

```
2020-01-31 19:42:08,287 INFO org.apache.hadoop.yarn.client.RMPProxy (main): Connecting to
ResourceManager at ip-172-31-74-240.ec2.internal/172.31.74.240:8032
2020-01-31 19:42:10,050 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat (main):
Total input files to process : 1
2020-01-31 19:42:10,058 INFO com.hadoop.compression.lzo.GPLNativeCodeLoader (main): Loaded
native gpl library
2020-01-31 19:42:10,060 INFO com.hadoop.compression.lzo.LzoCodec (main): Successfully loaded
& initialized native-lzo library [hadoop-lzo rev 5f788d5e8f90539ee331702c753fa250727128f4]
2020-01-31 19:42:10,372 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): number of
splits:20
2020-01-31 19:42:10,466 INFO org.apache.hadoop.mapreduce.JobSubmitter (main): Submitting
tokens for job: job_1580499650191_0001
2020-01-31 19:42:10,948 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl (main):
Submitted application application_1580499650191_0001
2020-01-31 19:42:11,037 INFO org.apache.hadoop.mapreduce.Job (main): The url to track the
job: http://ip-172-31-74-240.ec2.internal:20888/proxy/application_1580499650191_0001/
2020-01-31 19:42:11,038 INFO org.apache.hadoop.mapreduce.Job (main): Running job:
job_1580499650191_0001
2020-01-31 19:42:18,159 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1580499650191_0001 running in uber mode : false
2020-01-31 19:42:18,160 INFO org.apache.hadoop.mapreduce.Job (main): map 0% reduce 0%
2020-01-31 19:42:35,256 INFO org.apache.hadoop.mapreduce.Job (main): map 7% reduce 0%
2020-01-31 19:42:36,268 INFO org.apache.hadoop.mapreduce.Job (main): map 17% reduce 0%
2020-01-31 19:42:38,283 INFO org.apache.hadoop.mapreduce.Job (main): map 20% reduce 0%
2020-01-31 19:42:39,287 INFO org.apache.hadoop.mapreduce.Job (main): map 44% reduce 0%
2020-01-31 19:42:42,297 INFO org.apache.hadoop.mapreduce.Job (main): map 49% reduce 0%
2020-01-31 19:42:44,306 INFO org.apache.hadoop.mapreduce.Job (main): map 50% reduce 0%
2020-01-31 19:42:45,310 INFO org.apache.hadoop.mapreduce.Job (main): map 77% reduce 0%
2020-01-31 19:42:47,316 INFO org.apache.hadoop.mapreduce.Job (main): map 82% reduce 0%
2020-01-31 19:42:48,319 INFO org.apache.hadoop.mapreduce.Job (main): map 83% reduce 0%
2020-01-31 19:42:49,322 INFO org.apache.hadoop.mapreduce.Job (main): map 85% reduce 0%
2020-01-31 19:42:50,326 INFO org.apache.hadoop.mapreduce.Job (main): map 88% reduce 0%
2020-01-31 19:42:51,329 INFO org.apache.hadoop.mapreduce.Job (main): map 93% reduce 0%
2020-01-31 19:42:52,333 INFO org.apache.hadoop.mapreduce.Job (main): map 98% reduce 0%
2020-01-31 19:42:53,336 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 0%
2020-01-31 19:42:54,340 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 6%
2020-01-31 19:42:55,343 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 19%
2020-01-31 19:42:56,346 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 36%
2020-01-31 19:42:57,349 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 55%
2020-01-31 19:42:59,355 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 64%
2020-01-31 19:43:00,359 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 82%
2020-01-31 19:43:01,362 INFO org.apache.hadoop.mapreduce.Job (main): map 100% reduce 100%
2020-01-31 19:43:02,370 INFO org.apache.hadoop.mapreduce.Job (main): Job
job_1580499650191_0001 completed successfully
2020-01-31 19:43:02,476 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 56
    File System Counters
        FILE: Number of bytes read=63297557
        FILE: Number of bytes written=162979310
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=2100
        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=1319484634
        S3: Number of bytes written=67641452
        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=11
    Data-local map tasks=20
    Total time spent by all maps in occupied slots (ms)=50336736
    Total time spent by all reduces in occupied slots (ms)=21086976
    Total time spent by all map tasks (ms)=524341
    Total time spent by all reduce tasks (ms)=109828
    Total vcore-milliseconds taken by all map tasks=524341
    Total vcore-milliseconds taken by all reduce tasks=109828
    Total megabyte-milliseconds taken by all map tasks=1610775552
    Total megabyte-milliseconds taken by all reduce tasks=674783232
Map-Reduce Framework
    Map input records=85331845
    Map output records=85331845
    Map output bytes=961483442
    Map output materialized bytes=94415552
    Input split bytes=2100
    Combine input records=85331845
    Combine output records=15362582
    Reduce input groups=6626985
    Reduce shuffle bytes=94415552
    Reduce input records=15362582
    Reduce output records=6626985
    Spilled Records=30725164
    Shuffled Maps =220
    Failed Shuffles=0
    Merged Map outputs=220
    GC time elapsed (ms)=13785
    CPU time spent (ms)=407610
    Physical memory (bytes) snapshot=28363223040
    Virtual memory (bytes) snapshot=173637443584
    Total committed heap usage (bytes)=28430041088
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=1319484634
File Output Format Counters
    Bytes Written=67641452
```