Project 1 Documentation

Emily Hao, [eyhao@wpi.edu](mailto:eyhao@wpi.edu)

Daojun Liang, [dliang2@wpi.edu](mailto:dliang2@wpi.edu)

**Create datasets**

We attached the java code for creating datasets: customer.txt, transactions.txt

Note: transactions.txt takes a bit time to create.

**Uploading data into Hadoop**

We used the command

$ CD into where the datasets located in the folder

$ Hadoop fs –put customer.txt /user/Hadoop/input

$ Hadoop fs –put transactions.txt /user/Hadoop/input

**MapReduce Jobs**

Query 1

Query 2

Query 3

Query 4

Query 5

**Apache-Pig**

Note:

All of my inputs are loaded from the path ‘input/\_\_\_\_\_\_’

All of my outputs are stored to the path ‘output/\_\_\_\_\_’

Feel free to change the path if needed.

I attached the 4 Apache-Pig scripts in the folder.

To run the scripts,

$ Sudo start-all.sh

$ Pig

you can either paste the whole script into grunt command.

Or

$ Sudo start-all.sh

$ cd <to the folder>

$ Pig PigQuery1.pig

$ Pig PigQuery2.pig

$ Pig PigQuery3.pig

$ Pig PigQuery4.pig