**ACD\_BDDOF\_Session\_1\_Assignment\_6\_Main**

**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

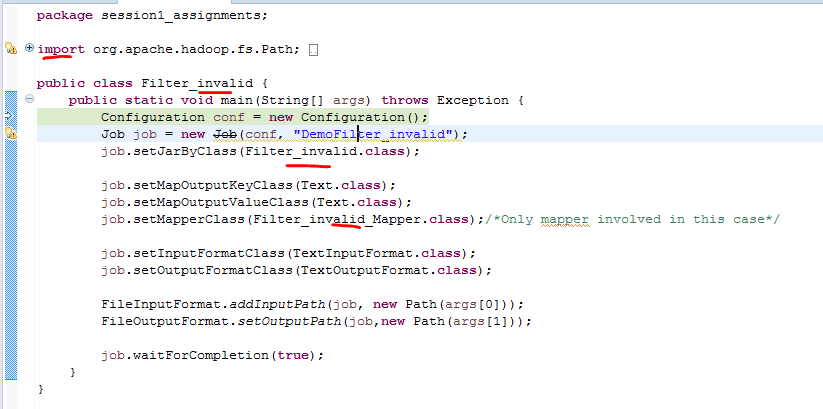
The fields are arranged like:Company Name|Product Name|Size in inches|State|Pin Code|Price

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.

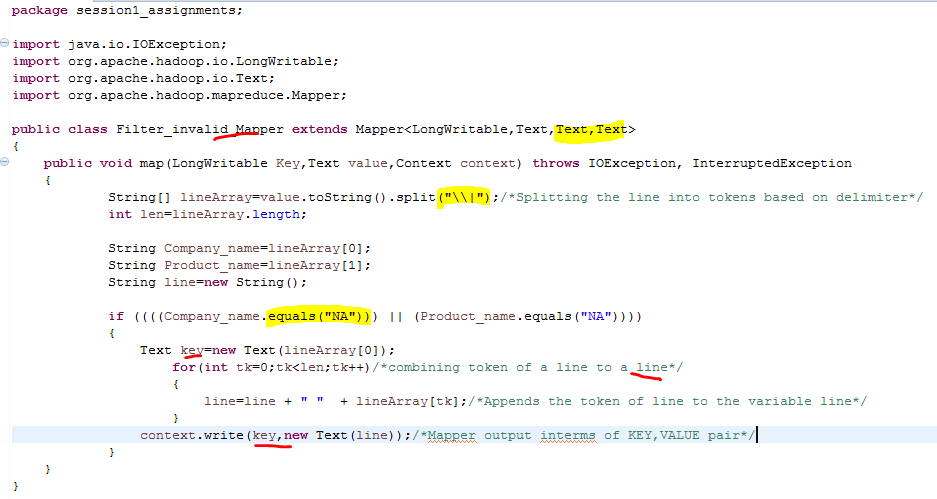
**Driver code:**



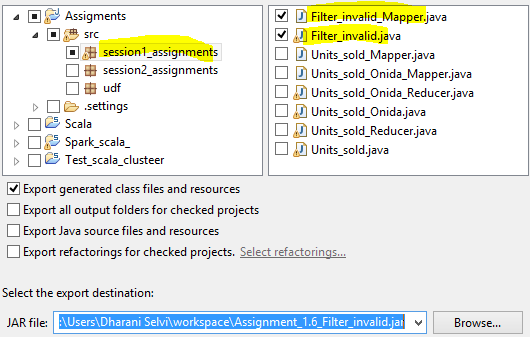
**Mapper code:**

Splits the input based on “|” and filter out only invalid records which contain 'NA' in either Company Name or Product Name.

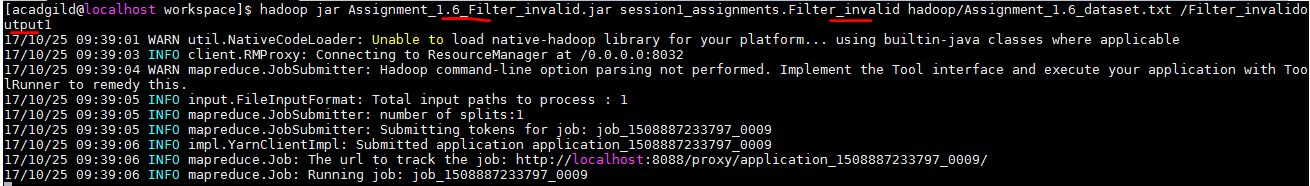
Outputs key – Company\_name,value – whole line to the console.



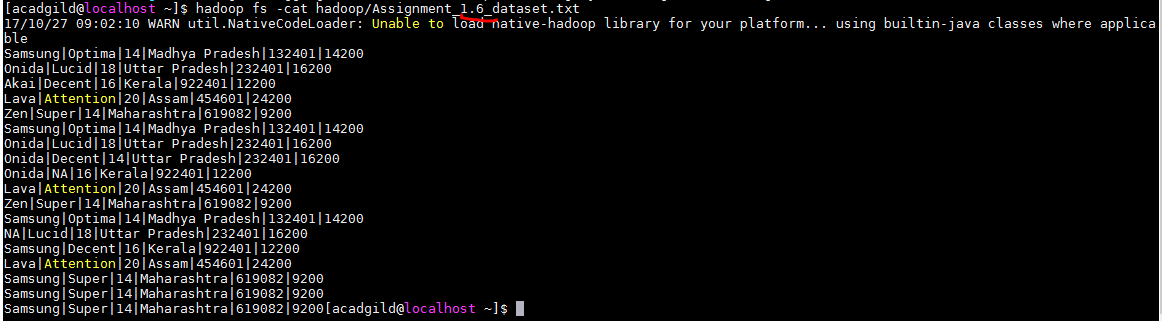
**Code export as JAR:**



**Job execution in cluster:**



**Input:**



**Output:**

Displays only the records has ‘NA’.

