1BM22CS157_BDA_lab7

```java
 1⊕ import java.io.IOException;
12  public class WCDriver extends Configured implements Tool {
13⊖ public int run(String args[]) throws IOException
14  {
15      if (args.length < 2)
16      {
17          System.out.println("Please give valid inputs");
18          return -1;
19      }
20      JobConf conf = new JobConf(WCDriver.class);
21      FileInputFormat.setInputPaths(conf, new Path(args[0]));
22      FileOutputFormat.setOutputPath(conf, new Path(args[1]));
23      conf.setMapperClass(WCMapper.class);
24      conf.setReducerClass(WCReducer.class);
25      conf.setMapOutputKeyClass(Text.class);
26      conf.setMapOutputValueClass(IntWritable.class);
27      conf.setOutputKeyClass(Text.class);conf.setOutputValueClass(IntWritable.class);
28      JobClient.runJob(conf);
29      return 0;
30      }
31      // Main Method
32⊖  public static void main(String args[]) throws Exception
33      {
34          int exitCode = ToolRunner.run(new WCDriver(), args);
35          System.out.println(exitCode);
36      }
37  }
```

Console ✕   Problems   Debug Shell   Coverage

```java
1⊕ import java.io.IOException;
9  public class WCMapper extends MapReduceBase implements Mapper<LongWritable,Text,Text,IntWritable> {
10 // Map function
11⊖     public void map(LongWritable key, Text value, OutputCollector<Text,IntWritable> output, Reporter rep) throws IOException
12     {
13         String line = value.toString();
14         // Splitting the line on spaces
15         for (String word : line.split(" "))
16         {
17             if (word.length() > 0)
18             {
19                 output.collect(new Text(word), new IntWritable(1));
20             }
21         }
22     }
23 }
```

```java
import java.io.IOException;
public class WCReducer extends MapReduceBase implements Reducer<Text,
IntWritable, Text, IntWritable> {
// Reduce function
    public void reduce(Text key, Iterator<IntWritable> value, OutputCollector<Text, IntWritable> output, Reporter rep) throws IOException
    {
        int count = 0;
        // Counting the frequency of each words
        while (value.hasNext())
        {
            IntWritable i = value.next();
            count += i.get();
        }
        output.collect(key, new IntWritable(count));
    }
}
```

```
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop jar WordCount.jar WCDriver /shuraih/test.txt /shuraih/out.txt
2025-04-29 15:11:42,145 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-04-29 15:11:42,185 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-04-29 15:11:42,185 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-04-29 15:11:42,191 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-04-29 15:11:42,254 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-04-29 15:11:42,303 INFO mapred.FileInputFormat: Total input files to process : 1
2025-04-29 15:11:42,328 INFO mapreduce.JobSubmitter: number of splits:1
2025-04-29 15:11:42,390 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1048437123_0001
2025-04-29 15:11:42,390 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-29 15:11:42,448 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-04-29 15:11:42,449 INFO mapreduce.Job: Running job: job_local1048437123_0001
2025-04-29 15:11:42,449 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-04-29 15:11:42,451 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2025-04-29 15:11:42,453 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:11:42,453 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:11:42,507 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-04-29 15:11:42,508 INFO mapred.LocalJobRunner: Starting task: attempt_local1048437123_0001_m_000000_0
2025-04-29 15:11:42,519 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-04-29 15:11:42,519 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-04-29 15:11:42,526 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2025-04-29 15:11:42,529 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/shuraih/test.txt:0+89
2025-04-29 15:11:42,538 INFO mapred.MapTask: numReduceTasks: 1
2025-04-29 15:11:42,569 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2025-04-29 15:11:42,569 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2025-04-29 15:11:42,569 INFO mapred.MapTask: soft limit at 83886080
2025-04-29 15:11:42,569 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2025-04-29 15:11:42,569 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2025-04-29 15:11:42,571 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2025-04-29 15:11:42,608 INFO mapred.LocalJobRunner:
2025-04-29 15:11:42,608 INFO mapred.MapTask: Starting flush of map output
2025-04-29 15:11:42,608 INFO mapred.MapTask: Spilling map output
2025-04-29 15:11:42,608 INFO mapred.MapTask: bufstart = 0; bufend = 169; bufvoid = 104857600
2025-04-29 15:11:42,608 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214320(104857280); length = 77/6553600
2025-04-29 15:11:42,611 INFO mapred.MapTask: Finished spill 0
2025-04-29 15:11:42,616 INFO mapred.Task: Task:attempt_local1048437123_0001_m_000000_0 is done. And is in the process of committing
2025-04-29 15:11:42,618 INFO mapred.LocalJobRunner: hdfs://localhost:9000/shuraih/test.txt:0+89
2025-04-29 15:11:42,618 INFO mapred.Task: Task 'attempt_local1048437123_0001_m_000000_0' done.
2025-04-29 15:11:42,621 INFO mapred.Task: Final Counters for attempt_local1048437123_0001_m_000000_0: Counters: 23
        File System Counters
                FILE: Number of bytes read=4049
                FILE: Number of bytes written=645512
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
```

```
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -ls /shuraih/out.txt
Found 2 items
-rw-r--r--   1 hadoop supergroup          0 2025-04-29 15:11 /shuraih/out.txt/_SUCCESS
-rw-r--r--   1 hadoop supergroup         69 2025-04-29 15:11 /shuraih/out.txt/part-00000
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$ hadoop fs -cat /shuraih/out.txt/part-00000
are     1
brother 1
family  1
hi      1
how     5
is      4
job     1
sister  1
you     1
your    4
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~/Desktop$
```

```
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [bmscecse-HP-Elite-Tower-800-G9-Desktop-PC]
Starting resourcemanager
Starting nodemanagers
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ su hduser
su: user hduser does not exist or the user entry does not contain all the required fields
hadoop@bmscecse-HP-Elite-Tower-800-G9-Desktop-PC:~$ jps
7890 ResourceManager
8054 NodeManager
8488 Jps
7259 DataNode
5468 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
7117 NameNode
7517 SecondaryNameNode
```