

# 분산시스템 과제 #4

박종현, 2025-04-17

공과대학 컴퓨터정보통신공학과

## 과제 목표

1. 2-8 자료 : 7, 8, 9, 14, 15쪽
2. 2-9 자료 : 11, 12, 19, 20쪽
3. 2-10 자료 : 8, 9, 17, 18쪽

[Ch 2-8 p.7-9]

```
Windows PowerShell
hduser@DESKTOP-63506IV: ~\hadoop_test\trigramcount$ javac -classpath $(hadoop classpath) -d bin src/*.java
jar -cvf trigramcount.jar -C bin .
hadoop jar trigramcount.jar CountTrigram input/under_sea tc_out1
Note: Some input files use or override a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
Note: src/TopN.java uses unchecked or unsafe operations.
Note: Recompile with -Xlint:unchecked for details.
added manifest
adding: CountTrigram.class(in = 2510) (out= 1231)(deflated 50%)
adding: TopNSItemFreqComparator.class(in = 730) (out= 438)(deflated 40%)
adding: ItemFreq.class(in = 1106) (out= 543)(deflated 50%)
adding: TopN.class(in = 2493) (out= 1260)(deflated 49%)
adding: CountTrigramReduce.class(in = 1636) (out= 686)(deflated 58%)
adding: CountTrigramMap.class(in = 2264) (out= 1012)(deflated 55%)
adding: TopNSReduce.class(in = 3020) (out= 1246)(deflated 58%)
adding: TopNSMap.class(in = 2926) (out= 1176)(deflated 59%)
2025-04-17 01:33:02,469 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-17 01:33:03,238 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/
hduser/.staging/job_1744821131799_0001
2025-04-17 01:33:04,101 INFO input.FileInputFormat: Total input files to process : 1
2025-04-17 01:33:05,029 INFO mapreduce.JobSubmitter: number of splits:1
2025-04-17 01:33:05,226 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1744821131799_0001
2025-04-17 01:33:05,227 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-17 01:33:05,439 INFO conf.Configuration: resource-types.xml not found
2025-04-17 01:33:05,440 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-04-17 01:33:05,958 INFO impl.YarnClientImpl: Submitted application application_1744821131799_0001
2025-04-17 01:33:06,022 INFO mapreduce.Job: The url to track the job: http://DESKTOP-63506IV:50000/proxy/application_17
```

```
Windows PowerShell
hduser@DESKTOP-63506IV: ~\hadoop_test\trigramcount$ hdfs dfs -ls tc_out1
Found 3 items
-rw-r--r-- 1 hduser supergroup 0 2025-04-17 01:33 tc_out1/_SUCCESS
-rw-r--r-- 1 hduser supergroup 1398378 2025-04-17 01:33 tc_out1/part-r-00000
drwxr-xr-x - hduser supergroup 0 2025-04-17 01:33 tc_out1/topN
hduser@DESKTOP-63506IV: ~\hadoop_test\trigramcount$ hdfs dfs -ls tc_out1/topN
Found 2 items
-rw-r--r-- 1 hduser supergroup 0 2025-04-17 01:33 tc_out1/topN/_SUCCESS
-rw-r--r-- 1 hduser supergroup 167 2025-04-17 01:33 tc_out1/topN/part-r-00000
hduser@DESKTOP-63506IV: ~\hadoop_test\trigramcount$
```

```

Windows PowerShell
hduser@DESKTOP-635D6IV: ~
! 2
!!home 4
!!home\t 1
!!ht 1
!action 2
!coaches 1
!width=60 5
!" 1
" 487
"" 10
""bonus 1
""i 1
""it's 1
""marry 1
""mirror 1
""missin' 1
""moon" 1
""ok 1
""polish"-similarly 1
""rumor 1
""the 1
""we 2
""you're 1
"$50 1
"" 1
"" 1
More

```

## [Ch 2-8 p.14-15]

```

Windows PowerShell
hduser@DESKTOP-635D6IV: ~
hduser@DESKTOP-635D6IV:~/hadoop_test/citationcount$ javac -classpath $(hadoop classpath) -d bin src/*.java
Note: src/CountCitation.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
hduser@DESKTOP-635D6IV:~/hadoop_test/citationcount$ jar -cvf citationcount_auth.jar -C bin .
added manifest
adding: CountCitation.class(in = 1768) (out= 878)(deflated 50%)
adding: CountCitation$Reduce.class(in = 1639) (out= 688)(deflated 58%)
adding: CountCitation$Map.class(in = 1424) (out= 546)(deflated 61%)
hduser@DESKTOP-635D6IV:~/hadoop_test/citationcount$ hadoop jar citationcount_auth.jar CountCitation input/2M.SRCID.DSTIID
cc_out_auth1
2025-04-17 01:38:14,604 INFO client.DefaultHohARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-17 01:38:15,101 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/
hduser/.staging/job_1744821131799_0003
2025-04-17 01:38:15,375 INFO input.FileInputFormat: Total input files to process : 1
2025-04-17 01:38:16,294 INFO mapreduce.JobSubmitter: number of splits:2
2025-04-17 01:38:16,859 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1744821131799_0003
2025-04-17 01:38:16,859 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-17 01:38:17,052 INFO conf.Configuration: resource-types.xml not found
2025-04-17 01:38:17,052 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-04-17 01:38:17,159 INFO impl.YarnClientImpl: Submitted application application_1744821131799_0003
2025-04-17 01:38:17,221 INFO mapreduce.Job: The url to track the job: http://DESKTOP-635D6IV:50000/proxy/application_17
44821131799_0003/
2025-04-17 01:38:17,221 INFO mapreduce.Job: Running job: job_1744821131799_0003
2025-04-17 01:38:27,389 INFO mapreduce.Job: Job job_1744821131799_0003 running in uber mode : false
2025-04-17 01:38:27,391 INFO mapreduce.Job: map 0% reduce 0%
2025-04-17 01:38:37,568 INFO mapreduce.Job: map 50% reduce 0%

```

```

Windows PowerShell
hduser@DESKTOP-635D6IV: ~
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=631
CPU time spent (ms)=32580
Physical memory (bytes) snapshot=1301839872
Virtual memory (bytes) snapshot=770021248
Total committed heap usage (bytes)=1278214144
Peak Map Physical memory (bytes)=616513536
Peak Map Virtual memory (bytes)=2570010624
Peak Reduce Physical memory (bytes)=335585280
Peak Reduce Virtual memory (bytes)=2567725056
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=151645466
File Output Format Counters
Bytes Written=12476697
hduser@DESKTOP-635D6IV:~/hadoop_test/citationcount$ hdfs dfs -ls cc_out_auth1
Found 2 items
-rw-r--r-- 1 hduser supergroup 0 2025-04-17 01:38 cc_out_auth1/_SUCCESS
-rw-r--r-- 1 hduser supergroup 12476697 2025-04-17 01:38 cc_out_auth1/part-r-00000
hduser@DESKTOP-635D6IV:~/hadoop_test/citationcount$

```

```

Windows PowerShell
hduser@DESKTOP-635D6IV: ~
10000001 2
10000010 2
10000017 1
10000030 1
1000007 2
1000016 3
10000176 8
10000218 3
10000230 4
10000244 2
10000288 1
1000030 36
10000323 4
1000035 24
10000383 3
1000044 13
10000462 6
10000495 2
10000498 2
100005 112
10000514 1
10000530 1
10000548 8
10000591 4
10000649 1
10000705 4
More--

```

```

Windows PowerShell
hduser@DESKTOP-635D6IV: ~
*/
public class CountCitation {

    public static class Map extends Mapper<Text, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);

        public void map(Text key, Text value, Context context) throws IOException, InterruptedException {
            context.write(key, one);
        }
    }

    public static class Reduce extends Reducer<Text, IntWritable, Text, IntWritable> {

        public void reduce(Text key, Iterable<IntWritable> values, Context context)
            throws IOException, InterruptedException {
            int sum = 0;
            for (IntWritable val : values) {
                sum += val.get();
            }
            context.write(key, new IntWritable(sum));
        }
    }
}

src/CountCitation.java [+] 22,26 314
-- INSERT --

```

```

Windows PowerShell
hduser@DESKTOP-635D6IV: ~
at javax.security.auth.Subject.doAs(Subject.java:422)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1953)
at org.apache.hadoop.mapreduce.Job.submit(Job.java:1674)
at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1695)
at CountCitation.main(CountCitation.java:63)
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:330)
at org.apache.hadoop.util.RunJar.main(RunJar.java:245)
hduser@DESKTOP-635D6IV:~/hadoop_test/citationcount$ hadoop jar citationcount_hub.jar CountCitation input/2M.SRCID.DSTID
cc_out_hub1
2025-04-17 01:51:05,085 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-17 01:51:05,588 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/
hduser/staging/job_1744821131799_0004
2025-04-17 01:51:05,821 INFO input.FileInputFormat: Total input files to process : 1
2025-04-17 01:51:06,734 INFO mapreduce.JobSubmitter: number of splits:2
2025-04-17 01:51:07,310 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1744821131799_0004
2025-04-17 01:51:07,310 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-17 01:51:07,501 INFO conf.Configuration: resource-types.xml not found
2025-04-17 01:51:07,501 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-04-17 01:51:07,598 INFO impl.YarnClientImpl: Submitted application application_1744821131799_0004
2025-04-17 01:51:07,655 INFO mapreduce.Job: The url to track the job: http://DESKTOP-635D6IV:50000/proxy/application_17
44821131799_0004/
2025-04-17 01:51:07,656 INFO mapreduce.Job: Running job: job_1744821131799_0004

```

```

Windows PowerShell
hduser@DESKTOP-63506IV: ~
Reduce input records=1170227
Reduce output records=1170226
Spilled Records=3357782
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=706
CPU time spent (ms)=26420
Physical memory (bytes) snapshot=1404674048
Virtual memory (bytes) snapshot=7700877312
Total committed heap usage (bytes)=1468530688
Peak Map Physical memory (bytes)=614895616
Peak Map Virtual memory (bytes)=2569154560
Peak Reduce Physical memory (bytes)=281899008
Peak Reduce Virtual memory (bytes)=2568613888

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=151645466
File Output Format Counters
  Bytes Written=12476697
hduser@DESKTOP-63506IV:~/hadoop_test/citationcount$

```

```

Windows PowerShell
hduser@DESKTOP-63506IV: ~
hduser@DESKTOP-63506IV:~/hadoop_test/citationcount$ hdfs dfs -ls cc_out_hub1
Found 2 items
-rw-r--r-- 1 hduser supergroup          0 2025-04-17 01:51 cc_out_hub1/ SUCCESS
-rw-r--r-- 1 hduser supergroup 12476697 2025-04-17 01:51 cc_out_hub1/part-r-00000
hduser@DESKTOP-63506IV:~/hadoop_test/citationcount$

```

```

Windows PowerShell
hduser@DESKTOP-63506IV: ~
10000001 2
10000010 2
10000017 1
10000030 1
1000007 2
1000016 3
10000176 8
10000218 3
10000230 4
10000244 2
10000288 1
1000030 36
10000323 4
1000035 24
10000383 3
1000044 13
10000462 6
10000495 2
10000498 2
100005 112
10000514 1
10000530 1
10000548 8
10000591 4
10000649 1
10000705 4
--More--

```

## [Ch 2-9 p.11-12]

```

Windows PowerShell
hduser@DESKTOP-63506IV: ~
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle$ hadoop jar joinidtitle.jar JoinIDTitle input/2M.TITLE.ID cc_out_auth1/part-r-00000 join_out1
2025-04-17 02:07:53,330 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-17 02:07:53,681 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-04-17 02:07:53,718 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hduser/.staging/job_1744821131799_0011
2025-04-17 02:07:53,970 INFO input.FileInputFormat: Total input files to process : 1
2025-04-17 02:07:53,996 INFO input.FileInputFormat: Total input files to process : 1
2025-04-17 02:07:54,466 INFO mapreduce.JobSubmitter: number of splits:2
2025-04-17 02:07:54,588 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1744821131799_0011
2025-04-17 02:07:54,588 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-17 02:07:54,769 INFO conf.Configuration: resource-types.xml not found
2025-04-17 02:07:54,769 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-04-17 02:07:54,881 INFO impl.YarnClientImpl: Submitted application application_1744821131799_0011
2025-04-17 02:07:54,920 INFO mapreduce.Job: The url to track the job: http://DESKTOP-63506IV:50000/proxy/application_1744821131799_0011/
2025-04-17 02:07:54,920 INFO mapreduce.Job: Running job: job_1744821131799_0011
2025-04-17 02:08:01,026 INFO mapreduce.Job: Job job_1744821131799_0011 running in uber mode : false
2025-04-17 02:08:01,027 INFO mapreduce.Job: map 0% reduce 0%
2025-04-17 02:08:09,130 INFO mapreduce.Job: map 50% reduce 0%
2025-04-17 02:08:15,174 INFO mapreduce.Job: map 100% reduce 0%
2025-04-17 02:08:21,208 INFO mapreduce.Job: map 100% reduce 100%
2025-04-17 02:08:22,229 INFO mapreduce.Job: Job job_1744821131799_0011 completed successfully
2025-04-17 02:08:22,359 INFO mapreduce.Job: Counters: 58
File System Counters
FILE: Number of bytes read=181179111

```

```

Windows PowerShell
hduser@DESKTOP-63506IV: ~
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=661
CPU time spent (ms)=24420
Physical memory (bytes) snapshot=1594916864
Virtual memory (bytes) snapshot=7697276928
Total committed heap usage (bytes)=1498939392
Peak Map Physical memory (bytes)=604798976
Peak Map Virtual memory (bytes)=2566443008
Peak Reduce Physical memory (bytes)=491356160
Peak Reduce Virtual memory (bytes)=2573942784
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
Stats
DocID+Citation=1170226
Number of Title+DocID=2115063
The number of pairs of matches=1170226
File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=41485504
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle$

```

```

Windows PowerShell
hduser@DESKTOP-63506IV: ~
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle$ hdfs dfs -ls join_out1
Found 2 items
-rw-r--r-- 1 hduser supergroup 0 2025-04-17 02:08 join_out1/_SUCCESS
-rw-r--r-- 1 hduser supergroup 41485504 2025-04-17 02:08 join_out1/part-r-00000
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle$ hdfs dfs -cat join_out1/part-r-00000 | head -5
10000001 Juan que reia 2
10000010 Golden bear oil 2
10000017 La Isla de los dibujos 1
10000030 Los Hombres sólo piensan en eso 1
1000007 Ettercap 2
cat: Unable to write to output stream.
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle$

```

## [Ch 2-9 p.19-20]

```

Windows PowerShell
hduser@DESKTOP-63506IV: ~
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle2$ javac -classpath $(hadoop classpath) -d bin src/*.java
Note: src/JoinIDTitle2.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle2$ jar -cvf joinidtitle2.jar -C bin .
added manifest
adding: JoinIDTitle2$MyMapper.class(in = 3467) (out= 1466)(deflated 57%)
adding: JoinIDTitle2.class(in = 1801) (out= 964)(deflated 46%)
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle2$ hadoop jar joinidtitle2.jar JoinIDTitle2 input/2M.TITLE.ID cc_out_aut
h1/part-r-00000 join_out2
2025-04-17 02:17:33,793 INFO client.DefaultNoHARMPFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-17 02:17:34,156 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
2025-04-17 02:17:34,189 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/
hduser/.staging/job_1744821131799_0013
2025-04-17 02:17:34,448 INFO input.FileInputFormat: Total input files to process : 1
2025-04-17 02:17:35,376 INFO mapreduce.JobSubmitter: number of splits:1
2025-04-17 02:17:35,709 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1744821131799_0013
2025-04-17 02:17:35,709 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-17 02:17:35,931 INFO conf.Configuration: resource-types.xml not found
2025-04-17 02:17:35,931 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-04-17 02:17:36,027 INFO impl.YarnClientImpl: Submitted application application_1744821131799_0013
2025-04-17 02:17:36,078 INFO mapreduce.Job: The url to track the job: http://DESKTOP-63506IV:50000/proxy/application_17
44821131799_0013/
2025-04-17 02:17:36,079 INFO mapreduce.Job: Running job: job_1744821131799_0013

```

```

Windows PowerShell
hduser@DESKTOP-63506IV: ~
Spilled Records=2340452
Shuffled Maps=1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=1250
CPU time spent (ms)=23240
Physical memory (bytes) snapshot=1005408256
Virtual memory (bytes) snapshot=5133529088
Total committed heap usage (bytes)=959971328
Peak Map Physical memory (bytes)=707133440
Peak Map Virtual memory (bytes)=2563792896
Peak Reduce Physical memory (bytes)=298274816
Peak Reduce Virtual memory (bytes)=2569736192
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
Stats
Number of Title+DocID=1170226
File Input Format Counters
Bytes Read=73550219
File Output Format Counters
Bytes Written=41485504
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle2$

```

```

Windows PowerShell
hduser@DESKTOP-63506IV: ~
Peak Reduce Physical memory (bytes)=298274816
Peak Reduce Virtual memory (bytes)=2569736192
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
Stats
Number of Title+DocID=1170226
File Input Format Counters
Bytes Read=73550219
File Output Format Counters
Bytes Written=41485504
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle2$ hdfs dfs -ls join_out2
Found 2 items
-rw-r--r-- 1 hduser supergroup 0 2025-04-17 02:18 join_out2/_SUCCESS
-rw-r--r-- 1 hduser supergroup 41485504 2025-04-17 02:18 join_out2/part-r-00000
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle2$ hdfs dfs -cat join_out2/part-r-00000 | head -5
!!! 600744 33 16381751 2
!!Destroy-Oh-Boy!!
!Action Pact! 2978668 1
!Hero 1921683 3
"300" 9798511 2
cat: Unable to write to output stream.
hduser@DESKTOP-63506IV:~/hadoop_test/joinidtitle2$

```

## [Ch 2-10 p.8-9]

```

Windows PowerShell
hdsuser@DESKTOP-63506IV: ~
hdsuser@DESKTOP-63506IV:~/hadoop_test/invertedindex$ javac -classpath $(hadoop classpath) -d bin src/*.java
Note: src/InvertedIndex.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
hdsuser@DESKTOP-63506IV:~/hadoop_test/invertedindex$ jar -cvf invertedindex.jar -C bin .
added manifest
adding: InvertedIndex$Reduce.class(in = 2107) (out= 925)(deflated 56%)
adding: InvertedIndex$Map.class(in = 2133) (out= 980)(deflated 54%)
adding: InvertedIndex.class(in = 1516) (out= 773)(deflated 49%)
hdsuser@DESKTOP-63506IV:~/hadoop_test/invertedindex$ hadoop jar invertedindex.jar InvertedIndex input/10K.ID.CONTENTS ii_
out1
2025-04-17 02:19:22,806 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-17 02:19:23,170 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
2025-04-17 02:19:23,207 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/
hdsuser/.staging/job_1744821131799_0014
2025-04-17 02:19:23,457 INFO input.FileInputFormat: Total input files to process : 1
2025-04-17 02:19:23,944 INFO mapreduce.JobSubmitter: number of splits:1
2025-04-17 02:19:24,060 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1744821131799_0014
2025-04-17 02:19:24,060 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-17 02:19:24,232 INFO conf.Configuration: resource-types.xml not found
2025-04-17 02:19:24,232 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-04-17 02:19:24,321 INFO impl.YarnClientImpl: Submitted application application_1744821131799_0014
2025-04-17 02:19:24,374 INFO mapreduce.Job: The url to track the job: http://DESKTOP-63506IV:50000/proxy/application_17
44821131799_0014/
2025-04-17 02:19:24,375 INFO mapreduce.Job: Running job: job_1744821131799_0014

```

```

Windows PowerShell
hdsuser@DESKTOP-63506IV: ~
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=9735683
File Output Format Counters
  Bytes Written=6556893
hdsuser@DESKTOP-63506IV:~/hadoop_test/invertedindex$ hdfs dfs -ls ii_out1
Found 2 items
-rw-r--r-- 1 hdsuser supergroup 0 2025-04-17 02:19 ii_out1/_SUCCESS
-rw-r--r-- 1 hdsuser supergroup 6556893 2025-04-17 02:19 ii_out1/part-r-000000
hdsuser@DESKTOP-63506IV:~/hadoop_test/invertedindex$ hdfs dfs -cat ii_out1/part-r-000000 | head -10
0
32485375,1157839,8857587,259953,29434133,3587831,107603,14431940,39406,119971,21447654,123128,6816176,17597291,2
439703,108801886,2185,14205946,19405277,110266,8276444,34875354,9752106,2554996,32722158,12072140,13737413,12242363,26148
720,11979629,23190042,13957506,922939,3120913,8678146,32097060,27401791,32109566,231291,107600,9007073,3509072,119565,27
940448,23667057,19638080,821165,131343,13592764,6717145,2030470,25236143,16804773,111115,23581408,116421,114604,19441805
,12092675,27083654,3869884,12072986,21536716,127550,135674,25476219,119591,26845973,18490501,873177,24887159,4064809,338
90412,259640,22736561,35276289,6704337,32081576,1083631,6916196,23955002,25679077,116152,20927257,139104,107570,139765,2
7707715,4037764,4095197,5721663,2522680,117402,19263470,4890927,736861,5662644,25018507,128976,114935,953399,17812392,26
582012,259374,20782214,5171379,4075106,21615135,21906554,13291495,24610237,134128,21326174,3174112,46953,22629002,150480
,1398193,15457257,1669676,13753904,27998014,3414795,1201310,3708222,112036,24683049,2447166,846960,19032211,20989204,253
23401,9538068,17718113,33892343,5655279,24468591,105290,15993881,119531,21799127,117317,126671,260233,32291506,5876585,2
5109246,23880171,6937695,991203,32895450,130269,19067812,22599011,15847691,1851995,18868579,14813030,6554636,28681660,33

```

## [Ch 2-10 p.17-18]

```

Windows PowerShell
hdsuser@DESKTOP-63506IV: ~
hdsuser@DESKTOP-63506IV:~/hadoop_test/invertedindex2$ javac -classpath $(hadoop classpath) -d bin src/*.java
Note: src/InvertedIndex2.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
hdsuser@DESKTOP-63506IV:~/hadoop_test/invertedindex2$ jar -cvf invertedindex2.jar -C bin .
added manifest
adding: WordIDSortComparator.class(in = 486) (out= 307)(deflated 36%)
adding: InvertedIndex2.class(in = 1837) (out= 927)(deflated 49%)
adding: WordIDGroupingComparator.class(in = 567) (out= 341)(deflated 39%)
adding: WordID.class(in = 2053) (out= 966)(deflated 52%)
adding: InvertedIndex2$Reduce.class(in = 2235) (out= 1044)(deflated 53%)
adding: InvertedIndex2$Map.class(in = 2674) (out= 1267)(deflated 52%)
adding: WordIDPartitioner.class(in = 712) (out= 406)(deflated 42%)
hdsuser@DESKTOP-63506IV:~/hadoop_test/invertedindex2$ hadoop jar invertedindex2.jar InvertedIndex2 input/10K.ID.CONTENTS
ii2_out1
2025-04-17 02:21:06,714 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-17 02:21:07,065 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
2025-04-17 02:21:07,086 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/
hdsuser/.staging/job_1744821131799_0015
2025-04-17 02:21:07,334 INFO input.FileInputFormat: Total input files to process : 1
2025-04-17 02:21:07,402 INFO mapreduce.JobSubmitter: number of splits:1
2025-04-17 02:21:07,534 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1744821131799_0015
2025-04-17 02:21:07,534 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-17 02:21:07,711 INFO conf.Configuration: resource-types.xml not found
2025-04-17 02:21:07,711 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-04-17 02:21:07,795 INFO impl.YarnClientImpl: Submitted application application_1744821131799_0015
2025-04-17 02:21:07,862 INFO mapreduce.Job: The url to track the job: http://DESKTOP-63506IV:50000/proxy/application_17

```

```

Windows PowerShell
hduser@DESKTOP-63506IV: ~
Peak Reduce Physical memory (bytes)=384638976
Peak Reduce Virtual memory (bytes)=2573369344
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=9735683
File Output Format Counters
Bytes Written=6556893
hduser@DESKTOP-63506IV:~/hadoop_test/invertedindex2$ hdfs dfs -ls ii2_out1
Found 2 items
-rw-r--r-- 1 hduser supergroup 0 2025-04-17 02:21 ii2_out1/_SUCCESS
-rw-r--r-- 1 hduser supergroup 6556893 2025-04-17 02:21 ii2_out1/part-r-00000
hduser@DESKTOP-63506IV:~/hadoop_test/invertedindex2$ hdfs dfs -cat ii2_out1/part-r-00000 | head -5
0
2185,39406,46953,56557,59739,105290,107570,107600,107603,110266,111115,112036,114463,114604,114935,115462,115647
,116152,116421,117317,117402,119358,119531,119565,119591,119971,120417,122885,123128,126671,127550,127824,128976,130031,
130269,131301,131343,134128,134896,135674,137366,139104,139765,150480,150962,213900,231291,259374,259640,259953,260233,7
36861,821165,846960,863472,873177,922939,953399,991203,1083631,1157839,1201310,1398193,1429917,1544374,1669676,1722590,1
851995,1888165,1932784,1999660,2030470,2333061,2439703,2447166,2522680,2554996,3038004,3120913,3159417,3174112,3414795,3
509072,3587831,3708222,3869884,4037764,4064809,4075106,4095197,4530540,4606153,4890927,5042786,5171379,5343488,5632253,5
655279,5662644,5721663,5876585,5901876,6255582,6262417,6554636,6554748,6704337,6717145,6816176,6916196,6930599,6937695,7
081129,7945204,8276444,8678146,8773592,8857587,8923503,9007073,9173011,9263305,9538068,9626096,9752106,9885025,10801886,
11979629,12072140,12072986,12092675,12242363,13291495,13592764,13737413,13753904,13957506,14160104,14205946,14431940,148

```