

Project1 for CS585

Team member :

Jingjun Zhang

Hongzhang Cheng

In this project, we at first used java to create two datasets. then used the command line to upload these two datasets into the hdfs system.

In order to do this, we at first created a dir in the hdfs using “hadoop fs -mkdir input” command. then used command” hadoop fs -put <path>” to upload the files to the system.

In the Query1, we only used mapper function.

in the Query 2, we used mapper, combiner and reducer. the combiner and reducer share the same reducer function.

In the Query 3, we used two mapper function and one reducer function.

In the Query 4, we only used one mapper function and one reducer function, but inside the mapper function we rewrite the setup function to preprocess the customer table and build the countryside and customerID relation.

In the Query 5, we make used 3 mapper function and 1 reducer function.