**BigData And Hadoop**

**Assignment 1 of Session 4**

**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 filter out the invalid records.

**Solution**

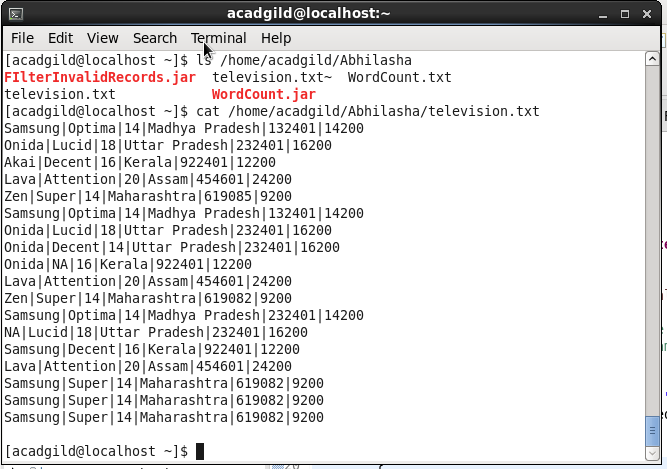
**Code files are as follows :**

Mapper class : FilterInvalidRecords\_Mapper.java

Driver class : FilterInvalidRecords.java

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

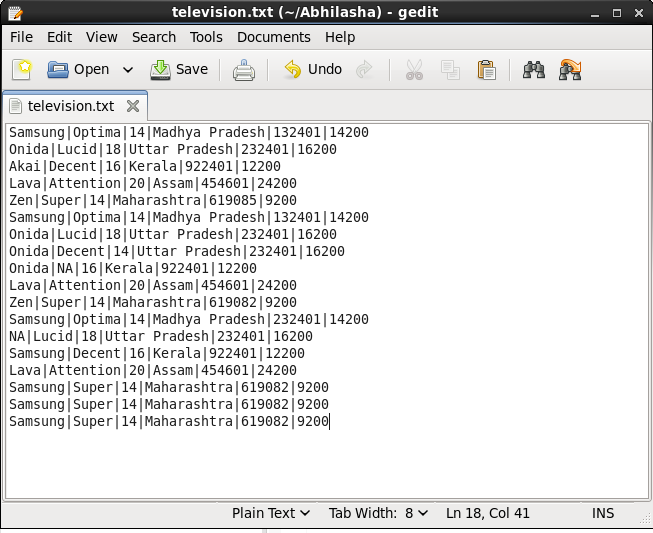
1. Input file present in ‘/home/acadgild/Abhilasha’. Its name is television.txt



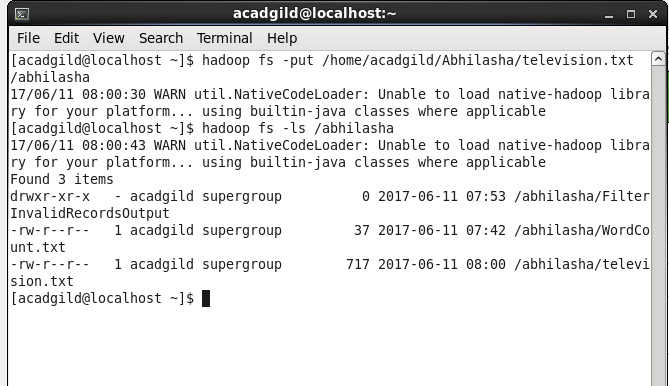
Command used to see the content of the input file

Path where input file is saved locally

1. Input file opened in gedit locally

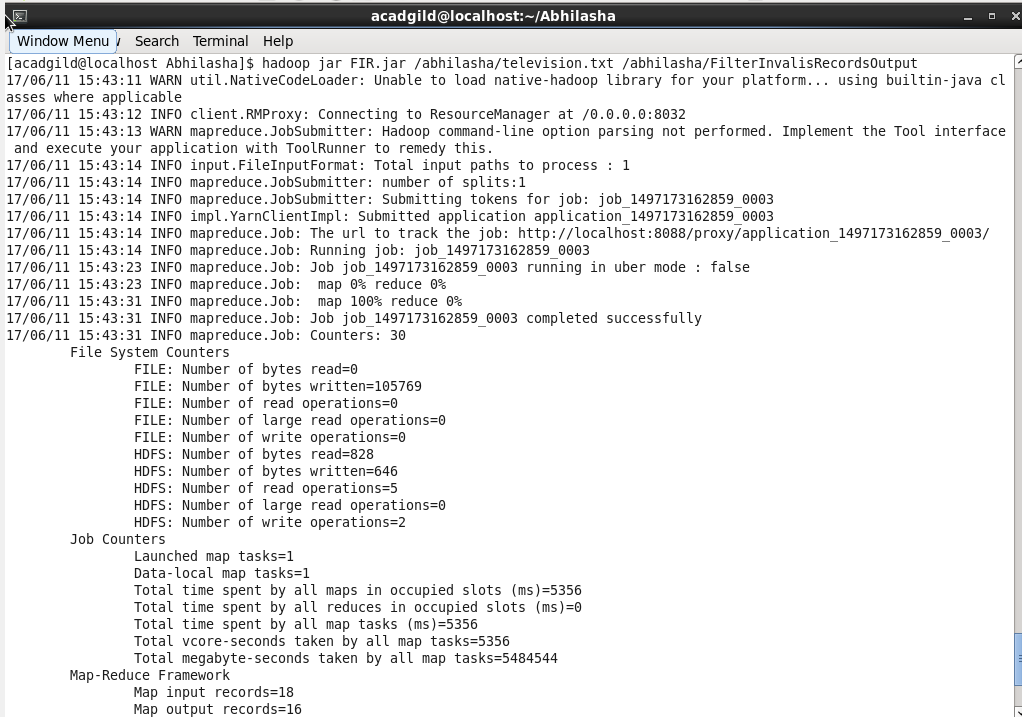


1. Command executed to put television.txt to hdfs system



television.txt stored on hdfs

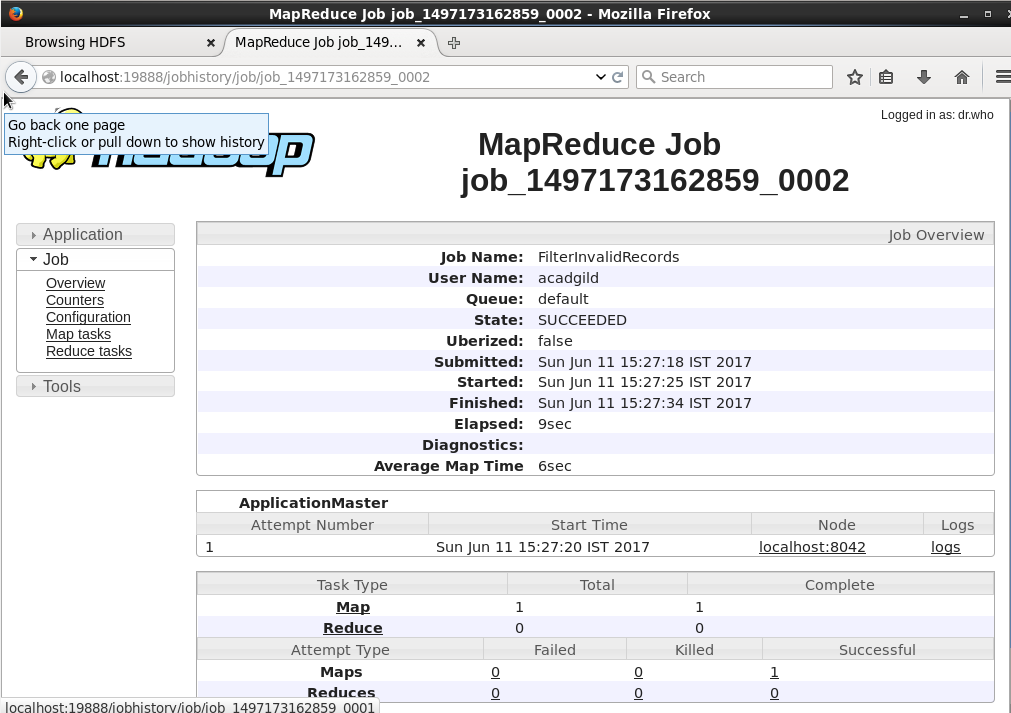
1. Executing the map-reduce program on Hadoop. Jar is placed locally at /home/acadgild/Abhilasha. Its name is FIR.jar



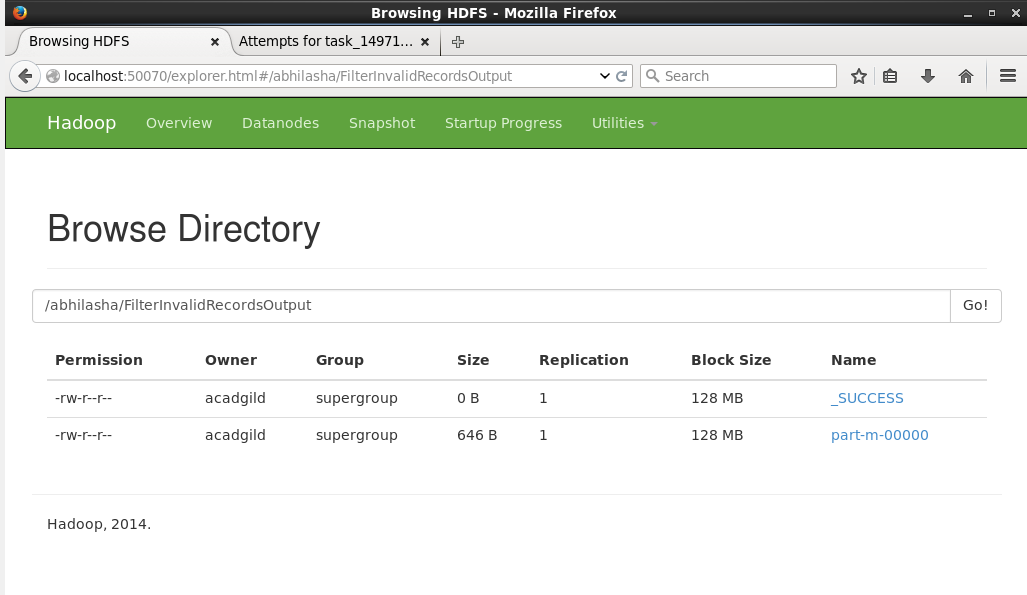
Map only job

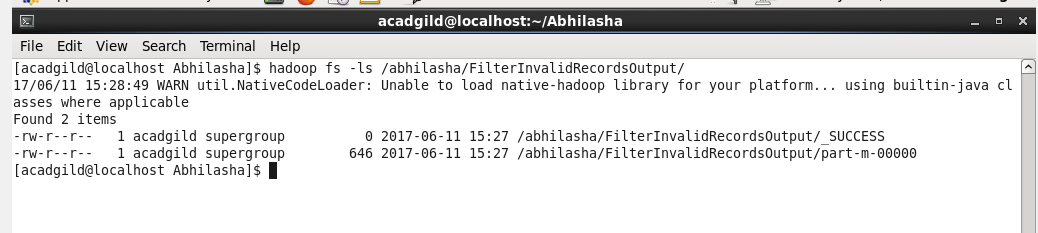
1. Job history server mentioning the success of the job





1. Output folder created on hdfs :





1. Content of output file part-m-00000 on hdfs

