

2020-03-27 06:08:02,340 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:08:03,240 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 06:08:03,247 INFO com.hadoop.compression.lzo.GPLNativeCodeLoader (main): Loaded native gpl library
2020-03-27 06:08:03,249 INFO com.hadoop.compression.lzo.LzoCodec (main): Successfully loaded & initialized native-lzo library [hadoop-lzo rev 29dbe82f1ecfd8384b89f1a32e5e2d9e5dfd7724]
2020-03-27 06:08:03,625 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:08:03,713 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0001
2020-03-27 06:08:04,218 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585289193100_0001
2020-03-27 06:08:04,274 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0001/
2020-03-27 06:08:04,275 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585289193100_0001
2020-03-27 06:08:10,341 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0001 running in uber mode : false
2020-03-27 06:08:10,342 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:08:23,420 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:08:28,442 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 15%
2020-03-27 06:08:29,445 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 38%
2020-03-27 06:08:30,450 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:08:30,456 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0001 completed successfully
2020-03-27 06:08:30,544 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5679225
FILE: Number of bytes written=13722591
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=104
HDFS: Number of bytes written=13875033
HDFS: Number of read operations=41
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=27661801
S3: Number of bytes written=0
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=995040
Total time spent by all reduces in occupied slots (ms)=9376320
Total time spent by all map tasks (ms)=10365
Total time spent by all reduce tasks (ms)=48835
Total vcore-milliseconds taken by all map tasks=10365
Total vcore-milliseconds taken by all reduce tasks=48835
Total megabyte-milliseconds taken by all map tasks=31841280
Total megabyte-milliseconds taken by all reduce tasks=300042240

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=13884009
Map output materialized bytes=5679344
Input split bytes=104
Combine input records=1000000

```
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5679344
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=2136
CPU time spent (ms)=38690
Physical memory (bytes) snapshot=4650102784
Virtual memory (bytes) snapshot=99373400064
Total committed heap usage (bytes)=4946657280

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=13772905
File Output Format Counters
  Bytes Written=13875033

2020-03-27 06:08:30,592 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:08:30,817 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 06:08:30,871 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 06:08:30,935 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585289193100_0002
2020-03-27 06:08:31,001 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585289193100_0002
2020-03-27 06:08:31,006 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0002/
2020-03-27 06:08:31,006 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585289193100_0002
2020-03-27 06:08:37,098 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0002 running in uber mode : false
2020-03-27 06:08:37,098 INFO org.apache.hadoop.mapreduce.Job (main):  map 0% reduce 0%
2020-03-27 06:08:48,145 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 0%
2020-03-27 06:08:56,187 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 8%
2020-03-27 06:08:57,191 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 46%
2020-03-27 06:08:58,197 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 92%
2020-03-27 06:08:59,201 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 100%
2020-03-27 06:08:59,210 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0002 completed successfully
2020-03-27 06:08:59,235 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters
  FILE: Number of bytes read=5724899
  FILE: Number of bytes written=13810590
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=13875138
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=16
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0
  S3: Number of bytes read=8888896
  S3: Number of bytes written=15888896
  S3: Number of read operations=0
  S3: Number of large read operations=0
```

S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=857664
Total time spent by all reduces in occupied slots (ms)=14489472
Total time spent by all map tasks (ms)=8934
Total time spent by all reduce tasks (ms)=75466
Total vcore-milliseconds taken by all map tasks=8934
Total vcore-milliseconds taken by all reduce tasks=75466
Total megabyte-milliseconds taken by all map tasks=27445248
Total megabyte-milliseconds taken by all reduce tasks=463663104

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5724847
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5724847
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=3842
CPU time spent (ms)=31700
Physical memory (bytes) snapshot=6823759872
Virtual memory (bytes) snapshot=99728142336
Total committed heap usage (bytes)=6909591552

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=8888896

File Output Format Counters

Bytes Written=15888896

2020-03-27 06:08:59,365 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:08:59,410 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 06:08:59,468 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:08:59,520 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0003
2020-03-27 06:08:59,550 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585289193100_0003
2020-03-27 06:08:59,574 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0003/
2020-03-27 06:08:59,574 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585289193100_0003
2020-03-27 06:09:04,625 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0003 running in uber mode : false
2020-03-27 06:09:04,625 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:09:16,671 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:09:22,707 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%

2020-03-27 06:09:23,712 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:09:23,718 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0003 completed successfully
2020-03-27 06:09:23,745 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5721565
FILE: Number of bytes written=13807454
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=104
HDFS: Number of bytes written=15873020
HDFS: Number of read operations=41
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=29661801
S3: Number of bytes written=0
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=982752
Total time spent by all reduces in occupied slots (ms)=7479552
Total time spent by all map tasks (ms)=10237
Total time spent by all reduce tasks (ms)=38956
Total vcore-milliseconds taken by all map tasks=10237
Total vcore-milliseconds taken by all reduce tasks=38956
Total megabyte-milliseconds taken by all map tasks=31448064
Total megabyte-milliseconds taken by all reduce tasks=239345664

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5721867
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5721867
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=2197
CPU time spent (ms)=35770
Physical memory (bytes) snapshot=4601802752
Virtual memory (bytes) snapshot=99376201728
Total committed heap usage (bytes)=4907335680

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=13772905

File Output Format Counters

Bytes Written=15873020

```
2020-03-27 06:09:23,772 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:09:23,918 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 06:09:23,964 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 06:09:24,001 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585289193100_0004
2020-03-27 06:09:24,031 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585289193100_0004
2020-03-27 06:09:24,035 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0004/
2020-03-27 06:09:24,036 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585289193100_0004
2020-03-27 06:09:31,092 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0004 running in uber mode : false
2020-03-27 06:09:31,092 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:09:42,152 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:09:48,172 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-03-27 06:09:49,175 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 15%
2020-03-27 06:09:50,178 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-03-27 06:09:51,181 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:09:51,185 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0004 completed successfully
2020-03-27 06:09:51,207 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

File System Counters

```
FILE: Number of bytes read=5724899
FILE: Number of bytes written=13810590
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=15873125
HDFS: Number of bytes written=0
HDFS: Number of read operations=16
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
S3: Number of bytes read=8888896
S3: Number of bytes written=15888896
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0
```

Job Counters

```
Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=815904
Total time spent by all reduces in occupied slots (ms)=14593536
Total time spent by all map tasks (ms)=8499
Total time spent by all reduce tasks (ms)=76008
Total vcore-milliseconds taken by all map tasks=8499
Total vcore-milliseconds taken by all reduce tasks=76008
Total megabyte-milliseconds taken by all map tasks=26108928
Total megabyte-milliseconds taken by all reduce tasks=466993152
```

Map-Reduce Framework

```
Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5724847
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5724847
Reduce input records=1000000
```

```
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=3585
CPU time spent (ms)=35420
Physical memory (bytes) snapshot=6882148352
Virtual memory (bytes) snapshot=99820908544
Total committed heap usage (bytes)=6917455872

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=8888896
File Output Format Counters
  Bytes Written=15888896
2020-03-27 06:09:51,230 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to
ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:09:51,327 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 06:09:51,370 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 06:09:51,394 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585289193100_0005
2020-03-27 06:09:51,416 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585289193100_0005
2020-03-27 06:09:51,423 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0005/
2020-03-27 06:09:51,423 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585289193100_0005
2020-03-27 06:09:56,469 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0005 running in uber mode : false
2020-03-27 06:09:56,469 INFO org.apache.hadoop.mapreduce.Job (main):  map 0% reduce 0%
2020-03-27 06:10:09,518 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 0%
2020-03-27 06:10:14,533 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 46%
2020-03-27 06:10:15,536 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 92%
2020-03-27 06:10:16,538 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 100%
2020-03-27 06:10:16,542 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0005 completed successfully
2020-03-27 06:10:16,564 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
File System Counters
  FILE: Number of bytes read=5721565
  FILE: Number of bytes written=13807454
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=104
  HDFS: Number of bytes written=15873020
  HDFS: Number of read operations=41
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=26
  S3: Number of bytes read=29661801
  S3: Number of bytes written=0
  S3: Number of read operations=0
  S3: Number of large read operations=0
  S3: Number of write operations=0
Job Counters
  Killed reduce tasks=1
  Launched map tasks=1
  Launched reduce tasks=13
```

Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=1028064
Total time spent by all reduces in occupied slots (ms)=7718016
Total time spent by all map tasks (ms)=10709
Total time spent by all reduce tasks (ms)=40198
Total vcore-milliseconds taken by all map tasks=10709
Total vcore-milliseconds taken by all reduce tasks=40198
Total megabyte-milliseconds taken by all map tasks=32898048
Total megabyte-milliseconds taken by all reduce tasks=246976512

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5721867
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5721867
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=2289
CPU time spent (ms)=36780
Physical memory (bytes) snapshot=4594176000
Virtual memory (bytes) snapshot=99349626880
Total committed heap usage (bytes)=4985978880

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=13772905

File Output Format Counters

Bytes Written=15873020

2020-03-27 06:10:16,584 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:10:16,987 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 06:10:17,020 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:10:17,036 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0006
2020-03-27 06:10:17,058 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585289193100_0006
2020-03-27 06:10:17,062 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0006/
2020-03-27 06:10:17,062 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585289193100_0006
2020-03-27 06:10:24,113 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0006 running in uber mode : false
2020-03-27 06:10:24,113 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:10:35,243 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:10:41,265 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-03-27 06:10:42,267 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 15%
2020-03-27 06:10:43,270 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 62%
2020-03-27 06:10:44,273 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:10:45,280 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0006 completed successfully

2020-03-27 06:10:45,301 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5724899
FILE: Number of bytes written=13810590
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=15873125
HDFS: Number of bytes written=0
HDFS: Number of read operations=16
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
S3: Number of bytes read=8888896
S3: Number of bytes written=15888896
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=834336
Total time spent by all reduces in occupied slots (ms)=14899584
Total time spent by all map tasks (ms)=8691
Total time spent by all reduce tasks (ms)=77602
Total vcore-milliseconds taken by all map tasks=8691
Total vcore-milliseconds taken by all reduce tasks=77602
Total megabyte-milliseconds taken by all map tasks=26698752
Total megabyte-milliseconds taken by all reduce tasks=476786688

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5724847
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5724847
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=3770
CPU time spent (ms)=34660
Physical memory (bytes) snapshot=6826090496
Virtual memory (bytes) snapshot=99822354432
Total committed heap usage (bytes)=6791626752

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=8888896

File Output Format Counters

Bytes Written=15888896

2020-03-27 06:10:45,322 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032

2020-03-27 06:10:45,349 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):

Total input files to process : 1
2020-03-27 06:10:45,384 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:10:45,402 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0007
2020-03-27 06:10:45,422 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585289193100_0007
2020-03-27 06:10:45,424 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0007/
2020-03-27 06:10:45,424 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585289193100_0007
2020-03-27 06:10:50,472 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0007 running in uber mode : false
2020-03-27 06:10:50,472 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:11:02,518 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:11:07,535 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-03-27 06:11:08,538 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 92%
2020-03-27 06:11:09,541 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:11:09,545 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0007 completed successfully
2020-03-27 06:11:09,599 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5721565
FILE: Number of bytes written=13807454
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=104
HDFS: Number of bytes written=15873020
HDFS: Number of read operations=41
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=29661801
S3: Number of bytes written=0
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=938304
Total time spent by all reduces in occupied slots (ms)=7748736
Total time spent by all map tasks (ms)=9774
Total time spent by all reduce tasks (ms)=40358
Total vcore-milliseconds taken by all map tasks=9774
Total vcore-milliseconds taken by all reduce tasks=40358
Total megabyte-milliseconds taken by all map tasks=30025728
Total megabyte-milliseconds taken by all reduce tasks=247959552

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5721867
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5721867
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =13
Failed Shuffles=0

```
Merged Map outputs=13
GC time elapsed (ms)=2222
CPU time spent (ms)=37100
Physical memory (bytes) snapshot=4628754432
Virtual memory (bytes) snapshot=99376680960
Total committed heap usage (bytes)=4991746048
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=13772905
File Output Format Counters
  Bytes Written=15873020
2020-03-27 06:11:09,618 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:11:09,718 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 06:11:10,157 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 06:11:10,176 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585289193100_0008
2020-03-27 06:11:10,393 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585289193100_0008
2020-03-27 06:11:10,395 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0008/
2020-03-27 06:11:10,395 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585289193100_0008
2020-03-27 06:11:17,441 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0008 running in uber mode : false
2020-03-27 06:11:17,441 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:11:28,510 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:11:34,528 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-03-27 06:11:36,534 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-03-27 06:11:37,537 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:11:38,543 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0008 completed successfully
2020-03-27 06:11:38,564 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
File System Counters
  FILE: Number of bytes read=5724899
  FILE: Number of bytes written=13810590
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=15873125
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=16
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0
  S3: Number of bytes read=8888896
  S3: Number of bytes written=15888896
  S3: Number of read operations=0
  S3: Number of large read operations=0
  S3: Number of write operations=0
Job Counters
  Killed reduce tasks=1
  Launched map tasks=1
  Launched reduce tasks=13
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=817824
  Total time spent by all reduces in occupied slots (ms)=14421120
  Total time spent by all map tasks (ms)=8519
```

Total time spent by all reduce tasks (ms)=75110
Total vcore-milliseconds taken by all map tasks=8519
Total vcore-milliseconds taken by all reduce tasks=75110
Total megabyte-milliseconds taken by all map tasks=26170368
Total megabyte-milliseconds taken by all reduce tasks=461475840

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5724847
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5724847
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=3809
CPU time spent (ms)=32390
Physical memory (bytes) snapshot=6902689792
Virtual memory (bytes) snapshot=99820285952
Total committed heap usage (bytes)=6937903104

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=8888896

File Output Format Counters

Bytes Written=15888896

2020-03-27 06:11:38,583 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:11:38,644 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 06:11:38,681 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:11:38,702 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0009
2020-03-27 06:11:38,922 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585289193100_0009
2020-03-27 06:11:38,924 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0009/
2020-03-27 06:11:38,924 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585289193100_0009
2020-03-27 06:11:43,962 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0009 running in uber mode : false
2020-03-27 06:11:43,963 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:11:57,010 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:12:03,032 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-03-27 06:12:04,036 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:12:05,041 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0009 completed successfully
2020-03-27 06:12:05,067 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5721565
FILE: Number of bytes written=13807454
FILE: Number of read operations=0
FILE: Number of large read operations=0

FILE: Number of write operations=0
HDFS: Number of bytes read=104
HDFS: Number of bytes written=15873020
HDFS: Number of read operations=41
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=29661801
S3: Number of bytes written=0
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=1053600
Total time spent by all reduces in occupied slots (ms)=7523136
Total time spent by all map tasks (ms)=10975
Total time spent by all reduce tasks (ms)=39183
Total vcore-milliseconds taken by all map tasks=10975
Total vcore-milliseconds taken by all reduce tasks=39183
Total megabyte-milliseconds taken by all map tasks=33715200
Total megabyte-milliseconds taken by all reduce tasks=240740352

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5721867
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5721867
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=2235
CPU time spent (ms)=36090
Physical memory (bytes) snapshot=4599398400
Virtual memory (bytes) snapshot=99362713600
Total committed heap usage (bytes)=4861722624

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=13772905

File Output Format Counters

Bytes Written=15873020

2020-03-27 06:12:05,086 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:12:05,234 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 06:12:05,271 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:12:05,291 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0010
2020-03-27 06:12:05,308 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):

Submitted application application_1585289193100_0010
2020-03-27 06:12:05,312 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0010/
2020-03-27 06:12:05,312 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585289193100_0010
2020-03-27 06:12:12,359 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0010 running in uber mode : false
2020-03-27 06:12:12,360 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:12:23,431 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:12:29,448 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-03-27 06:12:31,453 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-03-27 06:12:32,458 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:12:33,464 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0010 completed successfully
2020-03-27 06:12:33,486 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5724899
FILE: Number of bytes written=13810590
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=15873125
HDFS: Number of bytes written=0
HDFS: Number of read operations=16
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
S3: Number of bytes read=8888896
S3: Number of bytes written=15888896
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=820032
Total time spent by all reduces in occupied slots (ms)=14744256
Total time spent by all map tasks (ms)=8542
Total time spent by all reduce tasks (ms)=76793
Total vcore-milliseconds taken by all map tasks=8542
Total vcore-milliseconds taken by all reduce tasks=76793
Total megabyte-milliseconds taken by all map tasks=26241024
Total megabyte-milliseconds taken by all reduce tasks=471816192

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5724847
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5724847
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=3704
CPU time spent (ms)=36360
Physical memory (bytes) snapshot=6875607040
Virtual memory (bytes) snapshot=99982426112
Total committed heap usage (bytes)=6903824384

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=8888896

File Output Format Counters

Bytes Written=15888896

2020-03-27 06:12:33,507 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:12:33,540 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 06:12:33,573 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:12:33,590 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0011
2020-03-27 06:12:33,603 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585289193100_0011
2020-03-27 06:12:33,605 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0011/
2020-03-27 06:12:33,605 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585289193100_0011
2020-03-27 06:12:38,647 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0011 running in uber mode : false
2020-03-27 06:12:38,647 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:12:50,688 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:12:55,703 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 46%
2020-03-27 06:12:56,706 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 92%
2020-03-27 06:12:57,710 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:12:57,713 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0011 completed successfully
2020-03-27 06:12:57,740 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5721565
FILE: Number of bytes written=13807454
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=104
HDFS: Number of bytes written=15873020
HDFS: Number of read operations=41
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=29661801
S3: Number of bytes written=0
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=940032
Total time spent by all reduces in occupied slots (ms)=7923648
Total time spent by all map tasks (ms)=9792
Total time spent by all reduce tasks (ms)=41269
Total vcore-milliseconds taken by all map tasks=9792
Total vcore-milliseconds taken by all reduce tasks=41269
Total megabyte-milliseconds taken by all map tasks=30081024
Total megabyte-milliseconds taken by all reduce tasks=253556736

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5721867
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5721867
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=2163
CPU time spent (ms)=37170
Physical memory (bytes) snapshot=4539781120
Virtual memory (bytes) snapshot=99354468352
Total committed heap usage (bytes)=4853858304

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=13772905

File Output Format Counters

Bytes Written=15873020

2020-03-27 06:12:57,758 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:12:57,835 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 06:12:57,866 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:12:57,880 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0012
2020-03-27 06:12:58,090 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585289193100_0012
2020-03-27 06:12:58,092 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0012/
2020-03-27 06:12:58,092 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585289193100_0012
2020-03-27 06:13:04,192 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0012 running in uber mode : false
2020-03-27 06:13:04,192 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:13:15,338 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:13:21,356 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-03-27 06:13:23,360 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-03-27 06:13:24,364 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:13:25,370 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0012 completed successfully
2020-03-27 06:13:25,392 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5724899
FILE: Number of bytes written=13810590
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=15873125
HDFS: Number of bytes written=0
HDFS: Number of read operations=16
HDFS: Number of large read operations=0

HDFS: Number of write operations=0
S3: Number of bytes read=88888896
S3: Number of bytes written=158888896
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=824832
Total time spent by all reduces in occupied slots (ms)=14240640
Total time spent by all map tasks (ms)=8592
Total time spent by all reduce tasks (ms)=74170
Total vcore-milliseconds taken by all map tasks=8592
Total vcore-milliseconds taken by all reduce tasks=74170
Total megabyte-milliseconds taken by all map tasks=26394624
Total megabyte-milliseconds taken by all reduce tasks=455700480

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=158888896
Map output materialized bytes=5724847
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5724847
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=3547
CPU time spent (ms)=31610
Physical memory (bytes) snapshot=6712074240
Virtual memory (bytes) snapshot=99856928768
Total committed heap usage (bytes)=6658457600

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=88888896

File Output Format Counters

Bytes Written=158888896

2020-03-27 06:13:25,413 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:13:25,472 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 06:13:25,508 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:13:25,524 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0013
2020-03-27 06:13:25,734 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585289193100_0013
2020-03-27 06:13:25,736 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0013/
2020-03-27 06:13:25,736 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585289193100_0013


```
2020-03-27 06:13:30,778 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0013 running in uber mode : false
2020-03-27 06:13:30,779 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:13:42,824 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:13:46,838 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-03-27 06:13:47,841 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-03-27 06:13:48,844 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:13:48,849 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0013 completed successfully
2020-03-27 06:13:48,874 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

File System Counters

```
FILE: Number of bytes read=5721565
FILE: Number of bytes written=13807454
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=104
HDFS: Number of bytes written=15873020
HDFS: Number of read operations=41
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=29661801
S3: Number of bytes written=0
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0
```

Job Counters

```
Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=941280
Total time spent by all reduces in occupied slots (ms)=7480128
Total time spent by all map tasks (ms)=9805
Total time spent by all reduce tasks (ms)=38959
Total vcore-milliseconds taken by all map tasks=9805
Total vcore-milliseconds taken by all reduce tasks=38959
Total megabyte-milliseconds taken by all map tasks=30120960
Total megabyte-milliseconds taken by all reduce tasks=239364096
```

Map-Reduce Framework

```
Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5721867
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5721867
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=2070
CPU time spent (ms)=36830
Physical memory (bytes) snapshot=4490305536
Virtual memory (bytes) snapshot=99369320448
Total committed heap usage (bytes)=4647813120
```

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=13772905
File Output Format Counters
  Bytes Written=15873020
2020-03-27 06:13:48,895 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:13:49,017 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 06:13:49,057 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 06:13:49,076 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585289193100_0014
2020-03-27 06:13:49,287 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585289193100_0014
2020-03-27 06:13:49,288 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0014/
2020-03-27 06:13:49,288 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585289193100_0014
2020-03-27 06:13:56,334 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0014 running in uber mode : false
2020-03-27 06:13:56,334 INFO org.apache.hadoop.mapreduce.Job (main):  map 0% reduce 0%
2020-03-27 06:14:06,373 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 0%
2020-03-27 06:14:14,396 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 8%
2020-03-27 06:14:15,398 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 54%
2020-03-27 06:14:16,401 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 100%
2020-03-27 06:14:17,406 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0014 completed successfully
2020-03-27 06:14:17,427 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
File System Counters
  FILE: Number of bytes read=5724899
  FILE: Number of bytes written=13810590
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=15873125
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=16
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0
  S3: Number of bytes read=8888896
  S3: Number of bytes written=15888896
  S3: Number of read operations=0
  S3: Number of large read operations=0
  S3: Number of write operations=0
Job Counters
  Killed reduce tasks=1
  Launched map tasks=1
  Launched reduce tasks=13
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=800736
  Total time spent by all reduces in occupied slots (ms)=14632512
  Total time spent by all map tasks (ms)=8341
  Total time spent by all reduce tasks (ms)=76211
  Total vcore-milliseconds taken by all map tasks=8341
  Total vcore-milliseconds taken by all reduce tasks=76211
  Total megabyte-milliseconds taken by all map tasks=25623552
  Total megabyte-milliseconds taken by all reduce tasks=468240384
Map-Reduce Framework
  Map input records=1000000
  Map output records=1000000
  Map output bytes=15888896
  Map output materialized bytes=5724847
  Input split bytes=105
```

```
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5724847
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=3762
CPU time spent (ms)=33290
Physical memory (bytes) snapshot=7079776256
Virtual memory (bytes) snapshot=99871137792
Total committed heap usage (bytes)=7024410624

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=8888896

File Output Format Counters
  Bytes Written=15888896

2020-03-27 06:14:17,448 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:14:17,467 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 06:14:17,496 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 06:14:17,517 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585289193100_0015
2020-03-27 06:14:17,536 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585289193100_0015
2020-03-27 06:14:17,537 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0015/
2020-03-27 06:14:17,537 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585289193100_0015
2020-03-27 06:14:22,578 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0015 running in uber mode : false
2020-03-27 06:14:22,578 INFO org.apache.hadoop.mapreduce.Job (main):  map 0% reduce 0%
2020-03-27 06:14:34,618 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 0%
2020-03-27 06:14:40,633 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 54%
2020-03-27 06:14:41,636 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 100%
2020-03-27 06:14:41,639 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0015 completed successfully
2020-03-27 06:14:41,664 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters
  FILE: Number of bytes read=5721565
  FILE: Number of bytes written=13807454
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=104
  HDFS: Number of bytes written=15873020
  HDFS: Number of read operations=41
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=26
  S3: Number of bytes read=29661801
  S3: Number of bytes written=0
  S3: Number of read operations=0
  S3: Number of large read operations=0
  S3: Number of write operations=0
```

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=973152
Total time spent by all reduces in occupied slots (ms)=7600512
Total time spent by all map tasks (ms)=10137
Total time spent by all reduce tasks (ms)=39586
Total vcore-milliseconds taken by all map tasks=10137
Total vcore-milliseconds taken by all reduce tasks=39586
Total megabyte-milliseconds taken by all map tasks=31140864
Total megabyte-milliseconds taken by all reduce tasks=243216384

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5721867
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5721867
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=2116
CPU time spent (ms)=35670
Physical memory (bytes) snapshot=4404314112
Virtual memory (bytes) snapshot=99355488256
Total committed heap usage (bytes)=4679794688

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=13772905

File Output Format Counters

Bytes Written=15873020

2020-03-27 06:14:41,684 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to
ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:14:41,955 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 06:14:41,985 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 06:14:42,000 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585289193100_0016
2020-03-27 06:14:42,012 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585289193100_0016
2020-03-27 06:14:42,015 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0016/
2020-03-27 06:14:42,015 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585289193100_0016
2020-03-27 06:14:49,059 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0016 running in uber mode : false
2020-03-27 06:14:49,059 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:14:59,099 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:15:06,123 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-03-27 06:15:07,125 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 23%

2020-03-27 06:15:08,129 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 92%
2020-03-27 06:15:09,131 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:15:09,135 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0016 completed successfully
2020-03-27 06:15:09,157 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5724899
FILE: Number of bytes written=13810590
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=15873125
HDFS: Number of bytes written=0
HDFS: Number of read operations=16
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
S3: Number of bytes read=8888896
S3: Number of bytes written=15888896
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=808416
Total time spent by all reduces in occupied slots (ms)=14552640
Total time spent by all map tasks (ms)=8421
Total time spent by all reduce tasks (ms)=75795
Total vcore-milliseconds taken by all map tasks=8421
Total vcore-milliseconds taken by all reduce tasks=75795
Total megabyte-milliseconds taken by all map tasks=25869312
Total megabyte-milliseconds taken by all reduce tasks=465684480

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5724847
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5724847
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=3699
CPU time spent (ms)=31810
Physical memory (bytes) snapshot=6906945536
Virtual memory (bytes) snapshot=99814805504
Total committed heap usage (bytes)=6931087360

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=8888896

File Output Format Counters

```
Bytes Written=15888896
2020-03-27 06:15:09,177 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:15:09,242 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 06:15:09,274 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 06:15:09,289 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585289193100_0017
2020-03-27 06:15:09,302 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585289193100_0017
2020-03-27 06:15:09,314 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0017/
2020-03-27 06:15:09,314 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585289193100_0017
2020-03-27 06:15:14,355 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0017 running in uber mode : false
2020-03-27 06:15:14,355 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:15:26,394 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:15:31,410 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-03-27 06:15:32,413 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-03-27 06:15:33,415 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:15:33,419 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0017 completed successfully
2020-03-27 06:15:33,440 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

File System Counters

```
FILE: Number of bytes read=5721565
FILE: Number of bytes written=13807454
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=104
HDFS: Number of bytes written=15873020
HDFS: Number of read operations=41
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=29661801
S3: Number of bytes written=0
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0
```

Job Counters

```
Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=973056
Total time spent by all reduces in occupied slots (ms)=7389120
Total time spent by all map tasks (ms)=10136
Total time spent by all reduce tasks (ms)=38485
Total vcore-milliseconds taken by all map tasks=10136
Total vcore-milliseconds taken by all reduce tasks=38485
Total megabyte-milliseconds taken by all map tasks=31137792
Total megabyte-milliseconds taken by all reduce tasks=236451840
```

Map-Reduce Framework

```
Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5721867
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5721867
Reduce input records=999001
```

```
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=2135
CPU time spent (ms)=35810
Physical memory (bytes) snapshot=4599222272
Virtual memory (bytes) snapshot=99359002624
Total committed heap usage (bytes)=5285347328

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=13772905
File Output Format Counters
  Bytes Written=15873020

2020-03-27 06:15:33,457 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to
ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:15:33,544 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 06:15:33,577 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 06:15:33,591 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585289193100_0018
2020-03-27 06:15:33,600 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585289193100_0018
2020-03-27 06:15:33,602 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0018/
2020-03-27 06:15:33,602 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585289193100_0018
2020-03-27 06:15:40,648 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0018 running in uber mode : false
2020-03-27 06:15:40,648 INFO org.apache.hadoop.mapreduce.Job (main):  map 0% reduce 0%
2020-03-27 06:15:50,682 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 0%
2020-03-27 06:15:57,704 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 8%
2020-03-27 06:15:58,706 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 15%
2020-03-27 06:15:59,709 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 54%
2020-03-27 06:16:00,712 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 100%
2020-03-27 06:16:00,716 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585289193100_0018 completed successfully
2020-03-27 06:16:00,737 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters
  FILE: Number of bytes read=5724899
  FILE: Number of bytes written=13810590
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=15873125
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=16
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0
  S3: Number of bytes read=8888896
  S3: Number of bytes written=15888896
  S3: Number of read operations=0
  S3: Number of large read operations=0
  S3: Number of write operations=0

Job Counters
  Killed reduce tasks=1
  Launched map tasks=1
```

Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=789312
Total time spent by all reduces in occupied slots (ms)=14472192
Total time spent by all map tasks (ms)=8222
Total time spent by all reduce tasks (ms)=75376
Total vcore-milliseconds taken by all map tasks=8222
Total vcore-milliseconds taken by all reduce tasks=75376
Total megabyte-milliseconds taken by all map tasks=25257984
Total megabyte-milliseconds taken by all reduce tasks=463110144

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5724847
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5724847
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=3722
CPU time spent (ms)=33760
Physical memory (bytes) snapshot=7052791808
Virtual memory (bytes) snapshot=99836538880
Total committed heap usage (bytes)=7117733888

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=8888896

File Output Format Counters

Bytes Written=15888896

2020-03-27 06:16:00,755 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032
2020-03-27 06:16:00,775 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 06:16:00,809 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:16:00,824 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0019
2020-03-27 06:16:00,835 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585289193100_0019
2020-03-27 06:16:00,837 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0019/
2020-03-27 06:16:00,837 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585289193100_0019
2020-03-27 06:16:05,883 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0019 running in uber mode : false
2020-03-27 06:16:05,883 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:16:18,932 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:16:23,948 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-03-27 06:16:24,951 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 92%
2020-03-27 06:16:25,954 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:16:25,959 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0019 completed successfully

2020-03-27 06:16:25,980 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5721565
FILE: Number of bytes written=13807454
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=104
HDFS: Number of bytes written=15873020
HDFS: Number of read operations=41
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=29661801
S3: Number of bytes written=0
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=1021920
Total time spent by all reduces in occupied slots (ms)=7447680
Total time spent by all map tasks (ms)=10645
Total time spent by all reduce tasks (ms)=38790
Total vcore-milliseconds taken by all map tasks=10645
Total vcore-milliseconds taken by all reduce tasks=38790
Total megabyte-milliseconds taken by all map tasks=32701440
Total megabyte-milliseconds taken by all reduce tasks=238325760

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5721867
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5721867
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =13
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=2214
CPU time spent (ms)=37080
Physical memory (bytes) snapshot=4452012032
Virtual memory (bytes) snapshot=99381178368
Total committed heap usage (bytes)=4668784640

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=13772905

File Output Format Counters

Bytes Written=15873020

2020-03-27 06:16:25,997 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to ResourceManager at ip-172-31-8-239.ec2.internal/172.31.8.239:8032

2020-03-27 06:16:26,079 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):

Total input files to process : 1
2020-03-27 06:16:26,116 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 06:16:26,132 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585289193100_0020
2020-03-27 06:16:26,148 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585289193100_0020
2020-03-27 06:16:26,150 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-8-239.ec2.internal:20888/proxy/application_1585289193100_0020/
2020-03-27 06:16:26,150 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585289193100_0020
2020-03-27 06:16:33,195 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0020 running in uber mode : false
2020-03-27 06:16:33,195 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 06:16:44,230 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 06:16:50,252 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-03-27 06:16:51,254 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 31%
2020-03-27 06:16:52,257 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-03-27 06:16:53,259 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 06:16:53,263 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585289193100_0020 completed successfully
2020-03-27 06:16:53,283 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5724899
FILE: Number of bytes written=13810618
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=15873125
HDFS: Number of bytes written=0
HDFS: Number of read operations=16
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
S3: Number of bytes read=8888896
S3: Number of bytes written=15888896
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed reduce tasks=1
Launched map tasks=1
Launched reduce tasks=13
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=845280
Total time spent by all reduces in occupied slots (ms)=14368512
Total time spent by all map tasks (ms)=8805
Total time spent by all reduce tasks (ms)=74836
Total vcore-milliseconds taken by all map tasks=8805
Total vcore-milliseconds taken by all reduce tasks=74836
Total megabyte-milliseconds taken by all map tasks=27048960
Total megabyte-milliseconds taken by all reduce tasks=459792384

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5724847
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5724847
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =13

```
Failed Shuffles=0
Merged Map outputs=13
GC time elapsed (ms)=3812
CPU time spent (ms)=34420
Physical memory (bytes) snapshot=6915952640
Virtual memory (bytes) snapshot=99811737600
Total committed heap usage (bytes)=6982467584
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=8888896
File Output Format Counters
  Bytes Written=15888896
```