Project 3:

p1:

1. To get the data set for p1, please compile the java file dbCreator.java and run it.
2. To run the program in scala, please change the file path for variable transaction\_path value to your path for the transaction.csv
3. For the output file, please change the file path in line

t2.write.csv("file:///home/mqp/Desktop/Project3/CSV/t2.csv")

t6.write.csv("file:///home/mqp/Desktop/Project3/CSV/t6.csv")

to your own path

1. Then the program is ready to run

p2:

1. To get the data set for p2, please compile the java file dbCreatorForP.java and run it.
2. Upload data set P.csv to Hadoop
3. Change hdfs path in p2-step2.scala and p2-step3.scala

point\_path = "hdfs://localhost:9000/project-3/p2/P.csv"

1. Run in scala-shell

p3:

1. To get the dataset for p3, please compile the java file dbCreatorForP.java and run it.
2. To run the program in scala please change the file path in lane
   1. val n\_path = "file:///home/mqp/Desktop/Project3/p3/N.csv"
   2. val m\_path = "file:///home/mqp/Desktop/Project3/p3/M.csv"

To your own path for N.csv and M.csv

1. Then the program is ready to run