

```
(base) student@student:~$ su - hadoop
Password:
hadoop@student:~$ cd hadoop
hadoop@student:~/hadoop$ nano wordCount.txt
hadoop@student:~/hadoop$ hdfs dfs -mkdir /TA
hadoop@student:~/hadoop$ hdfs dfs -put /home/hadoop/wordCount.txt /TA
put: '/home/hadoop/wordCount.txt': No such file or directory
hadoop@student:~/hadoop$ hdfs dfs -put /home/hadoop/hadoop/wordCount.txt /TA
hadoop@student:~/hadoop$ nano WC_Mapper.java
hadoop@student:~/hadoop$ nano WC_Reducer.java
hadoop@student:~/hadoop$ nano WC_Runner.java
hadoop@student:~/hadoop$ javac -classpath "$(hadoop classpath)" -d WC_Mapper.java WC_Reducer.java WC_Runner.java
error: not a directory: WC_Mapper.java
hadoop@student:~/hadoop$ javac -classpath "$(hadoop classpath)" -d WC_Mapper.java WC_Reducer.java WC_Runner.java
ERROR: Classpath is not COMMAND nor fully qualified CLASSNAME.
error: not a directory: WC_Mapper.java
hadoop@student:~/hadoop$ javac -classpath "$(hadoop classpath)" -d WC_Mapper.java WC_Reducer.java WC_Runner.java
error: not a directory: WC_Mapper.java
hadoop@student:~/hadoop$ nano WC_Reducer.java
hadoop@student:~/hadoop$ javac -classpath "$(hadoop classpath)" -d . WC_Mapper.java WC_Reducer.java WC_Runner.java
hadoop@student:~/hadoop$ jar -cvf wordCount.jar com
added manifest
adding: com/(in = 0) (out= 0)(stored 0%)
adding: com/javatpoint/(in = 0) (out= 0)(stored 0%)
adding: com/javatpoint/WC_Runner.class(in = 1485) (out= 728)(deflated 50%)
adding: com/javatpoint/WC_Mapper.class(in = 1874) (out= 768)(deflated 59%)
adding: com/javatpoint/WC_Reducer.class(in = 1546) (out= 613)(deflated 60%)
hadoop@student:~/hadoop$ hadoop jar /home/hadoop/hadoop/wordCount.jar com.javatpoint.WC_Runner/TA/wordCount.txt/r_output
Exception in thread "main" java.lang.ClassNotFoundException: com.javatpoint.WC_Runner.TA.wordCount.txt.r_output
    at java.base/java.net.URLClassLoader.findClass(URLClassLoader.java:476)
    at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:594)
    at java.base/java.lang.ClassLoader.loadClass(ClassLoader.java:527)
    at java.base/java.lang.Class.forName0(Native Method)
    at java.base/java.lang.Class.forName(Class.java:398)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:321)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:241)
hadoop@student:~/hadoop$ nano WC_Reducer.java
hadoop@student:~/hadoop$ hadoop jar /home/hadoop/hadoop/wordCount.jar com.javatpoint.WC_Runner /TA/wordCount.txt/r_output
```

```
hadoop@student:~/hadoop$ nano WC_Reducer.java
hadoop@student:~/hadoop$ hadoop jar /home/hadoop/hadoop/wordCount.jar com.javatpoint.WC_Runner /TA/wordCount.txt/r_output
Exception in thread "main" java.lang.ArrayIndexOutOfBoundsException: Index 1 out of bounds for length 1
    at com.javatpoint.WC_Runner.main(WC_Runner.java:25)
    at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.base/java.lang.reflect.Method.invoke(Method.java:566)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:328)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:241)
hadoop@student:~/hadoop$ hadoop jar /home/hadoop/hadoop/wordCount.jar com.javatpoint.WC_Runner /TA/wordCount.txt/ r_output
2024-04-15 16:36:00,606 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-04-15 16:36:00,772 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-04-15 16:36:00,772 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-04-15 16:36:00,804 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-04-15 16:36:01,175 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-04-15 16:36:01,297 INFO mapred.FileInputFormat: Total input files to process : 1
2024-04-15 16:36:01,420 INFO mapreduce.JobSubmitter: number of splits:1
2024-04-15 16:36:01,775 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1802326078_0001
2024-04-15 16:36:01,776 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-04-15 16:36:01,957 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-04-15 16:36:01,959 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-04-15 16:36:01,960 INFO mapreduce.Job: Running job: job_local1802326078_0001
2024-04-15 16:36:01,961 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapred.FileOutputCommitter
2024-04-15 16:36:01,970 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-04-15 16:36:01,970 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-04-15 16:36:02,048 INFO mapred.LocalJobRunner: Waiting for map tasks
2024-04-15 16:36:02,051 INFO mapred.LocalJobRunner: Starting task: attempt_local1802326078_0001_m_000000_0
2024-04-15 16:36:02,071 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-04-15 16:36:02,071 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2024-04-15 16:36:02,089 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-04-15 16:36:02,097 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/TA/wordCount.txt:0+43
2024-04-15 16:36:02,113 INFO mapred.MapTask: numReduceTasks: 1
2024-04-15 16:36:02,168 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2024-04-15 16:36:02,168 INFO mapred.MapTask: mapreduce.task.sort.mb: 100
2024-04-15 16:36:02,168 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
```



```
Activities Terminal Apr 15 16:40 hadoop@student: ~/hadoop

2024-04-15 16:36:02,168 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2024-04-15 16:36:02,168 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2024-04-15 16:36:02,168 INFO mapred.MapTask: soft limit at 83886080
2024-04-15 16:36:02,168 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2024-04-15 16:36:02,168 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2024-04-15 16:36:02,172 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2024-04-15 16:36:02,271 INFO mapred.LocalJobRunner:
2024-04-15 16:36:02,271 INFO mapred.MapTask: Starting flush of map output
2024-04-15 16:36:02,272 INFO mapred.MapTask: Spilling map output
2024-04-15 16:36:02,272 INFO mapred.MapTask: bufstart = 0; bufend = 75; bufvoid = 104857600
2024-04-15 16:36:02,272 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214368(104857472); length = 29/6553600
2024-04-15 16:36:02,278 INFO mapred.MapTask: Finished spill 0
2024-04-15 16:36:02,288 INFO mapred.Task: Task:attempt_local1802326078_0001_m_000000_0 is done. And is in the process of committing
2024-04-15 16:36:02,303 INFO mapred.LocalJobRunner: hdfs://localhost:9000/TA/wordCount.txt:0+43
2024-04-15 16:36:02,303 INFO mapred.Task: Task 'attempt_local1802326078_0001_m_000000_0' done.
2024-04-15 16:36:02,308 INFO mapred.Task: Final Counters for attempt_local1802326078_0001_m_000000_0: Counters: 24
File System Counters
  FILE: Number of bytes read=3227
  FILE: Number of bytes written=645151
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=43
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=5
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=1
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=1
  Map output records=8
  Map output bytes=75
  Map output materialized bytes=67
  Input split bytes=90
  Combine input records=8
  Combine output records=5
  Spilled Records=5
  Failed Shuffles=0
  Hired WritableBytes=0
```

Activities  Terminal

Apr 15 16:40

```
hadoop@student: ~/hadoop
```

```
Spilled Records=5
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=7
Total committed heap usage (bytes)=228589568
```

```
File Input Format Counters
  Bytes Read=43
```

```

2024-04-15 16:36:02,308 INFO napred.LocalJobRunner: Finishing task: attempt_local1802326078_0001_m_000000_0
2024-04-15 16:36:02,309 INFO napred.LocalJobRunner: map task executor complete.
2024-04-15 16:36:02,312 INFO napred.LocalJobRunner: Waiting for reduce tasks
2024-04-15 16:36:02,312 INFO napred.LocalJobRunner: Starting task: attempt_local1802326078_0001_r_000000_0
2024-04-15 16:36:02,318 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-04-15 16:36:02,318 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false
2024-04-15 16:36:02,318 INFO napred.Task: Using ResourceCalculatorProcessTree : [ ]
2024-04-15 16:36:02,321 INFO napred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@50510d33
2024-04-15 16:36:02,322 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2024-04-15 16:36:02,336 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=706111040, maxSingleShuffleLimit=176527760, mergeThreshold=466033312, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2024-04-15 16:36:02,338 INFO reduce.EventFetcher: attempt_local1802326078_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2024-04-15 16:36:02,368 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local1802326078_0001_m_000000_0 decompressed 63 len: 67 to MEMORY
2024-04-15 16:36:02,375 INFO reduce.InMemoryMapOutput: Read 63 bytes from map-output for attempt_local1802326078_0001_m_000000_0
2024-04-15 16:36:02,378 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 63, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory -> 63
2024-04-15 16:36:02,387 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2024-04-15 16:36:02,387 INFO napred.LocalJobRunner: 1 / 1 copied.
2024-04-15 16:36:02,387 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2024-04-15 16:36:02,392 INFO napred.Merger: Merging 1 sorted segments
2024-04-15 16:36:02,392 INFO napred.Merger: Down to the last merge-pass, with 1 segments left of total size: 57 bytes
2024-04-15 16:36:02,394 INFO reduce.MergeManagerImpl: Merged 1 segments, 63 bytes to disk to satisfy reduce memory limit
2024-04-15 16:36:02,394 INFO reduce.MergeManagerImpl: Merging 1 files, 67 bytes from disk
2024-04-15 16:36:02,395 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2024-04-15 16:36:02,395 INFO napred.Merger: Merging 1 sorted segments
2024-04-15 16:36:02,396 INFO napred.Merger: Down to the last merge-pass, with 1 segments left of total size: 57 bytes
2024-04-15 16:36:02,397 INFO napred.LocalJobRunner: 1 / 1 copied.
2024-04-15 16:36:02,522 INFO napred.Task: Task:attempt_local1802326078_0001_r_000000_0 is done. And is in the process of committing

```

```
Activities Terminal Apr 15 16:40 hadoop@student: ~/hadoop

FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=86
HDFS: Number of bytes written=41
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=4
HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=1
  Map output records=8
  Map output bytes=75
  Map output materialized bytes=67
  Input split bytes=90
  Combine input records=8
  Combine output records=5
  Reduce input groups=5
  Reduce shuffle bytes=67
  Reduce input records=5
  Reduce output records=5
  Spilled Records=10
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=7
  Total committed heap usage (bytes)=457179136
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=43
File Output Format Counters
  Bytes Written=41
```

```
Activities Terminal Apr 15 16:40 hadoop@student: ~/hadoop

Spilled Records=10
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=7
Total committed heap usage (bytes)=457179136

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=43
File Output Format Counters
Bytes Written=41

hadoop@student:~/hadoop$ hdfs dfs -cat /r_output/part-00000
cat: '/r_output/part-00000': No such file or directory
hadoop@student:~/hadoop$ hdfs dfs -cat /output/part-00000
:) 1
Fine. 1
Good 1
Hello! 1
How 1
I 1
My 1
Sia. 1
What 1
am 1
are 1
is 2
morning 1
name 1
name? 1
you? 1
your 1
hadoop@student:~/hadoop$
```