Assignment : 11

Name: Shivraj Chaudar
Div: A **Class:** T.E
Roll no. : 3101018

#pom.xml

```xml
<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="http://maven.apache.org/POM/4.0.0"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/
maven-4.0.0.xsd">
  <modelVersion>4.0.0</modelVersion>

  <groupId>org.shivraj</groupId>
  <artifactId>WordCount</artifactId>
  <version>1.0-SNAPSHOT</version>

  <properties>
    <maven.compiler.source>8</maven.compiler.source>
    <maven.compiler.target>8</maven.compiler.target>
    <project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>
  </properties>

  <dependencies>
    <dependency>
      <groupId>org.apache.hadoop</groupId>
      <artifactId>hadoop-common</artifactId>
      <version>3.3.3</version>
    </dependency>
    <dependency>
      <groupId>org.apache.hadoop</groupId>
      <artifactId>hadoop-mapreduce-client-core</artifactId>
      <version>3.3.3</version>
    </dependency>
  </dependencies>
</project>
```

#WC_Runner.java

```java
package org.shivraj;

import java.io.IOException;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapred.FileInputFormat;
```

```java
import org.apache.hadoop.mapred.FileOutputFormat;

import org.apache.hadoop.mapred.JobClient;

import org.apache.hadoop.mapred.JobConf;

import org.apache.hadoop.mapred.TextInputFormat;

import org.apache.hadoop.mapred.TextOutputFormat;

public class WC_Runner {

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

        JobConf conf = new JobConf(WC_Runner.class);

        conf.setJobName("WordCount");

        conf.setOutputKeyClass(Text.class);

        conf.setOutputValueClass(IntWritable.class);

        conf.setMapperClass(WC_Mapper.class);

        conf.setCombinerClass(WC_Reducer.class);

        conf.setReducerClass(WC_Reducer.class);

        conf.setInputFormat(TextInputFormat.class);

        conf.setOutputFormat(TextOutputFormat.class);

        FileInputFormat.setInputPaths(conf,new Path(args[0]));

        FileOutputFormat.setOutputPath(conf,new Path(args[1]));

        JobClient.runJob(conf);

    }

}
```

#WC_Mapper.java

```java
package org.shivraj;

import java.io.IOException;

import java.util.StringTokenizer;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.LongWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapred.MapReduceBase;

import org.apache.hadoop.mapred.Mapper;

import org.apache.hadoop.mapred.OutputCollector;

import org.apache.hadoop.mapred.Reporter;
```

```java
public class WC_Mapper extends MapReduceBase implements
Mapper<LongWritable,Text,Text,IntWritable>{

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

    private Text word = new Text();

    public void map(LongWritable key, Text value,OutputCollector<Text,IntWritable>
output,Reporter reporter) throws IOException{

        String line = value.toString();

        StringTokenizer  tokenizer = new StringTokenizer(line);

        while (tokenizer.hasMoreTokens()){

            word.set(tokenizer.nextToken());

            output.collect(word, one);

        }

    }

}
```

#WC_Reducer.java

```java
package org.shivraj;

import java.io.IOException;

import java.util.Iterator;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapred.MapReduceBase;

import org.apache.hadoop.mapred.OutputCollector;

import org.apache.hadoop.mapred.Reducer;

import org.apache.hadoop.mapred.Reporter;


public class WC_Reducer  extends MapReduceBase implements
Reducer<Text,IntWritable,Text,IntWritable> {

    public void reduce(Text key, Iterator<IntWritable> values,OutputCollector<Text,IntWritable>
output,Reporter reporter) throws IOException {

        int sum=0;

        while (values.hasNext()) {

            sum+=values.next().get();

        }
```

```
        output.collect(key,new IntWritable(sum));

    }

}
```

#Input.txt

Hello wordcount MapReduce Hadoop program.

This is my first MapReduce program.

#Output

(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ % hdfs dfs -mkdir -p /user/process
2025-03-31 12:25:41,823 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ % hdfs dfs -put /Users/Shivrajchaudar/Desktop/input.txt /user/process
2025-03-31 12:26:49,178 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ % hadoop jar /Users/Shivrajchaudar/hadoop-3.4.1/share/hadoop/hdfs/WordCount-1.0-SNAPSHOT.jar org.shivraj.WC_Runner /user/process/input.txt /user/out
2025-03-31 12:35:46,244 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2025-03-31 12:35:46,623 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2025-03-31 12:35:46,762 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2025-03-31 12:35:46,827 INFO mapreduce.JobSubmissionFiles: Permissions on staging directory /tmp/hadoop-yarn/staging/shivrajchaudar/.staging are incorrect: rwxrwxrwx. Fixing permissions to correct value rwx------
2025-03-31 12:35:46,981 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-31 12:35:46,991 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/shivrajchaudar/.staging/job_1743357063034_0003
2025-03-31 12:35:47,160 INFO mapred.FileInputFormat: Total input files to process : 1
2025-03-31 12:35:48,025 INFO mapreduce.JobSubmitter: number of splits:2
2025-03-31 12:35:48,538 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1743357063034_0003
2025-03-31 12:35:48,538 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-31 12:35:48,675 INFO conf.Configuration: resource-types.xml not found
2025-03-31 12:35:48,675 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-03-31 12:35:48,779 INFO impl.YarnClientImpl: Submitted application application_1743357063034_0003
2025-03-31 12:35:48,799 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1743357063034_0003/
2025-03-31 12:35:48,799 INFO mapreduce.Job: Running job: job_1743357063034_0003

2025-03-31 12:35:53,962 INFO mapreduce.Job: Job job_1743357063034_0003 running in uber mode : false

2025-03-31 12:35:53,964 INFO mapreduce.Job: map 0% reduce 0%

2025-03-31 12:35:58,068 INFO mapreduce.Job: map 100% reduce 0%

2025-03-31 12:36:01,100 INFO mapreduce.Job: map 100% reduce 100%

2025-03-31 12:36:03,135 INFO mapreduce.Job: Job job_1743357063034_0003 completed successfully

2025-03-31 12:36:03,248 INFO mapreduce.Job: Counters: 50

    File System Counters

        FILE: Number of bytes read=150

        FILE: Number of bytes written=928700

        FILE: Number of read operations=0

        FILE: Number of large read operations=0

        FILE: Number of write operations=0

        HDFS: Number of bytes read=311

        HDFS: Number of bytes written=77

        HDFS: Number of read operations=11

        HDFS: Number of large read operations=0

        HDFS: Number of write operations=2

        HDFS: Number of bytes read erasure-coded=0

    Job Counters

        Launched map tasks=2

        Launched reduce tasks=1

        Data-local map tasks=2

        Total time spent by all maps in occupied slots (ms)=6600

        Total time spent by all reduces in occupied slots (ms)=3034

        Total time spent by all map tasks (ms)=3300

        Total time spent by all reduce tasks (ms)=1517

        Total vcore-milliseconds taken by all map tasks=3300

        Total vcore-milliseconds taken by all reduce tasks=1517

        Total megabyte-milliseconds taken by all map tasks=3379200

        Total megabyte-milliseconds taken by all reduce tasks=1553408

    Map-Reduce Framework

        Map input records=2

        Map output records=11

        Map output bytes=122

        Map output materialized bytes=156

        Input split bytes=192

        Combine input records=11

        Combine output records=11

        Reduce input groups=9

        Reduce shuffle bytes=156

        Reduce input records=11

        Reduce output records=9

        Spilled Records=22

        Shuffled Maps =2

        Failed Shuffles=0

        Merged Map outputs=2

GC time elapsed (ms)=96
                    CPU time spent (ms)=0
                    Physical memory (bytes) snapshot=0
                    Virtual memory (bytes) snapshot=0
                    Total committed heap usage (bytes)=724041728
            Shuffle Errors
                    BAD_ID=0
                    CONNECTION=0
                    IO_ERROR=0
                    WRONG_LENGTH=0
                    WRONG_MAP=0
                    WRONG_REDUCE=0
            File Input Format Counters
                    Bytes Read=119
            File Output Format Counters
                    Bytes Written=77

(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ % hdfs dfs -ls /user/out
2025-03-31 12:36:28,522 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   3 shivrajchaudar sharedgroup   0 2025-03-31 12:35 /user/out/_SUCCESS
-rw-r--r--   3 shivrajchaudar sharedgroup   77 2025-03-31 12:35 /user/out/part-00000

(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ % hdfs dfs -cat /user/out/part-00000
2025-03-31 12:37:04,184 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
Hadoop   1
Hello 1
MapReduce   2
This   1
first   1
is   1
my   1
program.   2
wordcount 1
(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ %

```
(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ % hdfs dfs -mkdir -p /user/process
2025-03-31 12:25:41,823 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ % hdfs dfs -put /Users/Shivrajchaudar/Desktop/input.txt /user/process
2025-03-31 12:26:49,178 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ % hadoop jar /Users/Shivrajchaudar/hadoop-3.4.1/share/hadoop/hdfs/WordCount-1.0-SNAPSHOT.jar org.shivraj.WC_Runner /user/process/input.txt /user/out
2025-03-31 12:35:46,244 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2025-03-31 12:35:46,623 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2025-03-31 12:35:46,762 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2025-03-31 12:35:46,827 INFO mapreduce.JobSubmissionFiles: Permissions on staging directory /tmp/hadoop-yarn/staging/shivrajchaudar/.staging are incorrect: rwxrwxrwx. Fixing permissions to correct value rwx------
2025-03-31 12:35:46,981 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-03-31 12:35:46,991 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/shivrajchaudar/.staging/job_1743357063034_0003
2025-03-31 12:35:47,160 INFO mapred.FileInputFormat: Total input files to process : 1
2025-03-31 12:35:48,025 INFO mapreduce.JobSubmitter: number of splits:2
2025-03-31 12:35:48,538 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1743357063034_0003
2025-03-31 12:35:48,538 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-03-31 12:35:48,675 INFO conf.Configuration: resource-types.xml not found
2025-03-31 12:35:48,675 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-03-31 12:35:48,779 INFO impl.YarnClientImpl: Submitted application application_1743357063034_0003
2025-03-31 12:35:48,799 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1743357063034_0003/
2025-03-31 12:35:48,799 INFO mapreduce.Job: Running job: job_1743357063034_0003
2025-03-31 12:35:53,962 INFO mapreduce.Job: Job job_1743357063034_0003 running in uber mode : false
2025-03-31 12:35:53,964 INFO mapreduce.Job:  map 0% reduce 0%
2025-03-31 12:35:58,068 INFO mapreduce.Job:  map 100% reduce 0%
2025-03-31 12:36:01,100 INFO mapreduce.Job:  map 100% reduce 100%
2025-03-31 12:36:03,135 INFO mapreduce.Job: Job job_1743357063034_0003 completed successfully
2025-03-31 12:36:03,248 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=150
                FILE: Number of bytes written=928700
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=311
                HDFS: Number of bytes written=77
                HDFS: Number of read operations=11
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=2
                Launched reduce tasks=1
                Data-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=6600
                Total time spent by all reduces in occupied slots (ms)=3034
                Total time spent by all map tasks (ms)=3300
                Total time spent by all reduce tasks (ms)=1517
                Total vcore-milliseconds taken by all map tasks=3300
                Total vcore-milliseconds taken by all reduce tasks=1517
                Total megabyte-milliseconds taken by all map tasks=3379200
                Total megabyte-milliseconds taken by all reduce tasks=1553408
        Map-Reduce Framework
                Map input records=2
                Map output records=11
                Map output bytes=122
                Map output materialized bytes=156
                Input split bytes=192
                Combine input records=11
                Combine output records=11
```

```
                Launched reduce tasks=1
                Data-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=6600
                Total time spent by all reduces in occupied slots (ms)=3034
                Total time spent by all map tasks (ms)=3300
                Total time spent by all reduce tasks (ms)=1517
                Total vcore-milliseconds taken by all map tasks=3300
                Total vcore-milliseconds taken by all reduce tasks=1517
                Total megabyte-milliseconds taken by all map tasks=3379200
                Total megabyte-milliseconds taken by all reduce tasks=1553408
        Map-Reduce Framework
                Map input records=2
                Map output records=11
                Map output bytes=122
                Map output materialized bytes=156
                Input split bytes=192
                Combine input records=11
                Combine output records=11
                Reduce input groups=9
                Reduce shuffle bytes=156
                Reduce input records=11
                Reduce output records=9
                Spilled Records=22
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=96
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=724041728
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=119
        File Output Format Counters
                Bytes Written=77
(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ % hdfs dfs -ls /user/out
2025-03-31 12:36:28,522 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   3 shivrajchaudar sharedgroup          0 2025-03-31 12:35 /user/out/_SUCCESS
-rw-r--r--   3 shivrajchaudar sharedgroup         77 2025-03-31 12:35 /user/out/part-00000
(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ % hdfs dfs -cat /user/out/part-00000
2025-03-31 12:37:04,184 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Hadoop  1
Hello   1
MapReduce       2
This    1
first   1
is      1
my      1
program.        2
wordcount       1
(base) shivrajchaudar@Shivrajs-Macbook-Pro ~ %
```

Hadoop   Overview   Datanodes   Datanode Volume Failures   Snapshot   Startup Progress   Utilities

Browse Directory

/user/process   Go!

Show 25 entries   Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	-rw-r--r--	shivrajchaudar	sharedgroup	79 B	Mar 31 12:26	3	128 MB	input.txt	🗑

Showing 1 to 1 of 1 entries

Previous 1 Next

Hadoop, 2024.

Hadoop   Overview   Datanodes   Datanode Volume Failures   Snapshot   Startup Progress   Utilities

Browse Directory

/user/out   Go!

Show 25 entries   Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	-rw-r--r--	shivrajchaudar	sharedgroup	0 B	Mar 31 12:35	3	128 MB	_SUCCESS	🗑
☐	-rw-r--r--	shivrajchaudar	sharedgroup	77 B	Mar 31 12:35	3	128 MB	part-00000	🗑

Showing 1 to 2 of 2 entries

Previous 1 Next

Hadoop, 2024.

localhost:9870/explorer.html#/user/out

Hadoop   Overview   Datanodes

Browse Directory

/user/out

Show 25 entries

	Permission	Owner		Block Size	Name	
☐	-rw-r--r--	shivrajch...		B MB	_SUCCESS	🗑
☐	-rw-r--r--	shivrajch...		B MB	part-00000	🗑

Showing 1 to 2 of 2 entries

Previous 1 Next

Hadoop, 2024.

File information - part-00000

Download   Head the file (first 32K)   Tail the file (last 32K)

Block information -- [Block 0 ▾]

Block ID: 1073741857

Block Pool ID: BP-2095407043-127.0.0.1-1743357293385

Generation Stamp: 1033

Size: 77

Availability:

- localhost

File contents

```
Hadoop    1
Hello     1
MapReduce  2
This 1
first 1
is    1
my    1
program. 2
wordcount     1
```

Close