

Yufei Lin, Jinhong Yu  
Sep 22nd, 2021  
Project 1

Project 1 Read Me

Question	Source	Author	Status	Comments
1	CreateDataset.java	Yufei Lin, Jinhong Yu	Fully Working	Yufei Lin for create transaction dataset, Jinhong Yu for create customer dataset. args[0] is path to save customer.txt, args[1] is path for save transaction.txt
2		Yufei Lin	Fully Working	Picture shown in the following
3.1	java/Query1.java	Yufei Lin	Fully Working	args[0] is customer.txt path, args[1] is output path.
3.2	java/Query2.java	Jinhong Yu	Fully Working	args[0] is customer.txt path, args[1] is transaction.txt path, args[2] is output path.
3.3	java/Query3.java	Yufei Lin	Fully Working	args[0] is customer.txt path, args[1] is transaction.txt path, args[2] is output path.
3.4	java/Query4.java	Jinhong Yu	Fully Working	args[0] is customer.txt path, args[1] is transaction.txt path, args[2] is output path.
3.5	java/Query5.java	Yufei Lin	Fully Working	args[0] is customer.txt path, args[1] is transaction.txt path, args[2] is output path.
4.1	pig/Query1.pig	Yufei Lin	Fully Working	\$customer is customer.txt path, \$transaction is transaction.txt path, \$out is output path. There is a .sh script to run Query1.pig
4.2	pig/Query2.pig	Jinhong Yu	Fully Working	\$customer is customer.txt path, \$transaction is transaction.txt path, \$out is output path. There is a .sh script to run Query2.pig
4.3	pig/Query3.pig	Yufei Lin	Fully Working	\$customer is customer.txt path, \$out is output path. There is a .sh script to run Query3.pig

4.4	pig/Query4.pig	Jinhong Yu	Fully Working	\$customer is customer.txt path, \$transaction is transaction.txt path, \$out is output path. There is a .sh script to run Query4.pig
-----	----------------	------------	---------------	---

#### Question 2 Result

```

➔ sbin hdfs dfs -copyFromLocal /home/ds503/shared_folder/Project1/data/Customer.txt /input
➔ sbin hdfs dfs -copyFromLocal /home/ds503/shared_folder/Project1/data/Transaction.txt /input
➔ sbin hdfs dfs -ls /input
Found 2 items
-rw-r--r-- 1 ds503 supergroup 2289085 2021-09-22 17:30 /input/Customer.txt
-rw-r--r-- 1 ds503 supergroup 295344034 2021-09-22 17:30 /input/Transaction.txt
➔ sbin █

```