

```
2020-02-15 01:51:25,320 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-79-141.ec2.internal/172.31.79.141:8032
2020-02-15 01:51:27,177 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-02-15 01:51:27,189 INFO com.hadoop.compression.lzo.GPLNativeCodeLoader (main): Loaded
native gpl library
2020-02-15 01:51:27,192 INFO com.hadoop.compression.lzo.LzoCodec (main): Successfully loaded
& initialized native-lzo library [hadoop-lzo rev 5f788d5e8f90539ee331702c753fa250727128f4]
2020-02-15 01:51:27,496 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:20
2020-02-15 01:51:27,616 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1581731370893_0001
2020-02-15 01:51:28,167 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1581731370893_0001
2020-02-15 01:51:28,252 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-79-141.ec2.internal:20888/proxy/application_1581731370893_0001/
2020-02-15 01:51:28,253 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1581731370893_0001
2020-02-15 01:51:34,332 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581731370893_0001 running in uber mode : false
2020-02-15 01:51:34,333 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-02-15 01:51:46,428 INFO org.apache.hadoop.mapreduce.Job (main): map 5% reduce 0%
2020-02-15 01:51:47,435 INFO org.apache.hadoop.mapreduce.Job (main): map 20% reduce 0%
2020-02-15 01:51:53,457 INFO org.apache.hadoop.mapreduce.Job (main): map 25% reduce 0%
2020-02-15 01:51:54,460 INFO org.apache.hadoop.mapreduce.Job (main): map 31% reduce 0%
2020-02-15 01:51:55,465 INFO org.apache.hadoop.mapreduce.Job (main): map 65% reduce 0%
2020-02-15 01:51:56,471 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-02-15 01:51:59,484 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 18%
2020-02-15 01:52:01,490 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 27%
2020-02-15 01:52:02,494 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 64%
2020-02-15 01:52:04,500 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 73%
2020-02-15 01:52:07,510 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 82%
2020-02-15 01:52:08,513 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 91%
2020-02-15 01:52:11,521 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 99%
2020-02-15 01:52:14,532 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 01:52:14,537 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581731370893_0001 completed successfully
2020-02-15 01:52:14,638 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 56
File System Counters
    FILE: Number of bytes read=1585832
    FILE: Number of bytes written=8547769
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=2240
    HDFS: Number of bytes written=1092272771
    HDFS: Number of read operations=73
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=22
    S3: Number of bytes read=1319498021
    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 map tasks=1
    Killed reduce tasks=1
    Launched map tasks=20
    Launched reduce tasks=12
    Data-local map tasks=20
    Total time spent by all maps in occupied slots (ms)=32944896
    Total time spent by all reduces in occupied slots (ms)=19432896
    Total time spent by all map tasks (ms)=343176
    Total time spent by all reduce tasks (ms)=101213
```

Total vcore-milliseconds taken by all map tasks=343176  
Total vcore-milliseconds taken by all reduce tasks=101213  
Total megabyte-milliseconds taken by all map tasks=1054236672  
Total megabyte-milliseconds taken by all reduce tasks=621852672

#### Map-Reduce Framework

Map input records=85331845  
Map output records=258402  
Map output bytes=3809721  
Map output materialized bytes=1696355  
Input split bytes=2240  
Combine input records=0  
Combine output records=0  
Reduce input groups=9999  
Reduce shuffle bytes=1696355  
Reduce input records=258402  
Reduce output records=73697208  
Spilled Records=516804  
Shuffled Maps =220  
Failed Shuffles=0  
Merged Map outputs=220  
GC time elapsed (ms)=11932  
CPU time spent (ms)=320970  
Physical memory (bytes) snapshot=29458456576  
Virtual memory (bytes) snapshot=173154861056  
Total committed heap usage (bytes)=29601824768

#### 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=1319498021

#### File Output Format Counters

Bytes Written=1092272771

2020-02-15 01:52:14,695 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-79-141.ec2.internal/172.31.79.141:8032  
2020-02-15 01:52:14,892 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 12  
2020-02-15 01:52:14,947 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:34  
2020-02-15 01:52:14,984 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job\_1581731370893\_0002  
2020-02-15 01:52:15,207 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application\_1581731370893\_0002  
2020-02-15 01:52:15,210 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-79-141.ec2.internal:20888/proxy/application\_1581731370893\_0002/  
2020-02-15 01:52:15,210 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job\_1581731370893\_0002  
2020-02-15 01:52:20,260 INFO org.apache.hadoop.mapreduce.Job (main): Job job\_1581731370893\_0002 running in uber mode : false  
2020-02-15 01:52:20,260 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%  
2020-02-15 01:52:31,328 INFO org.apache.hadoop.mapreduce.Job (main): map 3% reduce 0%  
2020-02-15 01:52:37,359 INFO org.apache.hadoop.mapreduce.Job (main): map 4% reduce 0%  
2020-02-15 01:52:39,366 INFO org.apache.hadoop.mapreduce.Job (main): map 14% reduce 0%  
2020-02-15 01:52:40,370 INFO org.apache.hadoop.mapreduce.Job (main): map 47% reduce 0%  
2020-02-15 01:52:41,374 INFO org.apache.hadoop.mapreduce.Job (main): map 61% reduce 0%  
2020-02-15 01:52:46,391 INFO org.apache.hadoop.mapreduce.Job (main): map 62% reduce 0%  
2020-02-15 01:52:49,400 INFO org.apache.hadoop.mapreduce.Job (main): map 63% reduce 0%  
2020-02-15 01:52:52,410 INFO org.apache.hadoop.mapreduce.Job (main): map 66% reduce 0%  
2020-02-15 01:52:53,414 INFO org.apache.hadoop.mapreduce.Job (main): map 69% reduce 0%  
2020-02-15 01:52:55,421 INFO org.apache.hadoop.mapreduce.Job (main): map 70% reduce 0%  
2020-02-15 01:52:56,424 INFO org.apache.hadoop.mapreduce.Job (main): map 75% reduce 0%  
2020-02-15 01:52:57,426 INFO org.apache.hadoop.mapreduce.Job (main): map 75% reduce 6%

```

2020-02-15 01:52:58,429 INFO org.apache.hadoop.mapreduce.Job (main): map 82% reduce 8%
2020-02-15 01:52:59,433 INFO org.apache.hadoop.mapreduce.Job (main): map 86% reduce 8%
2020-02-15 01:53:02,441 INFO org.apache.hadoop.mapreduce.Job (main): map 87% reduce 8%
2020-02-15 01:53:03,443 INFO org.apache.hadoop.mapreduce.Job (main): map 87% reduce 9%
2020-02-15 01:53:04,448 INFO org.apache.hadoop.mapreduce.Job (main): map 90% reduce 9%
2020-02-15 01:53:07,456 INFO org.apache.hadoop.mapreduce.Job (main): map 91% reduce 9%
2020-02-15 01:53:08,460 INFO org.apache.hadoop.mapreduce.Job (main): map 93% reduce 11%
2020-02-15 01:53:09,463 INFO org.apache.hadoop.mapreduce.Job (main): map 93% reduce 12%
2020-02-15 01:53:10,466 INFO org.apache.hadoop.mapreduce.Job (main): map 94% reduce 13%
2020-02-15 01:53:11,468 INFO org.apache.hadoop.mapreduce.Job (main): map 95% reduce 13%
2020-02-15 01:53:16,479 INFO org.apache.hadoop.mapreduce.Job (main): map 96% reduce 13%
2020-02-15 01:53:19,486 INFO org.apache.hadoop.mapreduce.Job (main): map 97% reduce 16%
2020-02-15 01:53:21,491 INFO org.apache.hadoop.mapreduce.Job (main): map 98% reduce 16%
2020-02-15 01:53:22,494 INFO org.apache.hadoop.mapreduce.Job (main): map 99% reduce 19%
2020-02-15 01:53:24,499 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 22%
2020-02-15 01:53:26,504 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 28%
2020-02-15 01:53:27,507 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 31%
2020-02-15 01:53:28,510 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 40%
2020-02-15 01:53:30,515 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 44%
2020-02-15 01:53:31,518 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 47%
2020-02-15 01:53:32,520 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 53%
2020-02-15 01:53:33,523 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 58%
2020-02-15 01:53:34,526 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 63%
2020-02-15 01:53:36,531 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 65%
2020-02-15 01:53:37,533 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 68%
2020-02-15 01:53:38,537 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 86%
2020-02-15 01:53:39,540 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 97%
2020-02-15 01:53:40,547 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 98%
2020-02-15 01:53:41,551 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 99%
2020-02-15 01:53:43,557 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 01:53:43,560 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581731370893_0002 completed successfully
2020-02-15 01:53:43,585 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 57

```

#### File System Counters

```

FILE: Number of bytes read=1788326016
FILE: Number of bytes written=2700957447
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=1092473593
HDFS: Number of bytes written=20291544
HDFS: Number of read operations=115
HDFS: Number of large read operations=0
HDFS: Number of write operations=22
S3: Number of bytes read=1319526205
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 map tasks=2
Killed reduce tasks=1
Launched map tasks=34
Launched reduce tasks=12
Data-local map tasks=20
Rack-local map tasks=14
Total time spent by all maps in occupied slots (ms)=72334176
Total time spent by all reduces in occupied slots (ms)=82231680
Total time spent by all map tasks (ms)=753481
Total time spent by all reduce tasks (ms)=428290
Total vcore-milliseconds taken by all map tasks=753481
Total vcore-milliseconds taken by all reduce tasks=428290
Total megabyte-milliseconds taken by all map tasks=2314693632
Total megabyte-milliseconds taken by all reduce tasks=2631413760

```

#### Map-Reduce Framework

Map input records=159029053  
Map output records=147652818  
Map output bytes=2916174874  
Map output materialized bytes=916661949  
Input split bytes=4214  
Combine input records=0  
Combine output records=0  
Reduce input groups=9999  
Reduce shuffle bytes=916661949  
Reduce input records=147652818  
Reduce output records=1560888  
Spilled Records=433927460  
Shuffled Maps =374  
Failed Shuffles=0  
Merged Map outputs=374  
GC time elapsed (ms)=19101  
CPU time spent (ms)=815270  
Physical memory (bytes) snapshot=55166971904  
Virtual memory (bytes) snapshot=238177067008  
Total committed heap usage (bytes)=55252090880

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=2411995584

File Output Format Counters

Bytes Written=20291544

2020-02-15 01:53:43,645 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-79-141.ec2.internal/172.31.79.141:8032  
2020-02-15 01:53:43,773 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 11  
2020-02-15 01:53:43,816 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:11  
2020-02-15 01:53:43,836 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job\_1581731370893\_0003  
2020-02-15 01:53:43,847 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application\_1581731370893\_0003  
2020-02-15 01:53:43,864 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-79-141.ec2.internal:20888/proxy/application\_1581731370893\_0003/  
2020-02-15 01:53:43,865 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job\_1581731370893\_0003  
2020-02-15 01:53:49,989 INFO org.apache.hadoop.mapreduce.Job (main): Job job\_1581731370893\_0003 running in uber mode : false  
2020-02-15 01:53:49,990 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%  
2020-02-15 01:53:58,042 INFO org.apache.hadoop.mapreduce.Job (main): map 18% reduce 0%  
2020-02-15 01:53:59,045 INFO org.apache.hadoop.mapreduce.Job (main): map 36% reduce 0%  
2020-02-15 01:54:01,053 INFO org.apache.hadoop.mapreduce.Job (main): map 64% reduce 0%  
2020-02-15 01:54:03,059 INFO org.apache.hadoop.mapreduce.Job (main): map 73% reduce 0%  
2020-02-15 01:54:04,062 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%  
2020-02-15 01:54:05,065 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 9%  
2020-02-15 01:54:06,068 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 55%  
2020-02-15 01:54:07,072 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 73%  
2020-02-15 01:54:08,074 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 82%  
2020-02-15 01:54:10,080 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 91%  
2020-02-15 01:54:11,082 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%  
2020-02-15 01:54:11,092 INFO org.apache.hadoop.mapreduce.Job (main): Job job\_1581731370893\_0003 completed successfully  
2020-02-15 01:54:11,115 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 56

File System Counters

FILE: Number of bytes read=1111001  
FILE: Number of bytes written=5959616

FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=20293183  
HDFS: Number of bytes written=0  
HDFS: Number of read operations=22  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=0  
S3: Number of bytes read=0  
S3: Number of bytes written=44  
S3: Number of read operations=0  
S3: Number of large read operations=0  
S3: Number of write operations=0

#### Job Counters

Killed map tasks=1  
Launched map tasks=11  
Launched reduce tasks=11  
Data-local map tasks=9  
Rack-local map tasks=2  
Total time spent by all maps in occupied slots (ms)=8832960  
Total time spent by all reduces in occupied slots (ms)=11947584  
Total time spent by all map tasks (ms)=92010  
Total time spent by all reduce tasks (ms)=62227  
Total vcore-milliseconds taken by all map tasks=92010  
Total vcore-milliseconds taken by all reduce tasks=62227  
Total megabyte-milliseconds taken by all map tasks=282654720  
Total megabyte-milliseconds taken by all reduce tasks=382322688

#### Map-Reduce Framework

Map input records=1560888  
Map output records=1560888  
Map output bytes=20291544  
Map output materialized bytes=1112749  
Input split bytes=1639  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=1112749  
Reduce input records=1560888  
Reduce output records=1  
Spilled Records=3121776  
Shuffled Maps =121  
Failed Shuffles=0  
Merged Map outputs=121  
GC time elapsed (ms)=7309  
CPU time spent (ms)=39460  
Physical memory (bytes) snapshot=12904456192  
Virtual memory (bytes) snapshot=131592196096  
Total committed heap usage (bytes)=12528386048

#### 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=20291544

#### File Output Format Counters

Bytes Written=44