

```
chosun@chosun-VirtualBox:~/hadoop$
/// ***** ///
>>> 콤바이너 클래스를 적용했을때와 적용하지 않았을때를 비교하려고 합니다.
>>> 5장의 예제를 비교하려고 하므로, HWEx5_5.jar을 생성했습니다.
>>> 입력 데이터는 과제3에서 생성했던 HW_0512에 저장되어있는 2008.csv를 사용합니다.
/// ***** ///
chosun@chosun-VirtualBox:~/hadoop$ ls
1996.csv.bz2  HWEx5_5.jar  HWEx8_3.jar  HomeWorkEX6_2_Seq.jar  README.txt  c.csv  contrib  hadoop-core-1.2.1.jar
hadoop-tools-1.2.1.jar  libexec  share  1997.csv  HWEx7_2MapSideJoin.jar  HomeWorkEX6_1.jar  HomeWorkEX6_3_Total.jar  bin
carriers.csv  docs  hadoop-examples-1.2.1.jar  ivy  logs  src  2008.csv  HWEx7_3ReduceSideJoin.jar  HomeWorkEX6_2_Map.jar
LICENSE.txt  build.xml  carriers_new2.csv  hadoop-ant-1.2.1.jar  hadoop-minicluster-1.2.1.jar  ivy.xml  pids  test.py
CHANGES.txt  HWEx8_2.jar  HomeWorkEX6_2_Search.jar  NOTICE.txt  c++  conf  hadoop-client-1.2.1.jar  hadoop-test-1.2.1.jar
lib  sbin  webapps
chosun@chosun-VirtualBox:~/hadoop$
/// ***** ///
>>> [Ex8.2] --- 콤바이너 클래스 적용 x
>>> jar 파일을 실행합니다.
>>> HW_0512 디렉토리의 파일(2008.csv)을 jar파일로 처리한 것을 HW_0515_delay1 디렉토리에 저장합니다.
>>> ArrivalDelayCount은 클래스 이름입니다.
>>> 이 콤바이너 클래스를 적용 전 실행 결과(5.5)와 적용 후 실행결과(8.2)를 비교할것입니다.
/// ***** ///
chosun@chosun-VirtualBox:~/hadoop$
chosun@chosun-VirtualBox:~/hadoop$ hadoop jar HWEx5_5.jar ArrivalDelayCount HW_0512 HW_0515_delay1
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/chosun/hadoop-1.2.1/hadoop-core-1.2.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
19/05/15 22:48:34 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
19/05/15 22:48:34 INFO input.FileInputFormat: Total input paths to process : 1
19/05/15 22:48:34 INFO util.NativeCodeLoader: Loaded the native-hadoop library
19/05/15 22:48:34 WARN snappy.LoadSnappy: Snappy native library not loaded
19/05/15 22:48:34 INFO mapred.JobClient: Running job: job_201905151856_0010 // job 시작
19/05/15 22:48:35 INFO mapred.JobClient: map 0% reduce 0% // 맵리듀스 시작
19/05/15 22:48:41 INFO mapred.JobClient: map 18% reduce 0% // ~~ ing
19/05/15 22:48:46 INFO mapred.JobClient: map 36% reduce 0%
19/05/15 22:48:49 INFO mapred.JobClient: map 36% reduce 12%
19/05/15 22:48:50 INFO mapred.JobClient: map 54% reduce 12%
19/05/15 22:48:53 INFO mapred.JobClient: map 63% reduce 12%
19/05/15 22:48:54 INFO mapred.JobClient: map 72% reduce 12%
19/05/15 22:48:57 INFO mapred.JobClient: map 90% reduce 12%
19/05/15 22:48:58 INFO mapred.JobClient: map 90% reduce 18%
19/05/15 22:49:00 INFO mapred.JobClient: map 100% reduce 18%
19/05/15 22:49:04 INFO mapred.JobClient: map 100% reduce 33% // ~~ ing
19/05/15 22:49:07 INFO mapred.JobClient: map 100% reduce 100% // 맵 끝
19/05/15 22:49:07 INFO mapred.JobClient: Job complete: job_201905151856_0010 // job 끝
19/05/15 22:49:07 INFO mapred.JobClient: Counters: 29
19/05/15 22:49:07 INFO mapred.JobClient: Map-Reduce Framework
19/05/15 22:49:07 INFO mapred.JobClient: Spilled Records=8401465 // 작업 중에 디스크에 기록 된 총 레코드 수
19/05/15 22:49:07 INFO mapred.JobClient: Map output materialized bytes=39379199 // 실제로 디스크에 기록된 맵 출력 바이트 수
19/05/15 22:49:07 INFO mapred.JobClient: Reduce input records=2979504 // 리듀스에 입력된 데이터의 수
19/05/15 22:49:07 INFO mapred.JobClient: Virtual memory (bytes) snapshot=25044484096 // 모든 작업에 사용된 총 가상 메모리
19/05/15 22:49:07 INFO mapred.JobClient: Map input records=7009728 // 맵에 입력되는 데이터의 수
19/05/15 22:49:07 INFO mapred.JobClient: SPLIT_RAW_BYTES=1265 // 맵에서 가져온 입력 분할 개체의 바이트 수
19/05/15 22:49:07 INFO mapred.JobClient: Map output bytes=33420125 // 모든 맵에 의해 생성된 압축되지 않은 출력 바이트수
19/05/15 22:49:07 INFO mapred.JobClient: Reduce shuffle bytes=39379199 // 리듀스 단계에 셔플된 맵 출력의 바이트 수
19/05/15 22:49:07 INFO mapred.JobClient: Physical memory (bytes) snapshot=2803625984 // 모든 작업에 사용되는 총 물리적 메모리
19/05/15 22:49:07 INFO mapred.JobClient: Reduce input groups=12 // 리듀스 단계에서 처리한 개별 키 그룹의 수
19/05/15 22:49:07 INFO mapred.JobClient: Combine output records=0 // 출력 레코드 수
```

```
19/05/15 22:49:07 INFO mapred.JobClient: Reduce output records=12 // 리듀스의 출력으로 생성된 데이터의 수
19/05/15 22:49:07 INFO mapred.JobClient: Map output records=2979504 // 맵의 출력으로 생성된 데이터의 수
19/05/15 22:49:07 INFO mapred.JobClient: Combine input records=0 // 입력 레코드 수
19/05/15 22:49:07 INFO mapred.JobClient: CPU time spent (ms)=40260 // 소요된 cpu시간
19/05/15 22:49:07 INFO mapred.JobClient: Total committed heap usage (bytes)=1993342976 // JVM에 사용할 수 있는 총 메모리 양
19/05/15 22:49:07 INFO mapred.JobClient: File Input Format Counters
19/05/15 22:49:07 INFO mapred.JobClient: Bytes Read=689454004 // 모든 파일 시스템에 대해 모든 태스크에서 읽은 데이터 양
19/05/15 22:49:07 INFO mapred.JobClient: FileSystemCounters
19/05/15 22:49:07 INFO mapred.JobClient: HDFS_BYTES_READ=689455269 // HDFS에서 읽은 데이터 양
19/05/15 22:49:07 INFO mapred.JobClient: FILE_BYTES_WRITTEN=111502929 // 로컬 파일 시스템에 기록된 데이터 양
19/05/15 22:49:07 INFO mapred.JobClient: FILE_BYTES_READ=71433572 // 로컬 파일 시스템에서 읽은 데이터 양
19/05/15 22:49:07 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=171 // HDFS에 기록된 데이터 양
19/05/15 22:49:07 INFO mapred.JobClient: Job Counters
19/05/15 22:49:07 INFO mapred.JobClient: Launched map tasks=11 // 실행된 총 맵 작업 수
19/05/15 22:49:07 INFO mapred.JobClient: Launched reduce tasks=1 // 실행된 총 리듀스 작업 수
19/05/15 22:49:07 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=25784 // 모든 리듀스 작업에 의해 소요된 총 시간(ms)
19/05/15 22:49:07 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
19/05/15 22:49:07 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=42821 // 모든 맵 작업에 의해 사용한 총 시간(ms)
19/05/15 22:49:07 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0
19/05/15 22:49:07 INFO mapred.JobClient: Data-local map tasks=11 // 필요한 데이터가 포함된 노드에서 시작된 맵 태스크 수
19/05/15 22:49:07 INFO mapred.JobClient: File Output Format Counters
19/05/15 22:49:07 INFO mapred.JobClient: Bytes Written=171 // 모든 파일 시스템에 대해 모든 태스크에서 작성된 데이터 양
```

```
chosun@chosun-VirtualBox:~/hadoop$
/// ***** ///
```

>>> 콤바이너 클래스 적용 전 데이터를 출력합니다.

```
/// ***** ///
```

```
chosun@chosun-VirtualBox:~/hadoop$ hadoop fs -cat HW_0515_delay1/part-r-00000 | head -10
```

```
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/chosun/hadoop-1.2.1/hadoop-core-1.2.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
```

```
2008,1 279427
2008,10183582
2008,11181506
2008,12280493
2008,2 278902
2008,3 294556
2008,4 256142
2008,5 254673
2008,6 295897
2008,7 264630
```

```
chosun@chosun-VirtualBox:~/hadoop$ hadoop fs -cat HW_0515_delay1/part-r-00000 | tail -10
```

```
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/chosun/hadoop-1.2.1/hadoop-core-1.2.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
```

```
2008,11181506
2008,12280493
2008,2 278902
2008,3 294556
2008,4 256142
2008,5 254673
2008,6 295897
2008,7 264630
2008,8 239737
2008,9 169959
```

```
chosun@chosun-VirtualBox:~/hadoop$
```

```
/// ***** ///
>>> [Ex8.2] --- 콤바이너 클래스 적용 o
>>> jar 파일을 실행합니다.
>>> HW_0512 디렉토리의 파일(2008.csv)을 jar파일로 처리한 것을 HW_0515_delay2디렉토리에 저장합니다.
>>> ArrivalDelayCountWithCombiner은 클래스 이름입니다.
/// ***** ///
chosun@chosun-VirtualBox:~/hadoop$ hadoop jar HWEx8_2.jar ArrivalDelayCountWithCombiner HW_0512 HW_0515_delay2
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/chosun/hadoop-1.2.1/hadoop-core-1.2.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
19/05/15 22:50:23 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
19/05/15 22:50:23 INFO input.FileInputFormat: Total input paths to process : 1
19/05/15 22:50:23 INFO util.NativeCodeLoader: Loaded the native-hadoop library
19/05/15 22:50:23 WARN snappy.LoadSnappy: Snappy native library not loaded
19/05/15 22:50:24 INFO mapred.JobClient: Running job: job_201905151856_0011 // job 시작
19/05/15 22:50:25 INFO mapred.JobClient: map 0% reduce 0% // 맵리듀스 시작
19/05/15 22:50:31 INFO mapred.JobClient: map 18% reduce 0% // ~~ ing
19/05/15 22:50:35 INFO mapred.JobClient: map 36% reduce 0%
19/05/15 22:50:39 INFO mapred.JobClient: map 54% reduce 12%
19/05/15 22:50:43 INFO mapred.JobClient: map 72% reduce 12%
19/05/15 22:50:46 INFO mapred.JobClient: map 81% reduce 12%
19/05/15 22:50:47 INFO mapred.JobClient: map 90% reduce 12%
19/05/15 22:50:48 INFO mapred.JobClient: map 100% reduce 18%
19/05/15 22:50:54 INFO mapred.JobClient: map 100% reduce 33% // ~~ ing
19/05/15 22:50:55 INFO mapred.JobClient: map 100% reduce 100% // 맵 끝
19/05/15 22:50:55 INFO mapred.JobClient: Job complete: job_201905151856_0011 // job 끝
19/05/15 22:50:55 INFO mapred.JobClient: Counters: 29
19/05/15 22:50:55 INFO mapred.JobClient: Map-Reduce Framework
19/05/15 22:50:55 INFO mapred.JobClient: Spilled Records=84 // 작업 중에 디스크에 기록 된 총 레코드 수
19/05/15 22:50:55 INFO mapred.JobClient: Map output materialized bytes=463 // 실제로 디스크에 기록된 맵 출력 바이트 수
19/05/15 22:50:55 INFO mapred.JobClient: Reduce input records=30 // 리듀스에 입력된 데이터의 수
19/05/15 22:50:55 INFO mapred.JobClient: Virtual memory (bytes) snapshot=25025552384 // 모든 작업에 사용된 총 가상 메모리
19/05/15 22:50:55 INFO mapred.JobClient: Map input records=7009728 // 맵에 입력되는 데이터의 수
19/05/15 22:50:55 INFO mapred.JobClient: SPLIT_RAW_BYTES=1265 // 맵에서 가져온 입력 분할 개체의 바이트 수
19/05/15 22:50:55 INFO mapred.JobClient: Map output bytes=33420125 // 모든 맵에 의해 생성된 압축되지 않은 출력 바이트수
19/05/15 22:50:55 INFO mapred.JobClient: Reduce shuffle bytes=463 // 리듀스 단계에 셔플된 맵 출력의 바이트 수
19/05/15 22:50:55 INFO mapred.JobClient: Physical memory (bytes) snapshot=2752540672 // 모든 작업에 사용되는 총 물리적 메모리
19/05/15 22:50:55 INFO mapred.JobClient: Reduce input groups=12 // 리듀스 단계에서 처리한 개별 키 그룹의 수
19/05/15 22:50:55 INFO mapred.JobClient: Combine output records=30 // 출력 레코드 수
19/05/15 22:50:55 INFO mapred.JobClient: Reduce output records=12 // 리듀스의 출력으로 생성된 데이터의 수
19/05/15 22:50:55 INFO mapred.JobClient: Map output records=2979504 // 맵의 출력으로 생성된 데이터의 수
19/05/15 22:50:55 INFO mapred.JobClient: Combine input records=2979504 // 입력 레코드 수
19/05/15 22:50:55 INFO mapred.JobClient: CPU time spent (ms)=36520 // 소요된 cpu시간
19/05/15 22:50:55 INFO mapred.JobClient: Total committed heap usage (bytes)=1954545664 // JVM에 사용할 수 있는 총 메모리 양
19/05/15 22:50:55 INFO mapred.JobClient: File Input Format Counters
19/05/15 22:50:55 INFO mapred.JobClient: Bytes Read=689454004 // 모든 파일 시스템에 대해 모든 테스크에서 읽은 데이터 양
19/05/15 22:50:55 INFO mapred.JobClient: FileSystemCounters
19/05/15 22:50:55 INFO mapred.JobClient: HDFS_BYTES_READ=689455269 // HDFS에서 읽은 데이터 양
19/05/15 22:50:55 INFO mapred.JobClient: FILE_BYTES_WRITTEN=694027 // 로컬 파일 시스템에 기록된 데이터 양
19/05/15 22:50:55 INFO mapred.JobClient: FILE_BYTES_READ=814 // 로컬 파일 시스템에서 읽은 데이터 양
19/05/15 22:50:55 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=171 // HDFS에 기록된 데이터 양
19/05/15 22:50:55 INFO mapred.JobClient: Job Counters
19/05/15 22:50:55 INFO mapred.JobClient: Launched map tasks=11 // 실행된 총 맵 작업 수
19/05/15 22:50:55 INFO mapred.JobClient: Launched reduce tasks=1 // 실행된 총 리듀스 작업 수
19/05/15 22:50:55 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=24533 // 모든 리듀스 작업에 의해 소요된 총 시간(ms)
19/05/15 22:50:55 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
19/05/15 22:50:55 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=40502 // 모든 맵 작업에 의해 사용한 총 시간(ms)
```

19/05/15 22:50:55 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0  
19/05/15 22:50:55 INFO mapred.JobClient: Data-local map tasks=11 // 필요한 데이터가 포함된 노드에서 시작된 맵 태스크 수  
19/05/15 22:50:55 INFO mapred.JobClient: File Output Format Counters  
19/05/15 22:50:55 INFO mapred.JobClient: Bytes Written=171 // 모든 파일 시스템에 대해 모든 태스크에서 작성된 데이터 양  
/// \*\*\*\*\*  
>>> 콤바이너 클래스가 적용되기 전과 후의 실행 결과를 비교합니다.

카운터	콤바이너 클래스 적용 전	콤바이너 클래스 적용 후
01. FILE_BYTES_WRITTEN	111502929	694027
02. HDFS_BYTES_READ	689455269	689455269
03. FILE_BYTES_READ	71433572	814
04. HDFS_BYTES_WRITTEN	171	171
05. Reduce shuffle bytes	39379199	463
06. Spilled Records	8401465	84
07. CPU time spent (ms)	40260	36520
08. Total committed heap usage	1993342976	1954545664
09. Reduce input records	2979504	30
10. Physical memory (bytes) snapshot	2803625984	2752540672

입력 데이터와 출력 데이터가 동일하므로 02와 04는 값이 같습니다.  
콤바이너 클래스를 적용하고 수치가 개선된 카운터는 01, 03, 05, 06입니다.  
그밖에 카운터 07, 08, 10의 수치도 미미하게 줄었음을 알 수 있습니다.

>>> 콤바이너 클래스가 적용된 데이터를 출력합니다.  
>>> 출력되는 데이터는 예제 5.5와 동일합니다.  
/// \*\*\*\*\*

```
chosun@chosun-VirtualBox:~/hadoop$ hadoop fs -cat HW_0515_delay2/part-r-00000 | head -10
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/chosun/hadoop-1.2.1/hadoop-core-1.2.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
2008,1 279427
2008,10183582
2008,11181506
2008,12280493
2008,2 278902
2008,3 294556
2008,4 256142
2008,5 254673
2008,6 295897
2008,7 264630
chosun@chosun-VirtualBox:~/hadoop$ hadoop fs -cat HW_0515_delay2/part-r-00000 | tail -10
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/chosun/hadoop-1.2.1/hadoop-core-1.2.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
2008,11181506
2008,12280493
2008,2 278902
2008,3 294556
2008,4 256142
2008,5 254673
2008,6 295897
2008,7 264630
2008,8 239737
2008,9 169959
```

```
/// ***** ///
>>> [Ex8.3]
>>> jar 파일을 실행합니다.
>>> HW_0512 디렉토리의 파일(2008.csv)을 jar파일로 처리한 것을 HW_0515_delay3디렉토리에 저장합니다.
>>> ArrivalDelayCountWithGzip은 클래스 이름입니다.
>>> 스넬피를 아직 적용하지 않은 상태에서 jar 파일을 실행시켰습니다.
>>> 스넬피 다운로드 실습페이지에서 Ex8_3 스넬피 적용한 실행결과와 비교합니다.

/// ***** ///
chosun@chosun-VirtualBox:~/hadoop$ hadoop jar HWEx8_3.jar ArrivalDelayCountWithGzip HW_0512 HW_0515_delay3
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/chosun/hadoop-1.2.1/hadoop-core-1.2.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
19/05/15 22:51:40 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
19/05/15 22:51:40 INFO input.FileInputFormat: Total input paths to process : 1
19/05/15 22:51:40 INFO util.NativeCodeLoader: Loaded the native-hadoop library
19/05/15 22:51:40 WARN snappy.LoadSnappy: Snappy native library not loaded // snappy 적용안됨
19/05/15 22:51:40 INFO mapred.JobClient: Running job: job_201905151856_0012 // job 시작
19/05/15 22:51:41 INFO mapred.JobClient: map 0% reduce 0% // 맵리듀스 시작
19/05/15 22:51:47 INFO mapred.JobClient: map 9% reduce 0% // ~~ ing
19/05/15 22:51:48 INFO mapred.JobClient: map 18% reduce 0%
19/05/15 22:51:53 INFO mapred.JobClient: map 36% reduce 0%
19/05/15 22:51:56 INFO mapred.JobClient: map 36% reduce 12%
19/05/15 22:51:58 INFO mapred.JobClient: map 54% reduce 12%
19/05/15 22:52:03 INFO mapred.JobClient: map 63% reduce 12%
19/05/15 22:52:04 INFO mapred.JobClient: map 72% reduce 12%
19/05/15 22:52:05 INFO mapred.JobClient: map 72% reduce 18%
19/05/15 22:52:08 INFO mapred.JobClient: map 81% reduce 18%
19/05/15 22:52:09 INFO mapred.JobClient: map 90% reduce 18%
19/05/15 22:52:10 INFO mapred.JobClient: map 100% reduce 18%
19/05/15 22:52:11 INFO mapred.JobClient: map 100% reduce 30%
19/05/15 22:52:20 INFO mapred.JobClient: map 100% reduce 33% // ~~ ing
19/05/15 22:52:24 INFO mapred.JobClient: map 100% reduce 100% // 맵 끝
19/05/15 22:52:24 INFO mapred.JobClient: Job complete: job_201905151856_0012 // job 끝
19/05/15 22:52:24 INFO mapred.JobClient: Counters: 29
19/05/15 22:52:24 INFO mapred.JobClient: Map-Reduce Framework
19/05/15 22:52:24 INFO mapred.JobClient: Spilled Records=8401465 // 작업 중에 디스크에 기록 된 총 레코드 수
19/05/15 22:52:24 INFO mapred.JobClient: Map output materialized bytes=76990 // 실제로 디스크에 기록된 맵 출력 바이트 수
19/05/15 22:52:24 INFO mapred.JobClient: Reduce input records=2979504 // 리듀스에 입력된 데이터의 수
19/05/15 22:52:24 INFO mapred.JobClient: Virtual memory (bytes) snapshot=25036963840 // 모든 작업에 사용된 총 가상 메모리
19/05/15 22:52:24 INFO mapred.JobClient: Map input records=7009728 // 맵에 입력되는 데이터의 수
19/05/15 22:52:24 INFO mapred.JobClient: SPLIT_RAW_BYTES=1265 // 맵에서 가져온 입력 분할 개체의 바이트 수
19/05/15 22:52:24 INFO mapred.JobClient: Map output bytes=33420125 // 모든 맵에 의해 생성된 압축되지 않은 출력 바이트수
19/05/15 22:52:24 INFO mapred.JobClient: Reduce shuffle bytes=76990 // 리듀스 단계에 셔플된 맵 출력의 바이트 수
19/05/15 22:52:24 INFO mapred.JobClient: Physical memory (bytes) snapshot=2808045568 // 모든 작업에 사용되는 총 물리적 메모리
19/05/15 22:52:24 INFO mapred.JobClient: Reduce input groups=12 // 리듀스 단계에서 처리한 개별 키 그룹의 수
19/05/15 22:52:24 INFO mapred.JobClient: Combine output records=0 // 출력 레코드 수
19/05/15 22:52:24 INFO mapred.JobClient: Reduce output records=12 // 리듀스의 출력으로 생성된 데이터의 수
19/05/15 22:52:24 INFO mapred.JobClient: Map output records=2979504 // 맵의 출력으로 생성된 데이터의 수
19/05/15 22:52:24 INFO mapred.JobClient: Combine input records=0 // 입력 레코드 수
19/05/15 22:52:24 INFO mapred.JobClient: CPU time spent (ms)=56980 // 소요된 cpu시간
19/05/15 22:52:24 INFO mapred.JobClient: Total committed heap usage (bytes)=1995440128 // JVM에 사용할 수 있는 총 메모리 양
19/05/15 22:52:24 INFO mapred.JobClient: File Input Format Counters
19/05/15 22:52:24 INFO mapred.JobClient: Bytes Read=689454004 // 모든 파일 시스템에 대해 모든 테스크에서 읽은 데이터 양
19/05/15 22:52:24 INFO mapred.JobClient: FileSystemCounters
19/05/15 22:52:24 INFO mapred.JobClient: HDFS_BYTES_READ=689455269 // HDFS에서 읽은 데이터 양
19/05/15 22:52:24 INFO mapred.JobClient: FILE_BYTES_WRITTEN=906833 // 로컬 파일 시스템에 기록된 데이터 양
19/05/15 22:52:24 INFO mapred.JobClient: FILE_BYTES_READ=139601 // 로컬 파일 시스템에서 읽은 데이터 양
```

19/05/15 22:52:24 INFO mapred.JobClient: HDFS\_BYTES\_WRITTEN=324 // HDFS에 기록된 데이터 양  
19/05/15 22:52:24 INFO mapred.JobClient: Job Counters  
19/05/15 22:52:24 INFO mapred.JobClient: Launched map tasks=11 // 실행된 총 맵 작업 수  
19/05/15 22:52:24 INFO mapred.JobClient: Launched reduce tasks=1 // 실행된 총 리듀스 작업 수  
19/05/15 22:52:24 INFO mapred.JobClient: SLOTS\_MILLIS\_REDUCES=35839 // 모든 리듀스 작업에 의해 소요된 총 시간(ms)  
19/05/15 22:52:24 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0  
19/05/15 22:52:24 INFO mapred.JobClient: SLOTS\_MILLIS\_MAPS=51746 // 모든 맵 작업에 의해 사용한 총 시간(ms)  
19/05/15 22:52:24 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0  
19/05/15 22:52:24 INFO mapred.JobClient: Data-local map tasks=11 // 필요한 데이터가 포함된 노드에서 시작된 맵 태스크 수  
19/05/15 22:52:24 INFO mapred.JobClient: File Output Format Counters  
19/05/15 22:52:24 INFO mapred.JobClient: Bytes Written=324 // 모든 파일 시스템에 대해 모든 태스크에서 작성된 데이터 양  
chosun@chosun-VirtualBox:~/hadoop\$

/// \*\*\*\*\* ///

>>> Gzip이 적용되기 전과 후의 실행 결과를 비교합니다.

카운터	Gzip 적용 x	Gzip 적용 o
01. FILE_BYTES_WRITTEN	111502929	906833
02. HDFS_BYTES_READ	689455269	689455269
03. FILE_BYTES_READ	71433572	139601
04. HDFS_BYTES_WRITTEN	171	324
05. Reduce shuffle bytes	39379199	76990
06. Spilled Records	8401465	8401465
07. CPU time spent (ms)	40260	56980
08. Total committed heap usage	1993342976	1995440128
09. Reduce input records	2979504	2979504
10. Physical memory (bytes) snapshot	2803625984	2808045568

잡을 실행하면 파일 입출력의 수치와 셔플의 수치가 현저히 줄었음을 알 수 있습니다. (01, 03, 05)

>>> 출력되는 데이터를 확인합니다.

>>> 데이터는 압축파일(Gzip)형태로 저장됩니다.

>>> cat명령어는 텍스트 파일만 출력할 수 있기 때문에 파일 내용을 읽으려면 text명령어를 사용해야 합니다.

/// \*\*\*\*\* ///

chosun@chosun-VirtualBox:~/hadoop\$ **hadoop fs -ls HW\_0515\_delay3**

WARNING: An illegal reflective access operation has occurred  
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/chosun/hadoop-1.2.1/hadoop-core-1.2.1.jar) to method sun.security.krb5.Config.getInstance()  
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil  
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations  
WARNING: All illegal access operations will be denied in a future release  
Found 3 items

```
-rw-r--r--  1 chosun supergroup      0 2019-05-15 22:52 /user/chosun/HW_0515_delay3/_SUCCESS
drwxr-xr-x  - chosun supergroup      0 2019-05-15 22:51 /user/chosun/HW_0515_delay3/_logs
-rw-r--r--  1 chosun supergroup    324 2019-05-15 22:52 /user/chosun/HW_0515_delay3/part-r-00000
```

chosun@chosun-VirtualBox:~/hadoop\$

chosun@chosun-VirtualBox:~/hadoop\$ **hadoop fs -text HW\_0515\_delay3/part-r-00000 | head -10**

WARNING: An illegal reflective access operation has occurred  
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/chosun/hadoop-1.2.1/hadoop-core-1.2.1.jar) to method sun.security.krb5.Config.getInstance()  
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil  
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations  
WARNING: All illegal access operations will be denied in a future release  
19/05/15 22:54:43 INFO util.NativeCodeLoader: Loaded the native-hadoop library  
19/05/15 22:54:43 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library  
19/05/15 22:54:43 INFO compress.CodecPool: Got brand-new decompressor  
19/05/15 22:54:43 INFO compress.CodecPool: Got brand-new decompressor  
19/05/15 22:54:43 INFO compress.CodecPool: Got brand-new decompressor  
19/05/15 22:54:43 INFO compress.CodecPool: Got brand-new decompressor  
2008,1 279427  
2008,10183582

```
chosun@chosun-VirtualBox:~/hadoop$ hadoop fs -text HW_0515_delay3/part-r-00000 | tail -10
```

2008,11	181506
2008,12	280493
2008,2	278902
2008,3	294556
2008,4	256142
2008,5	254673
2008,6	295897
2008,7	264630
2008,8	239737
2008,9	169959

```
chosun@chosun-VirtualBox:~/hadoop$
```

```
chosun@chosun-VirtualBox:~/hadoop$ hadoop fs -cat HW_0515_delay3/part-r-00000 | tail -10
```

```
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/chosun/hadoop-1.2.1/hadoop-core-1.2.1.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
```

```
SEQorg.apache.hadoop.io.Text org.apache.hadoop.io.IntWritable'org.apache.hadoop.io.compress.GzipCodec$5/PvP`h5/PvPc`7{48c3201dPnP7bPBPB@(e!nQWccaAMsienc`qnf`:+t}~cK.d4d`sM0chosun@chosun-VirtualBox:~/hadoop$  
chosun@chosun-VirtualBox:~/hadoop$
```