Assignment 4.1

Associated Data Files:

<https://drive.google.com/file/d/0Bxr27gVaXO5sVjQ5QW0wQ3RCTUU/view?usp=sharing>

The fields are arranged like:

Company Name|Product Name|Size in inches|State|Pin Code|Price

There are some invalid records which contain 'NA' in either Company Name or Product Name.

1. Write a Map Reduce program to filter out the invalid records.

The output of this program will act as input for subsequent tasks.

Map only job will fit for this context.

Code: Refer attached java files

CleanUpTask.java

CleanUpMapper.java

Zipped in testjar4.jar

Command to run the jar >> hadoop jar testjar4.jar hadoop. CleanUpTask /user/ec2-user/ass1\_1\_in /user/ec2-user/ass1\_1\_out

Result:

