

```
2020-02-15 05:38:24,291 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to
ResourceManager at ip-172-31-70-110.ec2.internal/172.31.70.110:8032
2020-02-15 05:38:26,359 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-02-15 05:38:26,367 INFO com.hadoop.compression.lzo.GPLNativeCodeLoader (main): Loaded
native gpl library
2020-02-15 05:38:26,369 INFO com.hadoop.compression.lzo.LzoCodec (main): Successfully loaded
& initialized native-lzo library [hadoop-lzo rev 5f788d5e8f90539ee331702c753fa250727128f4]
2020-02-15 05:38:26,731 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:20
2020-02-15 05:38:26,826 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1581745008987_0001
2020-02-15 05:38:27,273 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1581745008987_0001
2020-02-15 05:38:27,362 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-70-110.ec2.internal:20888/proxy/application_1581745008987_0001/
2020-02-15 05:38:27,363 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1581745008987_0001
2020-02-15 05:38:34,488 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581745008987_0001 running in uber mode : false
2020-02-15 05:38:34,489 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-02-15 05:40:01,856 INFO org.apache.hadoop.mapreduce.Job (main): map 5% reduce 0%
2020-02-15 05:40:03,861 INFO org.apache.hadoop.mapreduce.Job (main): map 10% reduce 0%
2020-02-15 05:40:07,872 INFO org.apache.hadoop.mapreduce.Job (main): map 15% reduce 0%
2020-02-15 05:40:08,875 INFO org.apache.hadoop.mapreduce.Job (main): map 25% reduce 0%
2020-02-15 05:40:09,878 INFO org.apache.hadoop.mapreduce.Job (main): map 30% reduce 0%
2020-02-15 05:40:12,886 INFO org.apache.hadoop.mapreduce.Job (main): map 35% reduce 0%
2020-02-15 05:40:14,891 INFO org.apache.hadoop.mapreduce.Job (main): map 40% reduce 0%
2020-02-15 05:40:17,898 INFO org.apache.hadoop.mapreduce.Job (main): map 40% reduce 1%
2020-02-15 05:40:18,902 INFO org.apache.hadoop.mapreduce.Job (main): map 40% reduce 3%
2020-02-15 05:40:22,915 INFO org.apache.hadoop.mapreduce.Job (main): map 40% reduce 4%
2020-02-15 05:40:23,918 INFO org.apache.hadoop.mapreduce.Job (main): map 40% reduce 5%
2020-02-15 05:40:24,920 INFO org.apache.hadoop.mapreduce.Job (main): map 40% reduce 6%
2020-02-15 05:40:29,932 INFO org.apache.hadoop.mapreduce.Job (main): map 40% reduce 7%
2020-02-15 05:41:07,014 INFO org.apache.hadoop.mapreduce.Job (main): map 41% reduce 7%
2020-02-15 05:41:08,016 INFO org.apache.hadoop.mapreduce.Job (main): map 46% reduce 7%
2020-02-15 05:41:09,019 INFO org.apache.hadoop.mapreduce.Job (main): map 51% reduce 7%
2020-02-15 05:41:10,021 INFO org.apache.hadoop.mapreduce.Job (main): map 56% reduce 7%
2020-02-15 05:41:11,023 INFO org.apache.hadoop.mapreduce.Job (main): map 56% reduce 8%
2020-02-15 05:41:12,026 INFO org.apache.hadoop.mapreduce.Job (main): map 56% reduce 9%
2020-02-15 05:41:13,028 INFO org.apache.hadoop.mapreduce.Job (main): map 56% reduce 10%
2020-02-15 05:41:17,036 INFO org.apache.hadoop.mapreduce.Job (main): map 61% reduce 10%
2020-02-15 05:41:19,041 INFO org.apache.hadoop.mapreduce.Job (main): map 61% reduce 11%
2020-02-15 05:41:20,043 INFO org.apache.hadoop.mapreduce.Job (main): map 66% reduce 11%
2020-02-15 05:41:21,045 INFO org.apache.hadoop.mapreduce.Job (main): map 71% reduce 11%
2020-02-15 05:41:23,050 INFO org.apache.hadoop.mapreduce.Job (main): map 76% reduce 11%
2020-02-15 05:41:24,052 INFO org.apache.hadoop.mapreduce.Job (main): map 81% reduce 14%
2020-02-15 05:41:25,054 INFO org.apache.hadoop.mapreduce.Job (main): map 81% reduce 16%
2020-02-15 05:41:26,057 INFO org.apache.hadoop.mapreduce.Job (main): map 86% reduce 16%
2020-02-15 05:41:30,065 INFO org.apache.hadoop.mapreduce.Job (main): map 96% reduce 17%
2020-02-15 05:41:31,068 INFO org.apache.hadoop.mapreduce.Job (main): map 96% reduce 18%
2020-02-15 05:41:32,070 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 18%
2020-02-15 05:41:33,072 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 46%
2020-02-15 05:41:34,076 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 77%
2020-02-15 05:41:35,078 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 92%
2020-02-15 05:41:38,173 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 05:41:39,179 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581745008987_0001 completed successfully
2020-02-15 05:41:39,284 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 55
File System Counters
FILE: Number of bytes read=961609
FILE: Number of bytes written=7538815
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=0
HDFS: Number of read operations=20
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
S3: Number of bytes read=27705161521
S3: Number of bytes written=52
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed map tasks=4
Launched map tasks=20
Launched reduce tasks=13
Data-local map tasks=20
Total time spent by all maps in occupied slots (ms)=258258624
Total time spent by all reduces in occupied slots (ms)=127279104
Total time spent by all map tasks (ms)=2690194
Total time spent by all reduce tasks (ms)=662912
Total vcore-milliseconds taken by all map tasks=2690194
Total vcore-milliseconds taken by all reduce tasks=662912
Total megabyte-milliseconds taken by all map tasks=8264275968
Total megabyte-milliseconds taken by all reduce tasks=4072931328

Map-Reduce Framework

Map input records=85331845
Map output records=1560888
Map output bytes=17169768
Map output materialized bytes=965551
Input split bytes=2240
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=965551
Reduce input records=1560888
Reduce output records=1
Spilled Records=3121776
Shuffled Maps =260
Failed Shuffles=0
Merged Map outputs=260
GC time elapsed (ms)=49225
CPU time spent (ms)=2561120
Physical memory (bytes) snapshot=32773898240
Virtual memory (bytes) snapshot=188082458624
Total committed heap usage (bytes)=31210864640

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=1319525021

File Output Format Counters

Bytes Written=52