

```
2020-02-15 01:39:58,742 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-69-156.ec2.internal/172.31.69.156:8032
2020-02-15 01:40:00,528 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-02-15 01:40:00,539 INFO com.hadoop.compression.lzo.GPLNativeCodeLoader (main): Loaded
native gpl library
2020-02-15 01:40:00,541 INFO com.hadoop.compression.lzo.LzoCodec (main): Successfully loaded
& initialized native-lzo library [hadoop-lzo rev 5f788d5e8f90539ee331702c753fa250727128f4]
2020-02-15 01:40:00,862 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:20
2020-02-15 01:40:00,982 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1581730657860_0001
2020-02-15 01:40:01,491 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1581730657860_0001
2020-02-15 01:40:01,588 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-69-156.ec2.internal:20888/proxy/application_1581730657860_0001/
2020-02-15 01:40:01,588 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1581730657860_0001
2020-02-15 01:40:07,671 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581730657860_0001 running in uber mode : false
2020-02-15 01:40:07,672 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-02-15 01:40:19,778 INFO org.apache.hadoop.mapreduce.Job (main): map 5% reduce 0%
2020-02-15 01:40:20,782 INFO org.apache.hadoop.mapreduce.Job (main): map 20% reduce 0%
2020-02-15 01:40:26,804 INFO org.apache.hadoop.mapreduce.Job (main): map 25% reduce 0%
2020-02-15 01:40:27,807 INFO org.apache.hadoop.mapreduce.Job (main): map 53% reduce 0%
2020-02-15 01:40:28,811 INFO org.apache.hadoop.mapreduce.Job (main): map 90% reduce 0%
2020-02-15 01:40:29,816 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 18%
2020-02-15 01:40:30,819 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 27%
2020-02-15 01:40:32,825 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 73%
2020-02-15 01:40:33,829 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 01:40:33,834 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581730657860_0001 completed successfully
2020-02-15 01:40:33,938 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

#### File System Counters

```
FILE: Number of bytes read=38474
FILE: Number of bytes written=5349877
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=4451828
HDFS: Number of read operations=73
HDFS: Number of large read operations=0
HDFS: Number of write operations=22
S3: Number of bytes read=1319517120
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
Launched map tasks=20
Launched reduce tasks=11
Data-local map tasks=20
Total time spent by all maps in occupied slots (ms)=32913600
Total time spent by all reduces in occupied slots (ms)=9398016
Total time spent by all map tasks (ms)=342850
Total time spent by all reduce tasks (ms)=48948
Total vcore-milliseconds taken by all map tasks=342850
Total vcore-milliseconds taken by all reduce tasks=48948
Total megabyte-milliseconds taken by all map tasks=1053235200
Total megabyte-milliseconds taken by all reduce tasks=300736512
```

#### Map-Reduce Framework

```
Map input records=85331845
```

Map output records=5680  
Map output bytes=63486  
Map output materialized bytes=45858  
Input split bytes=2240  
Combine input records=0  
Combine output records=0  
Reduce input groups=999  
Reduce shuffle bytes=45858  
Reduce input records=5680  
Reduce output records=402114  
Spilled Records=11360  
Shuffled Maps =220  
Failed Shuffles=0  
Merged Map outputs=220  
GC time elapsed (ms)=10814  
CPU time spent (ms)=254070  
Physical memory (bytes) snapshot=26068049920  
Virtual memory (bytes) snapshot=173278744576  
Total committed heap usage (bytes)=26253197312

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=1319517120

File Output Format Counters

Bytes Written=4451828

2020-02-15 01:40:33,995 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-69-156.ec2.internal/172.31.69.156:8032  
2020-02-15 01:40:34,168 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 12  
2020-02-15 01:40:34,235 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:31  
2020-02-15 01:40:34,268 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job\_1581730657860\_0002  
2020-02-15 01:40:34,299 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application\_1581730657860\_0002  
2020-02-15 01:40:34,310 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-69-156.ec2.internal:20888/proxy/application\_1581730657860\_0002/  
2020-02-15 01:40:34,310 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job\_1581730657860\_0002  
2020-02-15 01:40:39,363 INFO org.apache.hadoop.mapreduce.Job (main): Job job\_1581730657860\_0002 running in uber mode : false  
2020-02-15 01:40:39,363 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%  
2020-02-15 01:40:50,420 INFO org.apache.hadoop.mapreduce.Job (main): map 3% reduce 0%  
2020-02-15 01:40:51,423 INFO org.apache.hadoop.mapreduce.Job (main): map 13% reduce 0%  
2020-02-15 01:40:56,453 INFO org.apache.hadoop.mapreduce.Job (main): map 19% reduce 0%  
2020-02-15 01:40:57,456 INFO org.apache.hadoop.mapreduce.Job (main): map 32% reduce 0%  
2020-02-15 01:40:58,460 INFO org.apache.hadoop.mapreduce.Job (main): map 42% reduce 0%  
2020-02-15 01:40:59,465 INFO org.apache.hadoop.mapreduce.Job (main): map 83% reduce 0%  
2020-02-15 01:41:00,469 INFO org.apache.hadoop.mapreduce.Job (main): map 84% reduce 0%  
2020-02-15 01:41:02,475 INFO org.apache.hadoop.mapreduce.Job (main): map 90% reduce 0%  
2020-02-15 01:41:03,478 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%  
2020-02-15 01:41:04,482 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 45%  
2020-02-15 01:41:05,485 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 73%  
2020-02-15 01:41:06,489 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%  
2020-02-15 01:41:06,493 INFO org.apache.hadoop.mapreduce.Job (main): Job job\_1581730657860\_0002 completed successfully  
2020-02-15 01:41:06,516 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 56

File System Counters

FILE: Number of bytes read=4323872  
FILE: Number of bytes written=15755464

FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=4455619  
HDFS: Number of bytes written=42861  
HDFS: Number of read operations=106  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=22  
S3: Number of bytes read=1319517205  
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=31  
Launched reduce tasks=11  
Data-local map tasks=20  
Rack-local map tasks=11  
Total time spent by all maps in occupied slots (ms)=36445056  
Total time spent by all reduces in occupied slots (ms)=7855104  
Total time spent by all map tasks (ms)=379636  
Total time spent by all reduce tasks (ms)=40912  
Total vcore-milliseconds taken by all map tasks=379636  
Total vcore-milliseconds taken by all reduce tasks=40912  
Total megabyte-milliseconds taken by all map tasks=1166241792  
Total megabyte-milliseconds taken by all reduce tasks=251363328

#### Map-Reduce Framework

Map input records=85733959  
Map output records=809908  
Map output bytes=12125611  
Map output materialized bytes=4293907  
Input split bytes=3791  
Combine input records=0  
Combine output records=0  
Reduce input groups=999  
Reduce shuffle bytes=4293907  
Reduce input records=809908  
Reduce output records=3297  
Spilled Records=1619816  
Shuffled Maps =341  
Failed Shuffles=0  
Merged Map outputs=341  
GC time elapsed (ms)=12742  
CPU time spent (ms)=263630  
Physical memory (bytes) snapshot=30279585792  
Virtual memory (bytes) snapshot=223658627072  
Total committed heap usage (bytes)=30770462720

#### 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=1323969033

#### File Output Format Counters

Bytes Written=42861

2020-02-15 01:41:06,568 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-69-156.ec2.internal/172.31.69.156:8032  
2020-02-15 01:41:06,683 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 11  
2020-02-15 01:41:06,729 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:11

```
2020-02-15 01:41:06,750 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1581730657860_0003
2020-02-15 01:41:06,761 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1581730657860_0003
2020-02-15 01:41:06,769 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-69-156.ec2.internal:20888/proxy/application_1581730657860_0003/
2020-02-15 01:41:06,769 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1581730657860_0003
2020-02-15 01:41:12,864 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581730657860_0003 running in uber mode : false
2020-02-15 01:41:12,865 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-02-15 01:41:20,898 INFO org.apache.hadoop.mapreduce.Job (main): map 36% reduce 0%
2020-02-15 01:41:22,905 INFO org.apache.hadoop.mapreduce.Job (main): map 64% reduce 0%
2020-02-15 01:41:24,912 INFO org.apache.hadoop.mapreduce.Job (main): map 91% reduce 0%
2020-02-15 01:41:25,916 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-02-15 01:41:26,921 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 9%
2020-02-15 01:41:28,927 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 55%
2020-02-15 01:41:29,930 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 82%
2020-02-15 01:41:31,937 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 01:41:32,943 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581730657860_0003 completed successfully
2020-02-15 01:41:32,965 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

#### File System Counters

```
FILE: Number of bytes read=2573
FILE: Number of bytes written=3742859
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=44500
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=42
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=11
Launched reduce tasks=11
Data-local map tasks=11
Total time spent by all maps in occupied slots (ms)=7898784
Total time spent by all reduces in occupied slots (ms)=12101760
Total time spent by all map tasks (ms)=82279
Total time spent by all reduce tasks (ms)=63030
Total vcore-milliseconds taken by all map tasks=82279
Total vcore-milliseconds taken by all reduce tasks=63030
Total megabyte-milliseconds taken by all map tasks=252761088
Total megabyte-milliseconds taken by all reduce tasks=387256320
```

#### Map-Reduce Framework

```
Map input records=3297
Map output records=3297
Map output bytes=42861
Map output materialized bytes=4448
Input split bytes=1639
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=4448
Reduce input records=3297
Reduce output records=1
Spilled Records=6594
Shuffled Maps =121
```

```
Failed Shuffles=0
Merged Map outputs=121
GC time elapsed (ms)=4829
CPU time spent (ms)=23990
Physical memory (bytes) snapshot=11248308224
Virtual memory (bytes) snapshot=131527860224
Total committed heap usage (bytes)=10716446720
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=42861
File Output Format Counters
  Bytes Written=42
```