**BigData And Hadoop**

**Assignment 2 of Session 6**

**Problem Statement :**

Enhance the Map Reduce program of Task 8(refer session 6, assignment 1)to use multiple reducers for sorting. The driver should accept three additional values: the minimum units sold, the maximum units sold and number of reducers to use.Use units sold as key and company as value.Write a custom partitioner to divide the keys on the basis of range.Take minimum to be 0 and maximum to be 10. Divide them across 2 reducers.

**Solution**

**Stage1 : Generating sequence file**

Code files to find number of units for every company, with output as a sequence file is as follows :

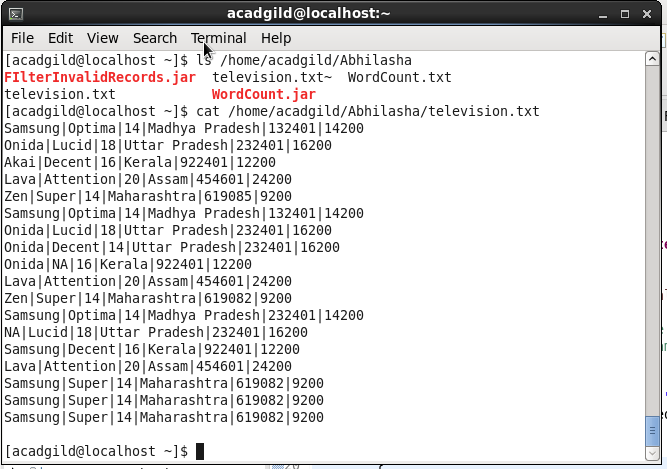
Mapper class : TelevisionSalesMapperC.java

Reducer class : TelevisionSalesReducerC.java

Driver class : TelevisionSalesC.java

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

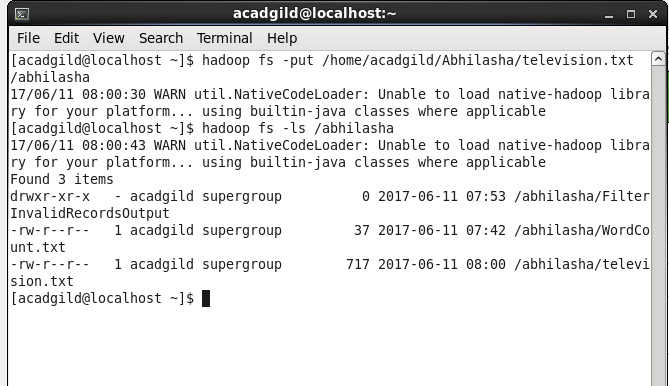
1. Input file present in ‘/home/acadgild/Abhilasha’. Its name is television.txt



Command used to see the content of the input file

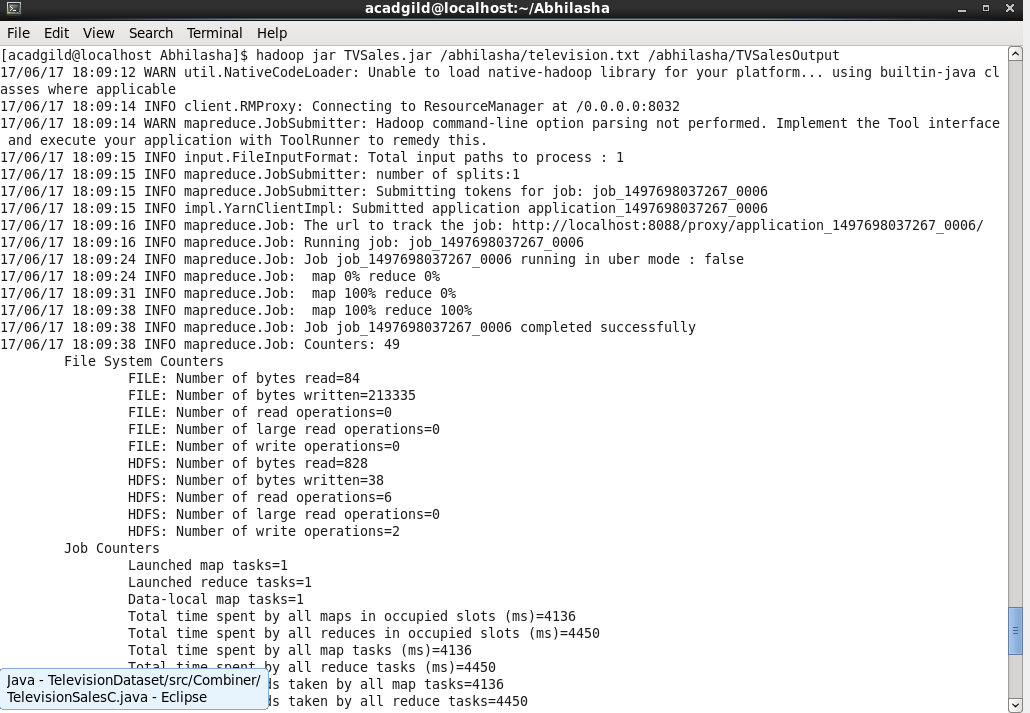
Path where input file is saved locally

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

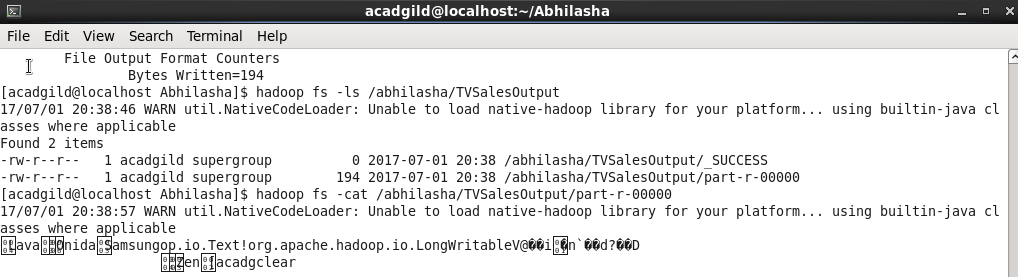


television.txt stored on hdfs

1. Executing the map-reduce program on Hadoop.



1. Output folder created on hdfs. The folder contains output in the form of sequence file, which on printing on console using ‘cat’ command was displayed as follows.



**Stage2 : Verifying the contents of sequence file through a java code**

Class Name : SequenceFileReader.java

It is a simple java code and not a map-reduce program.

Code is as follows :

package Assignment1;

import java.io.IOException;

import org.apache.hadoop.conf.Configuration;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.LongWritable;

import org.apache.hadoop.io.SequenceFile;

import org.apache.hadoop.io.Text;

public class SequenceFileReader

{

public static void main(String[] args) throws IOException

{

Path inputPath = new Path("/abhilasha/TVSalesOutput/part-r-00000");

SequenceFile.Reader reader = new SequenceFile.Reader(new Configuration(), SequenceFile.Reader.file(inputPath));

System.out.println("Is file compressed : "+reader.isCompressed());

Text key = new Text();

LongWritable value = new LongWritable();

while(reader.next(key, value))

{

System.out.println("key : "+key+"\t\tValue : "+value);

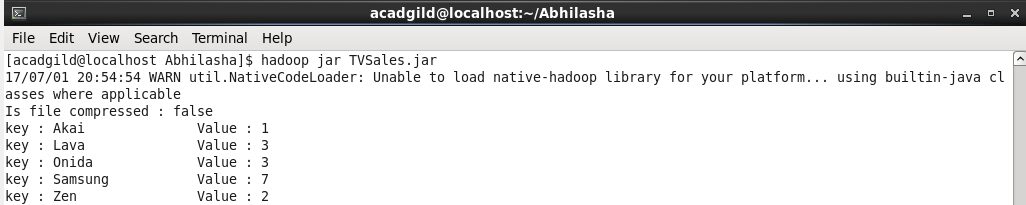
}

reader.close();

}

}

Output on executing this code is as follows :



**Stage3 : Sorting the data based on units sold**

Code files to find number of units for every company, with output as a sequence file is as follows :

Mapper class : TvMapper.java

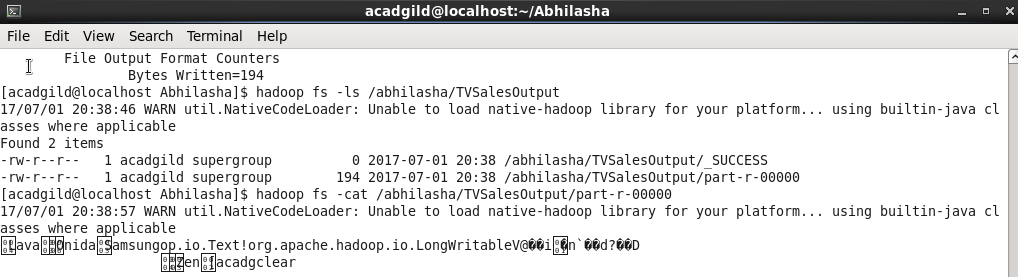
Reducer class : Default reducer used

Partitioner : TvPartitioner.class

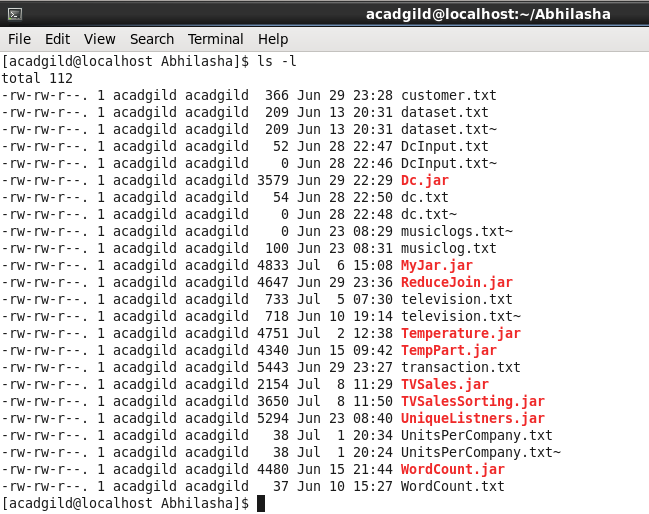
Driver class : TvDriver.java

Input Parameters :

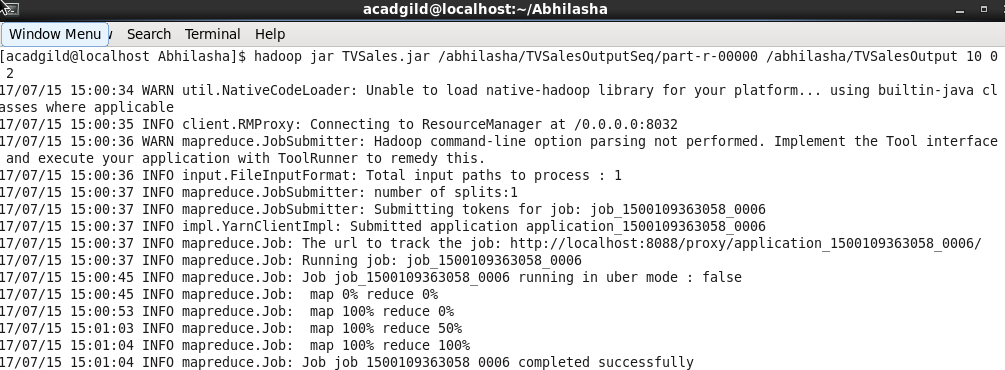
1. Input File path
2. Output directory path
3. Max Units Number
4. Min Units Number
5. Number of reducers
6. Input to this program is output of previous program and stored in /abhilasha/TVSalesOutput/part-r-00000. It is already present in hdfs.



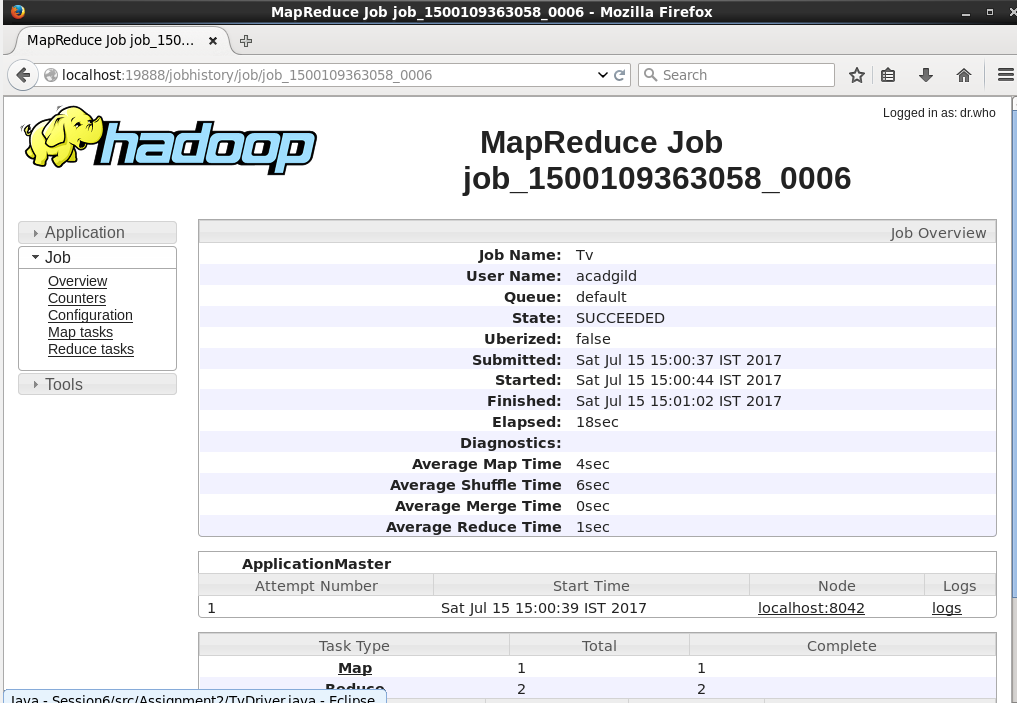
1. Jar TVSalesSorting.jar located at /home/acadgild/Abhilasha



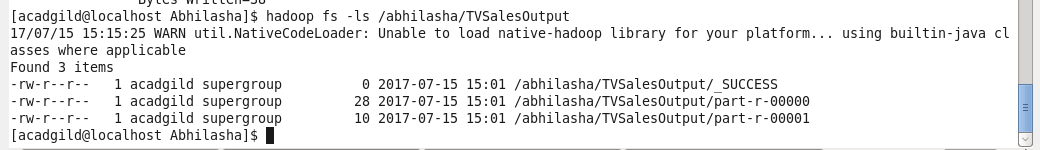
1. Executing the map-reduce program :

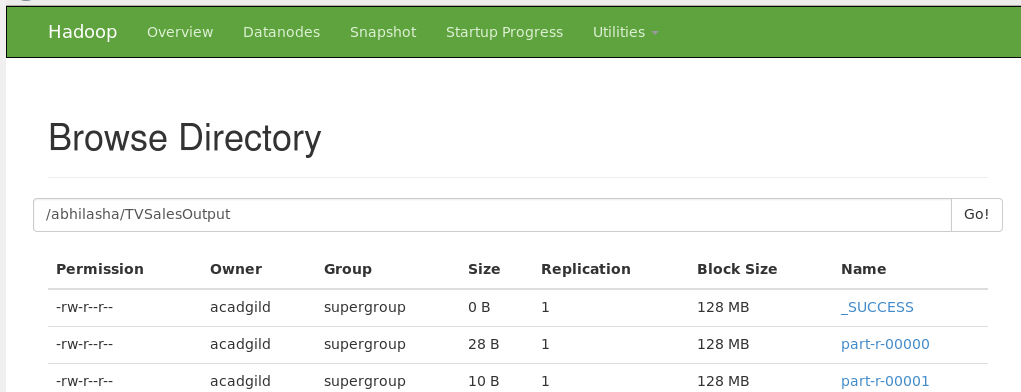


1. Job completed successfully : Number of Reducers mentioned is 2



1. Output files generated as follows:





1. Output content is as follows :

