4.Create Three Java Classes into the project. Name them WCDriver(having the main function),WCMapper,WCReducer.

Mapper Program:

```
// Importing libraries
import java.io.IOException;
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;
public class WCMapper extends MapReduceBase implements
Mapper&lt;LongWritable,Text, Text, IntWritable&gt; {
// Map function
public void map(LongWritable key, Text value,
OutputCollector&lt;Text,
IntWritable&gt; output, Reporter rep) throws IOException

{
String line = value.toString();
// Splitting the line on spaces
for (String word : line.split(&quot; &quot;))
{
if (word.length() &gt; 0)
{
output.collect(new Text(word), new IntWritable(1));
}
}
}
}
```

Reducer Code:

```
 // Importing libraries
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 WCReducer extends MapReduceBase implements
Reducer&lt;Text,IntWritable, Text, IntWritable&gt; {
// Reduce function
public void reduce(Text key, Iterator&lt;IntWritable&gt; value,
OutputCollector&lt;Text, IntWritable&gt; 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));
}
}
```

Driver Code:

```
// Importing libraries
import java.io.IOException;
import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;

import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;
public class WCDriver extends Configured implements Tool {
public int run(String args[]) throws IOException
{
if (args.length &lt; 2)
{
System.out.println(&quot;Please give valid inputs&quot;);
return -1;
}
JobConf conf = new JobConf(WCDriver.class);
FileInputFormat.setInputPaths(conf, new Path(args[0]));
FileOutputFormat.setOutputPath(conf, new Path(args[1]));
conf.setMapperClass(WCMapper.class);
conf.setReducerClass(WCReducer.class);
conf.setMapOutputKeyClass(Text.class);
conf.setMapOutputValueClass(IntWritable.class);
conf.setOutputKeyClass(Text.class);
conf.setOutputValueClass(IntWritable.class);
JobClient.runJob(conf);
return 0;
}
// Main Method
public static void main(String args[]) throws Exception
{
int exitCode = ToolRunner.run(new WCDriver(), args);
System.out.println(exitCode);
}
}
```

OUTPUT:

```
1 STARTING HADOOP
```

```
hduser@bmsce-Precision-T1700:~$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
hduser@localhost's password:
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-bmsce-Precision-T1700.out
hduser@localhost's password:
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-bmsce-Precision-T1700.out
Starting secondary namenodes [0.0.0.0]
hduser@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-secondarynamenode-bmsce-Precision-T1700.out
bmsce
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-bmsce-Precision-T1700.out
hduser@localhost's password:
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-bmsce-Precision-T1700.out
hduser@bmsce-Precision-T1700:~$ jps
8117 Jps
7285 DataNode
7994 NodeManager
7658 ResourceManager
7115 NameNode
7499 SecondaryNameNode
```

2 files in hadoop

```
hduser@bmsce-Precision-T1700:~$ hadoop fs -ls /
Found 4 items
drwxr-xr-x   - hduser supergroup          0 2019-10-23 16:07 /STUDENT_INFO
drwxr-xr-x   - hduser supergroup          0 2019-10-23 15:08 /arv
drwxrwxr-x   - hduser supergroup          0 2019-10-23 15:36 /tmp
drwxr-xr-x   - hduser supergroup          0 2019-08-01 16:03 /user
hduser@bmsce-Precision-T1700:~$ hadoop fs -mkdir /rgs
hduser@bmsce-Precision-T1700:~$ hadoop fs -copyFromLocal /home/bmsce/Desktop/sample.txt /rgs/test.txt
hduser@bmsce-Precision-T1700:~$ hadoop fs -ls /rgs/
Found 1 items
-rw-r--r--   1 hduser supergroup         89 2022-06-14 10:09 /rgs/test.txt
```

3 Running word count

```
hduser@bmsce-Precision-T1700:~$ hadoop jar /home/bmsce/eclipse-workspace/WordCount.jar WCDriver /rgs/test.txt /output/
22/06/14 10:14:29 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
22/06/14 10:14:29 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
22/06/14 10:14:29 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already
22/06/14 10:14:29 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool int
22/06/14 10:14:29 INFO mapred.FileInputFormat: Total input paths to process : 1
22/06/14 10:14:29 INFO mapreduce.JobSubmitter: number of splits:1
22/06/14 10:14:29 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1586040550_0001
22/06/14 10:14:29 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
22/06/14 10:14:29 INFO mapred.LocalJobRunner: OutputCommitter set in config null
22/06/14 10:14:29 INFO mapreduce.Job: Running job: job_local1586040550_0001
22/06/14 10:14:29 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
22/06/14 10:14:29 INFO mapred.LocalJobRunner: Waiting for map tasks
22/06/14 10:14:29 INFO mapred.LocalJobRunner: Starting task: attempt_local1586040550_0001_m_000000_0
22/06/14 10:14:29 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
```

```
22/06/14 10:14:29 INFO mapred.Task: Task:attempt_local1586040550_0001_r_000000_0 is done. And is in the process of committing
22/06/14 10:14:29 INFO mapred.LocalJobRunner: 1 / 1 copied.
22/06/14 10:14:29 INFO mapred.Task: Task attempt_local1586040550_0001_r_000000_0 is allowed to commit now
22/06/14 10:14:29 INFO output.FileOutputCommitter: Saved output of task 'attempt_local1586040550_0001_r_000000_0' to hdfs://loc
22/06/14 10:14:29 INFO mapred.LocalJobRunner: reduce > reduce
22/06/14 10:14:29 INFO mapred.Task: Task 'attempt_local1586040550_0001_r_000000_0' done.
22/06/14 10:14:29 INFO mapred.LocalJobRunner: Finishing task: attempt_local1586040550_0001_r_000000_0
22/06/14 10:14:29 INFO mapred.LocalJobRunner: reduce task executor complete.
22/06/14 10:14:30 INFO mapreduce.Job: Job job_local1586040550_0001 running in uber mode : false
22/06/14 10:14:30 INFO mapreduce.Job:  map 100% reduce 100%
22/06/14 10:14:30 INFO mapreduce.Job: Job job_local1586040550_0001 completed successfully
22/06/14 10:14:30 INFO mapreduce.Job: Counters: 38
        File System Counters
                FILE: Number of bytes read=8718
                FILE: Number of bytes written=510607
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=178
                HDFS: Number of bytes written=69
                HDFS: Number of read operations=13
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
        Map-Reduce Framework
                Map input records=5
                Map output records=20
                Map output bytes=169
                Map output materialized bytes=215
                Input split bytes=87
                Combine input records=0
                Combine output records=0
                Reduce input groups=10
                Reduce shuffle bytes=215
                Reduce input records=20
                Reduce output records=10
                Spilled Records=40
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=2
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=466616320
        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=89
        File Output Format Counters
                Bytes Written=69
0
```

4 MapReduce Output

```
hduser@bmsce-Precision-T1700:~$ hdfs dfs -cat /output/*
are     1
brother 1
family  1
hi      1
how     5
is      4
job     1
sister  1
you     1
your    4
hduser@bmsce-Precision-T1700:~$
```