**Assignment 3.3**

**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.

2. Write a Map Reduce program to calculate the total units sold for each Company.

3. Write a Map Reduce program to calculate the total units sold in each state for Onida

company.

**Input File**

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

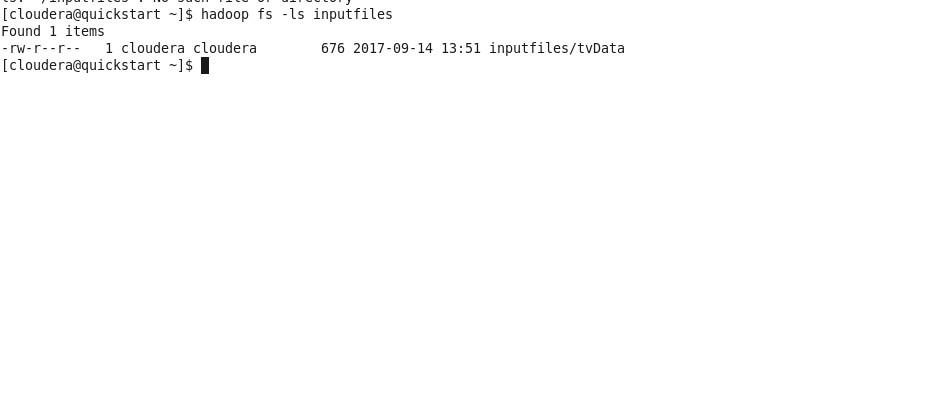
Lava|Attention|20|Assam|454601|24200

Samsung|Super|14|Maharashtra|619082|9200

Samsung|Super|14|Maharashtra|619082|9200

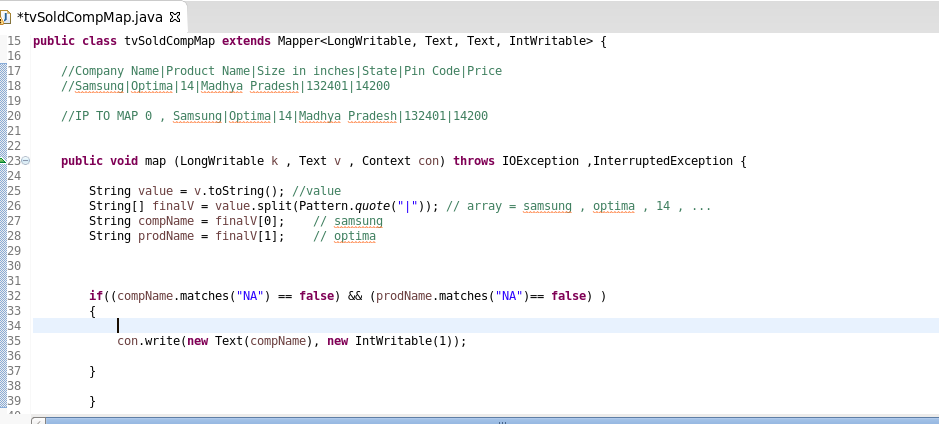
Samsung|Super|14|Maharashtra|619082|9200

**We have put the input file in the HDFS folder /inputfiles/tvData**



**2)** **Write a Map Reduce program to calculate the total units sold for each Company**

**Mapper Code :**



We have first of all filter records with bad records containing NA

We are writing CompName and How many occurences in the Input

So Output from Mapper will be

Samsung ,1

Onida ,1

Samsung ,1

And so on

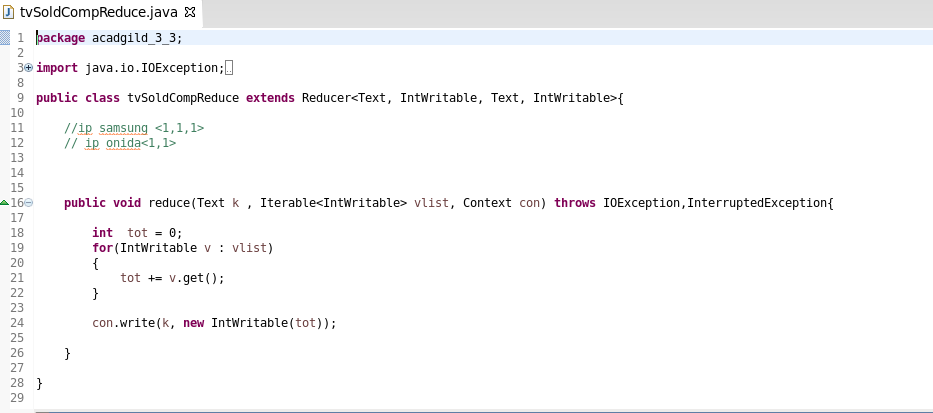
Shuffle And Sort will Happen and then the Records will go to Reducer Program.

**Reducer Code :**

Input to Reducer

Samsung <1,1,1>

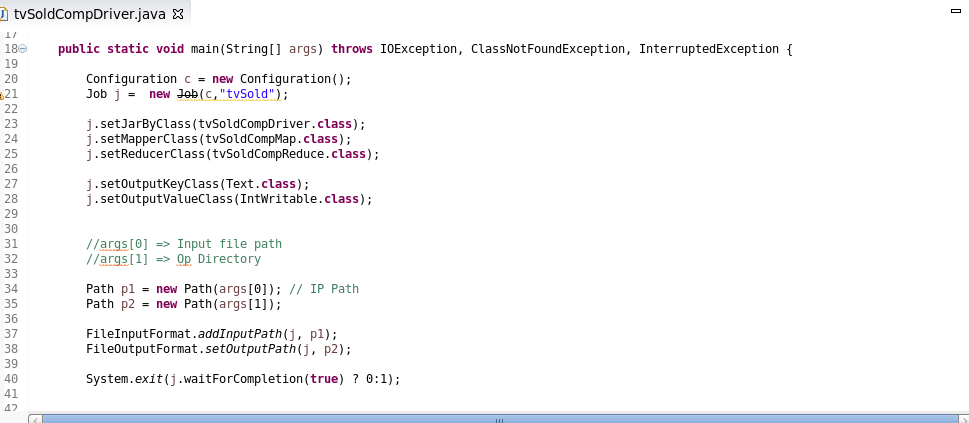
Onida <1,1> and so on



We have Aggregated Units Sold by Each Company.

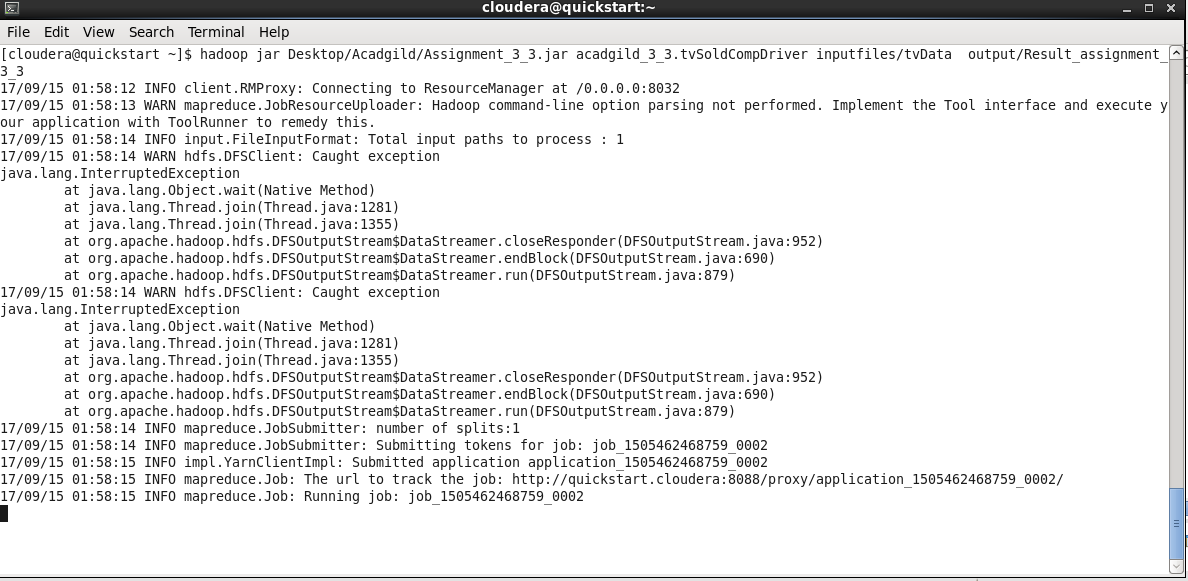
OP will be like Samsung ,3

**Driver Code :**



We have provided Mapper class , driver class and reducer class

**Execution Jar File in Cluster :**



**hadoop jar Desktop/Acadgild/Assignment\_3\_3.jar acadgild\_3\_3.tvSoldCompDriver inputfiles/tvData output/Result\_assignment\_3\_3**

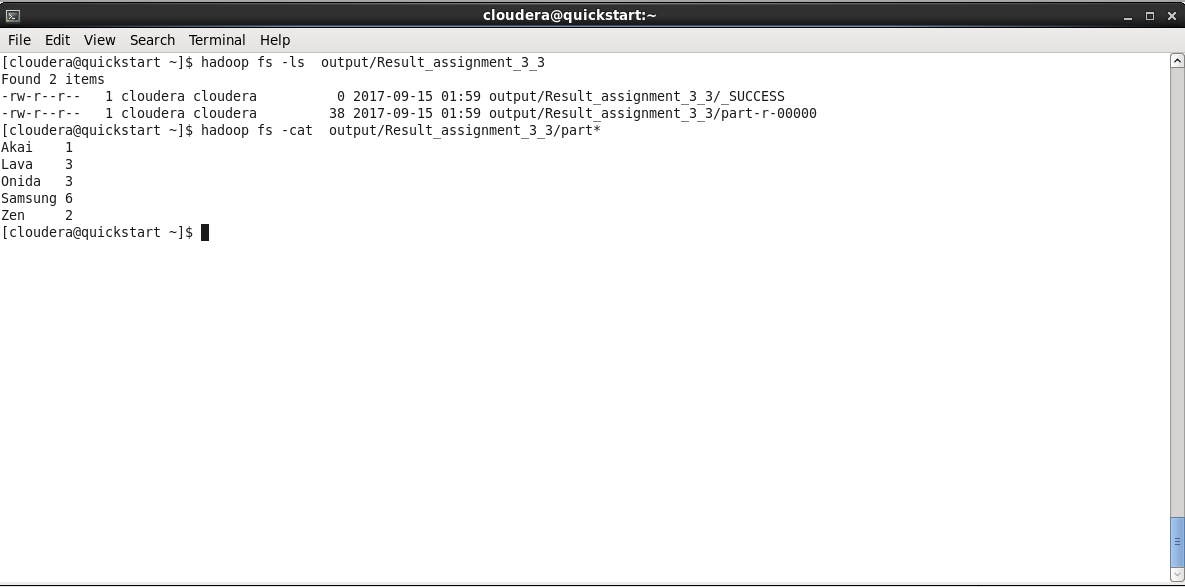
**Desktop/Acadgild/Assignment\_3\_3.jar :**  Jar file location which we want to run

**acadgild\_3\_3. tvSoldCompDriver:** Driver Class , where the main function resides

**inputfiles/tvData :** Input File

**output/Result\_assignment\_3\_3:** Output file or result Location

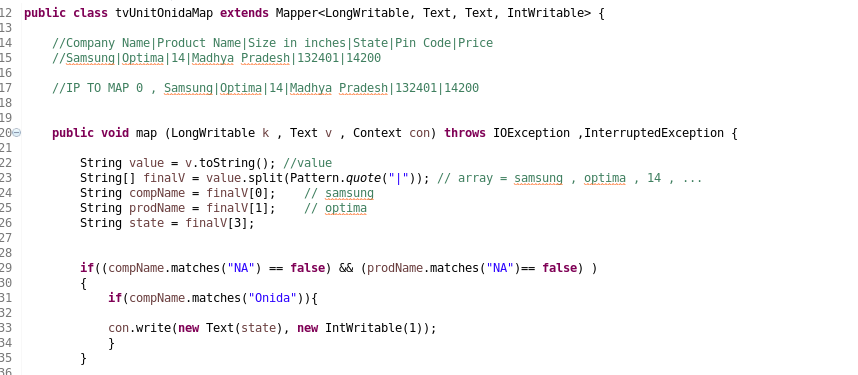
**Output Record**



**3) Write a Map Reduce program to calculate the total units sold in each state for Onida**

**company.**

**Mapper Code :**



In this Mapper code we are filtering the Records which have “NA” bad records then we are filtering the Column based on the Company .

So we are taking the records only which Compname is Onida

After Filtering we are sending the States with each occurrence Output of mapper will be something like this

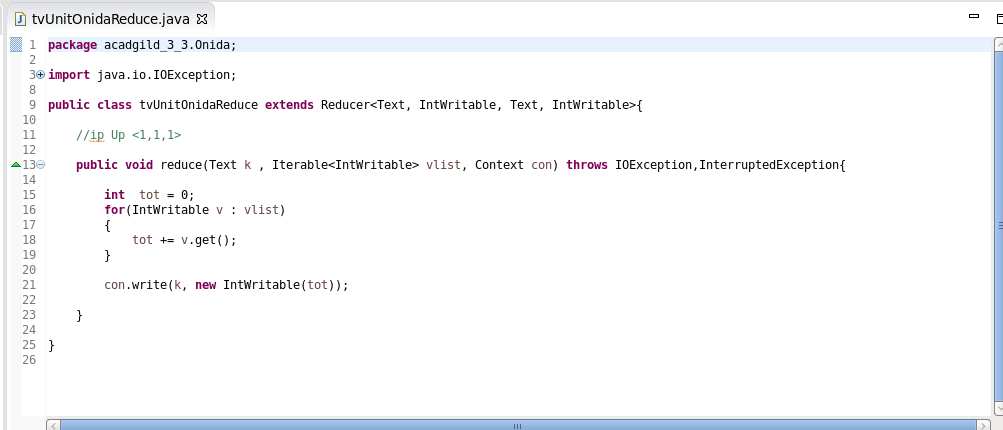
UP 1

UP 1  
 UP1

**Reducer Code :**

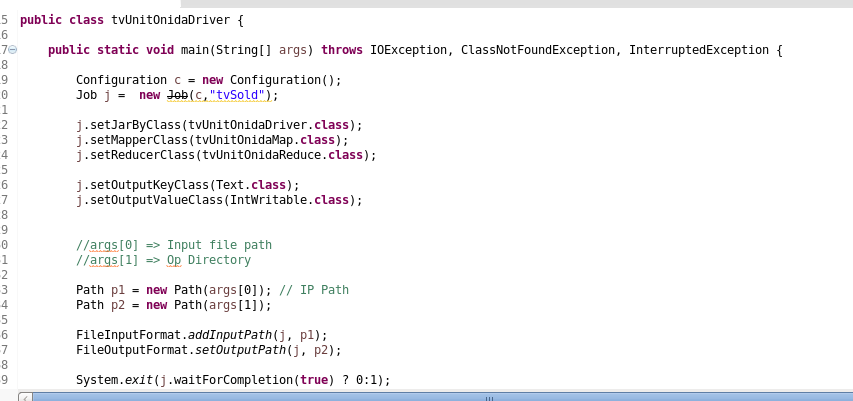
Input for Reducer Code after shuffle and sort

UP , <1,1,1>



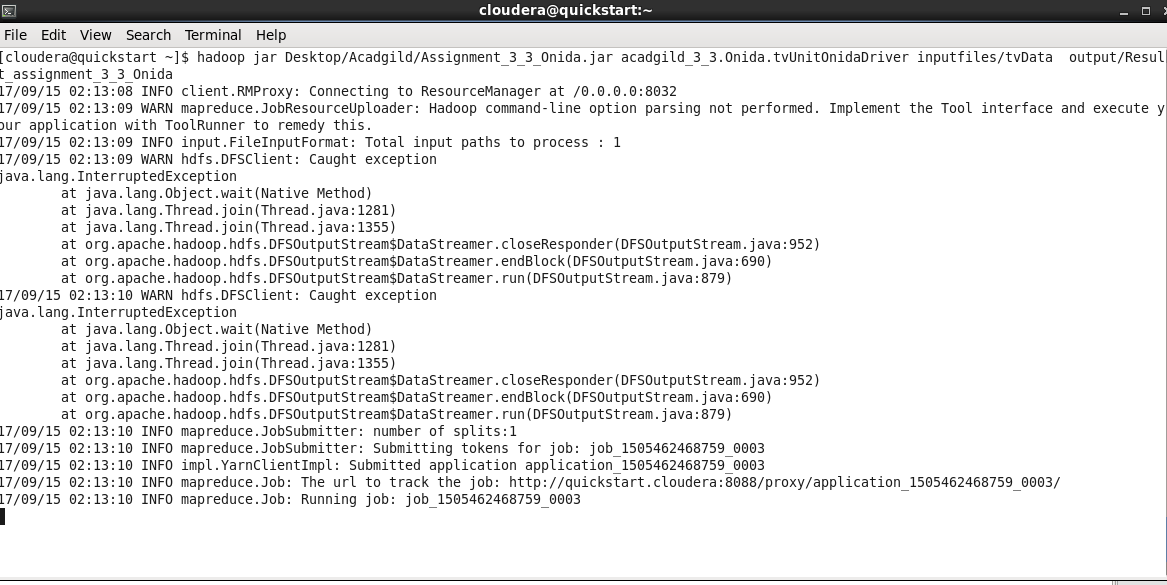
We are doing Sum on basis of the key State which we got from the Mapper

**Driver Code**



We have provided the Mapper , reducer and driver class

**Running the Jar**



h**adoop jar Desktop/Acadgild/Assignment\_3\_3\_Onida.jar acadgild\_3\_3.Onida.tvUnitOnidaDriver inputfiles/tvData output/Result\_assignment\_3\_3\_Onida**

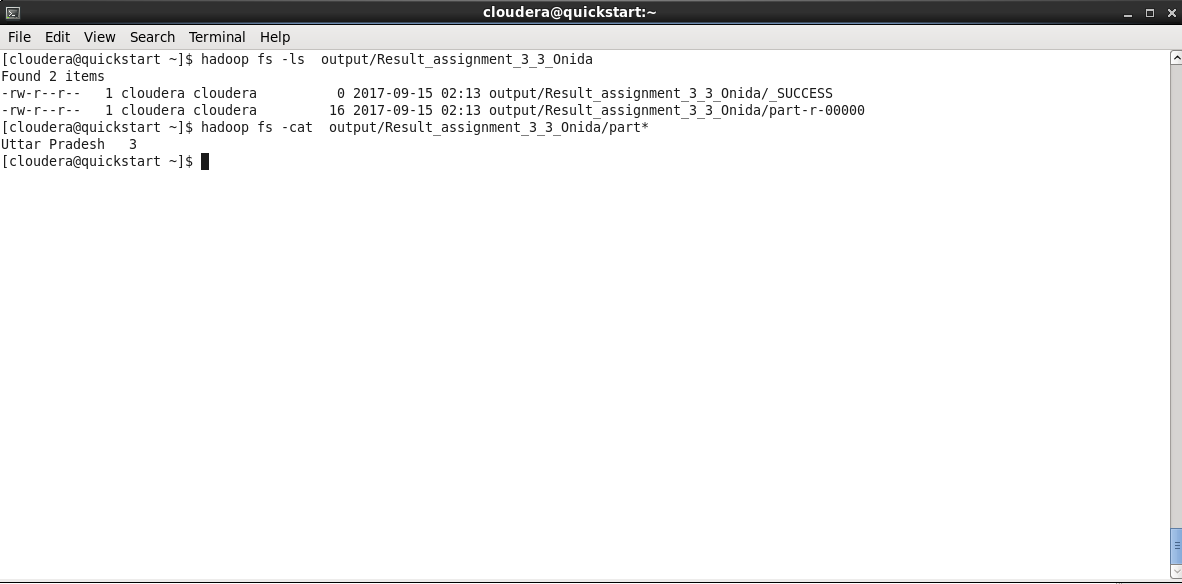
**Desktop/Acadgild/Assignment\_3\_3\_Onida.jar:**  Jar file location which we want to run

**acadgild\_3\_3.Onida.tvUnitOnidaDriver:** Driver Class , where the main function resides

**inputfiles/tvData :** Input File

**Result\_assignment\_3\_3\_Onida:** Output file or result Location

**Output :**



NOTES : Github will contain 2 jars

1. Two Jars
2. Document flow of exceution