

```
2020-02-15 05:55:02,004 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-77-138.ec2.internal/172.31.77.138:8032
2020-02-15 05:55:04,017 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-02-15 05:55:04,029 INFO com.hadoop.compression.lzo.GPLNativeCodeLoader (main): Loaded
native gpl library
2020-02-15 05:55:04,032 INFO com.hadoop.compression.lzo.LzoCodec (main): Successfully loaded
& initialized native-lzo library [hadoop-lzo rev 5f788d5e8f90539ee331702c753fa250727128f4]
2020-02-15 05:55:04,397 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:20
2020-02-15 05:55:04,513 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1581746036314_0001
2020-02-15 05:55:05,037 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1581746036314_0001
2020-02-15 05:55:05,135 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-77-138.ec2.internal:20888/proxy/application_1581746036314_0001/
2020-02-15 05:55:05,136 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1581746036314_0001
2020-02-15 05:55:11,215 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581746036314_0001 running in uber mode : false
2020-02-15 05:55:11,216 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-02-15 05:55:23,312 INFO org.apache.hadoop.mapreduce.Job (main): map 10% reduce 0%
2020-02-15 05:55:24,315 INFO org.apache.hadoop.mapreduce.Job (main): map 30% reduce 0%
2020-02-15 05:55:25,319 INFO org.apache.hadoop.mapreduce.Job (main): map 40% reduce 0%
2020-02-15 05:55:31,338 INFO org.apache.hadoop.mapreduce.Job (main): map 45% reduce 0%
2020-02-15 05:55:32,342 INFO org.apache.hadoop.mapreduce.Job (main): map 64% reduce 0%
2020-02-15 05:55:33,346 INFO org.apache.hadoop.mapreduce.Job (main): map 95% reduce 0%
2020-02-15 05:55:34,349 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 15%
2020-02-15 05:55:35,352 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-02-15 05:55:36,356 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 62%
2020-02-15 05:55:37,358 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 77%
2020-02-15 05:55:38,361 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 05:55:38,366 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581746036314_0001 completed successfully
2020-02-15 05:55:38,451 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
```

File System Counters

```
FILE: Number of bytes read=38608
FILE: Number of bytes written=5691176
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=79
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=1319478747
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=13
Data-local map tasks=20
Total time spent by all maps in occupied slots (ms)=29949888
Total time spent by all reduces in occupied slots (ms)=15733248
Total time spent by all map tasks (ms)=311978
Total time spent by all reduce tasks (ms)=81944
Total vcore-milliseconds taken by all map tasks=311978
Total vcore-milliseconds taken by all reduce tasks=81944
Total megabyte-milliseconds taken by all map tasks=958396416
Total megabyte-milliseconds taken by all reduce tasks=503463936
```

Map-Reduce Framework

Map input records=85331845
Map output records=5680
Map output bytes=63486
Map output materialized bytes=46726
Input split bytes=2240
Combine input records=0
Combine output records=0
Reduce input groups=999
Reduce shuffle bytes=46726
Reduce input records=5680
Reduce output records=402114
Spilled Records=11360
Shuffled Maps =260
Failed Shuffles=0
Merged Map outputs=260
GC time elapsed (ms)=10731
CPU time spent (ms)=247910
Physical memory (bytes) snapshot=26930008064
Virtual memory (bytes) snapshot=187635200000
Total committed heap usage (bytes)=27379892224

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=1319478747

File Output Format Counters

Bytes Written=4451828

2020-02-15 05:55:38,497 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to ResourceManager at ip-172-31-77-138.ec2.internal/172.31.77.138:8032
2020-02-15 05:55:38,716 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main): Total input files to process : 14
2020-02-15 05:55:38,785 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:33
2020-02-15 05:55:38,818 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1581746036314_0002
2020-02-15 05:55:38,847 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1581746036314_0002
2020-02-15 05:55:38,861 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-77-138.ec2.internal:20888/proxy/application_1581746036314_0002/
2020-02-15 05:55:38,861 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1581746036314_0002
2020-02-15 05:55:43,916 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1581746036314_0002 running in uber mode : false
2020-02-15 05:55:43,916 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-02-15 05:55:53,969 INFO org.apache.hadoop.mapreduce.Job (main): map 12% reduce 0%
2020-02-15 05:55:54,978 INFO org.apache.hadoop.mapreduce.Job (main): map 15% reduce 0%
2020-02-15 05:55:55,981 INFO org.apache.hadoop.mapreduce.Job (main): map 24% reduce 0%
2020-02-15 05:55:59,996 INFO org.apache.hadoop.mapreduce.Job (main): map 27% reduce 0%
2020-02-15 05:56:01,002 INFO org.apache.hadoop.mapreduce.Job (main): map 42% reduce 0%
2020-02-15 05:56:02,005 INFO org.apache.hadoop.mapreduce.Job (main): map 45% reduce 0%
2020-02-15 05:56:03,009 INFO org.apache.hadoop.mapreduce.Job (main): map 56% reduce 0%
2020-02-15 05:56:04,013 INFO org.apache.hadoop.mapreduce.Job (main): map 94% reduce 0%
2020-02-15 05:56:06,019 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-02-15 05:56:07,030 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-02-15 05:56:08,035 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 46%
2020-02-15 05:56:09,038 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 92%
2020-02-15 05:56:10,040 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 05:56:10,045 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1581746036314_0002 completed successfully
2020-02-15 05:56:10,070 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 56

File System Counters

FILE: Number of bytes read=4312152
FILE: Number of bytes written=16415314
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=4455901
HDFS: Number of bytes written=42861
HDFS: Number of read operations=118
HDFS: Number of large read operations=0
HDFS: Number of write operations=26
S3: Number of bytes read=1319513061
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=33
Launched reduce tasks=13
Data-local map tasks=20
Rack-local map tasks=13
Total time spent by all maps in occupied slots (ms)=38070720
Total time spent by all reduces in occupied slots (ms)=11844288
Total time spent by all map tasks (ms)=396570
Total time spent by all reduce tasks (ms)=61689
Total vcore-milliseconds taken by all map tasks=396570
Total vcore-milliseconds taken by all reduce tasks=61689
Total megabyte-milliseconds taken by all map tasks=1218263040
Total megabyte-milliseconds taken by all reduce tasks=379017216

Map-Reduce Framework

Map input records=85733959
Map output records=809908
Map output bytes=12125611
Map output materialized bytes=4284404
Input split bytes=4073
Combine input records=0
Combine output records=0
Reduce input groups=999
Reduce shuffle bytes=4284404
Reduce input records=809908
Reduce output records=3297
Spilled Records=1619816
Shuffled Maps =429
Failed Shuffles=0
Merged Map outputs=429
GC time elapsed (ms)=12583
CPU time spent (ms)=263150
Physical memory (bytes) snapshot=31122657280
Virtual memory (bytes) snapshot=247682895872
Total committed heap usage (bytes)=31520718848

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=1323964889

File Output Format Counters

Bytes Written=42861

2020-02-15 05:56:10,123 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to
ResourceManager at ip-172-31-77-138.ec2.internal/172.31.77.138:8032
2020-02-15 05:56:10,242 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):

Total input files to process : 13
2020-02-15 05:56:10,296 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of splits:13
2020-02-15 05:56:10,317 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting tokens for job: job_1581746036314_0003
2020-02-15 05:56:10,332 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main): Submitted application application_1581746036314_0003
2020-02-15 05:56:10,341 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the job: http://ip-172-31-77-138.ec2.internal:20888/proxy/application_1581746036314_0003/
2020-02-15 05:56:10,341 INFO org.apache.hadoop.mapreduce.Job (main): Running job: job_1581746036314_0003
2020-02-15 05:56:16,503 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1581746036314_0003 running in uber mode : false
2020-02-15 05:56:16,504 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-02-15 05:56:24,562 INFO org.apache.hadoop.mapreduce.Job (main): map 15% reduce 0%
2020-02-15 05:56:26,569 INFO org.apache.hadoop.mapreduce.Job (main): map 38% reduce 0%
2020-02-15 05:56:28,576 INFO org.apache.hadoop.mapreduce.Job (main): map 77% reduce 0%
2020-02-15 05:56:29,579 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-02-15 05:56:30,582 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 8%
2020-02-15 05:56:32,588 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 54%
2020-02-15 05:56:33,591 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 69%
2020-02-15 05:56:35,597 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 92%
2020-02-15 05:56:36,600 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 05:56:37,606 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1581746036314_0003 completed successfully
2020-02-15 05:56:37,628 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 56

File System Counters

FILE: Number of bytes read=2613
FILE: Number of bytes written=4423585
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=44798
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=42
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=9
Rack-local map tasks=4
Total time spent by all maps in occupied slots (ms)=10801728
Total time spent by all reduces in occupied slots (ms)=14115264
Total time spent by all map tasks (ms)=112518
Total time spent by all reduce tasks (ms)=73517
Total vcore-milliseconds taken by all map tasks=112518
Total vcore-milliseconds taken by all reduce tasks=73517
Total megabyte-milliseconds taken by all map tasks=345655296
Total megabyte-milliseconds taken by all reduce tasks=451688448

Map-Reduce Framework

Map input records=3297
Map output records=3297
Map output bytes=42861
Map output materialized bytes=5252
Input split bytes=1937
Combine input records=0
Combine output records=0
Reduce input groups=1

```
Reduce shuffle bytes=5252
Reduce input records=3297
Reduce output records=1
Spilled Records=6594
Shuffled Maps =169
Failed Shuffles=0
Merged Map outputs=169
GC time elapsed (ms)=7487
CPU time spent (ms)=32050
Physical memory (bytes) snapshot=13913632768
Virtual memory (bytes) snapshot=155680055296
Total committed heap usage (bytes)=13334740992

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
```