Q1:

pig -x local

tweets\_dt = LOAD 'tweet\_info.csv' USING PigStorage('\t') AS(id:long, tweet:chararray, username:chararray);

DUMP tweets\_dt;

users\_dt = LOAD ‘user\_info.csv' USING PigStorage('\t') AS(login\_info:long, username:chararray, state:chararray);

DUMP users\_dt;

Q2:

x = FILTER users\_dt BY (state == 'FM');

d = foreach x generate $0;

DUMP d;

Q3:

b = FILTER tweets\_dt BY (tweet matches '.\*fine.\*');

WORD = order b by $0;

DUMP WORD;

Q4:

joined = join tweets\_dt by username, users\_dt by username;

STORE joined INTO 'outputB';

Q5:

B = GROUP tweets\_dt BY username;

DESCRIBE B;

ILLUSTRATE B;

DUMP B;

data = FOREACH B GENERATE group, (int)COUNT(tweets\_dt.tweet) AS total;

DUMP data;

STORE data INTO 'outputC';

Q6:

X = FILTER data BY (int)total >= 5;

Q7: