


```

Total vcore-milliseconds taken by all reduce tasks=7036
Total megabyte-milliseconds taken by all map tasks=288761856
Total megabyte-milliseconds taken by all reduce tasks=21614592
Map-Reduce Framework
  Map input records=1405946
  Map output records=6067351
  Map output bytes=47130252
  Map output materialized bytes=8332
  Input split bytes=985
  Combine input records=6067351
  Combine output records=820
  Reduce input groups=82
  Reduce shuffle bytes=8332
  Reduce input records=820
  Reduce output records=82
  Spilled Records=1640
  Shuffled Maps =10
  Failed Shuffles=0
  Merged Map outputs=10
  GC time elapsed (ms)=7774
  CPU time spent (ms)=45100
  Physical memory (bytes) snapshot=7122997248
  Virtual memory (bytes) snapshot=37727354880
  Total committed heap usage (bytes)=6552027136
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=81613300
File Output Format Counters
  Bytes Written=891
[hadoop@ip-172-31-5-137 wordcount]$ rm WordCount3.jar
[hadoop@ip-172-31-5-137 wordcount]$ hadoop jar WordCount3.jar com.wordcount3.WordCount s3://biblewordcount/ s3://biblewordcount/output2/
18/09/30 06:05:53 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-5-137.us-east-2.compute.internal/172.31.5.137:8032
18/09/30 06:05:55 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/09/30 06:05:55 INFO input.FileInputFormat: Total input files to process : 10
18/09/30 06:05:56 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
18/09/30 06:05:56 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 29dbe82f1ecfd8384b89f1a32e5e2d9e5dfd7724]
18/09/30 06:05:56 INFO mapreduce.JobSubmitter: number of splits:10
18/09/30 06:05:56 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538278733852_0005
18/09/30 06:05:56 INFO impl.YarnClientImpl: Submitted application application_1538278733852_0005
18/09/30 06:05:56 INFO mapreduce.Job: The url to track the job: http://ip-172-31-5-137.us-east-2.compute.internal:20888/proxy/application_1538278733852_0005/
18/09/30 06:05:56 INFO mapreduce.Job: Running job: job_1538278733852_0005
18/09/30 06:06:07 INFO mapreduce.Job: Job job_1538278733852_0005 running in uber mode : false
18/09/30 06:06:07 INFO mapreduce.Job:  map 0% reduce 0%
18/09/30 06:06:24 INFO mapreduce.Job:  map 30% reduce 0%
18/09/30 06:06:25 INFO mapreduce.Job:  map 40% reduce 0%
18/09/30 06:06:29 INFO mapreduce.Job:  map 50% reduce 0%
18/09/30 06:06:31 INFO mapreduce.Job:  map 70% reduce 0%
18/09/30 06:06:32 INFO mapreduce.Job:  map 80% reduce 0%
18/09/30 06:06:33 INFO mapreduce.Job:  map 100% reduce 0%
18/09/30 06:06:34 INFO mapreduce.Job:  map 100% reduce 25%
18/09/30 06:06:37 INFO mapreduce.Job:  map 100% reduce 50%
18/09/30 06:06:38 INFO mapreduce.Job:  map 100% reduce 75%
18/09/30 06:06:39 INFO mapreduce.Job:  map 100% reduce 100%
18/09/30 06:06:39 INFO mapreduce.Job: Job job_1538278733852_0005 completed successfully
18/09/30 06:06:39 INFO mapreduce.Job: Counters: 55
File System Counters
  FILE: Number of bytes read=2023
  FILE: Number of bytes written=2398437
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=985
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=10
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0
  S3: Number of bytes read=81613300
  S3: Number of bytes written=891
  S3: Number of read operations=0
  S3: Number of large read operations=0
  S3: Number of write operations=0
Job Counters
  Killed map tasks=1
  Launched map tasks=10
  Launched reduce tasks=4

```

```

    Data-local map tasks=10
    Total time spent by all maps in occupied slots (ms)=9160848
    Total time spent by all reduces in occupied slots (ms)=3446592
    Total time spent by all map tasks (ms)=190851
    Total time spent by all reduce tasks (ms)=35902
    Total vcore-milliseconds taken by all map tasks=190851
    Total vcore-milliseconds taken by all reduce tasks=35902
    Total megabyte-milliseconds taken by all map tasks=293147136
    Total megabyte-milliseconds taken by all reduce tasks=110290944
Map-Reduce Framework
    Map input records=1405946
    Map output records=6067351
    Map output bytes=47130252
    Map output materialized bytes=9684
    Input split bytes=985
    Combine input records=6067351
    Combine output records=820
    Reduce input groups=82
    Reduce shuffle bytes=9684
    Reduce input records=820
    Reduce output records=82
    Spilled Records=1640
    Shuffled Maps =40
    Failed Shuffles=0
    Merged Map outputs=40
    GC time elapsed (ms)=7781
    CPU time spent (ms)=47820
    Physical memory (bytes) snapshot=8210546688
    Virtual memory (bytes) snapshot=51722678272
    Total committed heap usage (bytes)=7659323392
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=81613300
File Output Format Counters
    Bytes Written=891
[hadoop@ip-172-31-5-137 wordcount]$ clear

[hadoop@ip-172-31-5-137 wordcount]$ hadoop jar WordCount2.jar com.wordcount3.WordCount s3://biblewordcount/ s3://biblewordcount/ou
tput/
Exception in thread "main" java.lang.ClassNotFoundException: com.wordcount3.WordCount
    at java.net.URLClassLoader.findClass(URLClassLoader.java:381)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:424)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:357)
    at java.lang.Class.forName0(Native Method)
    at java.lang.Class.forName(Class.java:348)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:227)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:148)
[hadoop@ip-172-31-5-137 wordcount]$ hadoop jar WordCount2.jar com.wordcount2.WordCount s3://biblewordcount/ s3://biblewordcount/ou
tput/
18/09/30 06:19:55 INFO client.RMPProxy: Connecting to ResourceManager at ip-172-31-5-137.us-east-2.compute.internal/172.31.5.137:80
32
18/09/30 06:19:56 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interfa
ce and execute your application with ToolRunner to remedy this.
18/09/30 06:19:57 INFO input.FileInputFormat: Total input files to process : 11
18/09/30 06:19:57 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
18/09/30 06:19:57 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 29dbe82f1ecfd8384b89f1a3
2e5e2d9e5dfd7724]
18/09/30 06:19:57 INFO mapreduce.JobSubmitter: number of splits:11
18/09/30 06:19:57 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538278733852_0006
18/09/30 06:19:57 INFO impl.YarnClientImpl: Submitted application application_1538278733852_0006
18/09/30 06:19:57 INFO mapreduce.Job: The url to track the job: http://ip-172-31-5-137.us-east-2.compute.internal:20888/proxy/appl
ication_1538278733852_0006/
18/09/30 06:19:57 INFO mapreduce.Job: Running job: job_1538278733852_0006
^C[hadoop@ip-172-31-5-137 wordcount]$ hadoop jar WordCount2.jar com.wordcount2.WordCount s3://biblewordcount/ s3://biblewordcount/
tput/
18/09/30 06:20:45 INFO client.RMPProxy: Connecting to ResourceManager at ip-172-31-5-137.us-east-2.compute.internal/172.31.5.137:80
32
18/09/30 06:20:46 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interfa
ce and execute your application with ToolRunner to remedy this.
18/09/30 06:20:46 INFO input.FileInputFormat: Total input files to process : 9
18/09/30 06:20:46 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
18/09/30 06:20:46 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 29dbe82f1ecfd8384b89f1a3
2e5e2d9e5dfd7724]
18/09/30 06:20:46 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedExce
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1252)
    at java.lang.Thread.join(Thread.java:1326)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:973)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:624)

```

```
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:801)
18/09/30 06:20:46 INFO mapreduce.JobSubmitter: number of splits:9
18/09/30 06:20:47 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538278733852_0007
18/09/30 06:20:47 INFO impl.YarnClientImpl: Submitted application application_1538278733852_0007
18/09/30 06:20:47 INFO mapreduce.Job: The url to track the job: http://ip-172-31-5-137.us-east-2.compute.internal:20888/proxy/applic
ation_1538278733852_0007/
18/09/30 06:20:47 INFO mapreduce.Job: Running job: job_1538278733852_0007
18/09/30 06:21:13 INFO mapreduce.Job: Job job_1538278733852_0007 running in uber mode : false
18/09/30 06:21:13 INFO mapreduce.Job: map 0% reduce 0%
18/09/30 06:21:32 INFO mapreduce.Job: map 15% reduce 0%
18/09/30 06:21:33 INFO mapreduce.Job: map 22% reduce 0%
18/09/30 06:21:36 INFO mapreduce.Job: map 33% reduce 0%
18/09/30 06:21:37 INFO mapreduce.Job: map 44% reduce 0%
18/09/30 06:21:38 INFO mapreduce.Job: map 67% reduce 0%
18/09/30 06:21:39 INFO mapreduce.Job: map 74% reduce 0%
18/09/30 06:21:42 INFO mapreduce.Job: map 81% reduce 0%
18/09/30 06:21:43 INFO mapreduce.Job: map 89% reduce 0%
18/09/30 06:21:47 INFO mapreduce.Job: map 93% reduce 0%
18/09/30 06:21:50 INFO mapreduce.Job: map 100% reduce 0%
18/09/30 06:21:52 INFO mapreduce.Job: map 100% reduce 25%
18/09/30 06:21:53 INFO mapreduce.Job: map 100% reduce 50%
18/09/30 06:21:55 INFO mapreduce.Job: map 100% reduce 100%
18/09/30 06:21:56 INFO mapreduce.Job: Job job_1538278733852_0007 completed successfully
18/09/30 06:21:56 INFO mapreduce.Job: Counters: 56
  File System Counters
    FILE: Number of bytes read=7954111
    FILE: Number of bytes written=43348948
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=883
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=9
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=0
    S3: Number of bytes read=81612666
    S3: Number of bytes written=6175163
    S3: Number of read operations=0
    S3: Number of large read operations=0
    S3: Number of write operations=0
  Job Counters
    Killed map tasks=1
    Killed reduce tasks=1
    Launched map tasks=9
    Launched reduce tasks=4
    Data-local map tasks=9
    Total time spent by all maps in occupied slots (ms)=11008368
    Total time spent by all reduces in occupied slots (ms)=5668032
    Total time spent by all map tasks (ms)=229341
    Total time spent by all reduce tasks (ms)=59042
    Total vcore-milliseconds taken by all map tasks=229341
    Total vcore-milliseconds taken by all reduce tasks=59042
    Total megabyte-milliseconds taken by all map tasks=352267776
    Total megabyte-milliseconds taken by all reduce tasks=181377024
  Map-Reduce Framework
    Map input records=1405935
    Map output records=14203980
    Map output bytes=202111353
    Map output materialized bytes=33191208
    Input split bytes=883
    Combine input records=14203980
    Combine output records=3820158
    Reduce input groups=424462
    Reduce shuffle bytes=33191208
    Reduce input records=3820158
    Reduce output records=424462
    Spilled Records=7640316
    Shuffled Maps =36
    Failed Shuffles=0
    Merged Map outputs=36
    GC time elapsed (ms)=7339
    CPU time spent (ms)=122440
    Physical memory (bytes) snapshot=8967155712
    Virtual memory (bytes) snapshot=48418066432
    Total committed heap usage (bytes)=8195145728
  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=81612666
  File Output Format Counters
    Bytes Written=6175163
```

```
[hadoop@ip-172-31-5-137 wordcount]$ clear
```

```
[hadoop@ip-172-31-5-137 wordcount]$ hadoop jar WordCount.jar com.wordcount.WordCount s3://biblewordcount/ s3://biblewordcount/output/
18/09/30 06:29:51 INFO client.RMPProxy: Connecting to ResourceManager at ip-172-31-5-137.us-east-2.compute.internal/172.31.5.137:8032
18/09/30 06:29:52 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/09/30 06:29:53 INFO input.FileInputFormat: Total input files to process : 9
18/09/30 06:29:53 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
18/09/30 06:29:53 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 29dbe82f1ecfd8384b89f1a32e5e2d9e5dfd7724]
18/09/30 06:29:53 INFO mapreduce.JobSubmitter: number of splits:9
18/09/30 06:29:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1538278733852_0008
18/09/30 06:29:53 INFO impl.YarnClientImpl: Submitted application application_1538278733852_0008
18/09/30 06:29:53 INFO mapreduce.Job: The url to track the job: http://ip-172-31-5-137.us-east-2.compute.internal:20888/proxy/application_1538278733852_0008/
18/09/30 06:29:53 INFO mapreduce.Job: Running job: job_1538278733852_0008
18/09/30 06:30:03 INFO mapreduce.Job: Job job_1538278733852_0008 running in uber mode : false
18/09/30 06:30:03 INFO mapreduce.Job: map 0% reduce 0%
18/09/30 06:30:23 INFO mapreduce.Job: map 15% reduce 0%
18/09/30 06:30:24 INFO mapreduce.Job: map 22% reduce 0%
18/09/30 06:30:25 INFO mapreduce.Job: map 41% reduce 0%
18/09/30 06:30:26 INFO mapreduce.Job: map 67% reduce 0%
18/09/30 06:30:30 INFO mapreduce.Job: map 74% reduce 0%
18/09/30 06:30:32 INFO mapreduce.Job: map 81% reduce 0%
18/09/30 06:30:34 INFO mapreduce.Job: map 93% reduce 0%
18/09/30 06:30:35 INFO mapreduce.Job: map 96% reduce 0%
18/09/30 06:30:36 INFO mapreduce.Job: map 100% reduce 0%
18/09/30 06:30:37 INFO mapreduce.Job: map 100% reduce 50%
18/09/30 06:30:38 INFO mapreduce.Job: map 100% reduce 75%
18/09/30 06:30:40 INFO mapreduce.Job: map 100% reduce 100%
18/09/30 06:30:40 INFO mapreduce.Job: Job job_1538278733852_0008 completed successfully
18/09/30 06:30:41 INFO mapreduce.Job: Counters: 56
```

File System Counters

```
FILE: Number of bytes read=582882
FILE: Number of bytes written=5451386
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=883
HDFS: Number of bytes written=0
HDFS: Number of read operations=9
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
S3: Number of bytes read=81612666
S3: Number of bytes written=392595
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0
```

Job Counters

```
Killed map tasks=1
Killed reduce tasks=1
Launched map tasks=10
Launched reduce tasks=4
Data-local map tasks=10
Total time spent by all maps in occupied slots (ms)=9895104
Total time spent by all reduces in occupied slots (ms)=3971616
Total time spent by all map tasks (ms)=206148
Total time spent by all reduce tasks (ms)=41371
Total vcore-milliseconds taken by all map tasks=206148
Total vcore-milliseconds taken by all reduce tasks=41371
Total megabyte-milliseconds taken by all map tasks=316643328
Total megabyte-milliseconds taken by all reduce tasks=127091712
```

Map-Reduce Framework

```
Map input records=1405935
Map output records=8024985
Map output bytes=83136447
Map output materialized bytes=2664927
Input split bytes=883
Combine input records=8024985
Combine output records=310491
Reduce input groups=34499
Reduce shuffle bytes=2664927
Reduce input records=310491
Reduce output records=34499
Spilled Records=620982
Shuffled Maps =36
Failed Shuffles=0
Merged Map outputs=36
GC time elapsed (ms)=8861
CPU time spent (ms)=89610
Physical memory (bytes) snapshot=9417998336
Virtual memory (bytes) snapshot=48417718272
Total committed heap usage (bytes)=8585216000
```

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=81612666
File Output Format Counters
  Bytes Written=392595
[hadoop@ip-172-31-5-137 wordcount]$
```