Program 3 : Map reduce program for word count

```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{

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

{

 private IntWritable result = new IntWritable();

public void reduce(Text key, Iterable values,

 Context context ) throws IOException, InterruptedException {

int sum = 0;
```

```java
 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);

 }

 }
```
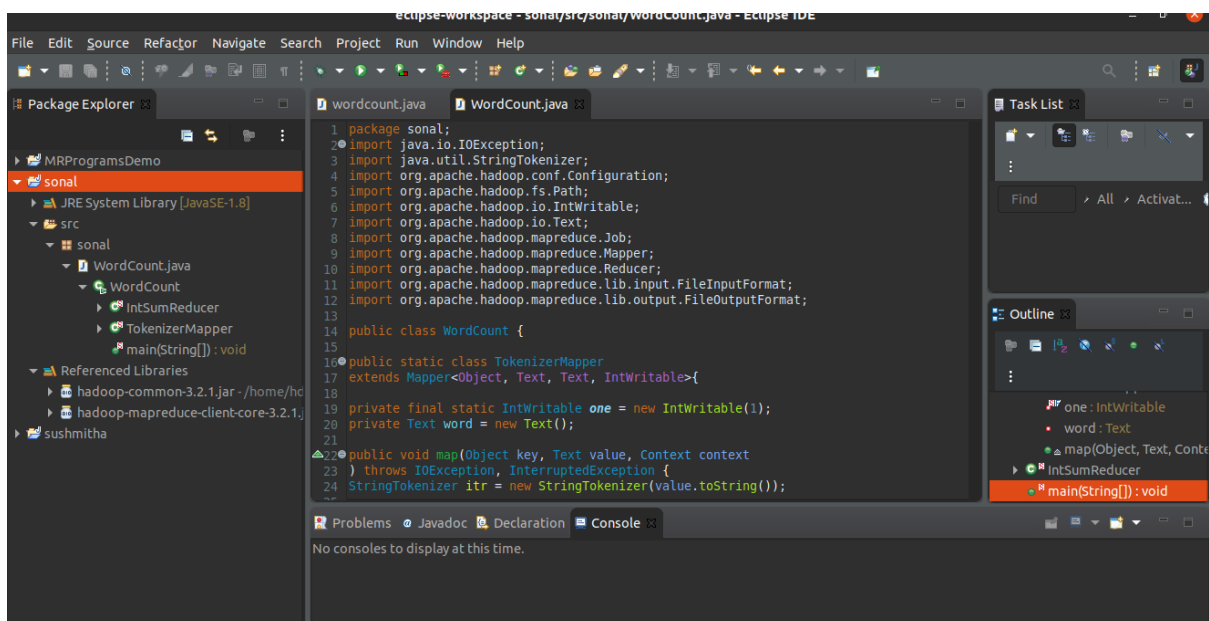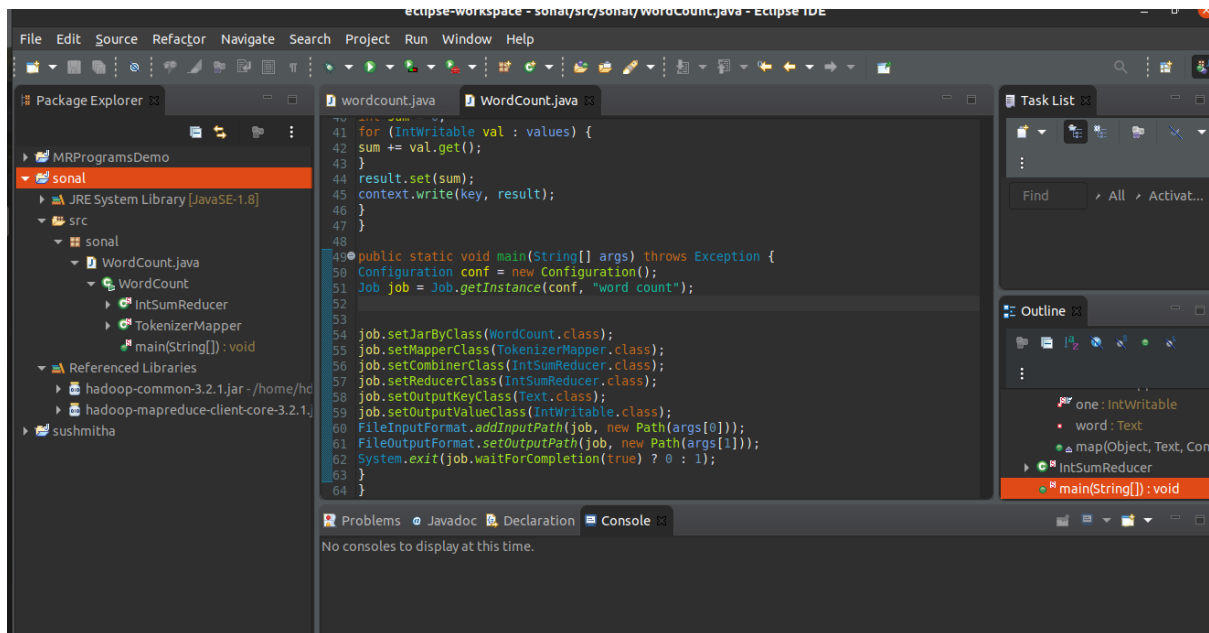
Eclipse IDE screenshot — WordCount.java

```java
          int sum = 0;
41        for (IntWritable val : values) {
42            sum += val.get();
43        }
44        result.set(sum);
45        context.write(key, result);
46    }
47 }
48
49  public static void main(String[] args) throws Exception {
50      Configuration conf = new Configuration();
51      Job job = Job.getInstance(conf, "word count");
52
53
54      job.setJarByClass(WordCount.class);
55      job.setMapperClass(TokenizerMapper.class);
56      job.setCombinerClass(IntSumReducer.class);
57      job.setReducerClass(IntSumReducer.class);
58      job.setOutputKeyClass(Text.class);
59      job.setOutputValueClass(IntWritable.class);
60      FileInputFormat.addInputPath(job, new Path(args[0]));
61      FileOutputFormat.setOutputPath(job, new Path(args[1]));
62      System.exit(job.waitForCompletion(true) ? 0 : 1);
63    }
64 }
```

Terminal output:

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1$ cd sbin
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
4017 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
5796 NodeManager
5063 NameNode
5641 ResourceManager
8475 Jps
5211 DataNode
5453 SecondaryNameNode
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hdoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
localhost: namenode is running as process 5063.  Stop it first.
Starting datanodes
localhost: datanode is running as process 5211.  Stop it first.
Starting secondary namenodes [admin1-HP-280-G4-MT-Business-PC]
admin1-HP-280-G4-MT-Business-PC: secondarynamenode is running as process 5453.  Stop it first.
Starting resourcemanager
resourcemanager is running as process 5641.  Stop it first.
Starting nodemanagers
localhost: nodemanager is running as process 5796.  Stop it first.
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p ~/input
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile - ~/input/test.txt
hy there how are you i am fine 2022-06-17 10:38:44,216 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, r
emoteHostTrusted = false
```

```
        at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hdoop/Desktop/sonal.jar ~/input ~/out
2022-06-17 10:46:34,549 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-17 10:46:34,781 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and ex
ecute your application with ToolRunner to remedy this.
2022-06-17 10:46:34,803 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_165544
0953471_0002
2022-06-17 10:46:34,884 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 10:46:34,982 INFO input.FileInputFormat: Total input files to process : 1
2022-06-17 10:46:35,095 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 10:46:35,553 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 10:46:35,569 INFO mapreduce.JobSubmitter: number of splits:1
2022-06-17 10:46:35,686 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 10:46:36,111 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655440953471_0002
2022-06-17 10:46:36,111 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-06-17 10:46:36,246 INFO conf.Configuration: resource-types.xml not found
2022-06-17 10:46:36,246 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-06-17 10:46:36,420 INFO impl.YarnClientImpl: Submitted application application_1655440953471_0002
2022-06-17 10:46:36,452 INFO mapreduce.Job: The url to track the job: http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1655440953471
_0002/
2022-06-17 10:46:36,452 INFO mapreduce.Job: Running job: job_1655440953471_0002
2022-06-17 10:46:42,549 INFO mapreduce.Job: Job job_1655440953471_0002 running in uber mode : false
2022-06-17 10:46:42,551 INFO mapreduce.Job:  map 0% reduce 0%
2022-06-17 10:46:45,602 INFO mapreduce.Job:  map 100% reduce 0%
2022-06-17 10:46:49,628 INFO mapreduce.Job:  map 100% reduce 100%
2022-06-17 10:46:49,650 INFO mapreduce.Job: Job job_1655440953471_0002 completed successfully
2022-06-17 10:46:49,730 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=98
                FILE: Number of bytes written=451079
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=168
                HDFS: Number of bytes written=56
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
```

```
        File System Counters
                FILE: Number of bytes read=98
                FILE: Number of bytes written=451079
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=168
                HDFS: Number of bytes written=56
                HDFS: Number of read operations=8
                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=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=1385
                Total time spent by all reduces in occupied slots (ms)=1497
                Total time spent by all map tasks (ms)=1385
                Total time spent by all reduce tasks (ms)=1497
                Total vcore-milliseconds taken by all map tasks=1385
                Total vcore-milliseconds taken by all reduce tasks=1497
                Total megabyte-milliseconds taken by all map tasks=1418240
                Total megabyte-milliseconds taken by all reduce tasks=1532928
        Map-Reduce Framework
                Map input records=1
                Map output records=13
                Map output bytes=108
                Map output materialized bytes=98
                Input split bytes=112
                Combine input records=13
                Combine output records=9
                Reduce input groups=9
                Reduce shuffle bytes=98
                Reduce input records=9
                Reduce output records=9
                Spilled Records=18
                Shuffled Maps =1
```

```
        Map-Reduce Framework
                Map input records=1
                Map output records=13
                Map output bytes=108
                Map output materialized bytes=98
                Input split bytes=112
                Combine input records=13
                Combine output records=9
                Reduce input groups=9
                Reduce shuffle bytes=98
                Reduce input records=9
                Reduce output records=9
                Spilled Records=18
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=64
                CPU time spent (ms)=710
                Physical memory (bytes) snapshot=485638144
                Virtual memory (bytes) snapshot=5080743936
                Total committed heap usage (bytes)=517996544
                Peak Map Physical memory (bytes)=294223872
                Peak Map Virtual memory (bytes)=2535649280
                Peak Reduce Physical memory (bytes)=191414272
                Peak Reduce Virtual memory (bytes)=2545094656
        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=56
        File Output Format Counters
                Bytes Written=56
```

output

```
                Bytes Read=56
        File Output Format Counters
                Bytes Written=56
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat ~/out/part*
2022-06-17 10:47:44,841 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
am      1
are     2
fine    1
helloo  1
how     2
hy      1
i       1
there   2
you     2
```

Program 4 : Map reduce program

```
import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;


public class TransactionCount {
//MAPPER CODE

public static class Map 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 myString = value.toString();
String[] userCount = myString.split(",");
output.collect(new Text(userCount[3]), one);


}
}

//REDUCER CODE
public static class Reduce extends MapReduceBase implements
Reducer<Text, IntWritable, Text, IntWritable> {
public void reduce(Text key, Iterator<IntWritable> values,
OutputCollector<Text, IntWritable> output, Reporter reporter) throws
IOException { //{little: {1,1}}
int finaluserCount = 0 ;
Text mykey = key ;
while(values.hasNext()) {
IntWritable value = values.next();
finaluserCount += value.get();
}
output.collect(mykey, new IntWritable(finaluserCount));
}
}
//DRIVER CODE
public static void main(String[] args) throws Exception {
JobConf conf = new JobConf(TransactionCount.class);
conf.setJobName("wordcount");
conf.setOutputKeyClass(Text.class);
conf.setOutputValueClass(IntWritable.class);
conf.setMapperClass(Map.class);
```

```
conf.setCombinerClass(Reduce.class);
conf.setReducerClass(Reduce.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);
}


}
```


Output