Zhenyu Hu and Kevin Heath
12 February 2017

<center>Project 1 Documentation:</center>

Create dataset:

      Use FileWriter and random data to generate dataset files

Java Queries:

1. Map-only query
2. Map-combine-reduce query. The combiner and reducer utilize the same reducer
3. Map-reduce replicated join query. We used the distributed cache to read in Customers.csv
4. Map-reduce with replicated join.
5. Twice Map-reduce by two jobs

Pig Queries:

      Query1 and query2 use join two dataset files and the query3 only need to read in Customers.csv