# E-Commerce Tasks Using PIG

Q1) Find all the transaction where amt>160.

```
step1 = LOAD '/user/cloudera/txns-large.dat' using PigStorage (',') as (tid, d, uid,
amt:double, cat, prod, city, state, pt);
step2 = FOREACH step1 generate uid, amt;
step3 = FILTER step2 by amt>160;
DUMP step3;
```

Q2) Count all the transaction where amount is between 175 to 200.

```
step1 = LOAD '/user/cloudera/txns-large.dat' using PigStorage (',') as (tid, d, uid,
amt:double, cat, prod, city, state, pt);
step2 = FOREACH step1 generate tid, amt, uid;
step3 = FILTER step2 by amt >175;
step4 = FILTER step3 by amt <200;
step5 = FOREACH step4 generate 1 as one;
step6 = GROUP step5 by one;
step7 = FOREACH step6 generate COUNT(step5.one);
DUMP step7;
```

Q3) 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, d, uid,
amt:double, cat, prod, city, state, pt);
step2 = FOREACH step1 generate tid, uid, amt;
step3 = GROUP step2 by uid;
step4 = FOREACH step3 GENERATE group, COUNT (step2.tid), SUM(step2.amt);
DUMP step4;
```

Q4)  Calculate total sales amt for each Month.

```
step1 = LOAD '/user/cloudera/txns-large.dat' using PigStorage (',') as (tid, d, uid,
amt:double, cat, prod, city, state, pt);
step2 = FOREACH step1 generate tid, SUBSTRING(d,0,2) as month, amt;
step3 = GROUP step2 by month;
step4 = FOREACH step3 GENERATE group, SUM(step2.amt);
DUMP step4;
```

Q5) 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, d, uid,
amt:double, cat, prod, city, state, pt);
step2 = FOREACH step1 generate SUBSTRING(d,0,2) as month;
step3 = GROUP step2 by month;
step4 = filter step3 by group=='01';
STORE step1 INTO '/user/cloudera/part-00001';
step4 = filter step3 by group=='02';
STORE step1 INTO '/user/cloudera/part-00002';
step4 = filter step3 by group=='03';
STORE step1 INTO '/user/cloudera/part-00003';
step4 = filter step3 by group=='04';
STORE step1 INTO '/user/cloudera/part-00004';
step4 = filter step3 by group=='05';
STORE step1 INTO '/user/cloudera/part-00005';
step4 = filter step3 by group=='06';
STORE step1 INTO '/user/cloudera/part-00006';
step4 = filter step3 by group=='07';
STORE step1 INTO '/user/cloudera/part-00007';
step4 = filter step3 by group=='08';
STORE step1 INTO '/user/cloudera/part-00008';
step4 = filter step3 by group=='09';
STORE step1 INTO '/user/cloudera/part-00009';
step4 = filter step3 by group=='10';
STORE step1 INTO '/user/cloudera/part-00010';
step4 = filter step3 by group=='11';
```

```
STORE step1 INTO '/user/cloudera/part-00011';
step4 = filter step3 by group=='12';
STORE step1 INTO '/user/cloudera/part-00012';
```

Q6) Find the profession of user who has spend the maximum amount

```
step1 = LOAD '/user/cloudera/txns-large.dat' using PigStorage (',') as (tid, d, uid,
amt:double, cat, prod, city, state, pt);
step2 = LOAD '/user/cloudera/Customer.dat' using PigStorage (',') as
(custid,fname,lname,age:double,prof);
step3 = JOIN step1 by uid, step2 by custid;
step4 = GROUP step3 by prof;
step5 = FOREACH step4 GENERATE group, SUM(step3.amt)as tamt;
step6 = ORDER step5 by tamt desc;
step7 = LIMIT step6 1;
dump step7;
```

Q7) Find the name of top 3 spenders.

```
step1 = LOAD '/user/cloudera/txns-large.dat' using PigStorage (',') as (tid, d, uid,
amt:double, cat, prod, city, state, pt);
step2 = LOAD '/user/cloudera/Customer.dat' using PigStorage (',') as
(custid,fname,lname,age:double,prof);
step3 = JOIN step1 by uid, step2 by custid;
step4 = GROUP step3 by fname;
step5 = FOREACH step4 GENERATE group, SUM(step3.amt)as tamt;
step6 = ORDER step5 by tamt desc;
step7 = LIMIT step6 3;
dump step7;
```

Q8) Find the user who has spent the max amount in July month.

```
step1 = LOAD '/user/cloudera/Transactional.dat' using PigStorage (',') as (tid, d, uid,
amt:double, cat, prod, city, state, pt);
step2 = LOAD '/user/cloudera/Customer.dat' using PigStorage (',') as (custid, fname,
lname, age:double, prof);
step3 = JOIN step1 by uid , step2 by custid;
step4 = FOREACH step3 GENERATE fname, SUBSTRING(d,0,2) as mon, amt;
step5 = FILTER step4 by mon=='07';
step6 = GROUP step5 by fname;
step7 = FOREACH step6 GENERATE group, SUM(step5.amt) as tcnt;
step8 = ORDER step7 by tcnt desc;
step9 = LIMIT step8 1;
dump step9;
```