**BigData And Hadoop**

**Task 4 of Assignment 1 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 for each Company. Make use of a custom partitioner with 4 reducers. Make sure that all records whose company name starts with A-F (upper or lower case) should go to 1st reducer, those starting with G-L to 2nd reducer, those starting with M-R to 3rd reducer and others to 4th reducer.

**Solution**

**Code files are as follows :**

Mapper class : TelevisionSalesPMapper.java

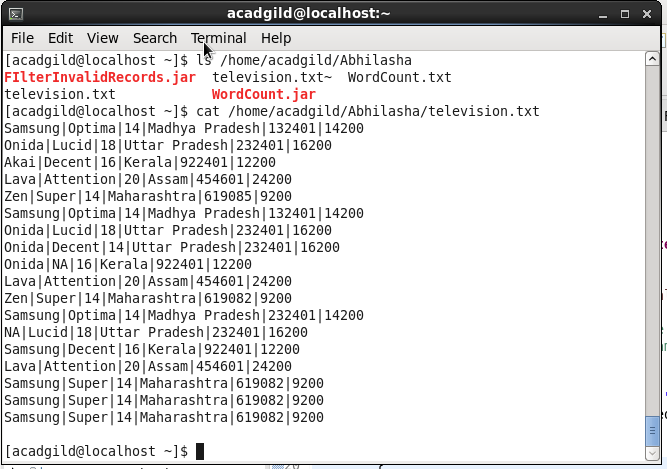
Partitioner class : TelevisionSalesPPartitioner.java

Reducer class : TelevisionSalesPReducer.java

Driver class : TelevisionSalesP.java

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

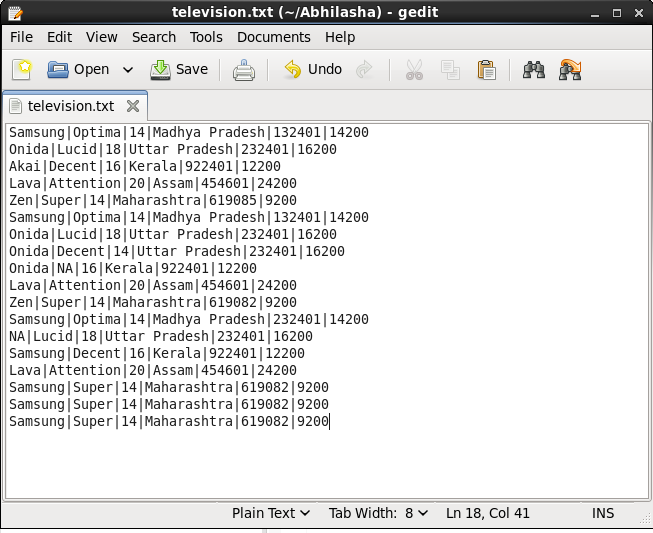
1. Input file present in ‘/home/acadgild/Abhilasha’. Its name is television.txt



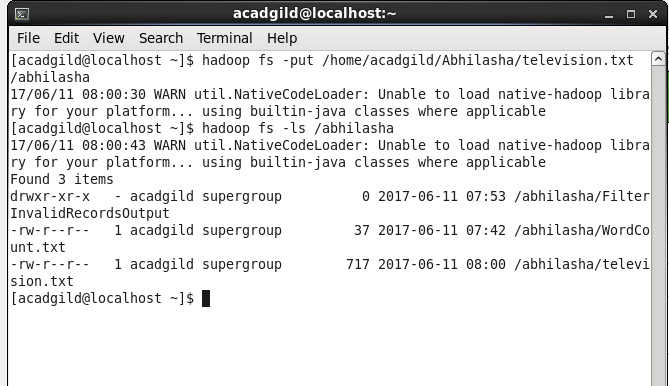
Command used to see the content of the input file

Path where input file is saved locally

1. Input file opened in gedit locally

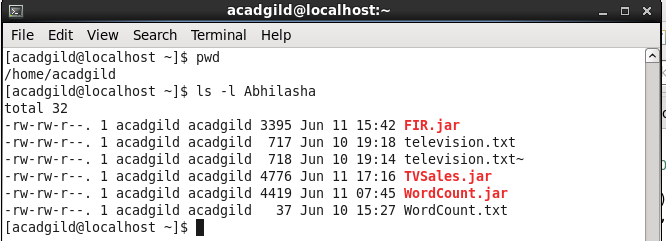


1. Command executed to put television.txt to hdfs system

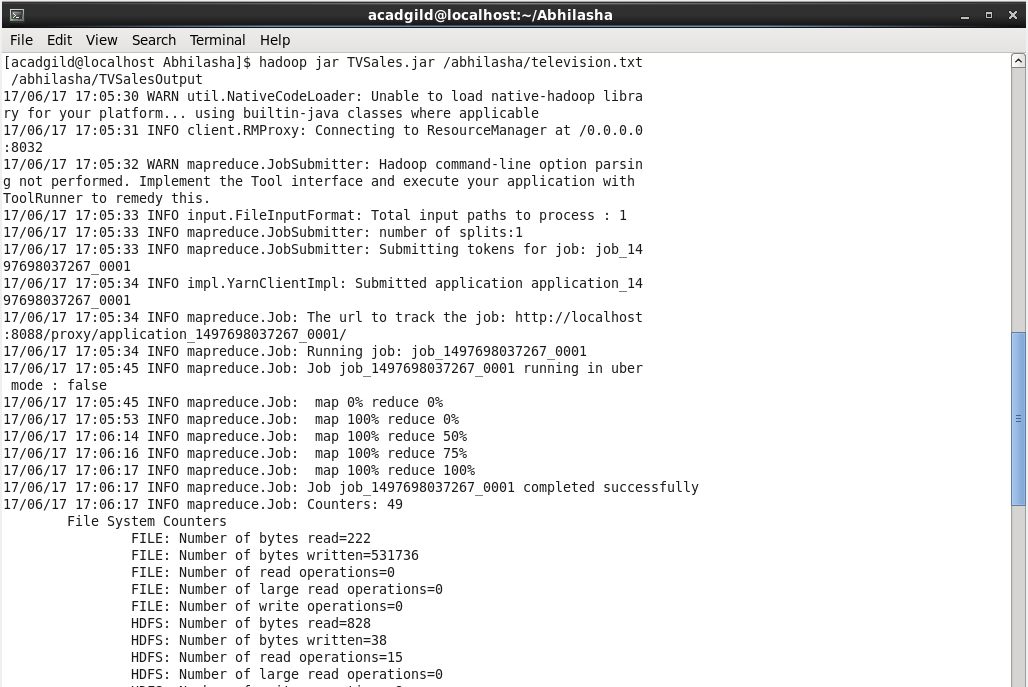


television.txt stored on hdfs

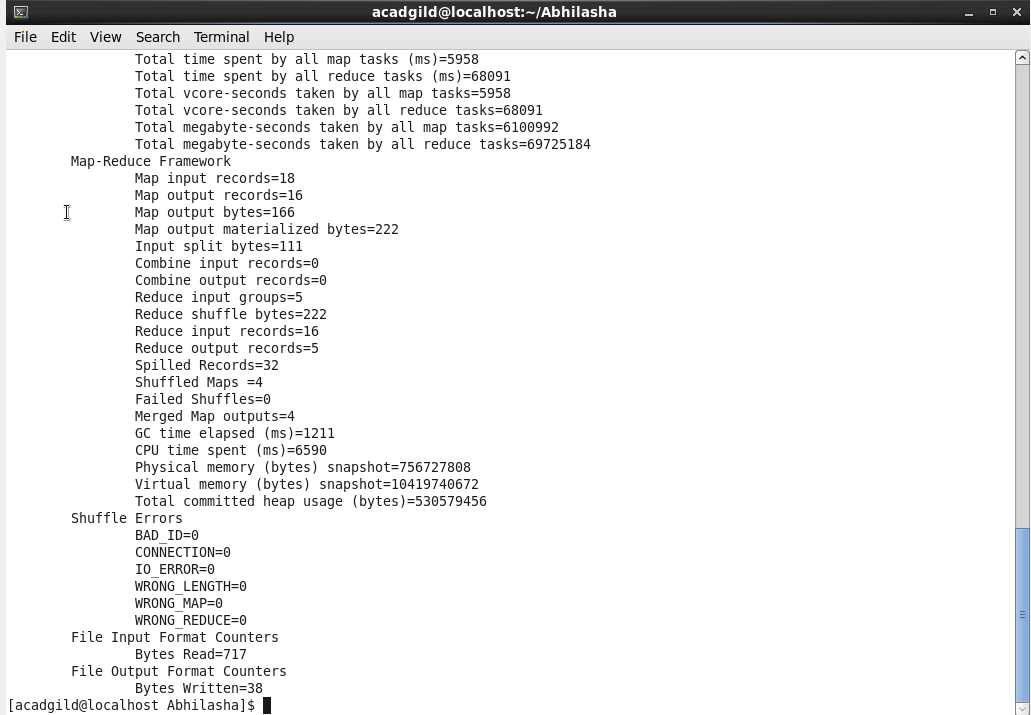
1. Jar is placed locally at /home/acadgild/Abhilasha. Its name is TVSales.jar



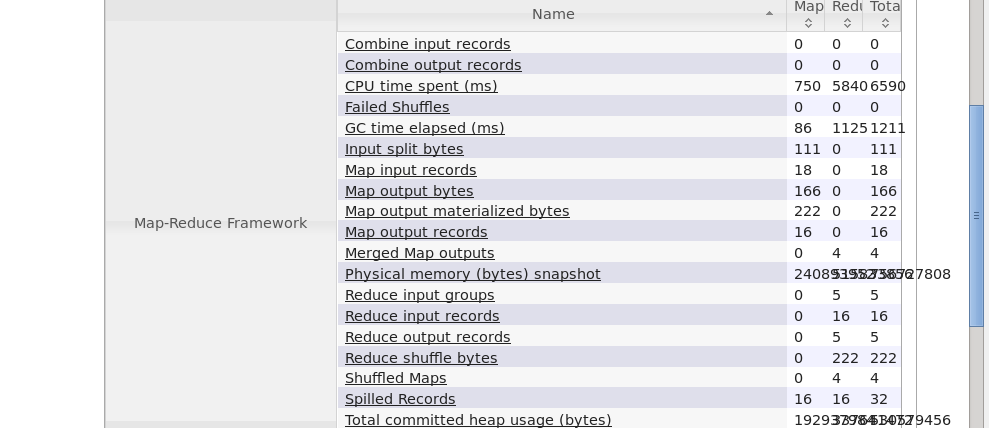
1. Executing the map-reduce program on Hadoop.



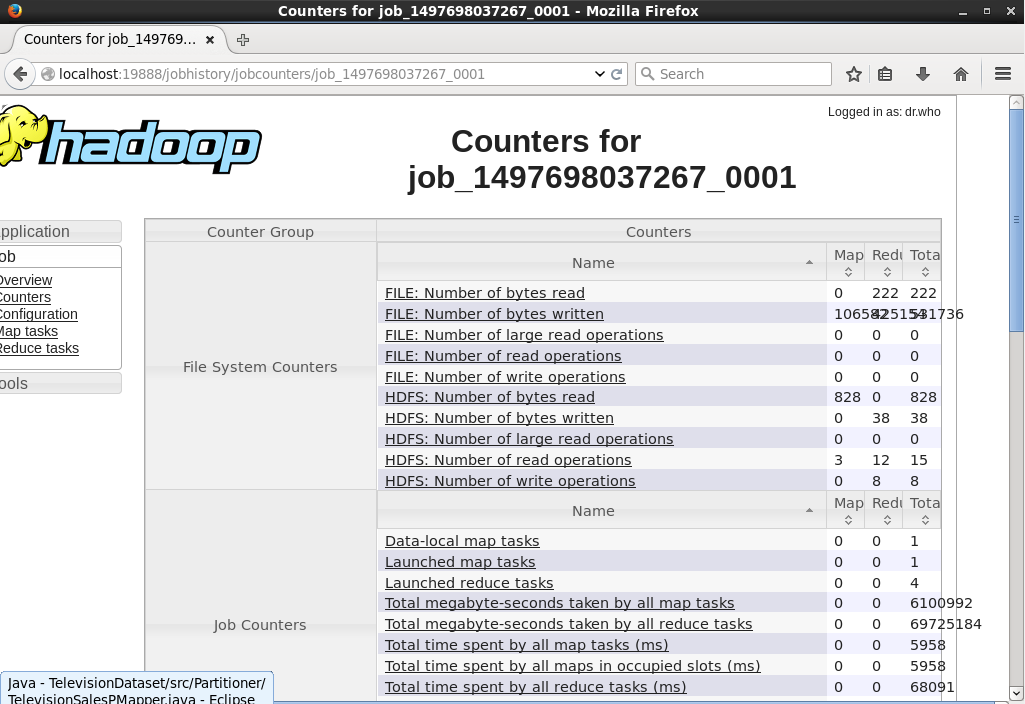
1. Logs showing partitioner was executed. Number of shuffled maps is 4



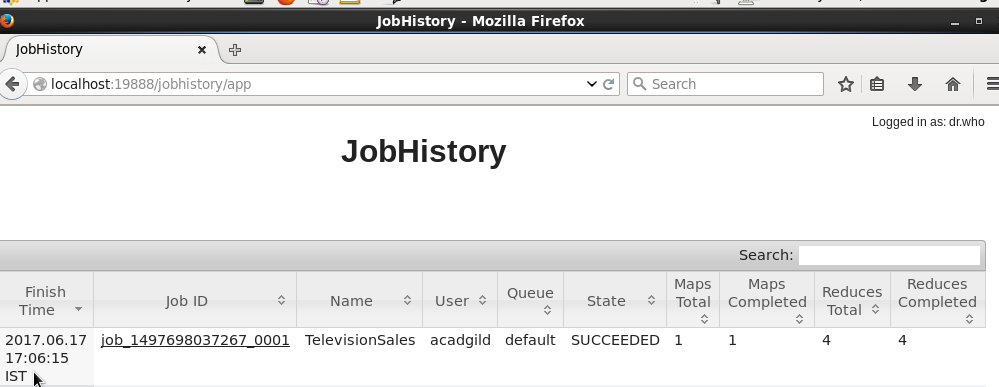
4 partitions were made.

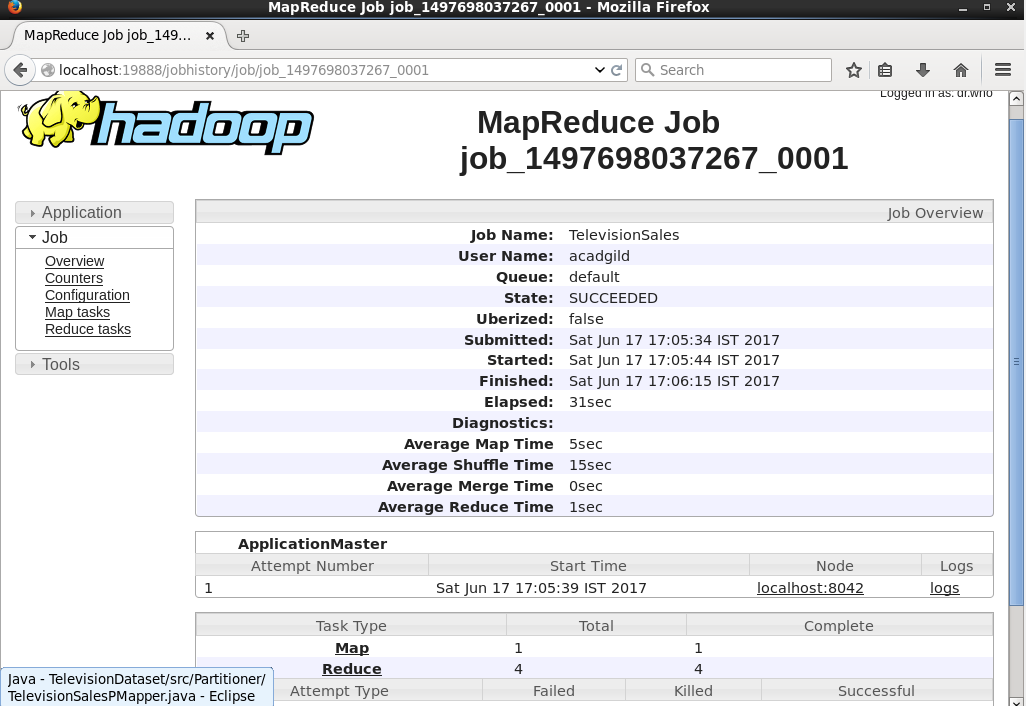


1. Logs showing number of reducers were 4

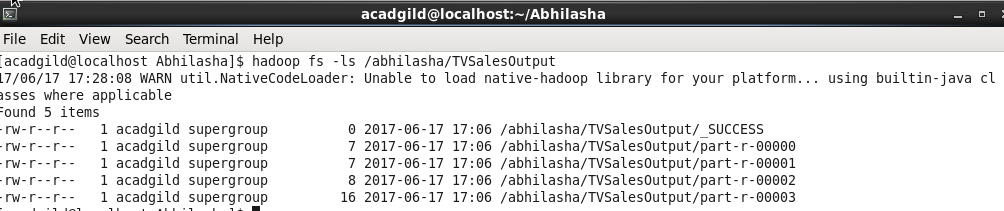


1. Job history server mentioning the success of the job with 1 map and 4 reducers

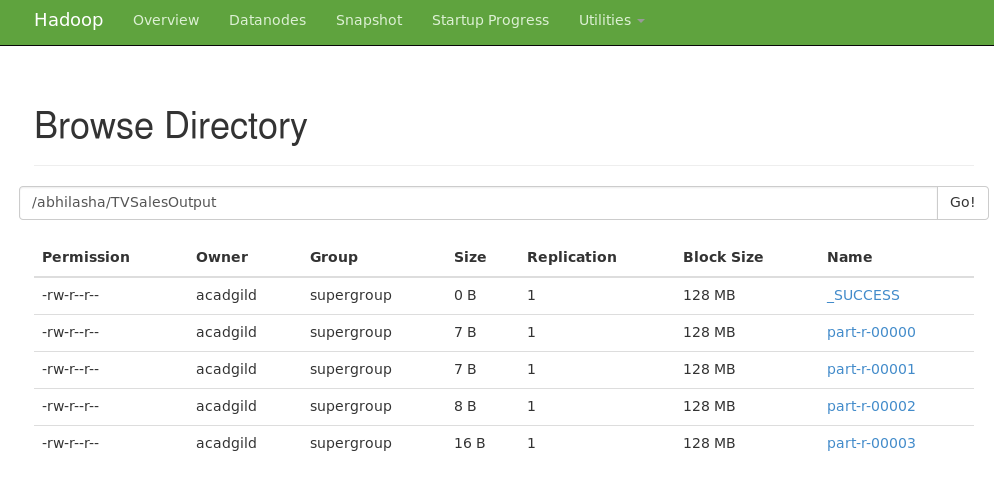




1. Output folder created on hdfs :



4 reducers, 4 output files



1. Content of output files part-r-00000, part-r-00001, part-r-00002, part-r-00003 on hdfs. Output verified

