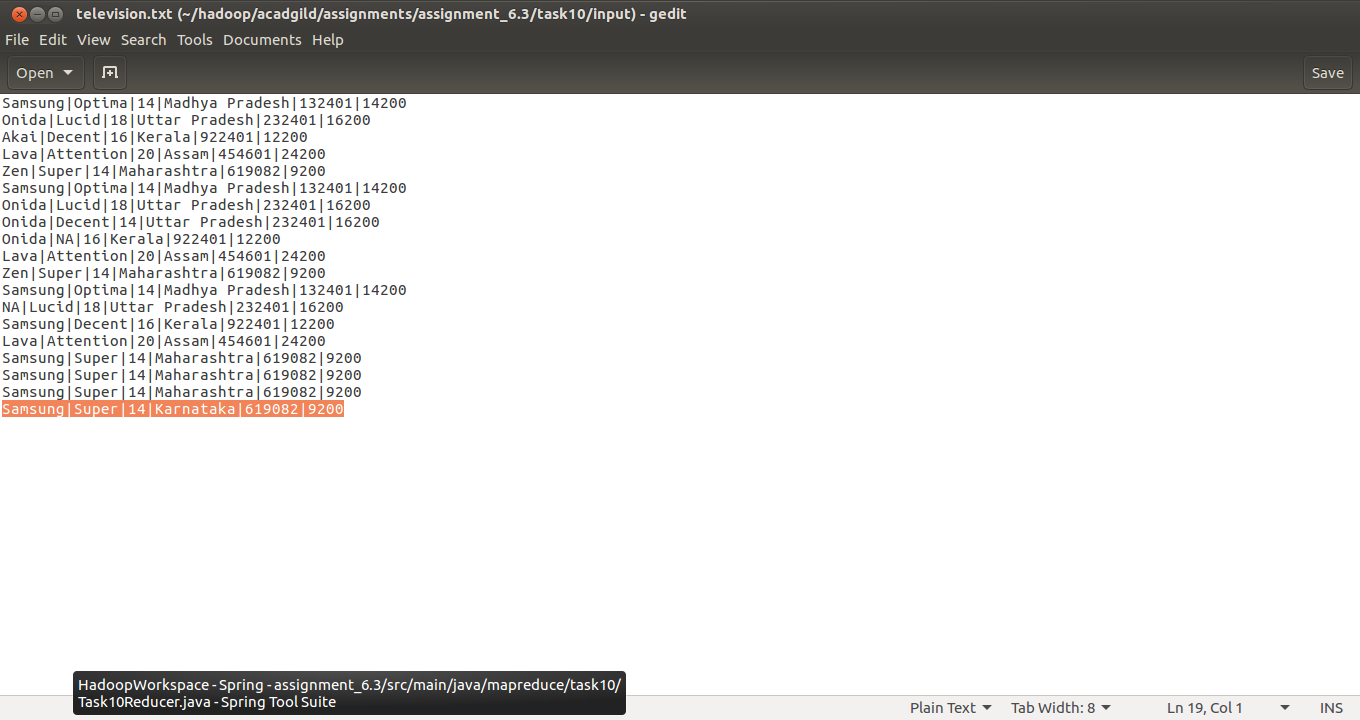
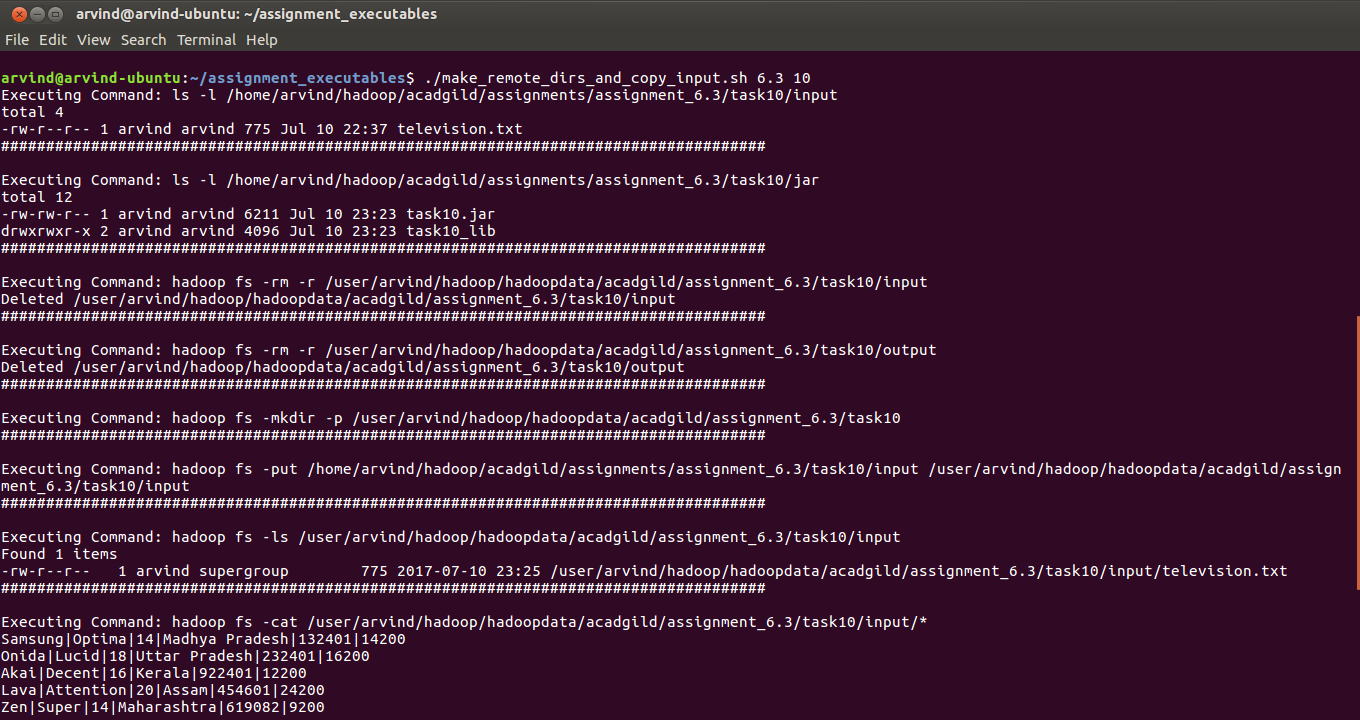
In this task, we are reading television sales data and calculating the top 3 statewise sales for each company

**make\_remote\_dirs\_and\_copy\_input.sh take 2 inputs, the assignment number and task number..**

1. It verifies if the input file and the jar for the assignment are kept in the appropriate location in the local file system using ls -l command.
2. Delete input and output directories in the hdfs for this assignment and task if already exists using hadoop fs -rm -r command(delete the folder and all its contents recursively)
3. Create a directory in HDFS for this assignment and task using hadoop fs -mkdir -p command
4. Copy the input from local file system to hdfs using hadoop fs -put command
5. Finally verify using hadoop fs -ls command if the input file has been copied into the HDFS

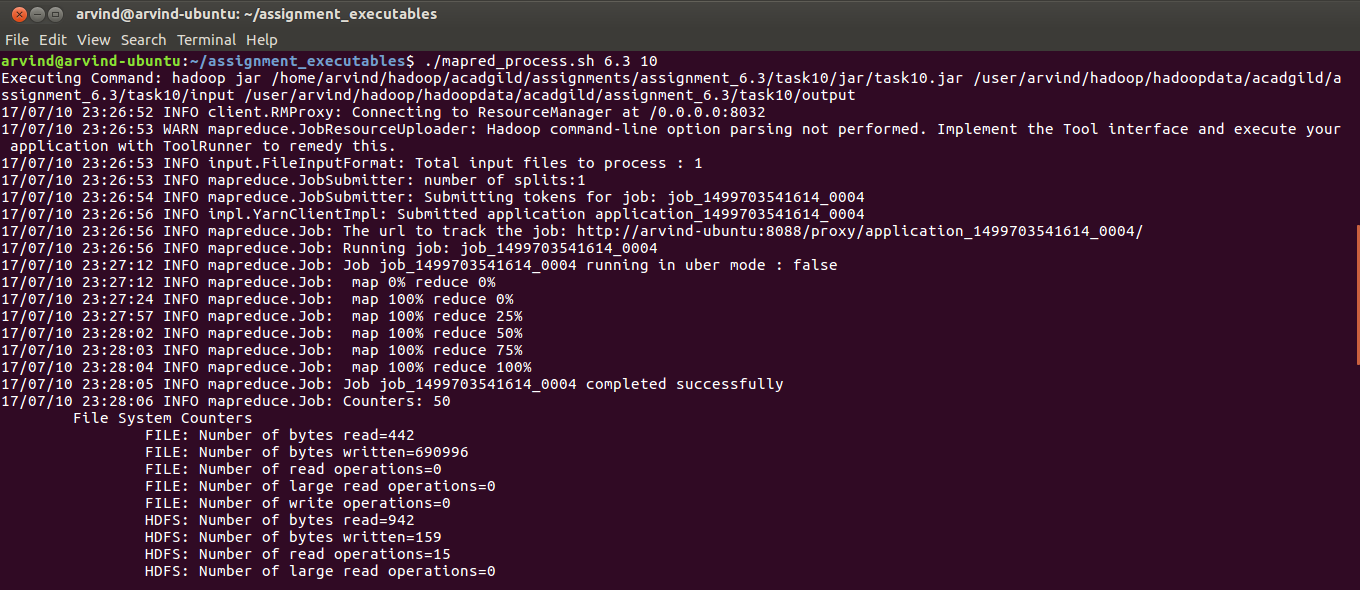
Note: As highlighted in the screenshot below all companies have sales data for a maximum of 3 states only. So, I introdcued a dummy row for Samsung. The state this time is Karnataka. So now Samsung has data for 4 states. We will later see in the output that the data for Karnataka is not displayed because we are displaying only the top 3 states for each company





**mapred\_process also takes assignment number and task number as input.**

1. It executes the JAR from the local file system using hadoop jar command

****

**output.sh also takes assignment number and task number as input**

1. It lists all the files in the output directory for this assignment and task job in HDFS. (hadoop fs -ls)
2. It displays the content of all the output files (hadoop fs -cat)
3. Finally it copies the output files from HDFS to local file system

Note: For Samsung data for only 3 states are displayed. That is because Karnataka is not among the top 3 states for Samsung sales.

