

Name : Vishal Gavali
Roll No. : TCOB38
Assignment No. : 11
Subject : DSBDA LAB

```
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;

import java.io.IOException;

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();

        @Override
        public void map(Object key, Text value, Context context) throws
IOException, InterruptedException {
            String line = value.toString();
            String[] words = line.split("\\s+");
            for (String w : words) {
                word.set(w);
                context.write(word, one);
            }
        }
    }

    public static class IntSumReducer extends Reducer<Text, IntWritable,
Text, IntWritable> {

        private IntWritable result = new IntWritable();

        @Override
        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 {
        if (args.length != 2) {
            System.err.println("Usage: WordCount <input path> <output
path>");
            System.exit(-1);
        }

        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);
    }
}

```

OUTPUT :

Openpart-r-00000Save-Desktop/Wordcountexp

11 For 1

12 Gibson 1

13 Magazine. 1

14 Map 1

15 Project 1

16 Reduce 1

17 Shadow 2

18 Shadow, 1

19 Smith 1

20 Smith's 1

21 Story 1

22 Street 2

23 The 4

24 Walter 1

25 WordCount 1

26 a 4

27 about 1

28 an 1

29 and 4

30 be 1

31 broadcasts 1

32 by 2

33 can 1

34 contained 1

35 contains 1

36 count 1

37 counts 1

38 crine-fighting 1

39 data 1

40 dated 1

41 different 1

42 each 1

43 evil 1

44 exanple 1

45 figure 1

46 files 2

47 files, 1

Plain TextTab Width: 8Ln 22, Col 10INS

Browse Directory

/WordCountTut   

Show entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	huser	supergroup	0 B	Apr 11 23:06	0	0 B	Input	
<input type="checkbox"/>	drwxr-xr-x	huser	supergroup	0 B	Apr 11 23:23	0	0 B	Output	

Showing 1 to 2 of 2 entries

Previous **1** Next

Hadoop, 2021.

Browse Directory

/WordCountTut/Output   

Show entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	huser	supergroup	0 B	Apr 11 23:23	1	128 MB	_SUCCESS	
<input type="checkbox"/>	-rw-r--r--	huser	supergroup	858 B	Apr 11 23:23	1	128 MB	part-r-00000	

Showing 1 to 2 of 2 entries

Previous **1** Next

File information - part-r-00000

[Download](#) [Head the file \(first 32K\)](#) [Tail the file \(last 32K\)](#)

Block information -- Block 0

Block ID: 1073741886

Block Pool ID: BP-1388353168-127.0.1.1-1647528100285

Generation Stamp: 1062

Size: 858

Availability:

- ubuntu-college

File contents

XXXXXXXXXX

a 4

about 1

an 1

and 4

be 1

broadcasts 1

by 2

can 1

contained 1

contains 1

count 1

counts 1

Go!

Search:

Block Size	Name
MB	_SUCCESS
MB	part-r-00000

Previous

1

Next