# Pig Queries:

The files are Query1.pig, Query2.pig, Query3.pig and Query4.pig are according to four Pig Questions in Project 2.

The command to run these pig scripts is following(take Query1 as a simple example),

pig Query1.pig

# Hadoop Streaming:

The mapper and reducer functions are developed by Python, which are according to mapper.py and reducer.py in the zip file.

The command to run them is following(there are two input directory, one for customers dataset and the other for transactions dataset),

hadoop jar /usr/share/hadoop/contrib/streaming/hadoop-streaming-1.2.1.jar -file /home/hadoop/Project2/Streaming/mapper.py -mapper /home/hadoop/Project2/Streaming/mapper.py -file /home/hadoop/Project2/Streaming/reducer.py -reducer /home/hadoop/Project2/Streaming/reducer.py -input /user/hadoop/c\_input/\* -input /user/hadoop/t\_input/\* -output /Project2/Q5\_output