING($3,1,8);

int\_A = foreach sub\_a generate $0 , (int)$1 ;

B =order int\_A by $1 ;

C = limit B 1;

D = foreach C generate $0;

dump D;

Output:

