Assignment- 1

AIM: Write a code in JAVA for a implement WordCount application that counts the number of occurrence of each word in a given input set using the Hadoop MapReduce framework on local-standalone set-up.

WordCount.java

```
import java.io.IOException;
import java.util.StringTokenizer;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WordCount {

 public static class TokenizerMapper
     extends Mapper<Object, Text, Text, IntWritable>{

  private final static IntWritable one = new IntWritable(1);
  private Text word = new Text();

  public void map(Object key, Text value, Context context
                  ) throws IOException, InterruptedException {
    StringTokenizer itr = new StringTokenizer(value.toString());
    while (itr.hasMoreTokens()) {
      word.set(itr.nextToken());
      context.write(word, one);
    }
  }
 }
 public static class IntSumReducer
     extends Reducer<Text,IntWritable,Text,IntWritable> {
  private IntWritable result = new IntWritable();

  public void reduce(Text key, Iterable<IntWritable> values,
                     Context context
                     ) throws IOException, InterruptedException {
    int sum = 0;
    for (IntWritable val : values) {
      sum += val.get();
    }
    result.set(sum);
    context.write(key, result);
  }
 }

 public static void main(String[] args) throws Exception {
```

```
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "word count");
    job.setJarByClass(WordCount.class);
    job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(IntSumReducer.class);
    job.setReducerClass(IntSumReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    System.exit(job.waitForCompletion(true) ? 0: 1);
  }
}
```

Input: input.txt

WordCount example reads text files and counts how often words occur. The input is text files, and the output is text files, each line of which contains a word and the count of how often it occurred, separated by a tab.

Map Reduce Project that works on weather data and process it, the final outcome of the project can be processed further to find similarities on different weather stations

The Shadow was an American pulp magazine published by Street & Smith from 1931 to 1949. Each issue contained a novel about The Shadow, a mysterious crime-fighting figure who spoke the line "Who knows what evil lurks in the hearts of men? The Shadow knows" in radio broadcasts of stories from Street & Smith's Detective Story Magazine. For the first issue, dated April 1931, Walter Gibson wrote the lead novel.

Output: output file (part-r-00000)

Wordcount Steps to run:

1. Starting Hadoop

   $ start-all.sh

2. Made A folder “wordcountexp” and write WordCount.java code.

3. Create new folder for input data.
4. Add input text file in the input data folder.
5. Create new folder to hold java class files.
6. Set HADOOP_CLASSPATH environment variable.

   $ export HADOOP_CLASSPATH=$(hadoop classpath)
7. Create a directory on HDFS

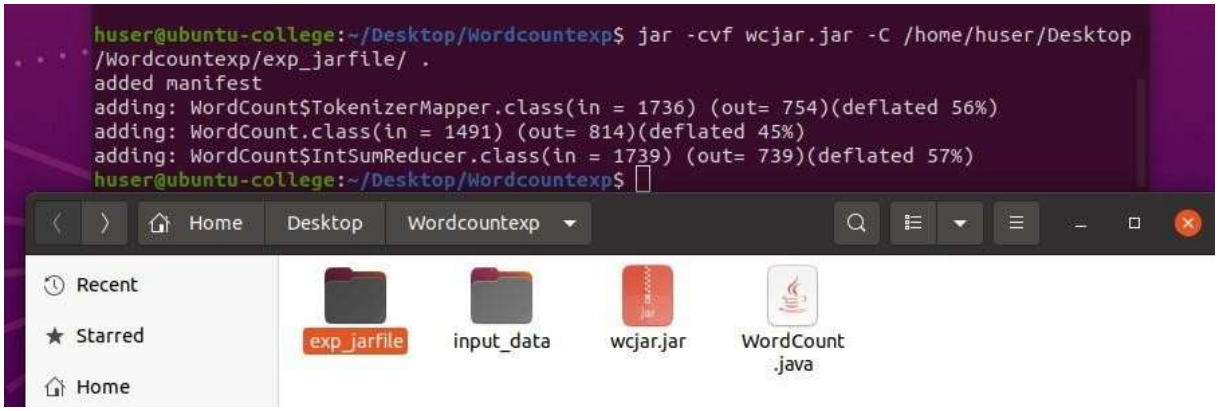
   $ hdfs dfs -mkdir /WordCountTut

   $ hdfs dfs -mkdir /WordCountTut/Input
8. Upload the input file (device) to that directory.

   $ hdfs dfs -put '/home/huser/Desktop/Wordcountexp/input_data/input.txt' /WordCountTut/Input
9. Compile the java code:

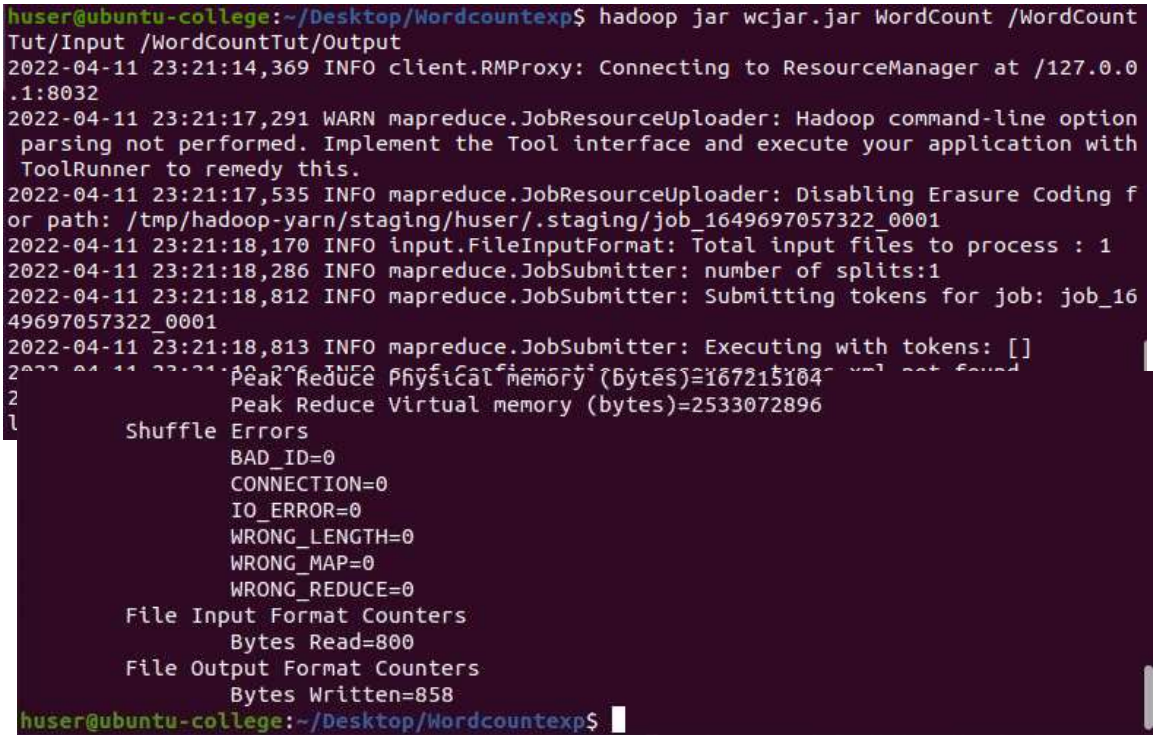
   $ javac -classpath $(HADOOP_CLASSPATH) -d '/home/huser/Desktop/Wordcountexp/exp_jarfile' /home/huser/Desktop/Wordcountexp/ .*java
10. Creation .jar file of classes:

    $ jar -cvf wcjar.jar -C '/home/huser/Desktop/Wordcountexp/exp_jarfile/ .



11. Running the jar file on Hadoop

    $ hadoop jar wcjar.jar WordCount /WordCountTut/Input /WordCountTut/Output



12. Check output on localhost:9870 /localhost:50070