Assignment 3.1

Introduction:

In this assignment, I’m going to write a **Map Only** java program to filter out the invalid records provided in the dataset file.

We are using **Map only** program since we are going to filter a few records.

Associated Data Files

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

Problem Statement

We have a dataset of sales of different TV sets across different locations.

Records look like:

Samsung|Optima|14|Madhya Pradesh|132401|14200

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. Map only job will fit for this

Context.



**MapReduce Java Code**



* To execute above program in hadoop mode firstly the source file is exported in jar file on to desktop or any desired location , then in the virtual machine from the location using the below command the jar file is given input to execute and the output is extracted as shown below

$ hadoop jar A3.jar /user/acadgild/shabbu/salesdetails1.txt /joutput



Input and Output Comparison

Input



Output

