**ASSIGNMENT 8.1**

**PROBVLEM STATEMENT :**

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

**STEP 1: loading of data**

s1=load ‘/home/acadgild/Desktop/petrol-dataset.txt’ USING PigStorage(‘,’) AS (ID:chararray, name:chararray, buy:chararray, sell:chararray, in:int ,year:int);

**Step 2 : grouping by name**

s2=group s1 by name;

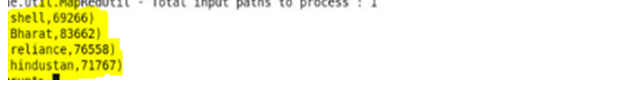
**Step 3: foreach case generate group based on sum**

s3= foreach s2 GENERATE group,SUM(s1.out)

**Step 4 : dump the data**

dump s3;

**Step 5 :output**



==========================================================

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

**STEP 1: loading of data**

a=load ‘/home/acadgild/Desktop/petrol-dataset.txt’ USING PigStorage(‘,’) AS (ID:chararray, name:chararray, buy:chararray, sell:chararray, in:int ,year:int);

**Step 2 : grouping by id**

b=group a by ID;

**Step 3: foreach case generate group based on sum**

c= foreach b GENERATE group,SUM(a.out) as sum;

**Step 4: order the data based on sum in desc order**

d= order c by sum DESC;

**Step 5 : limiting operation is done**

e=limit d 10;

**Step 6 : dump the data required**

dump e;



===========================================================

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

**STEP 1: loading of data**

a=load ‘/home/acadgild/Desktop/petrol-dataset.txt’ USING PigStorage(‘,’) AS (ID:chararray, name:chararray, buy:chararray, sell:chararray, in:int ,year:int);

**Step 2 : grouping by id and year**

b=group a (ID, year);

**Step 3: foreach case generate group based on sum and year**

c= foreach b GENERATE a.ID as id, a.year as year, SUM(a.in) as cons;

**Step 4: order the data based on sum in desc order**

d= order c by cons DESC;

**Step 5: foreach case generate group**

e=foreach d GENERATE id,year;

**Step 6: foreach case generate group by flatten**

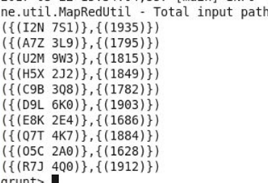
e=foreach d GENERATE flatten(id), flatten(year);

**Step 7 : limiting operation is done**

f= limit e 10;

**Step 8 : dump the data required**

dump f;



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

**STEP 1: loading of data**

a=load ‘/home/acadgild/Desktop/petrol-dataset.txt’ USING PigStorage(‘,’) AS (ID:chararray, name:chararray, buy:chararray, sell:chararray, in:int ,year:int);

**Step 2 : grouping by year**

b=group a by year;

**Step 3: foreach case generate group**

c= foreach b {

temp = foreach a GENERATE in-out;

generate a.name as name, a.year as year, sum(temp) as sale;};

**Step 4: order the data based on sale in desc order**

d=order c by sale DESC;

**Step 6: foreach case generate group by flatten**

e=foreach c generate flatten(name);

**Step 7 : limiting operation is done**

f=limit e 1;

**Step 8 : dump the data required**

dump f;

