**SESSION – 3**

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:

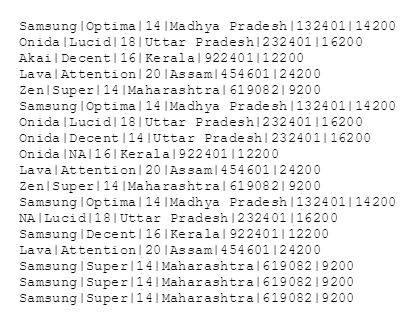
**Company Name|ProductName|Size in inches|State|PinCode|Price**

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

1. **Write a Map Reduce program to calculate the total units sold for each Company.**
2. **Write a Map Reduce program to calculate the total units sold in each state for Onida Company.**

**Data Files:**

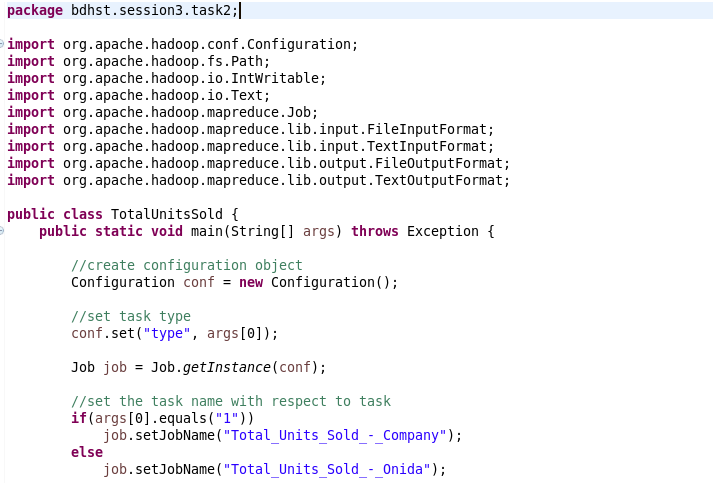
**Screenshot:**



**CODE:**

**Driver Class:**

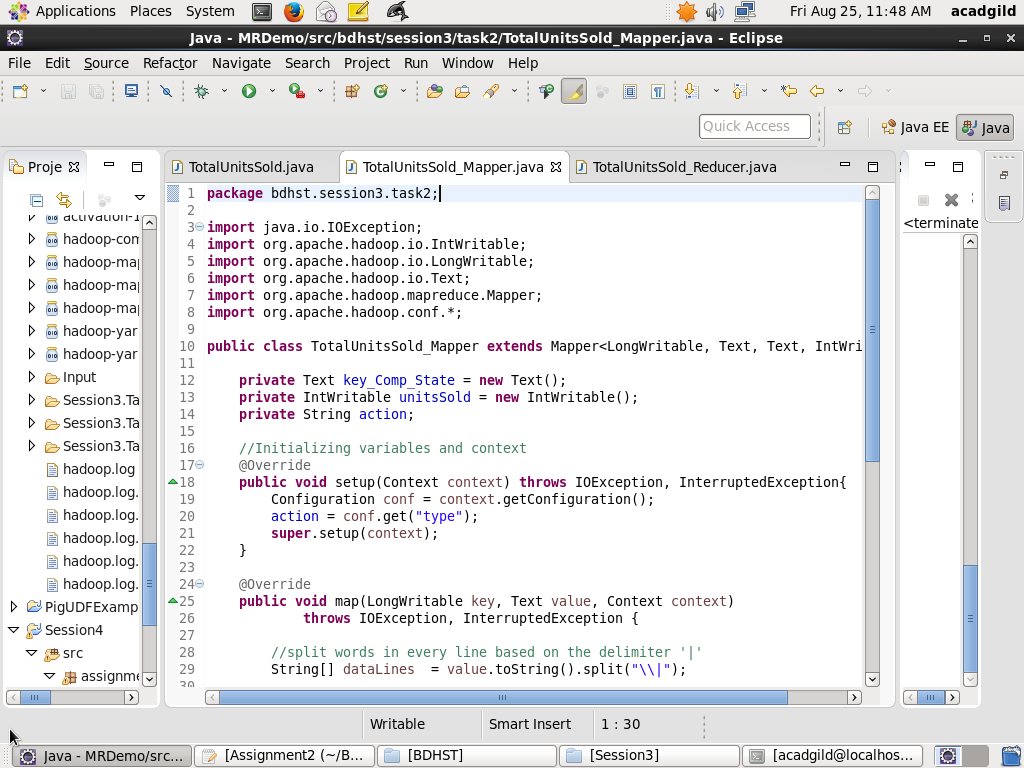
**TotalUnitsSold.java**

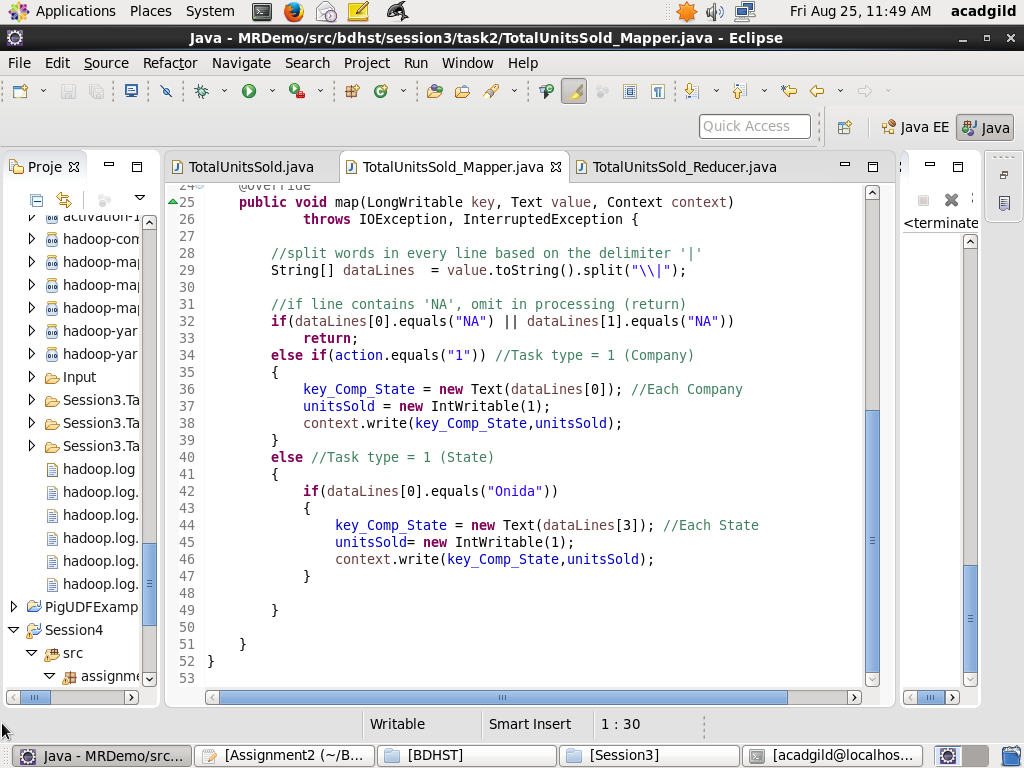


Txcvxcvxcv

**Mapper Class:**

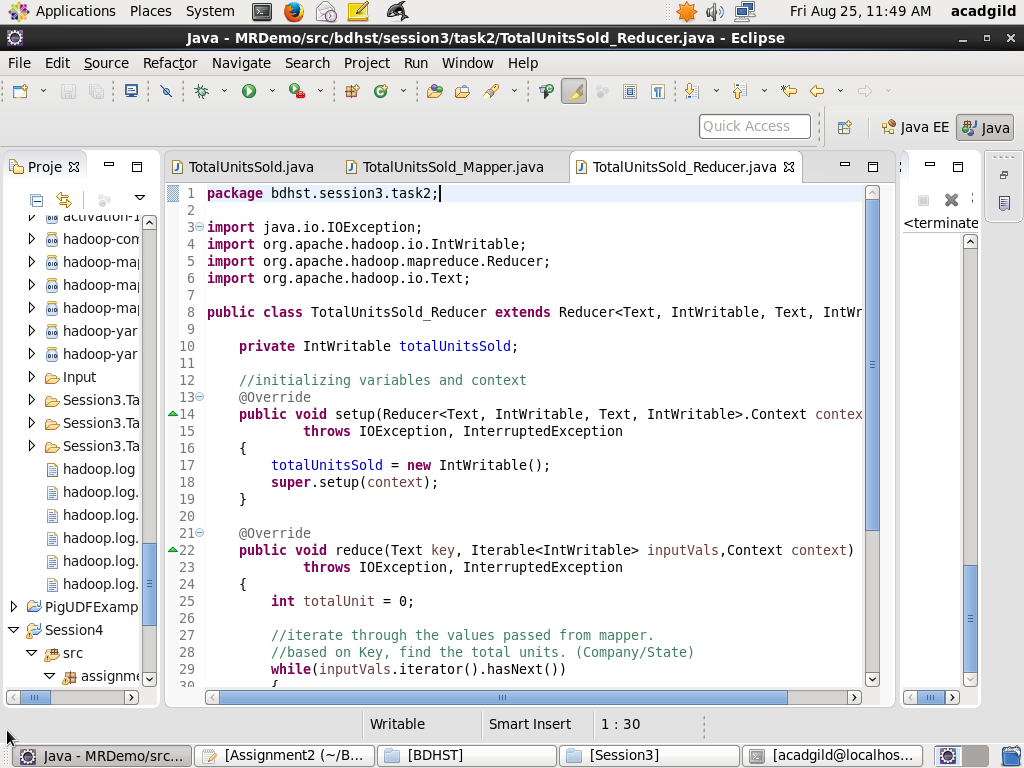
**TotalUnitsSold\_Mapper.java**

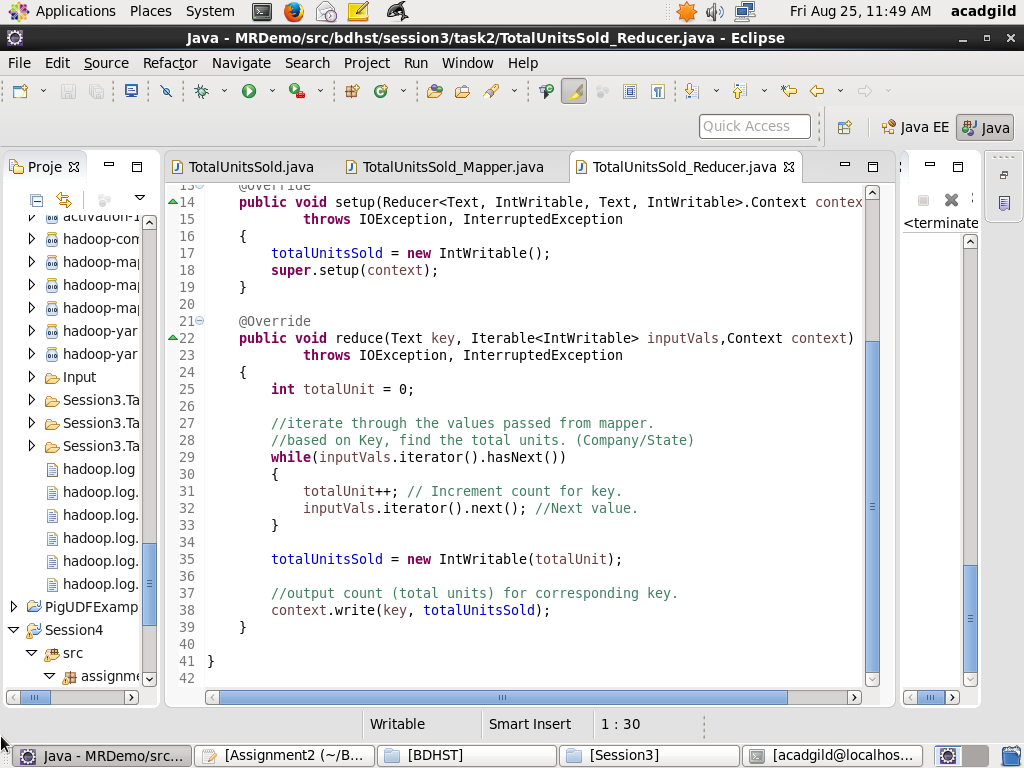


****

**Reducer Class:**

**TotalUnitsSold\_Reducer.java**

****

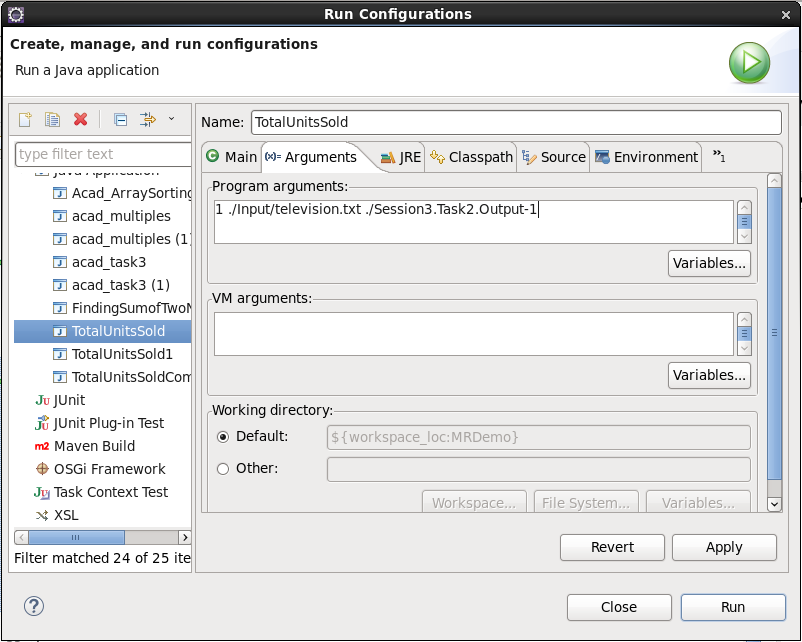
****

**IDE Execution and Export to .jar File.**

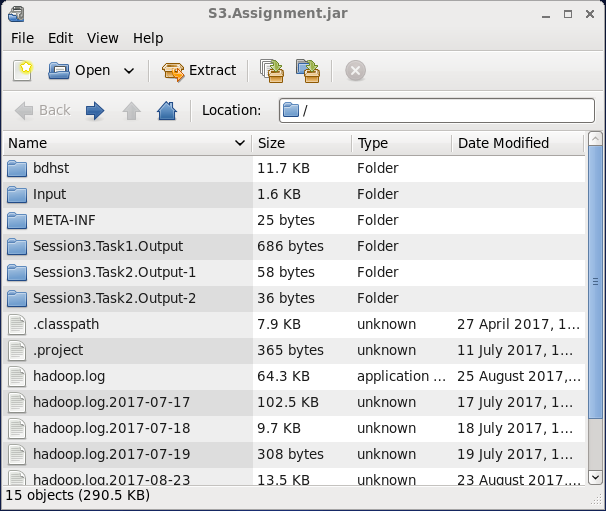
* Executing the program in the Eclipse IDE.

The Arguments list referred in the program is of the form:

*[Task to Execute (1/2)] [Input Dataset path in the IDE] [Output path in IDE]*

****

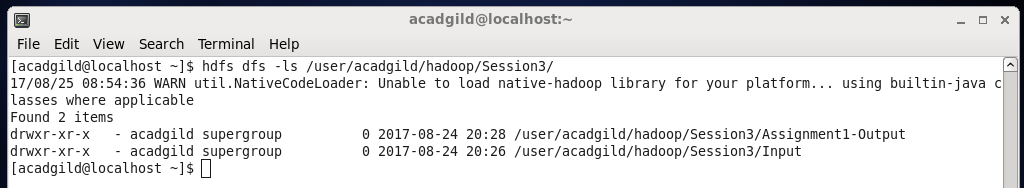
* The .jar file is exported from the IDE to a path in the local FS; to be moved to the HDFS.

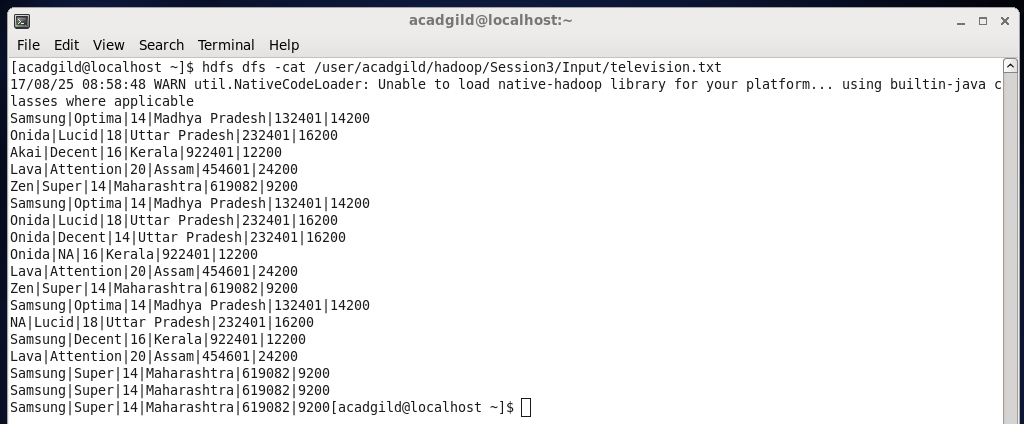
****

**OUTPUT for Task1:**Calculate the Total Units sold for each Company

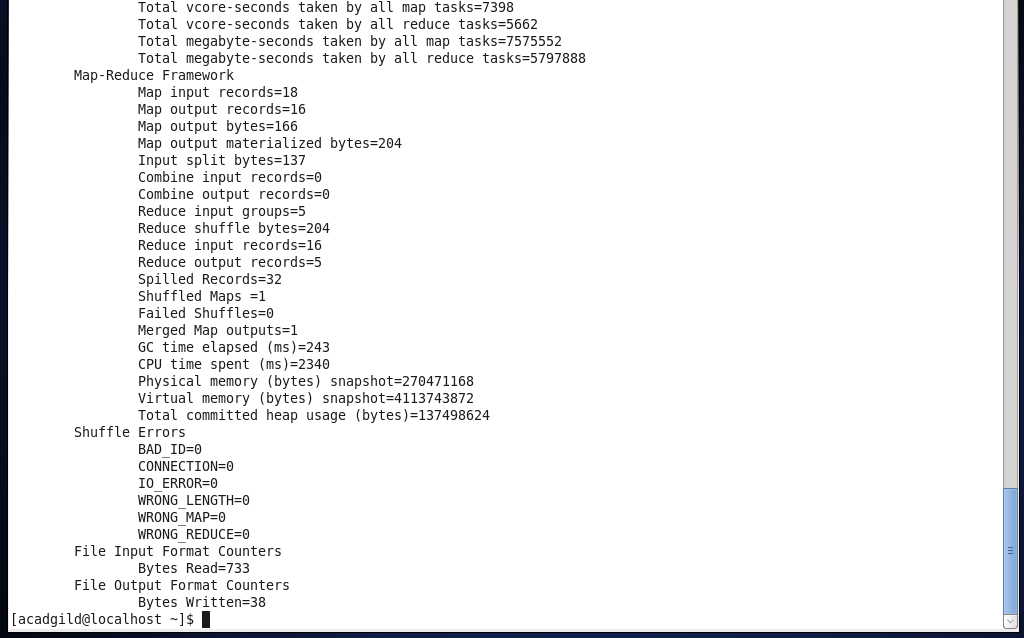
1. Listing contents of ~/Session3 that holds the Input Dataset to be used in the program.
2. Displaying the contents of the Input Dataset in the HDFS.
3. The .jar file containing the Program to find the Total Units sold for each Company and each state for Onida using the foll. format:

Hadoop jar *[Path of .jar file location in local FS] [Path of Main Class in jar] [Task to execute (1-Company, 2-Onida)] [Path of Input Dataset] [Path in HDFS for Output Storage]*

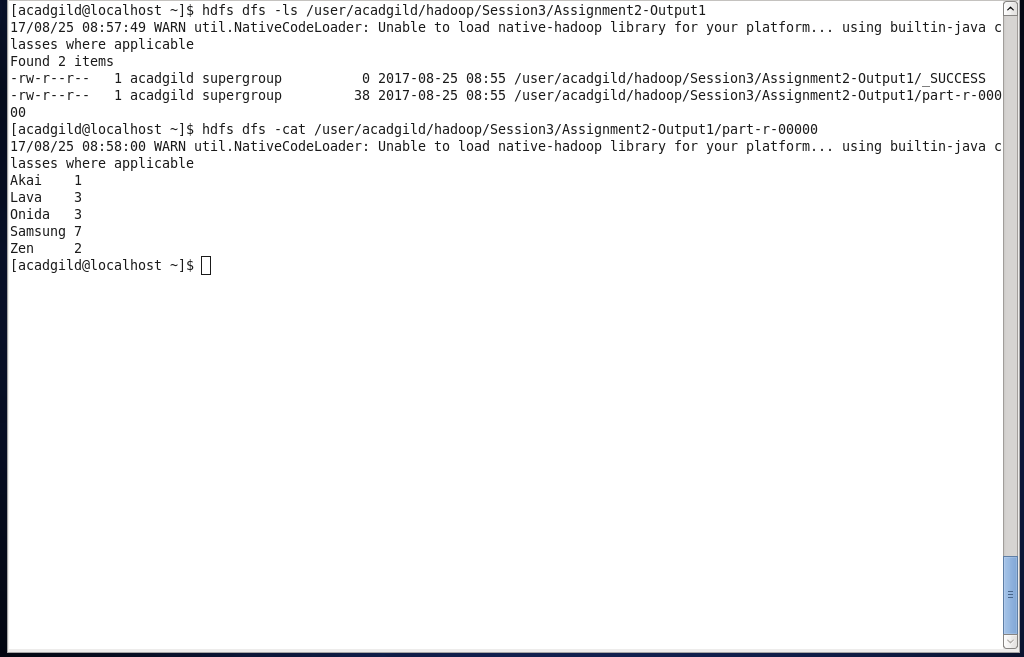
****

****

****

****

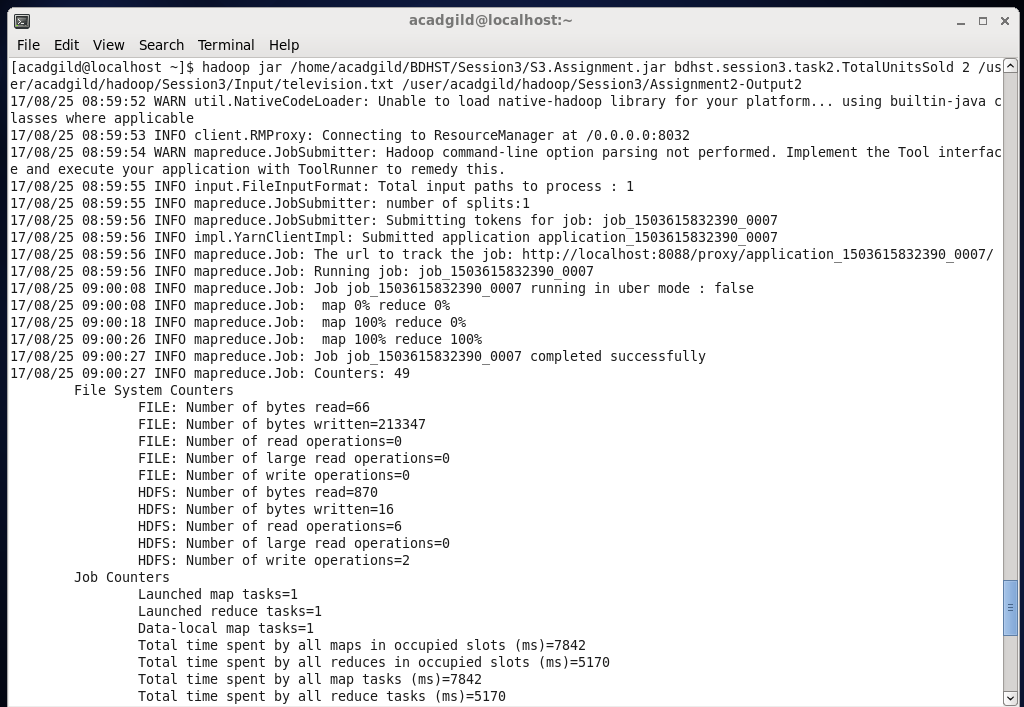
1. Displaying contents of Output folder (Assignent2-Output1).
2. Displaying contents of Result file (part-r-00000) in Assignent2-Output1.

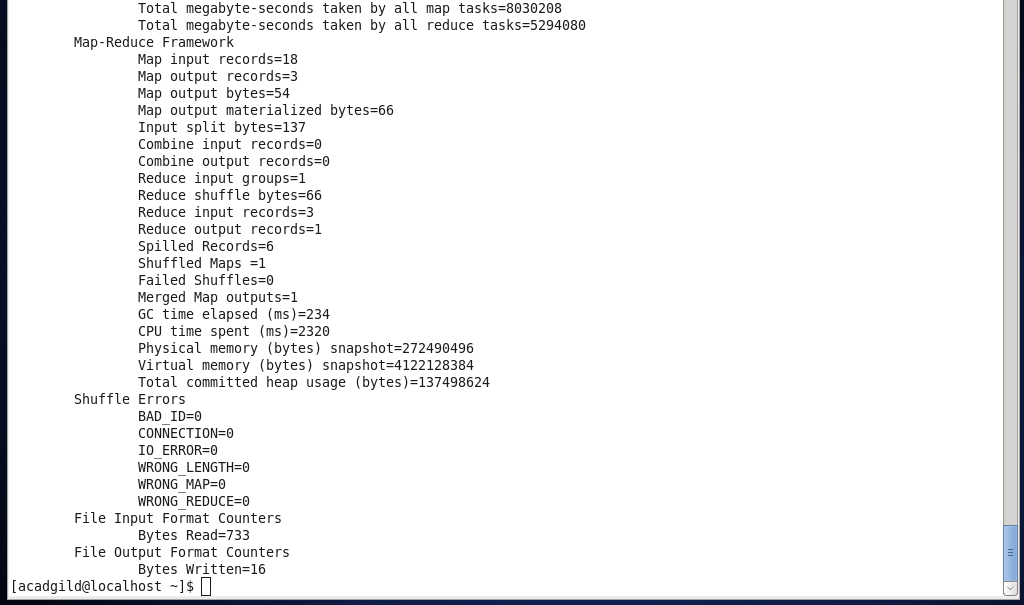
****

**OUTPUT for Task2:**Calculate the Total Units sold for Onida in each state.

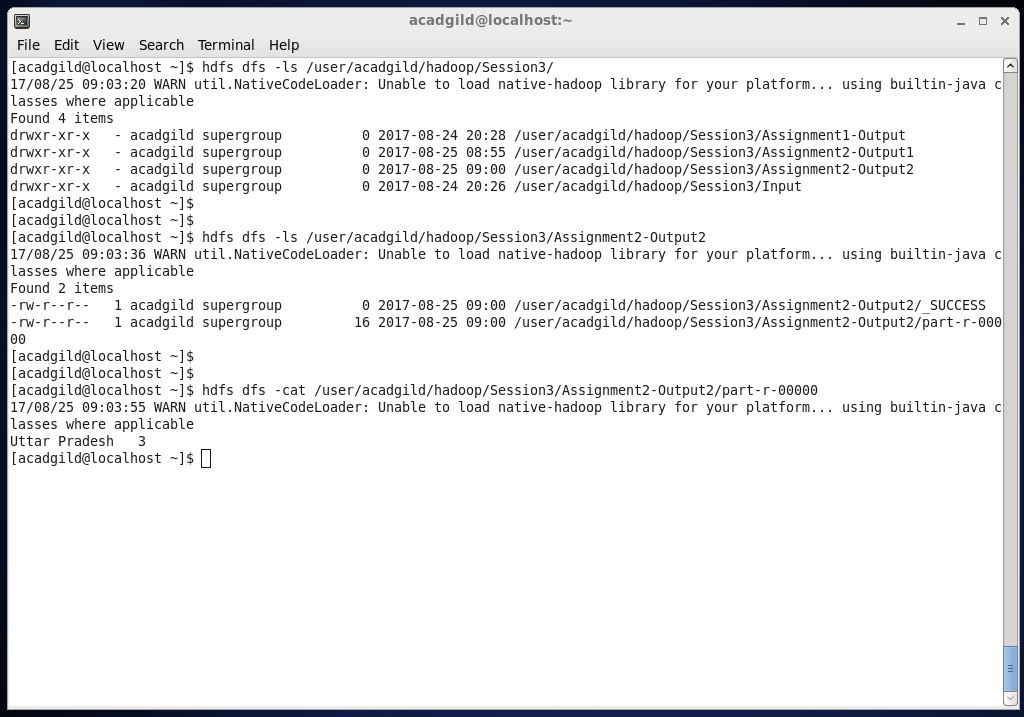
1. The .jar file containing the Program to find the Total Units sold for each Company and each state for Onida using the foll. format:

Hadoop jar *[Path of .jar file location in local FS] [Path of Main Class in jar] [Task to execute (1-Company, 2-Onida)] [Path of Input Dataset] [Path in HDFS for Output Storage]*

****

****

1. Listing contents of ~/Session3 that holds the Input Dataset to be used in the program.
2. Displaying contents of Output folder (Assignment2-Output2).
3. Displaying contents of Result file (part-r-00000) in Assignment2-Output2.

****