**BigData And Hadoop**

**Task 6 of Assignment 2 of Session 5**

**Problem Statement :**

Given a dataset of sales of different TV sets across different locations, with records in the format :

Company Name|Product Name|Size in inches|State|Pin Code|Price

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

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

Write a Map Reduce program to calculate the total units sold in each state for Onida company. Take advantage of the combiner.

**Solution**

**Code files are as follows :**

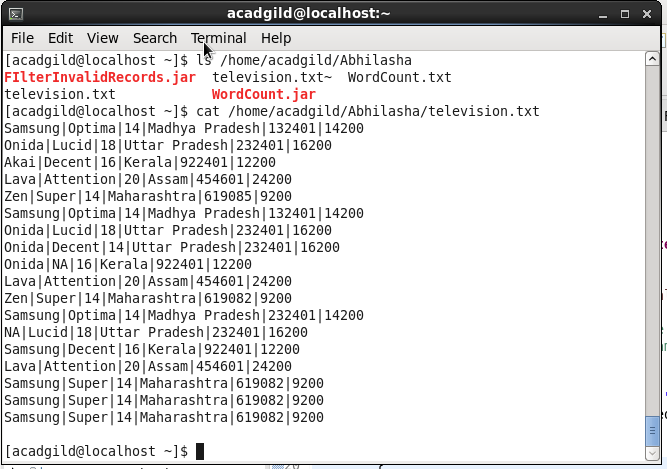
Mapper class : OnidaSalesMapper.java

Reducer class : OnidaSalesReducer.java

Driver class : OnidaSales.java

**Snapshots of the output are as follows :**

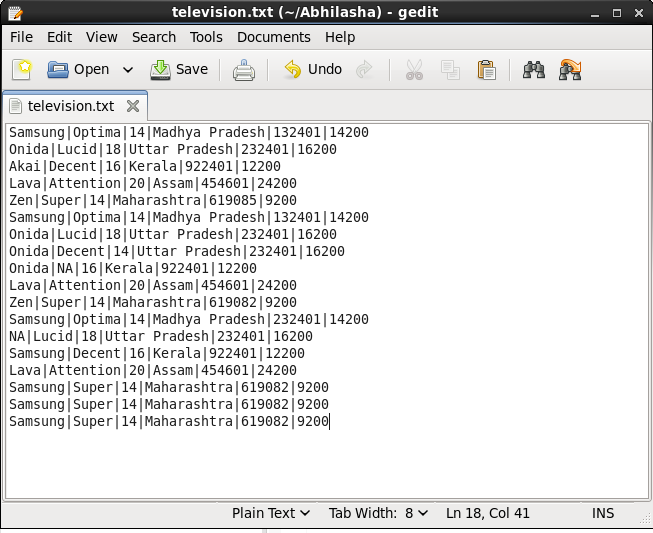
1. Input file present in ‘/home/acadgild/Abhilasha’. Its name is television.txt



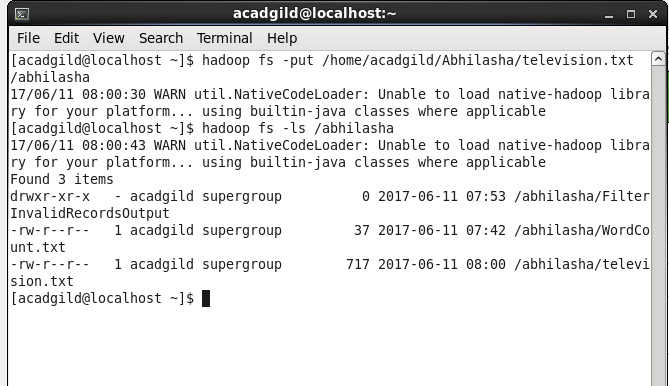
Command used to see the content of the input file

Path where input file is saved locally

1. Input file opened in gedit locally

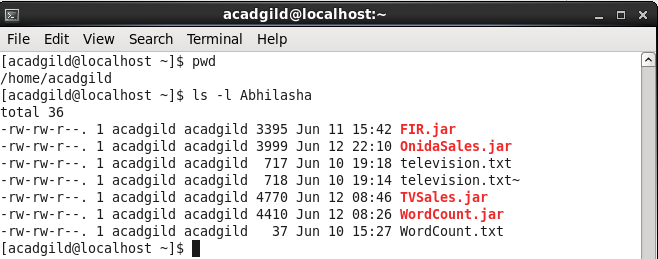


1. Command executed to put television.txt to hdfs system



television.txt stored on hdfs

1. Jar is placed locally at /home/acadgild/Abhilasha. Its name is TVSales.jar



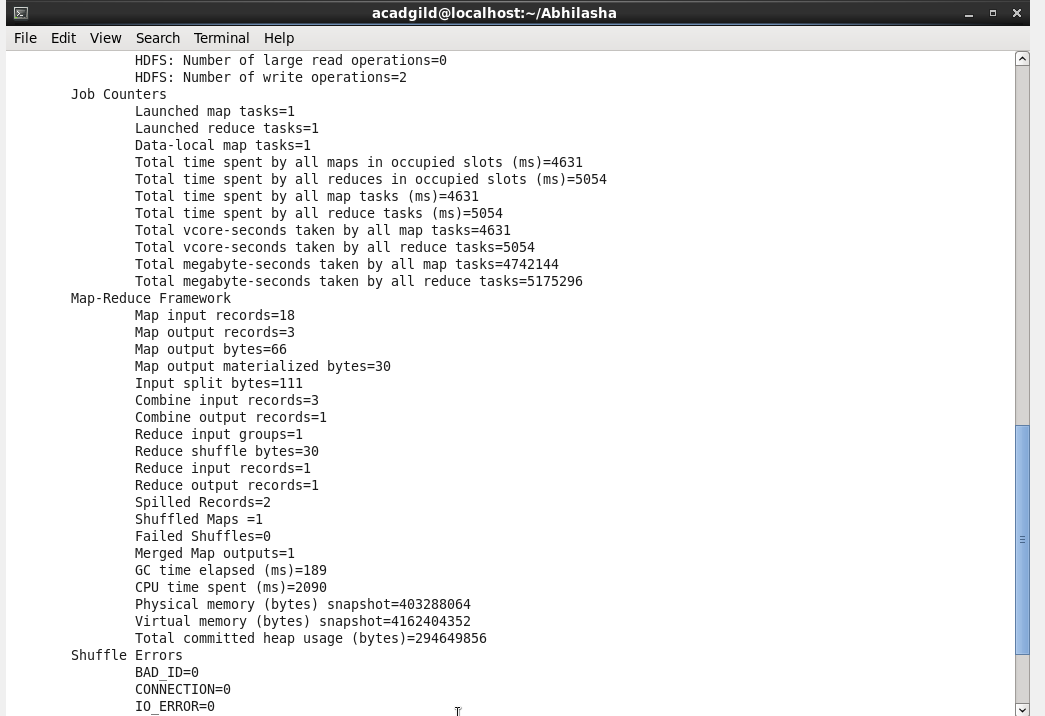
1. Executing the map-reduce program on Hadoop.

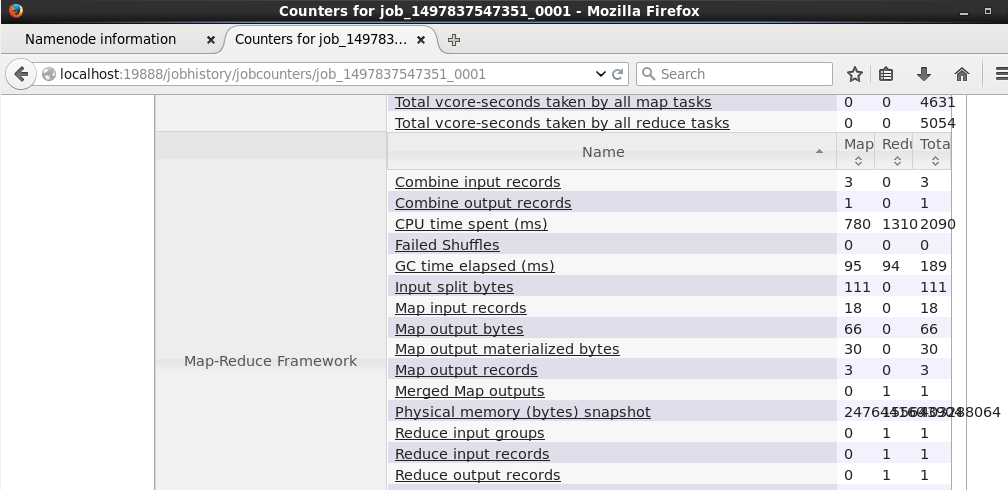


1. Logs mentioning the use of combiner.

If the combiner was not used, number of map output records would have been the number of reduce input records, i.e. 3.

However, the number of input records to the reducer is 1 as the combiner was executed, that performed aggregation of 3 records to give 1 record.



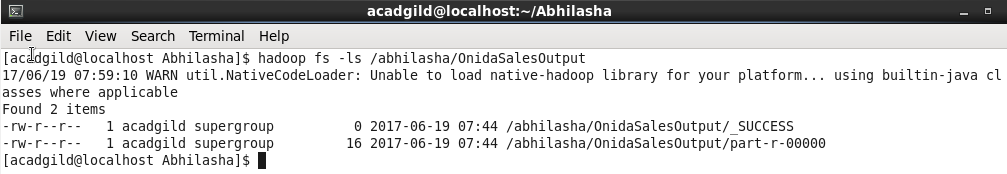


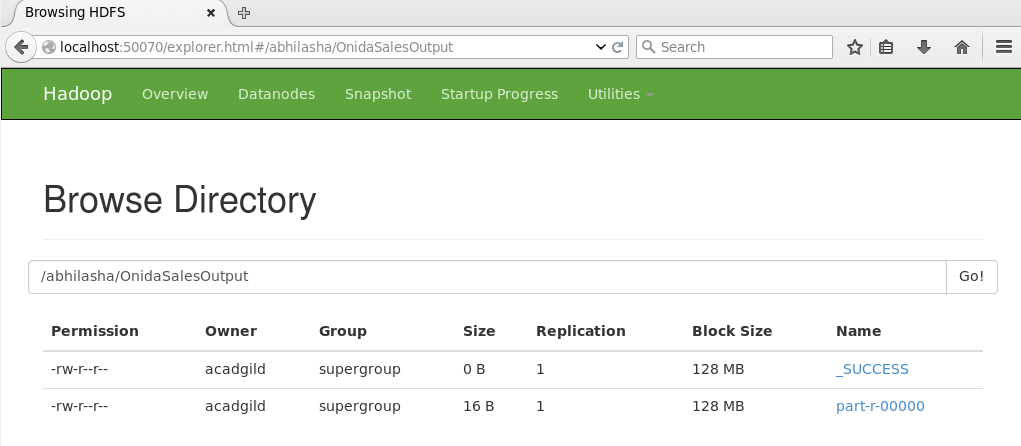
1. Job history server mentioning the success of the job





1. Output folder created on hdfs :





1. Content of output file part-r-00000 on hdfs. Output verified

