**E-Commerce Project using Pig**

**Task 1: Find all the transaction where amt>160.**

step1 = load '/user/cloudera/txns-large.dat' using PigStorage(',') as (tid, tdate, uid, amt:double, category, product, city, state, payment);

step2 = foreach step1 generate uid, amt;

step3 = filter step2 by amt>160;

dump step3;

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

**Task 2: Count all the transaction where amount is between 175 to 200.**

step1 = load '/user/cloudera/txns-large.dat' using PigStorage(',') as (tid, tdate, uid, amt:double, category, product, city, state, payment);

step2 = foreach step1 generate tid, amt;

step3 = filter step2 by (amt>170 and amt<200);

step4 = foreach step3 generate 1 as one;

step5 = group step4 by one;

step6 = foreach step5 generate COUNT(step4.one);

dump step6;

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

**Task 3: Calculate the total sum and total count of all the transaction for each user id.**

step1 = load '/user/cloudera/txns-large.dat' using PigStorage(',') as (tid, tdate, uid, amt:double, category, product, city, state, payment);

step2 = foreach step1 generate uid, amt;

step3 = group step2 by uid;

step4 = foreach step3 generate group, SUM(step2.amt);

dump step4;

**Calculate the average transaction value for each user id.**

step1 = load '/user/cloudera/txns-large.dat' using PigStorage(',') as (tid, tdate, uid, amt:double, category, product, city, state, payment);

step2 = foreach step1 generate $2 as uid,$3 as amt;

step3 = group step2 by uid;

step4 = foreach step3 generate group, SUM(step2.amt),COUNT(step2.amt),AVG(step2.amt);

dump step4;

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

**Task 4: Calculate total sales amt for each Month.**

step1 = load '/user/cloudera/txns-large.dat' using PigStorage(',') as (tid, tdate:chararray, uid, amt:double, category, product, city, state, payment);

step2 = foreach step1 generate SUBSTRING(tdate,0,2) as mon, amt;

step3 = group step2 by mon;

step4 = foreach step3 generate group, SUM(step2.amt) as sum;

dump step4;

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

**Task 5: Divide the file into 12 files, each file containing each month of data. For eg. file 1 should contain data of january txn, file 2 should contain data of feb txn.**

step1 = load '/user/cloudera/txns-large.dat' using PigStorage(',') as (tid, tdate:chararray, uid, amt:double, category, product, city, state, payment);

step2 = foreach step1 generate SUBSTRING(tdate,0,2) as tdate, tid, uid, amt, category, product, city, state, payment;

step3 = filter step2 by month=='01';

dump step3;

step4 = filter step2 by month=='02';

dump step4;

step5 = filter step2 by month=='03';

dump step5;

step6 = filter step2 by month=='04';

dump step6;

step7 = filter step2 by month=='05';

dump step7;

step8 = filter step2 by month=='06';

dump step8;

step9 = filter step2 by month=='07';

dump step9;

step10 = filter step2 by month=='08';

dump step10;

step11 = filter step2 by month=='09';

dump step11;

step12 = filter step2 by month=='10';

dump step12;

step13 = filter step2 by month=='11';

dump step13;

step14 = filter step2 by month=='12';

dump step14;

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

**Task 6: Sort the whole file on the basis of amt.**

step1 = load '/user/cloudera/txns-large.dat' using PigStorage(',') as (tid, tdate:chararray, uid, amt:double, category, product, city, state, payment);

step2 = foreach step1 generate tid, tdate, uid, amt, category, product, city, state, payment;

step3 = order step2 by amt;

dump step3;

**Task 7: Find the name of top 3 spenders**

step1 = load '/user/cloudera/txns-large.dat' using PigStorage(',') as (tid,tdate,uid:int,amt:double,cat,acc,city,state,pay);

step2 = load '/user/cloudera/custs-large.dat' using PigStorage(',') as

(uid:int,fname:chararray,lname,age,prof);

step3 = join step1 by uid, step2 by uid;

step4 = foreach step3 generate uid, amt, fname;

step5 = group step4 by (uid,fname);

step6 = foreach step5 generate group, SUM(step4.amt) as Total;

step7 = order step6 by Total DESC;

step8 = limit g 3;

dump step8;

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

**Task 8: Find the profession of user who has spend the maximum amount**

step1 = load '/user/cloudera/txns-large.dat' using PigStorage(',') as (tid,tdate,uid:int,amt:double,cat,acc,city,state,pay);

step2 = load '/user/cloudera/custs-large.dat' using PigStorage(',') as

(uid:int,fname:chararray,lname,age,prof:chararray);

step3 = join step1 by uid,step2 by uid;

step4 = foreach step3 generate $2 as ID,$3 as tamt,$13 as Prof;

step5 = group step4 by ($0,$2);

step6 = foreach step5 generate group,SUM(step4.tamt) as Res;

step7 = order step6 by Res DESC;

step8 = limit step7 1;

dump step8;