Spark version:2.2.1

Approach for SON algorithm:

- Use Apriori algorithm to get candidates in map phase 1
- Eliminate duplications in reduce phase 1
- Count each candidate in map phase 2
- Sum up the count for each candidate in reduce phase 2
- Determine if the candidate is the frequent item set in the final phase

Problem 1

Command line:

$bin/spark-submit --class Son Yi_Wei_SON.jar 1 small2.csv 3

$bin/spark-submit --class Son Yi_Wei_SON.jar 2 small2.csv 5

Problem 2

Command line:

$bin/spark-submit --class Son Yi_Wei_SON.jar 1 beauty.csv 50

$bin/spark-submit --class Son Yi_Wei_SON.jar 2 beauty.csv 40

$bin/spark-submit --class Son Yi_Wei_SON.jar 1 books.csv 1200

$bin/spark-submit --class Son Yi_Wei_SON.jar 2 books.csv 1500

File Name	Case Number	Support	Runtime (sec)
beauty.csv	1	50	183
beauty.csv	2	40	54
books.csv	1	1200	255
books.csv	2	1500	30