

```
2020-02-15 01:00:46,714 INFO org.apache.hadoop.yarn.client.RMProxy (main): Connecting to
ResourceManager at ip-172-31-2-22.ec2.internal/172.31.2.22:8032
2020-02-15 01:00:48,848 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-02-15 01:00:48,857 INFO com.hadoop.compression.lzo.GPLNativeCodeLoader (main): Loaded
native gpl library
2020-02-15 01:00:48,859 INFO com.hadoop.compression.lzo.LzoCodec (main): Successfully loaded
& initialized native-lzo library [hadoop-lzo rev 5f788d5e8f90539ee331702c753fa250727128f4]
2020-02-15 01:00:49,178 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:20
2020-02-15 01:00:49,279 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1581728349476_0001
2020-02-15 01:00:49,764 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1581728349476_0001
2020-02-15 01:00:49,861 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-2-22.ec2.internal:20888/proxy/application_1581728349476_0001/
2020-02-15 01:00:49,861 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1581728349476_0001
2020-02-15 01:00:57,061 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581728349476_0001 running in uber mode : false
2020-02-15 01:00:57,062 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-02-15 01:02:28,447 INFO org.apache.hadoop.mapreduce.Job (main): map 5% reduce 0%
2020-02-15 01:02:31,455 INFO org.apache.hadoop.mapreduce.Job (main): map 10% reduce 0%
2020-02-15 01:02:35,466 INFO org.apache.hadoop.mapreduce.Job (main): map 15% reduce 0%
2020-02-15 01:02:36,469 INFO org.apache.hadoop.mapreduce.Job (main): map 20% reduce 0%
2020-02-15 01:02:45,491 INFO org.apache.hadoop.mapreduce.Job (main): map 20% reduce 1%
2020-02-15 01:02:51,505 INFO org.apache.hadoop.mapreduce.Job (main): map 20% reduce 2%
2020-02-15 01:03:40,614 INFO org.apache.hadoop.mapreduce.Job (main): map 25% reduce 2%
2020-02-15 01:03:43,620 INFO org.apache.hadoop.mapreduce.Job (main): map 27% reduce 2%
2020-02-15 01:03:44,622 INFO org.apache.hadoop.mapreduce.Job (main): map 32% reduce 2%
2020-02-15 01:03:45,625 INFO org.apache.hadoop.mapreduce.Job (main): map 32% reduce 3%
2020-02-15 01:03:47,631 INFO org.apache.hadoop.mapreduce.Job (main): map 52% reduce 3%
2020-02-15 01:03:48,633 INFO org.apache.hadoop.mapreduce.Job (main): map 54% reduce 3%
2020-02-15 01:03:49,636 INFO org.apache.hadoop.mapreduce.Job (main): map 59% reduce 3%
2020-02-15 01:03:50,640 INFO org.apache.hadoop.mapreduce.Job (main): map 62% reduce 3%
2020-02-15 01:03:51,643 INFO org.apache.hadoop.mapreduce.Job (main): map 69% reduce 5%
2020-02-15 01:03:53,647 INFO org.apache.hadoop.mapreduce.Job (main): map 77% reduce 5%
2020-02-15 01:03:54,649 INFO org.apache.hadoop.mapreduce.Job (main): map 82% reduce 5%
2020-02-15 01:03:55,652 INFO org.apache.hadoop.mapreduce.Job (main): map 87% reduce 5%
2020-02-15 01:03:56,654 INFO org.apache.hadoop.mapreduce.Job (main): map 87% reduce 6%
2020-02-15 01:03:57,656 INFO org.apache.hadoop.mapreduce.Job (main): map 87% reduce 8%
2020-02-15 01:03:59,660 INFO org.apache.hadoop.mapreduce.Job (main): map 93% reduce 8%
2020-02-15 01:04:01,665 INFO org.apache.hadoop.mapreduce.Job (main): map 95% reduce 8%
2020-02-15 01:04:03,669 INFO org.apache.hadoop.mapreduce.Job (main): map 95% reduce 9%
2020-02-15 01:04:04,671 INFO org.apache.hadoop.mapreduce.Job (main): map 95% reduce 12%
2020-02-15 01:04:05,673 INFO org.apache.hadoop.mapreduce.Job (main): map 95% reduce 17%
2020-02-15 01:04:06,675 INFO org.apache.hadoop.mapreduce.Job (main): map 96% reduce 17%
2020-02-15 01:04:07,678 INFO org.apache.hadoop.mapreduce.Job (main): map 96% reduce 20%
2020-02-15 01:04:09,682 INFO org.apache.hadoop.mapreduce.Job (main): map 96% reduce 23%
2020-02-15 01:04:14,692 INFO org.apache.hadoop.mapreduce.Job (main): map 96% reduce 26%
2020-02-15 01:04:16,696 INFO org.apache.hadoop.mapreduce.Job (main): map 96% reduce 29%
2020-02-15 01:07:47,060 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 29%
2020-02-15 01:07:48,062 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 82%
2020-02-15 01:07:49,065 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 85%
2020-02-15 01:07:51,068 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 91%
2020-02-15 01:07:52,070 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-02-15 01:07:53,075 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1581728349476_0001 completed successfully
2020-02-15 01:07:53,173 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 56
File System Counters
FILE: Number of bytes read=4456142
FILE: Number of bytes written=14185548
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=27705144417
S3: Number of bytes written=53
S3: Number of read operations=0
S3: Number of large read operations=0
S3: Number of write operations=0

Job Counters

Killed map tasks=3
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)=322156320
Total time spent by all reduces in occupied slots (ms)=495527232
Total time spent by all map tasks (ms)=3355795
Total time spent by all reduce tasks (ms)=2580871
Total vcore-milliseconds taken by all map tasks=3355795
Total vcore-milliseconds taken by all reduce tasks=2580871
Total megabyte-milliseconds taken by all map tasks=10309002240
Total megabyte-milliseconds taken by all reduce tasks=15856871424

Map-Reduce Framework

Map input records=85331845
Map output records=7234833
Map output bytes=79583163
Map output materialized bytes=4459468
Input split bytes=2240
Combine input records=0
Combine output records=0
Reduce input groups=1
Reduce shuffle bytes=4459468
Reduce input records=7234833
Reduce output records=1
Spilled Records=14469666
Shuffled Maps =220
Failed Shuffles=0
Merged Map outputs=220
GC time elapsed (ms)=69260
CPU time spent (ms)=3182600
Physical memory (bytes) snapshot=31937728512
Virtual memory (bytes) snapshot=173287968768
Total committed heap usage (bytes)=30120345600

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=1319507917

File Output Format Counters

Bytes Written=53