Screen Shots of Execution
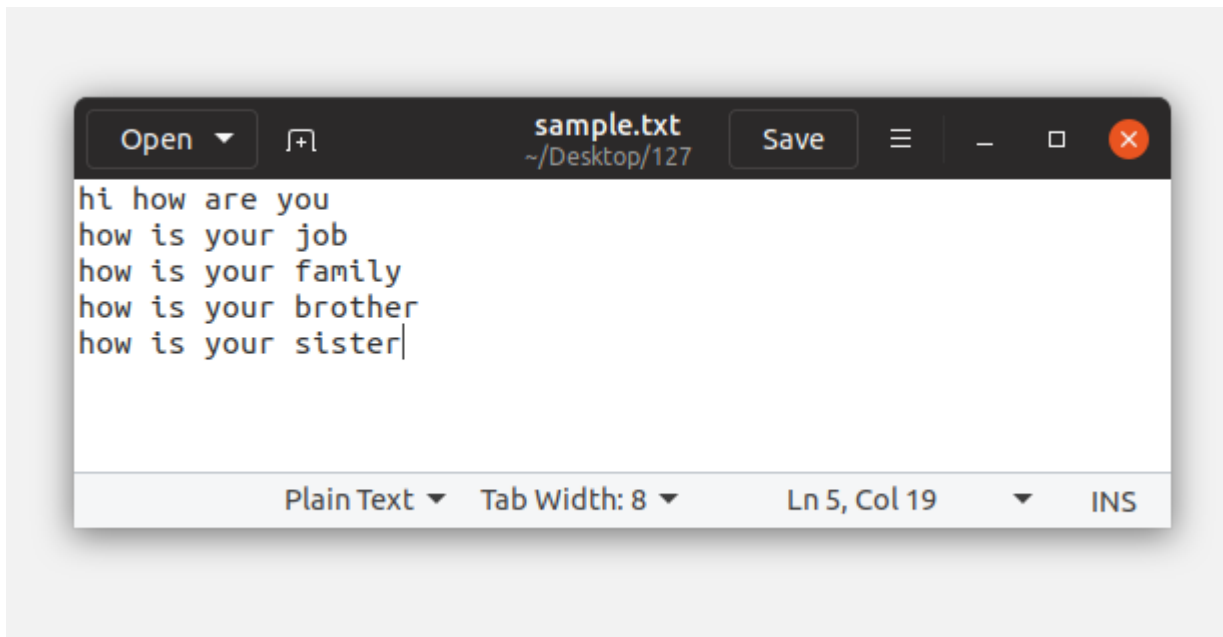


```
hduser@bmsce-Precision-T1700: ~                                    ×          bmsce@bmsce-Precision-T1700: ~/Desktop/127                    ×

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
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
6512 DataNode
7217 NodeManager
6884 ResourceManager
6725 SecondaryNameNode
6344 NameNode
7339 Jps
hduser@bmsce-Precision-T1700:~$ hadoop fs -ls /
Found 3 items
drwxr-xr-x   - hduser supergroup          0 2019-10-23 15:08 /dg
drwxrwxrwx   - hduser supergroup          0 2019-10-23 15:53 /tmp
drwxr-xr-x   - hduser supergroup          0 2019-08-01 16:03 /user
hduser@bmsce-Precision-T1700:~$ hadoop fs -ls /
Found 3 items
drwxr-xr-x   - hduser supergroup          0 2019-10-23 15:08 /dg
drwxrwxrwx   - hduser supergroup          0 2019-10-23 15:53 /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/127/sample.txt  /rgs/test.txt
hduser@bmsce-Precision-T1700:~$ hadoop jar /home/bmsce/Desktop/127/WordCount.jar WCDriver input output
22/06/14 10:00:28 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
22/06/14 10:00:28 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
22/06/14 10:00:28 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
22/06/14 10:00:28 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/06/14 10:00:28 INFO mapreduce.JobSubmitter: Cleaning up the staging area file:/app/hadoop/tmp/mapred/staging/hduser1428536112/.staging/job_local1428536112_0001
Exception in thread "main" org.apache.hadoop.mapred.InvalidInputException: Input path does not exist: hdfs://localhost:54310/user/hduser/input
        at org.apache.hadoop.mapred.FileInputFormat.singleThreadedListStatus(FileInputFormat.java:285)
        at org.apache.hadoop.mapred.FileInputFormat.listStatus(FileInputFormat.java:228)
        at org.apache.hadoop.mapred.FileInputFormat.getSplits(FileInputFormat.java:313)
        at org.apache.hadoop.mapreduce.JobSubmitter.writeOldSplits(JobSubmitter.java:624)
        at org.apache.hadoop.mapreduce.JobSubmitter.writeSplits(JobSubmitter.java:616)
        at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:492)
        at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1296)
        at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1293)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1628)
        at org.apache.hadoop.mapreduce.Job.submit(Job.java:1293)
        at org.apache.hadoop.mapred.JobClient$1.run(JobClient.java:562)
        at org.apache.hadoop.mapred.JobClient$1.run(JobClient.java:557)
```

```
        at org.apache.hadoop.mapred.JobClient$1.run(JobClient.java:557)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1628)
        at org.apache.hadoop.mapred.JobClient.submitJobInternal(JobClient.java:557)
        at org.apache.hadoop.mapred.JobClient.submitJob(JobClient.java:548)
        at org.apache.hadoop.mapred.JobClient.runJob(JobClient.java:833)
        at WCDriver.run(WCDriver.java:33)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)
        at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:84)
        at WCDriver.main(WCDriver.java:40)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
hduser@bmsce-Precision-T1700:~$ hadoop fs -ls /output
ls: `/output': No such file or directory
hduser@bmsce-Precision-T1700:~$ hadoop fs -ls /
Found 4 items
drwxr-xr-x   - hduser supergroup          0 2019-10-23 15:08 /dg
drwxr-xr-x   - hduser supergroup          0 2022-06-14 09:58 /rgs
drwxrwxrwx   - hduser supergroup          0 2019-10-23 15:53 /tmp
drwxr-xr-x   - hduser supergroup          0 2019-08-01 16:03 /user
hduser@bmsce-Precision-T1700:~$ hadoop fs -cat /output/part-00000
cat: `/output/part-00000': No such file or directory
hduser@bmsce-Precision-T1700:~$ cd rgs
bash: cd: rgs: No such file or directory
hduser@bmsce-Precision-T1700:~$ hadoop fs -ls /
Found 4 items
drwxr-xr-x   - hduser supergroup          0 2019-10-23 15:08 /dg
drwxr-xr-x   - hduser supergroup          0 2022-06-14 09:58 /rgs
drwxrwxrwx   - hduser supergroup          0 2019-10-23 15:53 /tmp
drwxr-xr-x   - hduser supergroup          0 2019-08-01 16:03 /user
hduser@bmsce-Precision-T1700:~$ cd /rgs
bash: cd: /rgs: No such file or directory
hduser@bmsce-Precision-T1700:~$ hdfs dfs /cd rgs
/cd: Unknown command
hduser@bmsce-Precision-T1700:~$ hdfs dfs -cd /rgs
-cd: Unknown command
hduser@bmsce-Precision-T1700:~$ hdfs dfs -cd /rgs/
-cd: Unknown command
hduser@bmsce-Precision-T1700:~$ hadoop fs -cd rgs
-cd: Unknown command
hduser@bmsce-Precision-T1700:~$ hdfs dfs -ls /rgs
Found 1 items
-rw-r--r--   1 hduser supergroup         89 2022-06-14 09:58 /rgs/test.txt
hduser@bmsce-Precision-T1700:~$ hdfs dfs -cat /rgs/test.txt
hi how are you
how is your job
how is your family
how is your brother
hduser@bmsce-Precision-T1700:~$ hadoop fs -copyFromLocal /home/bmsce/Deskthdusehduser@bmsce-Precision-T1700hduser@bmhdusehduser@bmshduser@bmshduser@hdusehdusehhduser@hhhduser@bmsce-Precision-T1700:~$ hado
```

```
hduser@bmsce-Precision-T1700:~$ hadoop fs -copyFromLocal /home/bmsce/Deskthdusehduser@bmsce-Precision-T1700hduser@bmhdusehduser@bmshdusehduser@bmshduser@hdusehdusehhduser@hhhduser@bmsce-Precision-T1700:~$ hado
op jar /home/bmsce/Desktop/127/WordCount.jar WCDriver /rgs/test.txt/output
Please give valid inputs
-1
hduser@bmsce-Precision-T1700:~$ hadoop jar /home/bmsce/Desktop/127/WordCount.jar WCDriver /rgs/test.txt /output
22/06/14 10:12:24 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
22/06/14 10:12:24 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
22/06/14 10:12:24 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
22/06/14 10:12:24 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/06/14 10:12:24 INFO mapred.FileInputFormat: Total input paths to process : 1
22/06/14 10:12:24 INFO mapreduce.JobSubmitter: number of splits:1
22/06/14 10:12:24 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local195599312_0001
22/06/14 10:12:24 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
22/06/14 10:12:24 INFO mapreduce.Job: Running job: job_local195599312_0001
22/06/14 10:12:24 INFO mapred.LocalJobRunner: OutputCommitter set in config null
22/06/14 10:12:24 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
22/06/14 10:12:24 INFO mapred.LocalJobRunner: Waiting for map tasks
22/06/14 10:12:24 INFO mapred.LocalJobRunner: Starting task: attempt_local195599312_0001_m_000000_0
22/06/14 10:12:24 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
22/06/14 10:12:24 INFO mapred.MapTask: Processing split: hdfs://localhost:54310/rgs/test.txt:0+89
22/06/14 10:12:24 INFO mapred.MapTask: numReduceTasks: 1
22/06/14 10:12:25 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
22/06/14 10:12:25 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
22/06/14 10:12:25 INFO mapred.MapTask: soft limit at 83886080
22/06/14 10:12:25 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
22/06/14 10:12:25 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
22/06/14 10:12:25 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
22/06/14 10:12:25 INFO mapred.LocalJobRunner:
22/06/14 10:12:25 INFO mapred.MapTask: Starting flush of map output
22/06/14 10:12:25 INFO mapred.MapTask: Spilling map output
22/06/14 10:12:25 INFO mapred.MapTask: bufstart = 0; bufend = 169; bufvoid = 104857600
22/06/14 10:12:25 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214320(104857280); length = 77/6553600
22/06/14 10:12:25 INFO mapred.MapTask: Finished spill 0
22/06/14 10:12:25 INFO mapred.Task: Task:attempt_local195599312_0001_m_000000_0 is done. And is in the process of committing
22/06/14 10:12:25 INFO mapred.LocalJobRunner: hdfs://localhost:54310/rgs/test.txt:0+89
22/06/14 10:12:25 INFO mapred.Task: Task 'attempt_local195599312_0001_m_000000_0' done.
22/06/14 10:12:25 INFO mapred.LocalJobRunner: Finishing task: attempt_local195599312_0001_m_000000_0
22/06/14 10:12:25 INFO mapred.LocalJobRunner: map task executor complete.
22/06/14 10:12:25 INFO mapred.LocalJobRunner: Waiting for reduce tasks
22/06/14 10:12:25 INFO mapred.LocalJobRunner: Starting task: attempt_local195599312_0001_r_000000_0
22/06/14 10:12:25 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
22/06/14 10:12:25 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@6ecbb28b
22/06/14 10:12:25 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=334338464, maxSingleShuffleLimit=83584616, mergeThreshold=220663392, ioSortFactor=10, memToMemMergeOutputsThreshold=10
22/06/14 10:12:25 INFO reduce.EventFetcher: attempt_local195599312_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
22/06/14 10:12:25 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local195599312_0001_m_000000_0 decomp: 211 len: 215 to MEMORY
22/06/14 10:12:25 INFO reduce.InMemoryMapOutput: Read 211 bytes from map-output for attempt_local195599312_0001_m_000000_0
22/06/14 10:12:25 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 211, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->211
22/06/14 10:12:25 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
22/06/14 10:12:25 INFO mapred.LocalJobRunner: 1 / 1 copied.
22/06/14 10:12:25 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
22/06/14 10:12:25 INFO mapred.Merger: Merging 1 sorted segments
22/06/14 10:12:25 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 205 bytes
22/06/14 10:12:25 INFO reduce.MergeManagerImpl: Merged 1 segments, 211 bytes to disk to satisfy reduce memory limit
22/06/14 10:12:25 INFO reduce.MergeManagerImpl: Merging 1 files, 215 bytes from disk
```

```
22/06/14 10:12:25 INFO reduce.MergeManagerImpl: Merged 1 segments, 211 bytes to disk to satisfy reduce memory limit
22/06/14 10:12:25 INFO reduce.MergeManagerImpl: Merging 1 files, 215 bytes from disk
22/06/14 10:12:25 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
22/06/14 10:12:25 INFO mapred.Merger: Merging 1 sorted segments
22/06/14 10:12:25 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 205 bytes
22/06/14 10:12:25 INFO mapred.LocalJobRunner: 1 / 1 copied.
22/06/14 10:12:25 INFO mapred.Task: Task:attempt_local195599312_0001_r_000000_0 is done. And is in the process of committing
22/06/14 10:12:25 INFO mapred.LocalJobRunner: 1 / 1 copied.
22/06/14 10:12:25 INFO mapred.Task: Task attempt_local195599312_0001_r_000000_0 is allowed to commit now
22/06/14 10:12:25 INFO output.FileOutputCommitter: Saved output of task 'attempt_local195599312_0001_r_000000_0' to hdfs://localhost:54310/output/_temporary/0/task_local195599312_0001_r_000000
22/06/14 10:12:25 INFO mapred.LocalJobRunner: reduce > reduce
22/06/14 10:12:25 INFO mapred.Task: Task 'attempt_local195599312_0001_r_000000_0' done.
22/06/14 10:12:25 INFO mapred.LocalJobRunner: Finishing task: attempt_local195599312_0001_r_000000_0
22/06/14 10:12:25 INFO mapred.LocalJobRunner: reduce task executor complete.
22/06/14 10:12:25 INFO mapreduce.Job: Job job_local195599312_0001 running in uber mode : false
22/06/14 10:12:25 INFO mapreduce.Job:  map 100% reduce 100%
22/06/14 10:12:25 INFO mapreduce.Job: Job job_local195599312_0001 completed successfully
22/06/14 10:12:25 INFO mapreduce.Job: Counters: 38
        File System Counters
                FILE: Number of bytes read=8630
                FILE: Number of bytes written=507803
                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)=1
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=468713472
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
```

```
                Total committed heap usage (bytes)=468713472
        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
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:~$
```

Input Sample file :


```
Open ▼    ⨆              sample.txt          Save    ≡   —   ▢   ✕
                        ~/Desktop/127
hi how are you
how is your job
how is your family
how is your brother
how is your sister
```
```
Plain Text ▼   Tab Width: 8 ▼        Ln 5, Col 19       ▼      INS
```

Final 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:~$ ▮
```

Program :

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

```java
// 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<LongWritable,Text,
Text, IntWritable> {

        // Map function
        public void map(LongWritable key, Text value, OutputCollector<Text,
                        IntWritable> output, Reporter rep) throws IOException
        {

                String line = value.toString();

                // Splitting the line on spaces
                for (String word : line.split(" "))
                {
                        if (word.length() > 0)
                        {
                                output.collect(new Text(word), new IntWritable(1));
                        }
                }
        }
}
```

Reducer Code: You have to copy paste this program into the WCReducer Java Class file.

```java
 // 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<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));
        }
}
```

Driver Code: You have to copy paste this program into the WCDriver Java Class file.

```java
// 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 < 2)
                {
                        System.out.println("Please give valid inputs");
                        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);
        }
}
```

Export as jar file

Mapreduce Program

1. hduser@bmsce-Precision-T1700:/$ su hduser
Password:
hduser@bmsce-Precision-T1700:/$ cd /
hduser@bmsce-Precision-T1700:/$ cd /usr/local/hadoop/sbin

2. hduser@bmsce-Precision-T1700:/usr/local/hadoop/sbin$ start-all.sh

3. hduser@bmsce-Precision-T1700:/usr/local/hadoop/sbin$ jps

4. create a file on Desktop(sample.txt) and type the below lines:

hi how are you
how is your job
how is your family
how is your brother
how is your sister

save your file

5. View the directory content
hadoop fs -ls /

6. Create a directory using the following command. If any directory existing, use the same
directory for the command
hadoop fs -mkdir /rgs

7. Copy the file into HDFS

hadoop fs -copyFromLocal D:/sample.txt /rgs/test.txt

8.Run the Map Reduce Program

hadoop fs -ls /output/

hadoop fs -copyFromLocal /home/hduser/Desktop/file1.txt /rgs/test.txt

hadoop jar /home/hduser/Desktop/MapReduceClient.jar WCDriver /rgs/test.txt /output/
hdfs dfs -cat /output/*

My File

WordCount   WordCount   sample.txt   WordCount
              101   .jar