

```
2020-03-27 05:41:01,290 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to
ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:41:02,228 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 05:41:02,236 INFO com.hadoop.compression.lzo.GPLNativeCodeLoader (main): Loaded
native gpl library
2020-03-27 05:41:02,238 INFO com.hadoop.compression.lzo.LzoCodec (main): Successfully loaded
& initialized native-lzo library [hadoop-lzo rev 29dbe82f1ecfd8384b89f1a32e5e2d9e5dfd7724]
2020-03-27 05:41:02,482 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 05:41:02,574 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585287572595_0001
2020-03-27 05:41:02,832 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585287572595_0001
2020-03-27 05:41:02,970 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0001/
2020-03-27 05:41:02,975 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585287572595_0001
2020-03-27 05:41:09,054 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0001 running in uber mode : false
2020-03-27 05:41:09,055 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:41:22,121 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:41:27,141 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 22%
2020-03-27 05:41:29,150 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:41:29,155 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0001 completed successfully
2020-03-27 05:41:29,236 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

File System Counters

```
FILE: Number of bytes read=5513856
FILE: Number of bytes written=12716560
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=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=18
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=991968
Total time spent by all reduces in occupied slots (ms)=6692928
Total time spent by all map tasks (ms)=10333
Total time spent by all reduce tasks (ms)=34859
Total vcore-milliseconds taken by all map tasks=10333
Total vcore-milliseconds taken by all reduce tasks=34859
Total megabyte-milliseconds taken by all map tasks=31742976
Total megabyte-milliseconds taken by all reduce tasks=214173696
```

Map-Reduce Framework

```
Map input records=1000000
Map output records=1000000
Map output bytes=13884009
Map output materialized bytes=5513947
Input split bytes=104
Combine input records=1000000
Combine output records=999001
```

```
Reduce input groups=999001
Reduce shuffle bytes=5513947
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=1816
CPU time spent (ms)=35650
Physical memory (bytes) snapshot=3826319360
Virtual memory (bytes) snapshot=70205849600
Total committed heap usage (bytes)=3875536896

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 05:41:29,275 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:41:29,542 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 05:41:29,587 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 05:41:29,634 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585287572595_0002
2020-03-27 05:41:29,673 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585287572595_0002
2020-03-27 05:41:29,692 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0002/
2020-03-27 05:41:29,692 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585287572595_0002
2020-03-27 05:41:36,749 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0002 running in uber mode : false
2020-03-27 05:41:36,750 INFO org.apache.hadoop.mapreduce.Job (main):  map 0% reduce 0%
2020-03-27 05:41:47,796 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 0%
2020-03-27 05:41:54,821 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 11%
2020-03-27 05:41:55,824 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 22%
2020-03-27 05:41:56,828 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 56%
2020-03-27 05:41:57,831 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 100%
2020-03-27 05:41:58,840 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0002 completed successfully
2020-03-27 05:41:58,869 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters
  FILE: Number of bytes read=5549512
  FILE: Number of bytes written=12785465
  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=12
  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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=841248
Total time spent by all reduces in occupied slots (ms)=10254144
Total time spent by all map tasks (ms)=8763
Total time spent by all reduce tasks (ms)=53407
Total vcore-milliseconds taken by all map tasks=8763
Total vcore-milliseconds taken by all reduce tasks=53407
Total megabyte-milliseconds taken by all map tasks=26919936
Total megabyte-milliseconds taken by all reduce tasks=328132608

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5549476
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5549476
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=2976
CPU time spent (ms)=28160
Physical memory (bytes) snapshot=5064511488
Virtual memory (bytes) snapshot=70616403968
Total committed heap usage (bytes)=5194121216

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 05:41:58,987 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to Resource Manager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:41:59,043 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:41:59,111 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:41:59,145 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0003
2020-03-27 05:41:59,169 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0003
2020-03-27 05:41:59,178 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0003/
2020-03-27 05:41:59,179 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0003
2020-03-27 05:42:04,240 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0003 running in uber mode : false
2020-03-27 05:42:04,240 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:42:16,299 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:42:21,317 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 22%
2020-03-27 05:42:22,322 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 56%

2020-03-27 05:42:23,326 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:42:23,330 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0003 completed successfully
2020-03-27 05:42:23,356 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5545864
FILE: Number of bytes written=12780803
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=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=18
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=940704
Total time spent by all reduces in occupied slots (ms)=5371584
Total time spent by all map tasks (ms)=9799
Total time spent by all reduce tasks (ms)=27977
Total vcore-milliseconds taken by all map tasks=9799
Total vcore-milliseconds taken by all reduce tasks=27977
Total megabyte-milliseconds taken by all map tasks=30102528
Total megabyte-milliseconds taken by all reduce tasks=171890688

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5546182
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5546182
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=1945
CPU time spent (ms)=31380
Physical memory (bytes) snapshot=3672252416
Virtual memory (bytes) snapshot=70276587520
Total committed heap usage (bytes)=4078960640

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 05:42:23,376 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:42:23,527 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 05:42:23,562 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 05:42:23,584 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585287572595_0004
2020-03-27 05:42:23,602 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585287572595_0004
2020-03-27 05:42:23,605 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0004/
2020-03-27 05:42:23,605 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585287572595_0004
2020-03-27 05:42:30,655 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0004 running in uber mode : false
2020-03-27 05:42:30,655 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:42:40,692 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:42:47,716 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 11%
2020-03-27 05:42:49,722 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 56%
2020-03-27 05:42:50,725 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:42:50,729 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0004 completed successfully
2020-03-27 05:42:50,751 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

File System Counters

```
FILE: Number of bytes read=5549512
FILE: Number of bytes written=12785465
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=12
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=809280
Total time spent by all reduces in occupied slots (ms)=10273536
Total time spent by all map tasks (ms)=8430
Total time spent by all reduce tasks (ms)=53508
Total vcore-milliseconds taken by all map tasks=8430
Total vcore-milliseconds taken by all reduce tasks=53508
Total megabyte-milliseconds taken by all map tasks=25896960
Total megabyte-milliseconds taken by all reduce tasks=328753152
```

Map-Reduce Framework

```
Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5549476
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5549476
Reduce input records=1000000
Reduce output records=1000000
```

Spilled Records=2000000
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=2891
CPU time spent (ms)=26620
Physical memory (bytes) snapshot=5137944576
Virtual memory (bytes) snapshot=70570098688
Total committed heap usage (bytes)=5238161408

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 05:42:50,777 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:42:50,910 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:42:50,944 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:42:51,367 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0005
2020-03-27 05:42:51,384 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0005
2020-03-27 05:42:51,386 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0005/
2020-03-27 05:42:51,386 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0005
2020-03-27 05:42:56,436 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0005 running in uber mode : false
2020-03-27 05:42:56,436 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:43:08,481 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:43:13,496 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 11%
2020-03-27 05:43:14,499 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 56%
2020-03-27 05:43:15,502 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:43:15,506 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0005 completed successfully
2020-03-27 05:43:15,527 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5545864
FILE: Number of bytes written=12780803
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=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=18
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=9
Data-local map tasks=1

Total time spent by all maps in occupied slots (ms)=958944
Total time spent by all reduces in occupied slots (ms)=5355264
Total time spent by all map tasks (ms)=9989
Total time spent by all reduce tasks (ms)=27892
Total vcore-milliseconds taken by all map tasks=9989
Total vcore-milliseconds taken by all reduce tasks=27892
Total megabyte-milliseconds taken by all map tasks=30686208
Total megabyte-milliseconds taken by all reduce tasks=171368448

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5546182
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5546182
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=1910
CPU time spent (ms)=30930
Physical memory (bytes) snapshot=3665022976
Virtual memory (bytes) snapshot=70210441216
Total committed heap usage (bytes)=4061134848

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 05:43:15,546 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:43:15,686 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:43:15,726 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:43:15,747 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0006
2020-03-27 05:43:15,769 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0006
2020-03-27 05:43:15,771 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0006/
2020-03-27 05:43:15,771 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0006
2020-03-27 05:43:21,875 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0006 running in uber mode : false
2020-03-27 05:43:21,876 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:43:32,954 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:43:39,977 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 11%
2020-03-27 05:43:40,980 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 56%
2020-03-27 05:43:41,982 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 78%
2020-03-27 05:43:42,985 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:43:42,989 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0006 completed successfully
2020-03-27 05:43:43,010 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5549512
FILE: Number of bytes written=12785465
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=12
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=819360
Total time spent by all reduces in occupied slots (ms)=9938496
Total time spent by all map tasks (ms)=8535
Total time spent by all reduce tasks (ms)=51763
Total vcore-milliseconds taken by all map tasks=8535
Total vcore-milliseconds taken by all reduce tasks=51763
Total megabyte-milliseconds taken by all map tasks=26219520
Total megabyte-milliseconds taken by all reduce tasks=318031872

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5549476
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5549476
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=3000
CPU time spent (ms)=26670
Physical memory (bytes) snapshot=5294559232
Virtual memory (bytes) snapshot=70492495872
Total committed heap usage (bytes)=5306843136

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 05:43:43,032 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:43:43,056 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process: 1

2020-03-27 05:43:43,091 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:43:43,110 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0007
2020-03-27 05:43:43,328 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0007
2020-03-27 05:43:43,331 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0007/
2020-03-27 05:43:43,331 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0007
2020-03-27 05:43:48,375 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0007 running in uber mode : false
2020-03-27 05:43:48,375 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:44:00,426 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:44:06,447 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 33%
2020-03-27 05:44:07,450 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:44:08,456 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0007 completed successfully
2020-03-27 05:44:08,478 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5545864
FILE: Number of bytes written=12780803
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=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=18
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=946176
Total time spent by all reduces in occupied slots (ms)=5567616
Total time spent by all map tasks (ms)=9856
Total time spent by all reduce tasks (ms)=28998
Total vcore-milliseconds taken by all map tasks=9856
Total vcore-milliseconds taken by all reduce tasks=28998
Total megabyte-milliseconds taken by all map tasks=30277632
Total megabyte-milliseconds taken by all reduce tasks=178163712

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5546182
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5546182
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=1945

```
CPU time spent (ms)=31900
Physical memory (bytes) snapshot=3672272896
Virtual memory (bytes) snapshot=7022363036
Total committed heap usage (bytes)=3880255488
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 05:44:08,496 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to
ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:44:08,598 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 05:44:08,628 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 05:44:08,644 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585287572595_0008
2020-03-27 05:44:08,860 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585287572595_0008
2020-03-27 05:44:08,863 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0008/
2020-03-27 05:44:08,863 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585287572595_0008
2020-03-27 05:44:15,912 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0008 running in uber mode : false
2020-03-27 05:44:15,912 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:44:25,948 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:44:32,974 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 11%
2020-03-27 05:44:33,977 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 44%
2020-03-27 05:44:34,980 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 78%
2020-03-27 05:44:35,983 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:44:35,987 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0008 completed successfully
2020-03-27 05:44:36,009 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
File System Counters
  FILE: Number of bytes read=5549512
  FILE: Number of bytes written=12785465
  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=12
  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=9
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=801408
  Total time spent by all reduces in occupied slots (ms)=10040832
  Total time spent by all map tasks (ms)=8348
  Total time spent by all reduce tasks (ms)=52296
```

Total vcore-milliseconds taken by all map tasks=8348
Total vcore-milliseconds taken by all reduce tasks=52296
Total megabyte-milliseconds taken by all map tasks=25645056
Total megabyte-milliseconds taken by all reduce tasks=321306624

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5549476
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5549476
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=3027
CPU time spent (ms)=27340
Physical memory (bytes) snapshot=5297930240
Virtual memory (bytes) snapshot=70599852032
Total committed heap usage (bytes)=5105516544

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 05:44:36,030 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:44:36,174 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:44:36,243 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:44:36,291 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0009
2020-03-27 05:44:36,307 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0009
2020-03-27 05:44:36,310 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0009/
2020-03-27 05:44:36,310 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0009
2020-03-27 05:44:41,353 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0009 running in uber mode : false
2020-03-27 05:44:41,353 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:44:53,393 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:44:58,408 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 11%
2020-03-27 05:44:59,411 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 56%
2020-03-27 05:45:00,414 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:45:01,420 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0009 completed successfully
2020-03-27 05:45:01,443 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5545864
FILE: Number of bytes written=12780803
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=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=18
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=983136
Total time spent by all reduces in occupied slots (ms)=5345856
Total time spent by all map tasks (ms)=10241
Total time spent by all reduce tasks (ms)=27843
Total vcore-milliseconds taken by all map tasks=10241
Total vcore-milliseconds taken by all reduce tasks=27843
Total megabyte-milliseconds taken by all map tasks=31460352
Total megabyte-milliseconds taken by all reduce tasks=171067392

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5546182
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5546182
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=2055
CPU time spent (ms)=32620
Physical memory (bytes) snapshot=3663405056
Virtual memory (bytes) snapshot=70241857536
Total committed heap usage (bytes)=3939500032

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 05:45:01,468 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:45:01,574 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:45:01,603 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:45:01,625 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0010
2020-03-27 05:45:01,651 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):

Submitted application application_1585287572595_0010
2020-03-27 05:45:01,653 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0010/
2020-03-27 05:45:01,653 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0010
2020-03-27 05:45:08,699 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0010 running in uber mode : false
2020-03-27 05:45:08,699 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:45:19,734 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:45:26,753 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 11%
2020-03-27 05:45:28,758 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 56%
2020-03-27 05:45:29,761 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:45:29,765 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0010 completed successfully
2020-03-27 05:45:29,785 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5549512
FILE: Number of bytes written=12785465
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=12
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=862752
Total time spent by all reduces in occupied slots (ms)=10178880
Total time spent by all map tasks (ms)=8987
Total time spent by all reduce tasks (ms)=53015
Total vcore-milliseconds taken by all map tasks=8987
Total vcore-milliseconds taken by all reduce tasks=53015
Total megabyte-milliseconds taken by all map tasks=27608064
Total megabyte-milliseconds taken by all reduce tasks=325724160

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5549476
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5549476
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=2954
CPU time spent (ms)=27690
Physical memory (bytes) snapshot=5101862912
Virtual memory (bytes) snapshot=70508498944
Total committed heap usage (bytes)=5225578496

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 05:45:29,806 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:45:29,829 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:45:29,861 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:45:29,874 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0011
2020-03-27 05:45:29,894 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0011
2020-03-27 05:45:29,897 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0011/
2020-03-27 05:45:29,897 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0011
2020-03-27 05:45:34,992 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0011 running in uber mode : false
2020-03-27 05:45:34,992 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:45:47,102 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:45:52,115 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 22%
2020-03-27 05:45:53,118 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 78%
2020-03-27 05:45:54,121 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:45:54,124 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0011 completed successfully
2020-03-27 05:45:54,149 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5545864
FILE: Number of bytes written=12780803
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=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=18
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=909792
Total time spent by all reduces in occupied slots (ms)=5429184
Total time spent by all map tasks (ms)=9477
Total time spent by all reduce tasks (ms)=28277
Total vcore-milliseconds taken by all map tasks=9477
Total vcore-milliseconds taken by all reduce tasks=28277
Total megabyte-milliseconds taken by all map tasks=29113344
Total megabyte-milliseconds taken by all reduce tasks=173733888

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5546182
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5546182
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=1984
CPU time spent (ms)=31140
Physical memory (bytes) snapshot=3638591488
Virtual memory (bytes) snapshot=70219849728
Total committed heap usage (bytes)=3864002560

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 05:45:54,167 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:45:54,341 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:45:54,369 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:45:54,385 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0012
2020-03-27 05:45:54,394 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0012
2020-03-27 05:45:54,402 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0012/
2020-03-27 05:45:54,402 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0012
2020-03-27 05:46:01,448 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0012 running in uber mode : false
2020-03-27 05:46:01,448 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:46:11,482 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:46:18,504 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 11%
2020-03-27 05:46:19,506 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 22%
2020-03-27 05:46:20,509 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 56%
2020-03-27 05:46:21,511 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:46:21,515 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0012 completed successfully
2020-03-27 05:46:21,535 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5549512
FILE: Number of bytes written=12785465
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=12

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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=804960
Total time spent by all reduces in occupied slots (ms)=10132608
Total time spent by all map tasks (ms)=8385
Total time spent by all reduce tasks (ms)=52774
Total vcore-milliseconds taken by all map tasks=8385
Total vcore-milliseconds taken by all reduce tasks=52774
Total megabyte-milliseconds taken by all map tasks=25758720
Total megabyte-milliseconds taken by all reduce tasks=324243456

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5549476
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5549476
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=2936
CPU time spent (ms)=26800
Physical memory (bytes) snapshot=5117046784
Virtual memory (bytes) snapshot=70551945216
Total committed heap usage (bytes)=5057282048

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 05:46:21,555 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:46:21,619 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:46:21,646 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:46:21,659 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0013
2020-03-27 05:46:21,667 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0013
2020-03-27 05:46:21,676 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0013/
2020-03-27 05:46:21,676 INFO org.apache.hadoop.mapreduce.Job (main): Running job:


```
job_1585287572595_0013
2020-03-27 05:46:26,728 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0013 running in uber mode : false
2020-03-27 05:46:26,728 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:46:38,769 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:46:43,785 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 44%
2020-03-27 05:46:44,789 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 78%
2020-03-27 05:46:45,791 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:46:45,795 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0013 completed successfully
2020-03-27 05:46:45,820 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

File System Counters

```
FILE: Number of bytes read=5545864
FILE: Number of bytes written=12780803
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=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=18
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=893664
Total time spent by all reduces in occupied slots (ms)=5232960
Total time spent by all map tasks (ms)=9309
Total time spent by all reduce tasks (ms)=27255
Total vcore-milliseconds taken by all map tasks=9309
Total vcore-milliseconds taken by all reduce tasks=27255
Total megabyte-milliseconds taken by all map tasks=28597248
Total megabyte-milliseconds taken by all reduce tasks=167454720
```

Map-Reduce Framework

```
Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5546182
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5546182
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=1877
CPU time spent (ms)=31020
Physical memory (bytes) snapshot=3620020224
Virtual memory (bytes) snapshot=70224977920
Total committed heap usage (bytes)=3913285632
```

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 05:46:45,839 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to
ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:46:45,958 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 05:46:45,988 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 05:46:46,002 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585287572595_0014
2020-03-27 05:46:46,024 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585287572595_0014
2020-03-27 05:46:46,026 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0014/
2020-03-27 05:46:46,026 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585287572595_0014
2020-03-27 05:46:52,119 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0014 running in uber mode : false
2020-03-27 05:46:52,119 INFO org.apache.hadoop.mapreduce.Job (main):  map 0% reduce 0%
2020-03-27 05:47:03,207 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 0%
2020-03-27 05:47:10,233 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 11%
2020-03-27 05:47:11,236 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 33%
2020-03-27 05:47:12,238 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 89%
2020-03-27 05:47:13,241 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 100%
2020-03-27 05:47:14,247 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0014 completed successfully
2020-03-27 05:47:14,271 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
File System Counters
  FILE: Number of bytes read=5549512
  FILE: Number of bytes written=12785465
  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=12
  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=9
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=802848
  Total time spent by all reduces in occupied slots (ms)=10304640
  Total time spent by all map tasks (ms)=8363
  Total time spent by all reduce tasks (ms)=53670
  Total vcore-milliseconds taken by all map tasks=8363
  Total vcore-milliseconds taken by all reduce tasks=53670
  Total megabyte-milliseconds taken by all map tasks=25691136
  Total megabyte-milliseconds taken by all reduce tasks=329748480
Map-Reduce Framework
  Map input records=1000000
  Map output records=1000000
  Map output bytes=15888896
```

Map output materialized bytes=5549476
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5549476
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=2970
CPU time spent (ms)=27810
Physical memory (bytes) snapshot=5261045760
Virtual memory (bytes) snapshot=70635339776
Total committed heap usage (bytes)=5170003968

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 05:47:14,288 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:47:14,310 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:47:14,340 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:47:14,353 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0015
2020-03-27 05:47:14,372 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0015
2020-03-27 05:47:14,374 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0015/
2020-03-27 05:47:14,374 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0015
2020-03-27 05:47:19,415 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0015 running in uber mode : false
2020-03-27 05:47:19,415 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:47:31,462 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:47:36,477 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 33%
2020-03-27 05:47:37,480 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:47:38,485 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0015 completed successfully
2020-03-27 05:47:38,508 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5545864
FILE: Number of bytes written=12780803
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=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=18
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=945600
Total time spent by all reduces in occupied slots (ms)=5512320
Total time spent by all map tasks (ms)=9850
Total time spent by all reduce tasks (ms)=28710
Total vcore-milliseconds taken by all map tasks=9850
Total vcore-milliseconds taken by all reduce tasks=28710
Total megabyte-milliseconds taken by all map tasks=30259200
Total megabyte-milliseconds taken by all reduce tasks=176394240

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5546182
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5546182
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=1934
CPU time spent (ms)=32530
Physical memory (bytes) snapshot=3608526848
Virtual memory (bytes) snapshot=70223769600
Total committed heap usage (bytes)=3857711104

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 05:47:38,526 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:47:38,698 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:47:38,720 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:47:38,734 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0016
2020-03-27 05:47:38,759 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0016
2020-03-27 05:47:38,760 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0016/
2020-03-27 05:47:38,761 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0016
2020-03-27 05:47:45,808 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0016 running in uber mode : false
2020-03-27 05:47:45,808 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:47:56,844 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%

```
2020-03-27 05:48:03,872 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 11%
2020-03-27 05:48:04,876 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 56%
2020-03-27 05:48:05,878 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:48:06,885 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0016 completed successfully
2020-03-27 05:48:06,907 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

File System Counters

```
FILE: Number of bytes read=5549512
FILE: Number of bytes written=12785465
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=12
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=818112
Total time spent by all reduces in occupied slots (ms)=9903168
Total time spent by all map tasks (ms)=8522
Total time spent by all reduce tasks (ms)=51579
Total vcore-milliseconds taken by all map tasks=8522
Total vcore-milliseconds taken by all reduce tasks=51579
Total megabyte-milliseconds taken by all map tasks=26179584
Total megabyte-milliseconds taken by all reduce tasks=316901376
```

Map-Reduce Framework

```
Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5549476
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5549476
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=2916
CPU time spent (ms)=27630
Physical memory (bytes) snapshot=5216301056
Virtual memory (bytes) snapshot=70518358016
Total committed heap usage (bytes)=5290590208
```

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 05:48:06,927 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:48:06,985 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:48:07,011 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:48:07,024 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0017
2020-03-27 05:48:07,036 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0017
2020-03-27 05:48:07,043 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0017/
2020-03-27 05:48:07,043 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0017
2020-03-27 05:48:13,084 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0017 running in uber mode : false
2020-03-27 05:48:13,084 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:48:25,136 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:48:31,154 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 56%
2020-03-27 05:48:32,157 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:48:32,160 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0017 completed successfully
2020-03-27 05:48:32,186 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5545864
FILE: Number of bytes written=12780803
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=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=18
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=994560
Total time spent by all reduces in occupied slots (ms)=5376384
Total time spent by all map tasks (ms)=10360
Total time spent by all reduce tasks (ms)=28002
Total vcore-milliseconds taken by all map tasks=10360
Total vcore-milliseconds taken by all reduce tasks=28002
Total megabyte-milliseconds taken by all map tasks=31825920
Total megabyte-milliseconds taken by all reduce tasks=172044288

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5546182
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5546182
Reduce input records=999001

```
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=1952
CPU time spent (ms)=32250
Physical memory (bytes) snapshot=3612024832
Virtual memory (bytes) snapshot=70199857152
Total committed heap usage (bytes)=3825729536

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 05:48:32,206 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to
ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:48:32,298 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-03-27 05:48:32,321 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 05:48:32,334 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585287572595_0018
2020-03-27 05:48:32,548 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585287572595_0018
2020-03-27 05:48:32,550 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0018/
2020-03-27 05:48:32,550 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585287572595_0018
2020-03-27 05:48:39,592 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0018 running in uber mode : false
2020-03-27 05:48:39,592 INFO org.apache.hadoop.mapreduce.Job (main):  map 0% reduce 0%
2020-03-27 05:48:49,625 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 0%
2020-03-27 05:48:56,644 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 11%
2020-03-27 05:48:58,650 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 56%
2020-03-27 05:48:59,652 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 100%
2020-03-27 05:48:59,656 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0018 completed successfully
2020-03-27 05:48:59,677 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
File System Counters
  FILE: Number of bytes read=5549512
  FILE: Number of bytes written=12785465
  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=12
  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=9
```

Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=812256
Total time spent by all reduces in occupied slots (ms)=10185024
Total time spent by all map tasks (ms)=8461
Total time spent by all reduce tasks (ms)=53047
Total vcore-milliseconds taken by all map tasks=8461
Total vcore-milliseconds taken by all reduce tasks=53047
Total megabyte-milliseconds taken by all map tasks=25992192
Total megabyte-milliseconds taken by all reduce tasks=325920768

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5549476
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5549476
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=2862
CPU time spent (ms)=27550
Physical memory (bytes) snapshot=5280112640
Virtual memory (bytes) snapshot=70548627456
Total committed heap usage (bytes)=5247598592

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 05:48:59,697 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:48:59,726 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 1
2020-03-27 05:48:59,759 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:1
2020-03-27 05:48:59,771 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1585287572595_0019
2020-03-27 05:48:59,781 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1585287572595_0019
2020-03-27 05:48:59,787 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0019/
2020-03-27 05:48:59,787 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1585287572595_0019
2020-03-27 05:49:04,828 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0019 running in uber mode : false
2020-03-27 05:49:04,828 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:49:16,880 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:49:22,898 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 44%
2020-03-27 05:49:23,901 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 89%
2020-03-27 05:49:24,903 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:49:24,907 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1585287572595_0019 completed successfully
2020-03-27 05:49:24,929 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55

File System Counters

FILE: Number of bytes read=5545864
FILE: Number of bytes written=12780803
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=29
HDFS: Number of large read operations=0
HDFS: Number of write operations=18
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=972288
Total time spent by all reduces in occupied slots (ms)=5309760
Total time spent by all map tasks (ms)=10128
Total time spent by all reduce tasks (ms)=27655
Total vcore-milliseconds taken by all map tasks=10128
Total vcore-milliseconds taken by all reduce tasks=27655
Total megabyte-milliseconds taken by all map tasks=31113216
Total megabyte-milliseconds taken by all reduce tasks=169912320

Map-Reduce Framework

Map input records=1000000
Map output records=1000000
Map output bytes=15884009
Map output materialized bytes=5546182
Input split bytes=104
Combine input records=1000000
Combine output records=999001
Reduce input groups=999001
Reduce shuffle bytes=5546182
Reduce input records=999001
Reduce output records=999001
Spilled Records=1998002
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
GC time elapsed (ms)=1988
CPU time spent (ms)=33170
Physical memory (bytes) snapshot=3584942080
Virtual memory (bytes) snapshot=70233837568
Total committed heap usage (bytes)=3603955712

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 05:49:24,953 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-12-118.ec2.internal/172.31.12.118:8032
2020-03-27 05:49:25,090 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process: 1

```
2020-03-27 05:49:25,126 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:1
2020-03-27 05:49:25,167 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1585287572595_0020
2020-03-27 05:49:26,810 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1585287572595_0020
2020-03-27 05:49:26,815 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-12-118.ec2.internal:20888/proxy/application_1585287572595_0020/
2020-03-27 05:49:26,815 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1585287572595_0020
2020-03-27 05:49:34,861 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0020 running in uber mode : false
2020-03-27 05:49:34,861 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-03-27 05:49:44,893 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-03-27 05:49:51,918 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 11%
2020-03-27 05:49:53,923 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 56%
2020-03-27 05:49:54,926 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-03-27 05:49:54,929 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1585287572595_0020 completed successfully
2020-03-27 05:49:54,950 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

File System Counters

```
FILE: Number of bytes read=5549512
FILE: Number of bytes written=12785485
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=12
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=9
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=772224
Total time spent by all reduces in occupied slots (ms)=10102656
Total time spent by all map tasks (ms)=8044
Total time spent by all reduce tasks (ms)=52618
Total vcore-milliseconds taken by all map tasks=8044
Total vcore-milliseconds taken by all reduce tasks=52618
Total megabyte-milliseconds taken by all map tasks=24711168
Total megabyte-milliseconds taken by all reduce tasks=323284992
```

Map-Reduce Framework

```
Map input records=1000000
Map output records=1000000
Map output bytes=15888896
Map output materialized bytes=5549476
Input split bytes=105
Combine input records=0
Combine output records=0
Reduce input groups=1000000
Reduce shuffle bytes=5549476
Reduce input records=1000000
Reduce output records=1000000
Spilled Records=2000000
Shuffled Maps =9
Failed Shuffles=0
Merged Map outputs=9
```

```
GC time elapsed (ms)=2804
CPU time spent (ms)=26980
Physical memory (bytes) snapshot=5262004224
Virtual memory (bytes) snapshot=70534561792
Total committed heap usage (bytes)=5155848192
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
```