Question 1)

a = load '/nxc161330/inputnikhil/business.csv/' as line;

b = foreach a generate FLATTEN((tuple(chararray,chararray,chararray))REGEX\_EXTRACT\_ALL(line,'(.\*)\\:\\:(.\*)\\:\\:(.\*)'))as(bid,address,categories);

x = FILTER b BY(address matches '.\*Palo Alto.\*');

c = load '/nxc161330/inputnikhil/review.csv/' as line2;

d = foreach c generate FLATTEN((tuple(chararray,chararray,chararray,float))REGEX\_EXTRACT\_ALL(line2,'(.\*)[\\:\\:(.\*)\\:\\:(.\*)\\:\\:(.\*)'))as(rid,uid,bid,rating)](file:///\\:\\:(.*)\\:\\:(.*)\\:\\:(.*)'))as(rid,uid,bid,rating));

e = join x by bid, d by bid;

f = GROUP e by(x::bid,x::address,x::categories);

g = foreach f generate group,AVG(e.(d::rating))as average;

h = order g by average desc;

i = limit h 10;

dump i;

Question 2)

a = load '/nxc161330/inputnikhil/business.csv/' as line;

b = foreach a generate FLATTEN((tuple(chararray,chararray,chararray))REGEX\_EXTRACT\_ALL(line,'(.\*)[\\:\\:(.\*)\\:\\:(.\*)'))as(bid,address,categories)](file:///\\:\\:(.*)\\:\\:(.*)'))as(bid,address,categories));

x = FILTER b BY(address matches '.\*CA.\*') AND NOT(address matches '.\*Palo Alto.\*');

c = load '/nxc161330/inputnikhil/review.csv/' as line2;

d = foreach c generate FLATTEN((tuple(chararray,chararray,chararray,float))REGEX\_EXTRACT\_ALL(line2,'(.\*)\\:\\:(.\*)\\:\\:(.\*)\\:\\:(.\*)'))as(rid,uid,bid,rating);

e = join x by bid, d by bid;

f = GROUP e by(x::bid,x::address,x::categories);

g = foreach f generate group,AVG(e.(d::rating))as average;

h = order g by average desc;

i = limit h 10;

dump i;

Question 3)

a = load '/nxc161330/inputnikhil/business.csv/' as line;

b = foreach a generate FLATTEN((tuple(chararray,chararray,chararray))REGEX\_EXTRACT\_ALL(line,'(.\*)[\\:\\:(.\*)\\:\\:(.\*)'))as(bid,address,categories)](file:///\\:\\:(.*)\\:\\:(.*)'))as(bid,address,categories));

c = load '/nxc161330/inputnikhil/review.csv/' as line2;

d = foreach c generate FLATTEN((tuple(chararray,chararray,chararray,float))REGEX\_EXTRACT\_ALL(line2,'(.\*)[\\:\\:(.\*)\\:\\:(.\*)\\:\\:(.\*)'))as(rid,uid,bid,rating)](file:///\\:\\:(.*)\\:\\:(.*)\\:\\:(.*)'))as(rid,uid,bid,rating));

e = cogroup b by bid, d by bid;

dump e;

Question 4)

a = load '/nxc161330/inputnikhil/business.csv/' as line;

b = foreach a generate FLATTEN((tuple(chararray,chararray,chararray))REGEX\_EXTRACT\_ALL(line,'(.\*)[\\:\\:(.\*)\\:\\:(.\*)'))as(bid,address,categories)](file:///\\:\\:(.*)\\:\\:(.*)'))as(bid,address,categories));

x = FILTER b BY(address matches '.\*Stanford.\*');

c = load '/nxc161330/inputnikhil/review.csv/' as line2;

d = foreach c generate FLATTEN((tuple(chararray,chararray,chararray,float))REGEX\_EXTRACT\_ALL(line2,'(.\*)[\\:\\:(.\*)\\:\\:(.\*)\\:\\:(.\*)'))as(rid,uid,bid,rating)](file:///\\:\\:(.*)\\:\\:(.*)\\:\\:(.*)'))as(rid,uid,bid,rating));

e = join x by bid, d by bid;

f = GROUP e by(x::uid,x::rating);

i = limit f 10;

dump i;