**Q1-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 sales\_differentTV : https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/sales\_differentTV.txt**

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

**Sales of different TV Task 1:**

**Write a MapReduce 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.**

Dataset:

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

Onida|Lucid|18|Uttar Pradesh|232401|16200

Akai|Decent|16|Kerala|922401|12200

Lava|Attention|20|Assam|454601|24200

Zen|Super|14|Maharashtra|619082|9200

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

Onida|Lucid|18|Uttar Pradesh|232401|16200

Onida|Decent|14|Uttar Pradesh|232401|16200

Onida|NA|16|Kerala|922401|12200

Lava|Attention|20|Assam|454601|24200

Zen|Super|14|Maharashtra|619082|9200

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

NA|Lucid|18|Uttar Pradesh|232401|16200

Samsung|Decent|16|Kerala|922401|12200

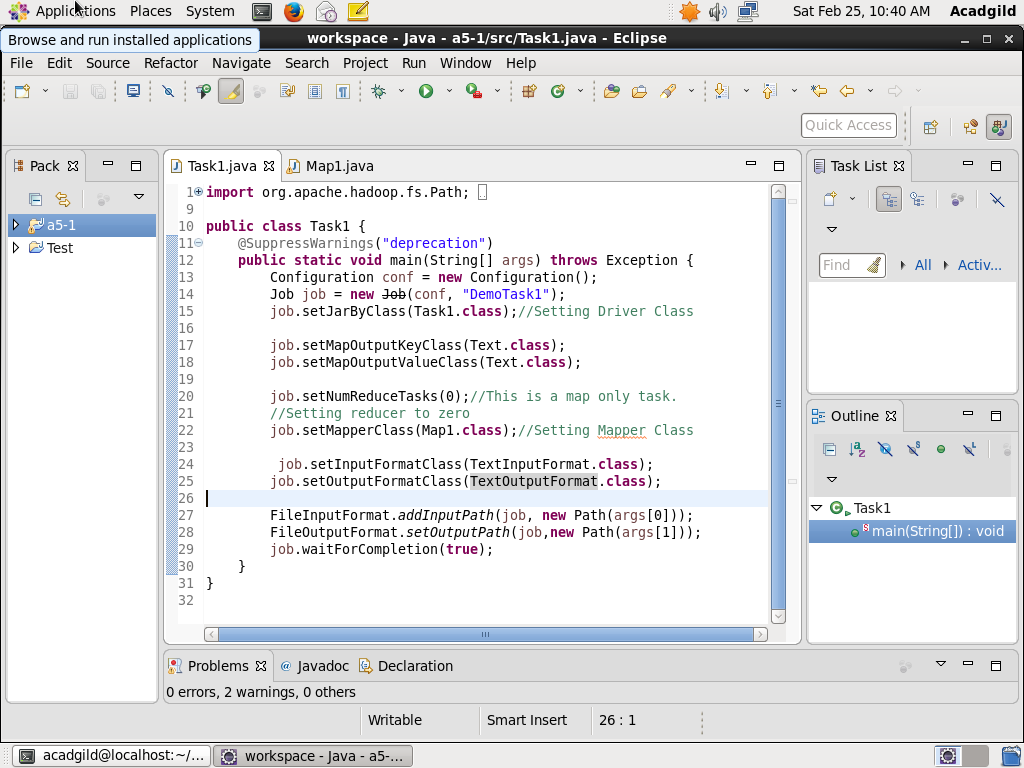
Lava|Attention|20|Assam|454601|24200

Samsung|Super|14|Maharashtra|619082|9200

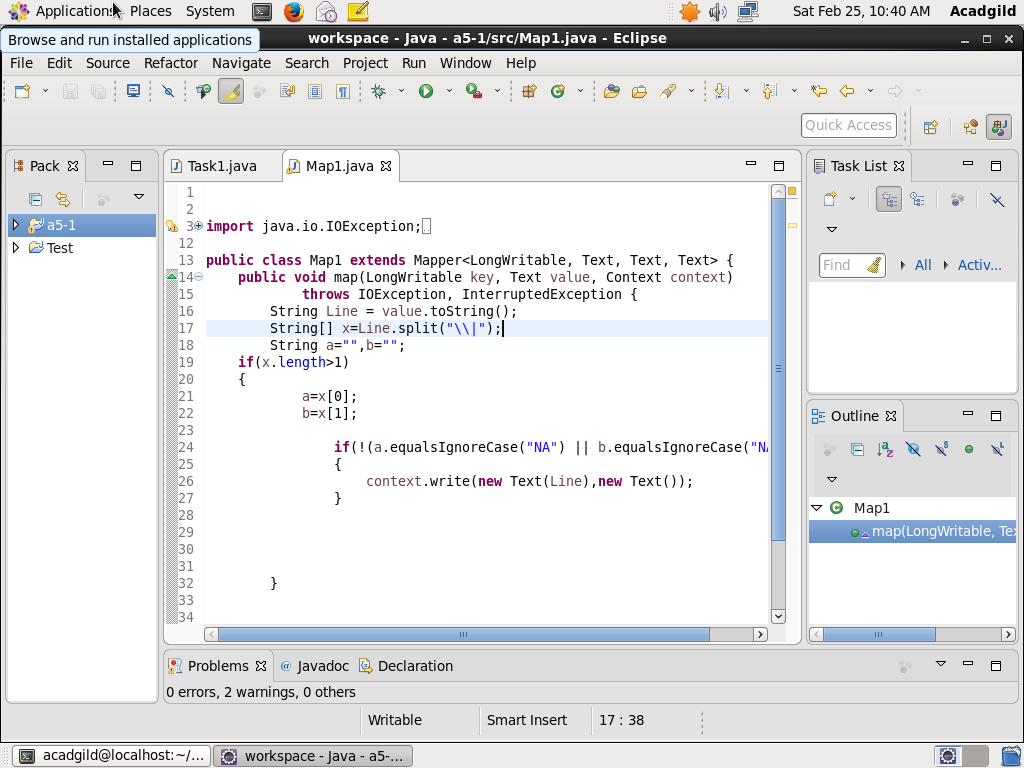
Samsung|Super|14|Maharashtra|619082|9200

Samsung|Super|14|Maharashtra|619082|9200

**PROGRAM:**

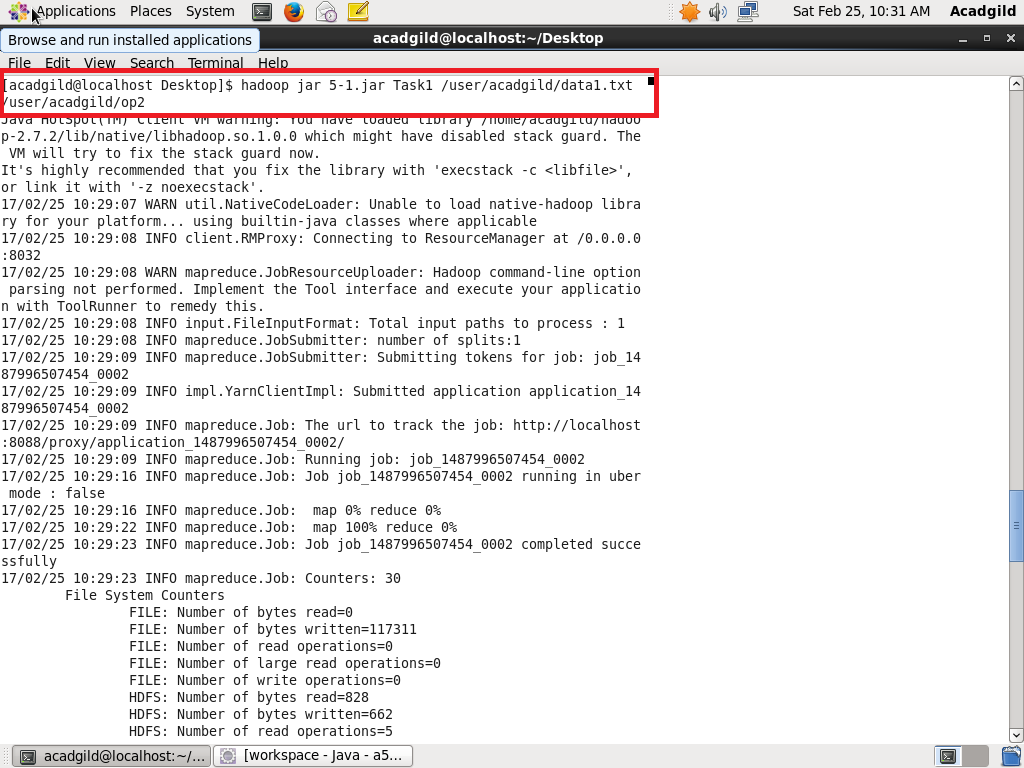


This is the driver class where all the configurations like setting the Mapper Class, Reducer Class , input format, output format etc. are set. Since this is a map only job, the reducer tasks is set to zero(line 20).

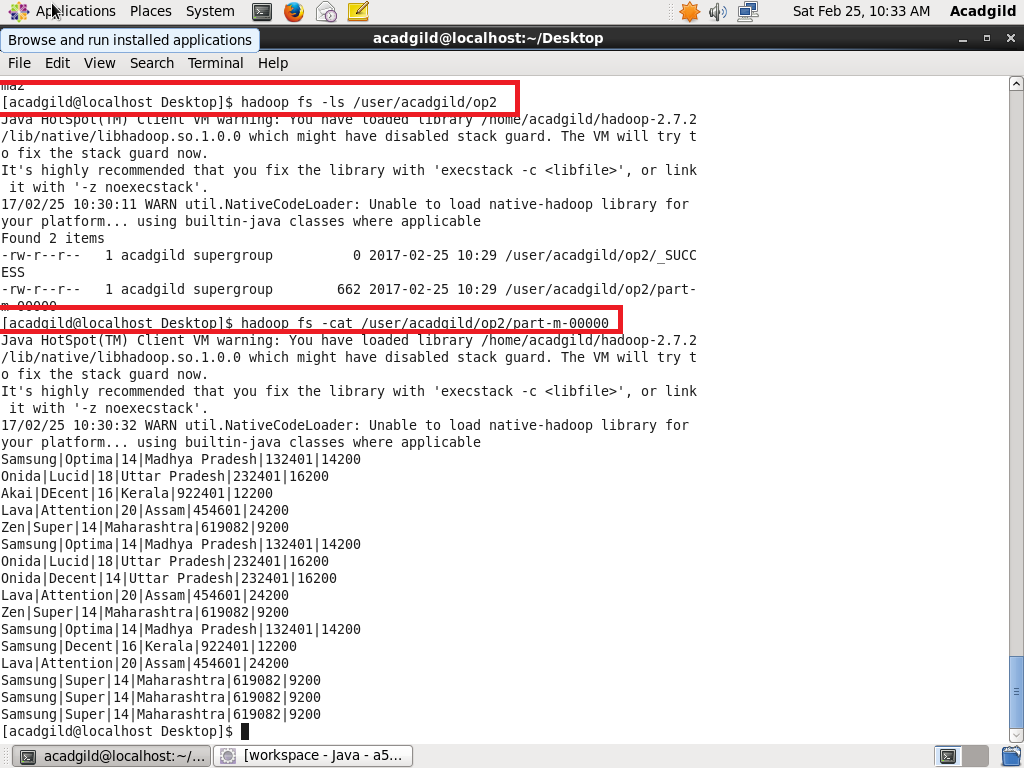


This is the map class which segregates the data. It extends the Mapper class and overrides the Map method.

**OUTPUT:**

****

Executing the jar file and storing the output in /user/acadgild/op2



1. Checking whether the directory /user/acadgild/op2 exists using ls command

2. Displaying the contents of file part-m-00000 which contains output using cat command. As seen in the image above, all records containing NA have been eliminated.