

```
2020-02-15 06:08:48,972 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-68-231.ec2.internal/172.31.68.231:8032
2020-02-15 06:08:51,072 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-02-15 06:08:51,084 INFO com.hadoop.compression.lzo.GPLNativeCodeLoader (main): Loaded
native gpl library
2020-02-15 06:08:51,087 INFO com.hadoop.compression.lzo.LzoCodec (main): Successfully loaded
& initialized native-lzo library [hadoop-lzo rev 5f788d5e8f90539ee331702c753fa250727128f4]
2020-02-15 06:08:51,467 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:20
2020-02-15 06:08:51,590 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1581746851051_0001
2020-02-15 06:08:52,051 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1581746851051_0001
2020-02-15 06:08:52,149 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-68-231.ec2.internal:20888/proxy/application_1581746851051_0001/
2020-02-15 06:08:52,150 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1581746851051_0001
2020-02-15 06:08:58,235 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581746851051_0001 running in uber mode : false
2020-02-15 06:08:58,236 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-02-15 06:09:11,346 INFO org.apache.hadoop.mapreduce.Job (main): map 40% reduce 0%
2020-02-15 06:09:18,370 INFO org.apache.hadoop.mapreduce.Job (main): map 51% reduce 0%
2020-02-15 06:09:19,374 INFO org.apache.hadoop.mapreduce.Job (main): map 68% reduce 0%
2020-02-15 06:09:20,380 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-02-15 06:09:22,388 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-02-15 06:09:23,392 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 23%
2020-02-15 06:09:24,395 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 46%
2020-02-15 06:09:25,399 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-02-15 06:09:26,402 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 62%
2020-02-15 06:09:27,406 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 77%
2020-02-15 06:09:28,409 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 85%
2020-02-15 06:09:29,412 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 92%
2020-02-15 06:09:36,433 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 99%
2020-02-15 06:09:38,439 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 06:09:38,443 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581746851051_0001 completed successfully
2020-02-15 06:09:38,542 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 56
File System Counters
    FILE: Number of bytes read=1586274
    FILE: Number of bytes written=8890472
    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=79
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=26
    S3: Number of bytes read=1319501573
    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=13
    Data-local map tasks=20
    Total time spent by all maps in occupied slots (ms)=29968416
    Total time spent by all reduces in occupied slots (ms)=23177472
    Total time spent by all map tasks (ms)=312171
    Total time spent by all reduce tasks (ms)=120716
```

Total vcore-milliseconds taken by all map tasks=312171  
Total vcore-milliseconds taken by all reduce tasks=120716  
Total megabyte-milliseconds taken by all map tasks=958989312  
Total megabyte-milliseconds taken by all reduce tasks=741679104

Map-Reduce Framework

Map input records=85331845  
Map output records=258402  
Map output bytes=3809721  
Map output materialized bytes=1698337  
Input split bytes=2240  
Combine input records=0  
Combine output records=0  
Reduce input groups=9999  
Reduce shuffle bytes=1698337  
Reduce input records=258402  
Reduce output records=73697208  
Spilled Records=516804  
Shuffled Maps =260  
Failed Shuffles=0  
Merged Map outputs=260  
GC time elapsed (ms)=10713  
CPU time spent (ms)=315940  
Physical memory (bytes) snapshot=31484452864  
Virtual memory (bytes) snapshot=187754172416  
Total committed heap usage (bytes)=31585206272

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=1319501573

File Output Format Counters

Bytes Written=1092272771

2020-02-15 06:09:38,585 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-68-231.ec2.internal/172.31.68.231:8032  
2020-02-15 06:09:38,877 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 14  
2020-02-15 06:09:38,931 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:35  
2020-02-15 06:09:38,964 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job\_1581746851051\_0002  
2020-02-15 06:09:39,001 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application\_1581746851051\_0002  
2020-02-15 06:09:39,012 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-68-231.ec2.internal:20888/proxy/application\_1581746851051\_0002/  
2020-02-15 06:09:39,013 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job\_1581746851051\_0002  
2020-02-15 06:09:44,068 INFO org.apache.hadoop.mapreduce.Job (main): Job job\_1581746851051\_0002 running in uber mode : false  
2020-02-15 06:09:44,068 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%  
2020-02-15 06:09:56,139 INFO org.apache.hadoop.mapreduce.Job (main): map 6% reduce 0%  
2020-02-15 06:10:02,164 INFO org.apache.hadoop.mapreduce.Job (main): map 14% reduce 0%  
2020-02-15 06:10:03,167 INFO org.apache.hadoop.mapreduce.Job (main): map 23% reduce 0%  
2020-02-15 06:10:04,170 INFO org.apache.hadoop.mapreduce.Job (main): map 39% reduce 0%  
2020-02-15 06:10:05,175 INFO org.apache.hadoop.mapreduce.Job (main): map 62% reduce 0%  
2020-02-15 06:10:10,190 INFO org.apache.hadoop.mapreduce.Job (main): map 64% reduce 0%  
2020-02-15 06:10:11,193 INFO org.apache.hadoop.mapreduce.Job (main): map 66% reduce 0%  
2020-02-15 06:10:12,196 INFO org.apache.hadoop.mapreduce.Job (main): map 70% reduce 0%  
2020-02-15 06:10:15,205 INFO org.apache.hadoop.mapreduce.Job (main): map 71% reduce 0%  
2020-02-15 06:10:16,208 INFO org.apache.hadoop.mapreduce.Job (main): map 77% reduce 0%  
2020-02-15 06:10:17,213 INFO org.apache.hadoop.mapreduce.Job (main): map 80% reduce 0%  
2020-02-15 06:10:19,220 INFO org.apache.hadoop.mapreduce.Job (main): map 82% reduce 0%

```
2020-02-15 06:10:20,222 INFO org.apache.hadoop.mapreduce.Job (main): map 84% reduce 4%
2020-02-15 06:10:21,225 INFO org.apache.hadoop.mapreduce.Job (main): map 88% reduce 6%
2020-02-15 06:10:22,228 INFO org.apache.hadoop.mapreduce.Job (main): map 92% reduce 6%
2020-02-15 06:10:23,230 INFO org.apache.hadoop.mapreduce.Job (main): map 93% reduce 8%
2020-02-15 06:10:24,232 INFO org.apache.hadoop.mapreduce.Job (main): map 93% reduce 10%
2020-02-15 06:10:25,235 INFO org.apache.hadoop.mapreduce.Job (main): map 94% reduce 10%
2020-02-15 06:10:27,240 INFO org.apache.hadoop.mapreduce.Job (main): map 95% reduce 11%
2020-02-15 06:10:29,245 INFO org.apache.hadoop.mapreduce.Job (main): map 96% reduce 11%
2020-02-15 06:10:31,250 INFO org.apache.hadoop.mapreduce.Job (main): map 96% reduce 13%
2020-02-15 06:10:33,255 INFO org.apache.hadoop.mapreduce.Job (main): map 97% reduce 13%
2020-02-15 06:10:35,260 INFO org.apache.hadoop.mapreduce.Job (main): map 98% reduce 18%
2020-02-15 06:10:37,264 INFO org.apache.hadoop.mapreduce.Job (main): map 98% reduce 21%
2020-02-15 06:10:40,271 INFO org.apache.hadoop.mapreduce.Job (main): map 99% reduce 24%
2020-02-15 06:10:42,275 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 24%
2020-02-15 06:10:44,280 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 27%
2020-02-15 06:10:45,282 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 28%
2020-02-15 06:10:46,285 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 30%
2020-02-15 06:10:47,288 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 37%
2020-02-15 06:10:48,290 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 40%
2020-02-15 06:10:49,293 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 46%
2020-02-15 06:10:50,295 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 52%
2020-02-15 06:10:51,298 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 64%
2020-02-15 06:10:52,303 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 67%
2020-02-15 06:10:53,306 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 73%
2020-02-15 06:10:55,310 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 77%
2020-02-15 06:10:56,313 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 89%
2020-02-15 06:10:57,316 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 92%
2020-02-15 06:10:58,319 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 99%
2020-02-15 06:11:00,324 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 06:11:00,329 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581746851051_0002 completed successfully
2020-02-15 06:11:00,351 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 57
```

#### File System Counters

```
FILE: Number of bytes read=1804018093
FILE: Number of bytes written=2715573747
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=1092408198
HDFS: Number of bytes written=20291544
HDFS: Number of read operations=124
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=1319523245
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=35
Launched reduce tasks=14
Data-local map tasks=20
Rack-local map tasks=15
Total time spent by all maps in occupied slots (ms)=71902944
Total time spent by all reduces in occupied slots (ms)=93371904
Total time spent by all map tasks (ms)=748989
Total time spent by all reduce tasks (ms)=486312
Total vcore-milliseconds taken by all map tasks=748989
Total vcore-milliseconds taken by all reduce tasks=486312
Total megabyte-milliseconds taken by all map tasks=2300894208
Total megabyte-milliseconds taken by all reduce tasks=2987900928
```

#### Map-Reduce Framework

```
Map input records=159029053
```

Map output records=147652818  
Map output bytes=2916174874  
Map output materialized bytes=916726979  
Input split bytes=4355  
Combine input records=0  
Combine output records=0  
Reduce input groups=9999  
Reduce shuffle bytes=916726979  
Reduce input records=147652818  
Reduce output records=1560888  
Spilled Records=435780540  
Shuffled Maps =455  
Failed Shuffles=0  
Merged Map outputs=455  
GC time elapsed (ms)=18793  
CPU time spent (ms)=820420  
Physical memory (bytes) snapshot=58329157632  
Virtual memory (bytes) snapshot=257183797248  
Total committed heap usage (bytes)=60731949056

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=2411927088

File Output Format Counters

Bytes Written=20291544

2020-02-15 06:11:00,410 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-68-231.ec2.internal/172.31.68.231:8032  
2020-02-15 06:11:00,555 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 13  
2020-02-15 06:11:00,602 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:13  
2020-02-15 06:11:00,621 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job\_1581746851051\_0003  
2020-02-15 06:11:00,633 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application\_1581746851051\_0003  
2020-02-15 06:11:00,638 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-68-231.ec2.internal:20888/proxy/application\_1581746851051\_0003/  
2020-02-15 06:11:00,638 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job\_1581746851051\_0003  
2020-02-15 06:11:06,694 INFO org.apache.hadoop.mapreduce.Job (main): Job job\_1581746851051\_0003 running in uber mode : false  
2020-02-15 06:11:06,694 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%  
2020-02-15 06:11:12,719 INFO org.apache.hadoop.mapreduce.Job (main): map 8% reduce 0%  
2020-02-15 06:11:19,744 INFO org.apache.hadoop.mapreduce.Job (main): map 77% reduce 0%  
2020-02-15 06:11:20,747 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%  
2020-02-15 06:11:21,750 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%  
2020-02-15 06:11:26,763 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 85%  
2020-02-15 06:11:27,766 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%  
2020-02-15 06:11:28,773 INFO org.apache.hadoop.mapreduce.Job (main): Job job\_1581746851051\_0003 completed successfully  
2020-02-15 06:11:28,796 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 56

File System Counters

FILE: Number of bytes read=1111041  
FILE: Number of bytes written=6640298  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=20293481  
HDFS: Number of bytes written=0  
HDFS: Number of read operations=26

```
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=13
  Launched reduce tasks=13
  Data-local map tasks=11
  Rack-local map tasks=2
  Total time spent by all maps in occupied slots (ms)=12898176
  Total time spent by all reduces in occupied slots (ms)=14774400
  Total time spent by all map tasks (ms)=134356
  Total time spent by all reduce tasks (ms)=76950
  Total vcore-milliseconds taken by all map tasks=134356
  Total vcore-milliseconds taken by all reduce tasks=76950
  Total megabyte-milliseconds taken by all map tasks=412741632
  Total megabyte-milliseconds taken by all reduce tasks=472780800
Map-Reduce Framework
  Map input records=1560888
  Map output records=1560888
  Map output bytes=20291544
  Map output materialized bytes=1113509
  Input split bytes=1937
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=1113509
  Reduce input records=1560888
  Reduce output records=1
  Spilled Records=3121776
  Shuffled Maps =169
  Failed Shuffles=0
  Merged Map outputs=169
  GC time elapsed (ms)=9802
  CPU time spent (ms)=48490
  Physical memory (bytes) snapshot=15426416640
  Virtual memory (bytes) snapshot=155457376256
  Total committed heap usage (bytes)=15148777472
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
```