



9/9/2018

Pig, spark Logs & Spark Commands

SPARKPROJECT-1

Name: RUKSANA BHANU SHEIK

SPARK COMMANDS:

SPARK COMMAND₁: LOOKUP

```
spark2-submit --class sparkAssignment.LookUp.SingleLookUp \  
  
--master yarn --deploy-mode client \  
  
--name program1 --conf "spark.app.id=program1_id" \  
  
LookUp-o.o.1-SNAPSHOT-jar-with-dependencies.jar /user/ec2-  
user/spark_assignment/input/yellow_tripdata* /user/ec2-  
user/spark_assignment/output/spark1_single_row_lookup
```

SPARK COMMAND₂: FILTER

```
spark2-submit --class sparkAssignment.Filter.Filter \  
  
--master yarn --deploy-mode client \  
  
--name program2 --conf "spark.app.id=program2_id" \  
  
Filter-o.o.1-SNAPSHOT-jar-with-dependencies.jar /user/ec2-  
user/spark_assignment/input/yellow_tripdata* /user/ec2-user/spark_assignment/output/spark2_filter
```

SPARK COMMAND₃: GROUPBY

```
spark2-submit --class sparkAssignment.GroupBy.GroupBy \  
  
--master yarn --deploy-mode client \  
  
--name program3 --conf "spark.app.id=program3_id" \  
  
GroupBy-o.o.1-SNAPSHOT-jar-with-dependencies.jar /user/ec2-  
user/spark_assignment/input/yellow_tripdata* /user/ec2-  
user/spark_assignment/output/spark3_groupby_orderby
```

PIG LOGS

PIG LOG₁: LOOKUP

```
ec2-user@ip-172-31-84-37:~$ pig
2018-09-08 13:06:32,177 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 13% complete
2018-09-08 13:06:57,555 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 37% complete
2018-09-08 13:07:21,147 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 41% complete
2018-09-08 13:07:44,309 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 45% complete
2018-09-08 13:08:06,101 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete
2018-09-08 13:08:11,997 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2018-09-08 13:08:11,998 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion: PigVersion: User: StartedAt: FinishedAt: Features:
2.6.0-cdh5.15.0 0.12.0-cdh5.15.0 ec2-user 2018-09-08 13:08:38 2018-09-08 13:08:11 FILTER

Summary:

Job Stats (time in seconds):
JobId: Maps: Reduces: MapMapTime: MinMapTime: AvgMapTime: MedianMapTime: MaxReduceTime: MinReduceTime: AvgReduceTime: MedianReduceTime:
1536418193861_0001 36 0 26 21 24 24 n/a n/a n/a n/a INPUT_DATA,INPUT_DATA_CLEAN MAP_ONLY /user/ec2-user/spark_assignment/output/pig1_single_row_lookup

Input(s):
Successfully read 54518003 records (4310013312 bytes) from: "/user/ec2-user/spark_assignment/input/yellow_tripdata*"

Output(s):
Successfully stored 1 records (194 bytes) in: "/user/ec2-user/spark_assignment/output/pig1_single_row_lookup"

Counters:
Total records written: 1
Total bytes written: 194
Spillable Memory Manager spill count: 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
Job 1536418193861_0001

2018-09-08 13:08:12,016 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TRY
E CONVERSION FAILED 30 times(s)
2018-09-08 13:08:12,016 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning ACCESSING_NON_EXIST
ENT_FIELD 36 times(s)
2018-09-08 13:08:12,016 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/pig1_single_row_lookup/part* | wc -l
1
[ec2-user@ip-172-31-84-37 ~]$
```

[ec2-user@ip-172-31-84-37 ~]\$ pig lookUp.pig

log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).

log4j:WARN Please initialize the log4j system properly.

log4j:WARN See <http://logging.apache.org/log4j/1.2/faq.html#noconfig> for more info.

2018-09-08 12:57:00,583 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.15.0 (rUnversioned directory) compiled May 24 2018, 04:18:02

2018-09-08 12:57:01,486 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/ec2-user/.pigbootup not found

2018-09-08 12:57:01,590 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address

2018-09-08 12:57:01,590 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 12:57:01,590 [main] INFO
org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file
system at: hdfs://ip-172-31-84-37.ec2.internal:8020

2018-09-08 12:57:02,068 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 12:57:02,109 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 12:57:02,143 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 12:57:02,184 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 12:57:02,216 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 12:57:02,248 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 12:57:02,281 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 12:57:02,323 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 12:57:03,362 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in
the script: FILTER

2018-09-08 12:57:03,398 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer
- {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite,
GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter,
MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter,
StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier,
PartitionFilterOptimizer]}

2018-09-08 12:57:03,520 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation
threshold: 100 optimistic? false

2018-09-08 12:57:03,541 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan
size before optimization: 1

2018-09-08 12:57:03,541 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan
size after optimization: 1

2018-09-08 12:57:03,646 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to
ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

2018-09-08 12:57:03,880 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are
added to the job

2018-09-08 12:57:03,973 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use
mapreduce.reduce.markreset.buffer.percent

2018-09-08 12:57:03,973 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-09-08 12:57:03,973 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress

2018-09-08 12:57:05,776 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar
file Job4084767596570883592.jar

2018-09-08 12:57:09,690 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file
Job4084767596570883592.jar created

2018-09-08 12:57:09,690 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.jar
is deprecated. Instead, use mapreduce.job.jar

2018-09-08 12:57:09,713 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up
single store job

2018-09-08 12:57:09,722 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key
[pig.schematuple] is false, will not generate code.

2018-09-08 12:57:09,722 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to
move generated code to distributed cache

2018-09-08 12:57:09,722 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []

2018-09-08 12:57:09,784 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-
reduce job(s) waiting for submission.

2018-09-08 12:57:09,786 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address

2018-09-08 12:57:09,793 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

2018-09-08 12:57:09,811 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 12:57:10,290 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 6

2018-09-08 12:57:10,291 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 6

2018-09-08 12:57:10,324 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 36

2018-09-08 12:57:10,797 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:36

2018-09-08 12:57:11,316 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1536410193861_0002

2018-09-08 12:57:11,595 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1536410193861_0002

2018-09-08 12:57:11,640 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0002/

2018-09-08 12:57:11,640 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1536410193861_0002

2018-09-08 12:57:11,640 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases INPUT_DATA,INPUT_DATA_CLEAN

2018-09-08 12:57:11,640 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: INPUT_DATA[3,13],INPUT_DATA[-1,-1],INPUT_DATA_CLEAN[6,19] C: R:

2018-09-08 12:57:11,677 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete

2018-09-08 12:57:47,793 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 4%
complete

2018-09-08 12:58:14,089 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 8%
complete

2018-09-08 12:58:39,939 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12%
complete

2018-09-08 12:59:04,905 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 16%
complete

2018-09-08 12:59:34,188 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 20%
complete

2018-09-08 12:59:58,639 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 24%
complete

2018-09-08 13:00:25,137 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 29%
complete

2018-09-08 13:00:55,919 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 33%
complete

2018-09-08 13:01:20,172 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 37%
complete

2018-09-08 13:01:45,020 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 41%
complete

2018-09-08 13:02:10,559 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 45%
complete

2018-09-08 13:02:30,071 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 49%
complete

2018-09-08 13:02:38,298 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces

2018-09-08 13:02:38,352 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete

2018-09-08 13:02:38,354 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.6.0-cdh5.15.0	0.12.0-cdh5.15.0	ec2-user	2018-09-08 12:57:03	2018-09-08 13:02:38	FILTER

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature	Outputs
job_1536410193861_0002	36	0	29	17	24	23	n/a	n/a	n/a	n/a			
INPUT_DATA,INPUT_DATA_CLEAN MAP_ONLY hdfs://ip-172-31-84-37.ec2.internal:8020/tmp/temp2107763893/tmp-2136203872,													

Input(s):

Successfully read 54518003 records (4810073732 bytes) from: "/user/ec2-user/spark_assignment/input/yellow_tripdata*"

Output(s):

Successfully stored 1 records (115 bytes) in: "hdfs://ip-172-31-84-37.ec2.internal:8020/tmp/temp2107763893/tmp-2136203872"

Counters:

Total records written : 1

Total bytes written : 115

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job_1536410193861_0002

2018-09-08 13:02:38,498 [main] WARN

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 70 time(s).

2018-09-08 13:02:38,498 [main] WARN

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning ACCESSING_NON_EXISTENT_FIELD 96 time(s).

2018-09-08 13:02:38,498 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

2018-09-08 13:02:38,501 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:02:38,502 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key
[pig.schematuple] was not set... will not generate code.

2018-09-08 13:02:38,517 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total
input paths to process : 36

2018-09-08 13:02:38,523 [main] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 36

(2,2017-10-01 00:15:30,2017-10-01 00:25:11,1,2.17,1,N,141,142,1,9.0,0.5,0.5,2.06,0.0,0.3,12.36)

2018-09-08 13:02:38,689 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in
the script: FILTER

2018-09-08 13:02:38,691 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer
- {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite,
GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter,

MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}

2018-09-08 13:02:38,698 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator

2018-09-08 13:02:38,704 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

2018-09-08 13:02:38,706 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1

2018-09-08 13:02:38,706 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1

2018-09-08 13:02:38,732 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

2018-09-08 13:02:38,734 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job

2018-09-08 13:02:38,751 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-09-08 13:02:40,468 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job6121571451769000810.jar

2018-09-08 13:02:44,467 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job6121571451769000810.jar created

2018-09-08 13:02:44,477 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2018-09-08 13:02:44,478 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.

2018-09-08 13:02:44,478 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache

2018-09-08 13:02:44,478 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

2018-09-08 13:02:44,502 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

2018-09-08 13:02:44,516 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

2018-09-08 13:02:44,524 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:02:44,680 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 6

2018-09-08 13:02:44,680 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 6

2018-09-08 13:02:44,696 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 36

2018-09-08 13:02:45,578 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:36

2018-09-08 13:02:46,016 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1536410193861_0003

2018-09-08 13:02:46,066 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1536410193861_0003

2018-09-08 13:02:46,069 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0003/

2018-09-08 13:02:46,069 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1536410193861_0003

2018-09-08 13:02:46,069 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases INPUT_DATA,INPUT_DATA_CLEAN

2018-09-08 13:02:46,069 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: INPUT_DATA[3,13],INPUT_DATA[-1,-1],INPUT_DATA_CLEAN[6,19] C: R:

2018-09-08 13:02:46,097 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0%
complete

2018-09-08 13:03:24,552 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 4%
complete

2018-09-08 13:03:48,990 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 8%
complete

2018-09-08 13:04:16,013 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12%
complete

2018-09-08 13:04:46,103 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 16%
complete

2018-09-08 13:05:15,923 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 20%
complete

2018-09-08 13:05:40,668 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 24%
complete

2018-09-08 13:06:06,197 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 29%
complete

2018-09-08 13:06:32,177 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 33%
complete

2018-09-08 13:06:57,555 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 37%
complete

2018-09-08 13:07:21,147 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 41%
complete

2018-09-08 13:07:44,309 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 45%
complete

2018-09-08 13:08:08,101 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50%
complete

2018-09-08 13:08:11,897 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100%
complete

2018-09-08 13:08:11,898 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.6.0-cdh5.15.0	0.12.0-cdh5.15.0	ec2-user	2018-09-08 13:02:38	2018-09-08 13:08:11	FILTER

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature	Outputs
job_1536410193861_0003	36	0	28	21	24	24	n/a	n/a	n/a	n/a	n/a	n/a	
INPUT_DATA,INPUT_DATA_CLEAN MAP_ONLY /user/ec2-user/spark_assignment/output/pig1_single_row_lookup,													

Input(s):

Successfully read 54518003 records (4810073732 bytes) from: "/user/ec2-user/spark_assignment/input/yellow_tripdata*"

Output(s):

Successfully stored 1 records (94 bytes) in: "/user/ec2-user/spark_assignment/output/pig1_single_row_lookup"

Counters:

Total records written : 1

Total bytes written : 94

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job_1536410193861_0003

2018-09-08 13:08:12,016 [main] WARN

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 70 time(s).

2018-09-08 13:08:12,016 [main] WARN

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning ACCESSING_NON_EXISTENT_FIELD 96 time(s).

2018-09-08 13:08:12,016 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

[ec2-user@ip-172-31-84-37 ~]\$ hadoop fs -cat /user/ec2-
user/spark_assignment/output/pig1_single_row_lookup/part* | wc -l

1

[ec2-user@ip-172-31-84-37 ~]\$

PIG LOG2: FILTER

```

ec2-user@ip-172-31-84-37:~$ cat /var/log/hadoop-hdfs/ec2-user/hadoop-hdfs-ec2-user-2018-09-08-13:46:14.150.log
2018-09-08 13:46:14,150 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - 17% complete
2018-09-08 13:46:14,150 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - 41% complete
2018-09-08 13:46:14,150 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - 45% complete
2018-09-08 13:46:14,150 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - 50% complete
2018-09-08 13:46:14,150 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - Application state is completed. FinalApplicationStatus=SUCCESS. Reducers
waiting for job history server
2018-09-08 13:46:14,150 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - 100% complete
2018-09-08 13:46:14,150 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - 100% complete
2018-09-08 13:46:14,150 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.0-cdh5.15.0 0.12.0-cdh5.15.0 ec2-user 2018-09-08 13:46:14 2018-09-08 13:46:14 FILTER

Success!

Job State (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime A
lias Feature Outputs
job_1536410193861_0113 36 0 49 36 42 42 n/a n/a n/a n/a INPUT_DATA,OUTPUT_DATA MAP_ONLY /user/ec2-user
r/spark_assignment/output/pig2_filter,

Input(s):
Successfully read 54815004 records (4010473732 bytes) from: "/user/ec2-user/spark_assignment/input"

Output(s):
Successfully stored 31029 records (2545544 bytes) in: "/user/ec2-user/spark_assignment/output/pig2_filter"

Counters:
Total records written: 31029
Total bytes written: 2545544
Spillable Memory Manager spill count: 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1536410193861_0113

2018-09-08 13:46:14,150 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - Encountered Warning ACCESSING_NON_EXIST
ENT_FIELD 98 times(s).
2018-09-08 13:46:14,150 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISTANCE_TY
P CONVERSION FAILED 84 times(s).
2018-09-08 13:46:14,150 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - Success!
[ec2-user@ip-172-31-84-37 ~]$

```

[ec2-user@ip-172-31-84-37 ~]\$ pig filter.pig

log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).

log4j:WARN Please initialize the log4j system properly.

log4j:WARN See <http://logging.apache.org/log4j/1.2/faq.html#noconfig> for more info.

2018-09-08 13:46:14,150 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.15.0 (rUnversioned directory) compiled May 24 2018, 04:18:02

2018-09-08 13:46:16,956 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/ec2-user/.pigbootup not found

2018-09-08 13:46:17,352 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address

2018-09-08 13:46:17,363 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:46:17,363 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://ip-172-31-84-37.ec2.internal:8020

2018-09-08 13:46:18,915 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:46:19,067 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:46:19,184 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:46:19,309 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:46:19,464 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:46:19,582 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:46:19,696 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:46:19,844 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:46:21,877 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER

2018-09-08 13:46:21,988 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}

2018-09-08 13:46:22,047 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator

2018-09-08 13:46:22,317 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

2018-09-08 13:46:22,380 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1

2018-09-08 13:46:22,381 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1

2018-09-08 13:46:22,709 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

2018-09-08 13:46:23,585 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job

2018-09-08 13:46:23,856 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent

2018-09-08 13:46:23,856 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-09-08 13:46:23,856 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress

2018-09-08 13:46:28,521 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job6640639193106956959.jar

2018-09-08 13:46:37,558 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job6640639193106956959.jar created

2018-09-08 13:46:37,558 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.jar is deprecated. Instead, use mapreduce.job.jar

2018-09-08 13:46:37,615 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2018-09-08 13:46:37,626 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.

2018-09-08 13:46:37,626 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache

2018-09-08 13:46:37,626 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

2018-09-08 13:46:37,764 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

2018-09-08 13:46:37,825 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address

2018-09-08 13:46:37,864 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

2018-09-08 13:46:38,005 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 13:46:39,912 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 6

2018-09-08 13:46:39,912 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 6

2018-09-08 13:46:39,948 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 36

2018-09-08 13:46:40,328 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:36

2018-09-08 13:46:41,036 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1536410193861_0113

2018-09-08 13:46:42,005 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1536410193861_0113

2018-09-08 13:46:42,084 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0113/

2018-09-08 13:46:42,094 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1536410193861_0113

2018-09-08 13:46:42,094 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases INPUT_DATA,OUTPUT_DATA

2018-09-08 13:46:42,094 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: INPUT_DATA[1,13],INPUT_DATA[-1,-1],OUTPUT_DATA[4,14] C: R:

2018-09-08 13:46:42,285 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete

2018-09-08 13:47:56,355 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 4% complete

2018-09-08 13:48:45,362 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 8%
complete

2018-09-08 13:49:27,736 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12%
complete

2018-09-08 13:50:13,787 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 16%
complete

2018-09-08 13:51:01,226 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 20%
complete

2018-09-08 13:51:44,895 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 24%
complete

2018-09-08 13:52:29,035 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 29%
complete

2018-09-08 13:53:11,532 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 33%
complete

2018-09-08 13:53:52,972 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 37%
complete

2018-09-08 13:54:37,060 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 41%
complete

2018-09-08 13:55:25,686 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 45%
complete

2018-09-08 13:56:09,269 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50%
complete

2018-09-08 13:56:19,300 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application
state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-09-08 13:56:19,611 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces

2018-09-08 13:56:19,802 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete

2018-09-08 13:56:19,804 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.6.0-cdh5.15.0	0.12.0-cdh5.15.0	ec2-user	2018-09-08 13:46:23	2018-09-08 13:56:19	FILTER

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature	Outputs
job_1536410193861_0113	36	0	49	36	42	42	n/a	n/a	n/a	n/a			
INPUT_DATA,OUTPUT_DATA MAP_ONLY /user/ec2-user/spark_assignment/output/pig2_filter,													

Input(s):

Successfully read 54518004 records (4810073732 bytes) from: "/user/ec2-user/spark_assignment/input"

Output(s):

Successfully stored 31029 records (2948843 bytes) in: "/user/ec2-user/spark_assignment/output/pig2_filter"

Counters:

Total records written : 31029

Total bytes written : 2948843

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job_1536410193861_0113

2018-09-08 13:56:20,306 [main] WARN

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning ACCESSING_NON_EXISTENT_FIELD 96 time(s).

2018-09-08 13:56:20,306 [main] WARN

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 84 time(s).

2018-09-08 13:56:20,306 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

[ec2-user@ip-172-31-84-37 ~]\$

PIG LOG3: GROUPBY

```
ec2-user@ip-172-31-84-37:~$
Successfully read 34518083 records (4810012332 bytes) from: "/user/ec2-user/spark_assignment/input/yellow_tripdata*"

Output(s):
Successfully stored 4 records (39 bytes) in: "/user/ec2-user/spark_assignment/output/pig3_groupby_orderby*"

Counters:
Total records written: 4
Total bytes written: 39
Spillable Memory Manager spill count: 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
Job_1536410193861_0354 -> Job_1536410193861_0374,
Job_1536410193861_0374 -> Job_1536410193861_0382,
Job_1536410193861_0382

2018-09-08 15:33:50,111 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:46450. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-09-08 15:33:51,113 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:46450. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-09-08 15:33:52,122 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:46450. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-09-08 15:33:52,236 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED, Redirecting to job history server
2018-09-08 15:33:53,603 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:36070. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-09-08 15:33:54,603 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:36070. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-09-08 15:33:55,625 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:36070. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-09-08 15:33:56,941 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED, Redirecting to job history server
2018-09-08 15:33:56,259 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED, Redirecting to job history server
2018-09-08 15:33:56,436 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TUPLE_CONVERSION_FAILED TO time(s)
2018-09-08 15:33:56,436 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning ACCESSING_NON_EXISTENT_FIELD 96 time(s)
2018-09-08 15:33:56,436 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/pig3_groupby_orderby/part* | wc -l
4
[ec2-user@ip-172-31-84-37 ~]$
```

[ec2-user@ip-172-31-84-37 ~]\$ pig groupBy.pig

log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).

log4j:WARN Please initialize the log4j system properly.

log4j:WARN See <http://logging.apache.org/log4j/1.2/faq.html#noconfig> for more info.

2018-09-08 15:18:07,225 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.15.0 (rUnversioned directory) compiled May 24 2018, 04:18:02

2018-09-08 15:18:09,790 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/ec2-user/.pigbootup not found

2018-09-08 15:18:10,242 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address

2018-09-08 15:18:10,242 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:18:10,242 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://ip-172-31-84-37.ec2.internal:8020

2018-09-08 15:18:11,504 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:18:11,586 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:18:11,716 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:18:11,964 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:18:12,119 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:18:12,217 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:18:12,385 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:18:12,475 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:18:15,211 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY,ORDER_BY,FILTER

2018-09-08 15:18:15,289 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}

2018-09-08 15:18:15,358 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator

2018-09-08 15:18:15,634 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

2018-09-08 15:18:15,771 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.CombinerOptimizer - Choosing to move algebraic foreach to combiner

2018-09-08 15:18:15,809 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan
size before optimization: 3

2018-09-08 15:18:15,810 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan
size after optimization: 3

2018-09-08 15:18:16,111 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to
ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

2018-09-08 15:18:16,689 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are
added to the job

2018-09-08 15:18:16,937 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use
mapreduce.reduce.markreset.buffer.percent

2018-09-08 15:18:16,937 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-09-08 15:18:16,938 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress

2018-09-08 15:18:16,942 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce
phase detected, estimating # of required reducers.

2018-09-08 15:18:16,943 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using
reducer estimator:
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator

2018-09-08 15:18:16,967 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator -
BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=4807895414

2018-09-08 15:18:16,968 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting
Parallelism to 5

2018-09-08 15:18:16,968 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces

2018-09-08 15:18:21,580 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar
file Job2235762098253081301.jar

2018-09-08 15:18:31,488 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file
Job2235762098253081301.jar created

2018-09-08 15:18:31,488 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.jar
is deprecated. Instead, use mapreduce.job.jar

2018-09-08 15:18:31,562 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up
single store job

2018-09-08 15:18:31,602 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key
[pig.schematuple] is false, will not generate code.

2018-09-08 15:18:31,602 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to
move generated code to distributed cache

2018-09-08 15:18:31,603 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []

2018-09-08 15:18:31,930 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-
reduce job(s) waiting for submission.

2018-09-08 15:18:31,931 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address

2018-09-08 15:18:31,988 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to
ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

2018-09-08 15:18:32,053 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:18:33,577 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat -
Total input paths to process : 6

2018-09-08 15:18:33,577 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 6

2018-09-08 15:18:33,738 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to
process : 36

2018-09-08 15:18:34,600 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:36

2018-09-08 15:18:35,447 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1536410193861_0358

2018-09-08 15:18:36,507 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1536410193861_0358

2018-09-08 15:18:36,622 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0358/

2018-09-08 15:18:36,622 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
HadoopJobId: job_1536410193861_0358

2018-09-08 15:18:36,622 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing
aliases GROUP_INPUT,INPUT_DATA,INPUT_DATA_CLEAN,PYMTTYPE_COUNT

2018-09-08 15:18:36,622 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed
locations: M: INPUT_DATA[3,13],INPUT_DATA[-1,-
1],INPUT_DATA_CLEAN[6,19],PYMTTYPE_COUNT[8,17],GROUP_INPUT[7,14] C:
PYMTTYPE_COUNT[8,17],GROUP_INPUT[7,14] R: PYMTTYPE_COUNT[8,17]

2018-09-08 15:18:36,716 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0%
complete

2018-09-08 15:22:02,841 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 4%
complete

2018-09-08 15:24:47,202 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 8%
complete

2018-09-08 15:27:34,901 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12%
complete

2018-09-08 15:29:47,752 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 16%
complete

2018-09-08 15:31:07,748 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 23% complete

2018-09-08 15:31:20,381 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 30% complete

2018-09-08 15:31:29,614 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job

2018-09-08 15:31:29,644 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-09-08 15:31:29,655 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.

2018-09-08 15:31:29,655 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator:
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator

2018-09-08 15:31:29,716 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=71

2018-09-08 15:31:29,718 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1

2018-09-08 15:31:32,828 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job282876690340771677.jar

2018-09-08 15:31:40,098 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job282876690340771677.jar created

2018-09-08 15:31:40,113 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2018-09-08 15:31:40,114 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.

2018-09-08 15:31:40,114 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache

2018-09-08 15:31:40,114 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

2018-09-08 15:31:40,148 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

2018-09-08 15:31:40,191 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

2018-09-08 15:31:40,196 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:31:41,219 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 5

2018-09-08 15:31:41,236 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 5

2018-09-08 15:31:41,237 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-09-08 15:31:42,431 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1

2018-09-08 15:31:42,672 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1536410193861_0374

2018-09-08 15:31:42,862 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1536410193861_0374

2018-09-08 15:31:42,868 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0374/

2018-09-08 15:31:42,868 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1536410193861_0374

2018-09-08 15:31:42,868 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases SORT_PYMTTYPE_COUNT

2018-09-08 15:31:42,868 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed
locations: M: SORT_PYMTTYPE_COUNT[9,22] C: R:

2018-09-08 15:32:22,385 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50%
complete

2018-09-08 15:32:36,256 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 66%
complete

2018-09-08 15:32:39,897 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are
added to the job

2018-09-08 15:32:39,925 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-09-08 15:32:39,936 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce
phase detected, estimating # of required reducers.

2018-09-08 15:32:39,936 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting
Parallelism to 1

2018-09-08 15:32:42,850 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar
file Job5476052953941751435.jar

2018-09-08 15:32:51,260 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file
Job5476052953941751435.jar created

2018-09-08 15:32:51,274 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up
single store job

2018-09-08 15:32:51,275 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key
[pig.schematuple] is false, will not generate code.

2018-09-08 15:32:51,275 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to
move generated code to distributed cache

2018-09-08 15:32:51,275 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []

2018-09-08 15:32:51,347 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-
reduce job(s) waiting for submission.

2018-09-08 15:32:51,388 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to
ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

2018-09-08 15:32:51,395 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation -
fs.default.name is deprecated. Instead, use fs.defaultFS

2018-09-08 15:32:51,971 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat -
Total input paths to process : 5

2018-09-08 15:32:51,971 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 5

2018-09-08 15:32:51,971 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to
process : 1

2018-09-08 15:32:52,289 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of
splits:1

2018-09-08 15:32:52,892 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -
Submitting tokens for job: job_1536410193861_0382

2018-09-08 15:32:53,308 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl -
Submitted application application_1536410193861_0382

2018-09-08 15:32:53,330 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the
job: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0382/

2018-09-08 15:32:53,334 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
HadoopJobId: job_1536410193861_0382

2018-09-08 15:32:53,334 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing
aliases OUTPUT_DATA,SORT_PYMTTYPE_COUNT

2018-09-08 15:32:53,334 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed
locations: M: SORT_PYMTTYPE_COUNT[9,22] C: R: OUTPUT_DATA[10,13]

2018-09-08 15:33:30,720 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 83%
complete

2018-09-08 15:33:48,934 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100%
complete

2018-09-08 15:33:49,097 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
---------------	------------	--------	-----------	------------	----------

2.6.0-cdh5.15.0	0.12.0-cdh5.15.0	ec2-user	2018-09-08 15:18:16	2018-09-08 15:33:49	GROUP_BY,ORDER_BY,FILTER
-----------------	------------------	----------	---------------------	---------------------	--------------------------

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature	Outputs
-------	------	---------	------------	------------	------------	---------------	---------------	---------------	---------------	-----------------	-------	---------	---------

job_1536410193861_0358	36	5	68	38	54	53	140	13	39	14			
GROUP_INPUT,INPUT_DATA,INPUT_DATA_CLEAN,PYMTTYPE_COUNT GROUP_BY,COMBINER													

job_1536410193861_0374	1	1	9	9	9	9	11	11	11	11			
SORT_PYMTTYPE_COUNT SAMPLER													

job_1536410193861_0382	1	1	9	9	9	9	12	12	12	12			
OUTPUT_DATA,SORT_PYMTTYPE_COUNT ORDER_BY /user/ec2-user/spark_assignment/output/pig3_groupby_orderby,													

Input(s):

Successfully read 54518003 records (4810073732 bytes) from: "/user/ec2-user/spark_assignment/input/yellow_tripdata*"

Output(s):

Successfully stored 4 records (39 bytes) in: "/user/ec2-user/spark_assignment/output/pig3_groupby_orderby"

Counters:

Total records written : 4

Total bytes written : 39

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job_1536410193861_0358 -> job_1536410193861_0374,

job_1536410193861_0374 -> job_1536410193861_0382,

job_1536410193861_0382

2018-09-08 15:33:50,111 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:46430. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-09-08 15:33:51,115 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:46430. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-09-08 15:33:52,122 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:46430. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-09-08 15:33:52,286 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-09-08 15:33:53,803 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:35070. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-09-08 15:33:54,803 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:35070. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-09-08 15:33:55,810 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-172-31-84-37.ec2.internal/172.31.84.37:35070. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-09-08 15:33:55,941 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-09-08 15:33:56,239 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-09-08 15:33:56,430 [main] WARN
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 70 time(s).

2018-09-08 15:33:56,430 [main] WARN
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning ACCESSING_NON_EXISTENT_FIELD 96 time(s).

2018-09-08 15:33:56,430 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!

[ec2-user@ip-172-31-84-37 ~]\$ hadoop fs -cat /user/ec2-user/spark_assignment/output/pig3_groupby_orderby/part* | wc -l

4

[ec2-user@ip-172-31-84-37 ~]\$

SPARK LOG₁: LOOKUP

```
[ec2-user@ip-172-31-84-37 ~]$ spark2-submit --class sparkAssignment.LookUp.SingleLookUp \
> --master yarn --deploy-mode client \
> --name program1 --conf "spark.app.id=program1_id" \
> LookUp-0.0.1-SNAPSHOT-jar-with-dependencies.jar /user/ec2-
user/spark_assignment/input/yellow_tripdata* /user/ec2-
user/spark_assignment/output/spark1_single_row_lookup
```

33

18/09/08 14:25:04 INFO spark.SecurityManager: Changing modify acls groups to:

18/09/08 14:25:04 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(ec2-user); groups with view permissions: Set(); users with modify permissions: Set(ec2-user); groups with modify permissions: Set()

18/09/08 14:25:05 INFO util.Utils: Successfully started service 'sparkDriver' on port 41041.

18/09/08 14:25:05 INFO spark.SparkEnv: Registering MapOutputTracker

18/09/08 14:25:05 INFO spark.SparkEnv: Registering BlockManagerMaster

18/09/08 14:25:05 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information

18/09/08 14:25:05 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up

18/09/08 14:25:05 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-c49db4dd-406f-4339-8913-83d46470a2cb

18/09/08 14:25:05 INFO memory.MemoryStore: MemoryStore started with capacity 366.3 MB

18/09/08 14:25:06 INFO spark.SparkEnv: Registering OutputCommitCoordinator

18/09/08 14:25:06 INFO util.log: Logging initialized @14223ms

18/09/08 14:25:06 INFO server.Server: jetty-9.3.z-SNAPSHOT

18/09/08 14:25:07 INFO server.Server: Started @14688ms

18/09/08 14:25:07 INFO server.AbstractConnector: Started ServerConnector@22ee2do{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}

18/09/08 14:25:07 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.

18/09/08 14:25:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5cad8b7d{/jobs,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@10567255{/jobs/json,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1c4ee95c{/jobs/job,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6548bb7d{/jobs/job/json,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@54336c81{/stages,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@35e52059{/stages/json,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@49bd54f7{/stages/stage,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@79ab3a71{/stages/stage/json,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3d829787{/stages/pool,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@51bde877{/stages/pool/json,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@492fc69e{/storage,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@2fb68ec6{/storage/json,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3add81c4{/storage/rdd,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@1c65121{/storage/rdd/json,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@57dc9128{/environment,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@17ae98d7{/environment/json,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@6ac4944a{/executors,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@39fc6b2c{/executors/json,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3ee39dao{/executors/threadDump,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@7cc9ce8{/executors/threadDump/json,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@cob41d6{/static,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@75201592{/null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@aa5455e{/api,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3b4ef7{/jobs/job/kill,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5987e932{/stages/stage/kill,null,AVAILABLE,@Spark}

18/09/08 14:25:07 INFO ui.SparkUI: Bound SparkUI to o.o.o.o, and started at http://ip-172-31-84-37.ec2.internal:4040

18/09/08 14:25:07 INFO spark.SparkContext: Added JAR file:/home/ec2-user/LookUp-o.o.1-SNAPSHOT-jar-with-dependencies.jar at spark://ip-172-31-84-37.ec2.internal:41041/jars/LookUp-o.o.1-SNAPSHOT-jar-with-dependencies.jar with timestamp 1536416707635

18/09/08 14:25:08 INFO util.Utils: Using initial executors = 0, max of
spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and
spark.executor.instances

18/09/08 14:25:12 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

18/09/08 14:25:13 INFO yarn.Client: Requesting a new application from cluster with 1 NodeManagers

18/09/08 14:25:13 INFO yarn.Client: Verifying our application has not requested more than the
maximum memory capability of the cluster (8192 MB per container)

18/09/08 14:25:13 INFO yarn.Client: Will allocate AM container, with 896 MB memory including 384 MB
overhead

18/09/08 14:25:13 INFO yarn.Client: Setting up container launch context for our AM

18/09/08 14:25:13 INFO yarn.Client: Setting up the launch environment for our AM container

18/09/08 14:25:13 INFO yarn.Client: Preparing resources for our AM container

18/09/08 14:25:14 INFO yarn.Client: Uploading resource file:/tmp/spark-33cd9f9e-ocd3-4b90-b6dd-90f834a4adf3/___spark_conf___2836258616572553936.zip -> hdfs://ip-172-31-84-37.ec2.internal:8020/user/ec2-user/.sparkStaging/application_1536410193861_0211/___spark_conf___.zip

18/09/08 14:25:15 INFO spark.SecurityManager: Changing view acls to: ec2-user

18/09/08 14:25:15 INFO spark.SecurityManager: Changing modify acls to: ec2-user

18/09/08 14:25:15 INFO spark.SecurityManager: Changing view acls groups to:

18/09/08 14:25:15 INFO spark.SecurityManager: Changing modify acls groups to:

18/09/08 14:25:15 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(ec2-user); groups with view permissions: Set(); users with modify permissions: Set(ec2-user); groups with modify permissions: Set()

18/09/08 14:25:19 INFO yarn.Client: Submitting application application_1536410193861_0211 to ResourceManager

18/09/08 14:25:20 INFO impl.YarnClientImpl: Submitted application application_1536410193861_0211

18/09/08 14:25:20 INFO cluster.SchedulerExtensionServices: Starting Yarn extension services with app application_1536410193861_0211 and attemptId None

18/09/08 14:25:21 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:21 INFO yarn.Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: N/A

ApplicationMaster RPC port: -1

queue: root.users.ec2-user

start time: 1536416719824

final status: UNDEFINED

tracking URL: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0211/

user: ec2-user

18/09/08 14:25:22 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:23 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:24 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:25 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:26 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:27 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:28 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:29 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:30 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:31 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:32 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:33 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:34 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:35 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:36 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:37 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:38 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:38 INFO cluster.YarnClientSchedulerBackend: Add WebUI Filter.
org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter, Map(PROXY_HOSTS -> ip-172-31-84-37.ec2.internal, PROXY_URI_BASES -> http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0211), /proxy/application_1536410193861_0211

18/09/08 14:25:38 INFO ui.JettyUtils: Adding filter:
org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter

18/09/08 14:25:39 INFO yarn.Client: Application report for application_1536410193861_0211 (state: ACCEPTED)

18/09/08 14:25:40 INFO cluster.YarnSchedulerBackend\$YarnSchedulerEndpoint: ApplicationMaster registered as NettyRpcEndpointRef(spark-client://YarnAM)

18/09/08 14:25:40 INFO yarn.Client: Application report for application_1536410193861_0211 (state: RUNNING)

18/09/08 14:25:40 INFO yarn.Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: 172.31.84.37

ApplicationMaster RPC port: 0

queue: root.users.ec2-user

start time: 1536416719824

final status: UNDEFINED

tracking URL: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0211/

user: ec2-user

18/09/08 14:25:40 INFO cluster.YarnClientSchedulerBackend: Application application_1536410193861_0211 has started running.

18/09/08 14:25:40 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 33027.

18/09/08 14:25:40 INFO netty.NettyBlockTransferService: Server created on ip-172-31-84-37.ec2.internal:33027

18/09/08 14:25:40 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy

18/09/08 14:25:40 INFO storage.BlockManagerMaster: Registering BlockManager BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 33027, None)

18/09/08 14:25:40 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:33027 with 366.3 MB RAM, BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 33027, None)

18/09/08 14:25:40 INFO storage.BlockManagerMaster: Registered BlockManager BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 33027, None)

18/09/08 14:25:40 INFO storage.BlockManager: external shuffle service port = 7337

18/09/08 14:25:40 INFO storage.BlockManager: Initialized BlockManager: BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 33027, None)

18/09/08 14:25:40 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@4aaecabd{/metrics/json,null,AVAILABLE,@Spark}

18/09/08 14:25:41 INFO scheduler.EventLoggingListener: Logging events to hdfs://ip-172-31-84-37.ec2.internal:8020/user/spark/spark2ApplicationHistory/application_1536410193861_0211

18/09/08 14:25:41 INFO util.Utils: Using initial executors = 0, max of
spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and
spark.executor.instances

18/09/08 14:25:41 INFO cluster.YarnClientSchedulerBackend: SchedulerBackend is ready for scheduling
beginning after reached minRegisteredResourcesRatio: 0.8

18/09/08 14:25:42 INFO memory.MemoryStore: Block broadcast_0 stored as values in memory
(estimated size 303.3 KB, free 366.0 MB)

18/09/08 14:25:43 INFO memory.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory
(estimated size 27.1 KB, free 366.0 MB)

18/09/08 14:25:43 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on ip-172-31-84-37.ec2.internal:33027 (size: 27.1 KB, free: 366.3 MB)

18/09/08 14:25:43 INFO spark.SparkContext: Created broadcast 0 from textFile at
SingleLookup.java:19

18/09/08 14:25:44 INFO mapred.FileInputFormat: Total input paths to process : 6

18/09/08 14:25:44 INFO spark.SparkContext: Starting job: collect at SingleLookup.java:50

18/09/08 14:25:44 INFO scheduler.DAGScheduler: Got job 0 (collect at SingleLookup.java:50) with 39
output partitions

18/09/08 14:25:44 INFO scheduler.DAGScheduler: Final stage: ResultStage 0 (collect at
SingleLookup.java:50)

18/09/08 14:25:44 INFO scheduler.DAGScheduler: Parents of final stage: List()

18/09/08 14:25:44 INFO scheduler.DAGScheduler: Missing parents: List()

18/09/08 14:25:44 INFO scheduler.DAGScheduler: Submitting ResultStage 0 (MapPartitionsRDD[3] at
filter at SingleLookup.java:25), which has no missing parents

18/09/08 14:25:44 INFO memory.MemoryStore: Block broadcast_1 stored as values in memory
(estimated size 4.3 KB, free 366.0 MB)

18/09/08 14:25:44 INFO memory.MemoryStore: Block broadcast_1_pieceo stored as bytes in memory (estimated size 2.4 KB, free 366.0 MB)

18/09/08 14:25:44 INFO storage.BlockManagerInfo: Added broadcast_1_pieceo in memory on ip-172-31-84-37.ec2.internal:33027 (size: 2.4 KB, free: 366.3 MB)

18/09/08 14:25:44 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1039

18/09/08 14:25:44 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 0 (MapPartitionsRDD[3] at filter at SingleLookup.java:25) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/09/08 14:25:44 INFO cluster.YarnScheduler: Adding task set 0.0 with 39 tasks

18/09/08 14:25:45 INFO spark.ExecutorAllocationManager: Requesting 1 new executor because tasks are backlogged (new desired total will be 1)

18/09/08 14:25:47 INFO spark.ExecutorAllocationManager: Requesting 2 new executors because tasks are backlogged (new desired total will be 3)

18/09/08 14:25:47 INFO spark.ExecutorAllocationManager: Requesting 4 new executors because tasks are backlogged (new desired total will be 7)

18/09/08 14:25:49 INFO spark.ExecutorAllocationManager: Requesting 8 new executors because tasks are backlogged (new desired total will be 15)

18/09/08 14:25:49 INFO spark.ExecutorAllocationManager: Requesting 16 new executors because tasks are backlogged (new desired total will be 31)

18/09/08 14:25:50 INFO spark.ExecutorAllocationManager: Requesting 8 new executors because tasks are backlogged (new desired total will be 39)

18/09/08 14:25:58 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (172.31.84.37:49286) with ID 1

18/09/08 14:25:58 INFO spark.ExecutorAllocationManager: New executor 1 has registered (new total is 1)

18/09/08 14:25:58 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, ip-172-31-84-37.ec2.internal, executor 1, partition 0, NODE_LOCAL, 7963 bytes)

18/09/08 14:25:58 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:35508 with 366.3 MB RAM, BlockManagerId(1, ip-172-31-84-37.ec2.internal, 35508, None)

18/09/08 14:25:58 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (172.31.84.37:49294) with ID 2

18/09/08 14:25:59 INFO spark.ExecutorAllocationManager: New executor 2 has registered (new total is 2)

18/09/08 14:25:58 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, ip-172-31-84-37.ec2.internal, executor 2, partition 1, NODE_LOCAL, 7963 bytes)

18/09/08 14:25:59 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:42678 with 366.3 MB RAM, BlockManagerId(2, ip-172-31-84-37.ec2.internal, 42678, None)

18/09/08 14:25:59 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (172.31.84.37:49292) with ID 3

18/09/08 14:25:59 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, ip-172-31-84-37.ec2.internal, executor 3, partition 2, NODE_LOCAL, 7963 bytes)

18/09/08 14:25:59 INFO spark.ExecutorAllocationManager: New executor 3 has registered (new total is 3)

18/09/08 14:25:59 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:40642 with 366.3 MB RAM, BlockManagerId(3, ip-172-31-84-37.ec2.internal, 40642, None)

18/09/08 14:26:01 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-84-37.ec2.internal:35508 (size: 2.4 KB, free: 366.3 MB)

18/09/08 14:26:01 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-84-37.ec2.internal:42678 (size: 2.4 KB, free: 366.3 MB)

18/09/08 14:26:02 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-84-37.ec2.internal:40642 (size: 2.4 KB, free: 366.3 MB)

18/09/08 14:26:03 INFO storage.BlockManagerInfo: Added broadcast_o_piece0 in memory on ip-172-31-84-37.ec2.internal:35508 (size: 27.1 KB, free: 366.3 MB)

18/09/08 14:26:03 INFO storage.BlockManagerInfo: Added broadcast_o_piece0 in memory on ip-172-31-84-37.ec2.internal:42678 (size: 27.1 KB, free: 366.3 MB)

18/09/08 14:26:04 INFO storage.BlockManagerInfo: Added broadcast_o_piece0 in memory on ip-172-31-84-37.ec2.internal:40642 (size: 27.1 KB, free: 366.3 MB)

18/09/08 14:26:33 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, ip-172-31-84-37.ec2.internal, executor 1, partition 3, NODE_LOCAL, 7963 bytes)

18/09/08 14:26:33 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 34670 ms on ip-172-31-84-37.ec2.internal (executor 1) (1/39)

18/09/08 14:26:36 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 0.0 (TID 4, ip-172-31-84-37.ec2.internal, executor 3, partition 4, NODE_LOCAL, 7963 bytes)

18/09/08 14:26:36 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 36925 ms on ip-172-31-84-37.ec2.internal (executor 3) (2/39)

18/09/08 14:26:37 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 0.0 (TID 5, ip-172-31-84-37.ec2.internal, executor 2, partition 5, NODE_LOCAL, 7963 bytes)

18/09/08 14:26:37 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 0.0 (TID 1) in 38162 ms on ip-172-31-84-37.ec2.internal (executor 2) (3/39)

18/09/08 14:26:49 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 0.0 (TID 6, ip-172-31-84-37.ec2.internal, executor 3, partition 6, NODE_LOCAL, 7963 bytes)

18/09/08 14:26:49 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 13240 ms on ip-172-31-84-37.ec2.internal (executor 3) (4/39)

18/09/08 14:26:56 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, ip-172-31-84-37.ec2.internal, executor 2, partition 7, NODE_LOCAL, 7963 bytes)

18/09/08 14:26:56 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 0.0 (TID 5) in 19552 ms on ip-172-31-84-37.ec2.internal (executor 2) (5/39)

18/09/08 14:27:01 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, ip-172-31-84-37.ec2.internal, executor 1, partition 8, NODE_LOCAL, 7963 bytes)

18/09/08 14:27:01 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 0.0 (TID 3) in 28417 ms on ip-172-31-84-37.ec2.internal (executor 1) (6/39)

18/09/08 14:27:15 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, ip-172-31-84-37.ec2.internal, executor 3, partition 9, NODE_LOCAL, 7963 bytes)

18/09/08 14:27:15 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 25910 ms on ip-172-31-84-37.ec2.internal (executor 3) (7/39)

18/09/08 14:27:17 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, ip-172-31-84-37.ec2.internal, executor 2, partition 10, NODE_LOCAL, 7963 bytes)

18/09/08 14:27:17 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 20436 ms on ip-172-31-84-37.ec2.internal (executor 2) (8/39)

18/09/08 14:27:26 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, ip-172-31-84-37.ec2.internal, executor 1, partition 11, NODE_LOCAL, 7963 bytes)

18/09/08 14:27:26 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 24481 ms on ip-172-31-84-37.ec2.internal (executor 1) (9/39)

18/09/08 14:27:33 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 0.0 (TID 12, ip-172-31-84-37.ec2.internal, executor 2, partition 12, NODE_LOCAL, 7963 bytes)

18/09/08 14:27:33 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 16020 ms on ip-172-31-84-37.ec2.internal (executor 2) (10/39)

18/09/08 14:27:37 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13, ip-172-31-84-37.ec2.internal, executor 3, partition 13, NODE_LOCAL, 7963 bytes)

18/09/08 14:27:37 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 21763 ms on ip-172-31-84-37.ec2.internal (executor 3) (11/39)

18/09/08 14:27:38 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, ip-172-31-84-37.ec2.internal, executor 1, partition 14, NODE_LOCAL, 7963 bytes)

18/09/08 14:27:38 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 0.0 (TID 11) in 12420 ms on ip-172-31-84-37.ec2.internal (executor 1) (12/39)

18/09/08 14:27:50 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, ip-172-31-84-37.ec2.internal, executor 2, partition 15, NODE_LOCAL, 7963 bytes)

18/09/08 14:27:50 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 17500 ms on ip-172-31-84-37.ec2.internal (executor 2) (13/39)

18/09/08 14:28:00 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, ip-172-31-84-37.ec2.internal, executor 1, partition 16, NODE_LOCAL, 7963 bytes)

18/09/08 14:28:00 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 0.0 (TID 14) in 22273 ms on ip-172-31-84-37.ec2.internal (executor 1) (14/39)

18/09/08 14:28:01 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 0.0 (TID 17, ip-172-31-84-37.ec2.internal, executor 3, partition 17, NODE_LOCAL, 7963 bytes)

18/09/08 14:28:01 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 0.0 (TID 13) in 24402 ms on ip-172-31-84-37.ec2.internal (executor 3) (15/39)

18/09/08 14:28:08 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 0.0 (TID 18, ip-172-31-84-37.ec2.internal, executor 2, partition 18, NODE_LOCAL, 7963 bytes)

18/09/08 14:28:08 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 18035 ms on ip-172-31-84-37.ec2.internal (executor 2) (16/39)

18/09/08 14:28:22 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 0.0 (TID 19, ip-172-31-84-37.ec2.internal, executor 1, partition 19, NODE_LOCAL, 7963 bytes)

18/09/08 14:28:22 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 21689 ms on ip-172-31-84-37.ec2.internal (executor 1) (17/39)

18/09/08 14:28:22 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 0.0 (TID 20, ip-172-31-84-37.ec2.internal, executor 3, partition 20, NODE_LOCAL, 7963 bytes)

18/09/08 14:28:22 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 0.0 (TID 17) in 20999 ms on ip-172-31-84-37.ec2.internal (executor 3) (18/39)

18/09/08 14:28:26 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 0.0 (TID 21, ip-172-31-84-37.ec2.internal, executor 2, partition 21, NODE_LOCAL, 7963 bytes)

18/09/08 14:28:26 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 0.0 (TID 18) in 18174 ms on ip-172-31-84-37.ec2.internal (executor 2) (19/39)

18/09/08 14:28:42 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 0.0 (TID 22, ip-172-31-84-37.ec2.internal, executor 1, partition 22, NODE_LOCAL, 7963 bytes)

18/09/08 14:28:42 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 0.0 (TID 19) in 20054 ms on ip-172-31-84-37.ec2.internal (executor 1) (20/39)

18/09/08 14:28:44 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 0.0 (TID 23, ip-172-31-84-37.ec2.internal, executor 3, partition 23, NODE_LOCAL, 7963 bytes)

18/09/08 14:28:44 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 0.0 (TID 20) in 22388 ms on ip-172-31-84-37.ec2.internal (executor 3) (21/39)

18/09/08 14:28:48 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 0.0 (TID 24, ip-172-31-84-37.ec2.internal, executor 2, partition 24, NODE_LOCAL, 7963 bytes)

18/09/08 14:28:48 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 0.0 (TID 21) in 22125 ms on ip-172-31-84-37.ec2.internal (executor 2) (22/39)

18/09/08 14:28:59 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 0.0 (TID 25, ip-172-31-84-37.ec2.internal, executor 2, partition 25, NODE_LOCAL, 7963 bytes)

18/09/08 14:28:59 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 0.0 (TID 24) in 10553 ms on ip-172-31-84-37.ec2.internal (executor 2) (23/39)

18/09/08 14:29:03 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 0.0 (TID 26, ip-172-31-84-37.ec2.internal, executor 1, partition 26, NODE_LOCAL, 7963 bytes)

18/09/08 14:29:03 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 0.0 (TID 22) in 21347 ms on ip-172-31-84-37.ec2.internal (executor 1) (24/39)

18/09/08 14:29:03 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 0.0 (TID 27, ip-172-31-84-37.ec2.internal, executor 3, partition 27, NODE_LOCAL, 7963 bytes)

18/09/08 14:29:03 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 0.0 (TID 23) in 18962 ms on ip-172-31-84-37.ec2.internal (executor 3) (25/39)

18/09/08 14:29:16 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 0.0 (TID 28, ip-172-31-84-37.ec2.internal, executor 3, partition 28, NODE_LOCAL, 7963 bytes)

18/09/08 14:29:16 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 0.0 (TID 27) in 12092 ms on ip-172-31-84-37.ec2.internal (executor 3) (26/39)

18/09/08 14:29:22 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 0.0 (TID 29, ip-172-31-84-37.ec2.internal, executor 2, partition 29, NODE_LOCAL, 7963 bytes)

18/09/08 14:29:22 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 0.0 (TID 25) in 22602 ms on ip-172-31-84-37.ec2.internal (executor 2) (27/39)

18/09/08 14:29:29 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 0.0 (TID 30, ip-172-31-84-37.ec2.internal, executor 1, partition 30, NODE_LOCAL, 7963 bytes)

18/09/08 14:29:29 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 0.0 (TID 26) in 25761 ms on ip-172-31-84-37.ec2.internal (executor 1) (28/39)

18/09/08 14:29:35 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 0.0 (TID 31, ip-172-31-84-37.ec2.internal, executor 3, partition 31, NODE_LOCAL, 7963 bytes)

18/09/08 14:29:35 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 0.0 (TID 28) in 19729 ms on ip-172-31-84-37.ec2.internal (executor 3) (29/39)

18/09/08 14:29:38 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 0.0 (TID 32, ip-172-31-84-37.ec2.internal, executor 3, partition 32, NODE_LOCAL, 7963 bytes)

18/09/08 14:29:38 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 0.0 (TID 31) in 3181 ms on ip-172-31-84-37.ec2.internal (executor 3) (30/39)

18/09/08 14:29:43 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 0.0 (TID 33, ip-172-31-84-37.ec2.internal, executor 2, partition 33, NODE_LOCAL, 7963 bytes)

18/09/08 14:29:43 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 0.0 (TID 29) in 21031 ms on ip-172-31-84-37.ec2.internal (executor 2) (31/39)

18/09/08 14:29:55 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 0.0 (TID 34, ip-172-31-84-37.ec2.internal, executor 1, partition 34, NODE_LOCAL, 7963 bytes)

18/09/08 14:29:55 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 0.0 (TID 30) in 25484 ms on ip-172-31-84-37.ec2.internal (executor 1) (32/39)

18/09/08 14:30:01 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 0.0 (TID 35, ip-172-31-84-37.ec2.internal, executor 3, partition 35, NODE_LOCAL, 7963 bytes)

18/09/08 14:30:01 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 0.0 (TID 32) in 22593 ms on ip-172-31-84-37.ec2.internal (executor 3) (33/39)

18/09/08 14:30:02 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 0.0 (TID 36, ip-172-31-84-37.ec2.internal, executor 2, partition 36, NODE_LOCAL, 7963 bytes)

18/09/08 14:30:02 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 0.0 (TID 33) in 19103 ms on ip-172-31-84-37.ec2.internal (executor 2) (34/39)

18/09/08 14:30:14 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 0.0 (TID 37, ip-172-31-84-37.ec2.internal, executor 1, partition 37, NODE_LOCAL, 7963 bytes)

18/09/08 14:30:14 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 0.0 (TID 34) in 19605 ms on ip-172-31-84-37.ec2.internal (executor 1) (35/39)

18/09/08 14:30:20 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 0.0 (TID 38, ip-172-31-84-37.ec2.internal, executor 2, partition 38, NODE_LOCAL, 7963 bytes)

18/09/08 14:30:20 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 0.0 (TID 36) in 18662 ms on ip-172-31-84-37.ec2.internal (executor 2) (36/39)

18/09/08 14:30:24 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 0.0 (TID 35) in 23059 ms on ip-172-31-84-37.ec2.internal (executor 3) (37/39)

18/09/08 14:30:27 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 0.0 (TID 38) in 6875 ms on ip-172-31-84-37.ec2.internal (executor 2) (38/39)

18/09/08 14:30:28 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 0.0 (TID 37) in 13436 ms on ip-172-31-84-37.ec2.internal (executor 1) (39/39)

18/09/08 14:30:28 INFO cluster.YarnScheduler: Removed TaskSet 0.0, whose tasks have all completed, from pool

18/09/08 14:30:28 INFO scheduler.DAGScheduler: ResultStage 0 (collect at SingleLookUp.java:50) finished in 283.753 s

18/09/08 14:30:28 INFO scheduler.DAGScheduler: Job 0 finished: collect at SingleLookUp.java:50, took 284.047485 s

18/09/08 14:30:28 INFO Configuration.deprecation: mapred.output.dir is deprecated. Instead, use mapreduce.output.fileoutputformat.outputdir

18/09/08 14:30:28 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/09/08 14:30:28 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/09/08 14:30:28 INFO spark.SparkContext: Starting job: runJob at SparkHadoopWriter.scala:78

18/09/08 14:30:28 INFO scheduler.DAGScheduler: Got job 1 (runJob at SparkHadoopWriter.scala:78) with 39 output partitions

18/09/08 14:30:28 INFO scheduler.DAGScheduler: Final stage: ResultStage 1 (runJob at SparkHadoopWriter.scala:78)

18/09/08 14:30:28 INFO scheduler.DAGScheduler: Parents of final stage: List()

18/09/08 14:30:28 INFO scheduler.DAGScheduler: Missing parents: List()

18/09/08 14:30:28 INFO scheduler.DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[4] at saveAsTextFile at SingleLookup.java:53), which has no missing parents

18/09/08 14:30:28 INFO memory.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 81.5 KB, free 365.9 MB)

18/09/08 14:30:28 INFO memory.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 30.3 KB, free 365.9 MB)

18/09/08 14:30:28 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-84-37.ec2.internal:33027 (size: 30.3 KB, free: 366.2 MB)

18/09/08 14:30:28 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1039

18/09/08 14:30:28 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 1 (MapPartitionsRDD[4] at saveAsTextFile at SingleLookup.java:53) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/09/08 14:30:28 INFO cluster.YarnScheduler: Adding task set 1.0 with 39 tasks

18/09/08 14:30:28 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 39, ip-172-31-84-37.ec2.internal, executor 1, partition 0, NODE_LOCAL, 7963 bytes)

18/09/08 14:30:28 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 40, ip-172-31-84-37.ec2.internal, executor 3, partition 1, NODE_LOCAL, 7963 bytes)

18/09/08 14:30:28 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 41, ip-172-31-84-37.ec2.internal, executor 2, partition 2, NODE_LOCAL, 7963 bytes)

18/09/08 14:30:28 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-84-37.ec2.internal:35508 (size: 30.3 KB, free: 366.2 MB)

18/09/08 14:30:28 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-84-37.ec2.internal:42678 (size: 30.3 KB, free: 366.2 MB)

18/09/08 14:30:28 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-84-37.ec2.internal:40642 (size: 30.3 KB, free: 366.2 MB)

18/09/08 14:30:29 INFO spark.ExecutorAllocationManager: Requesting 4 new executors because tasks are backlogged (new desired total will be 4)

18/09/08 14:30:30 INFO spark.ExecutorAllocationManager: Requesting 1 new executor because tasks are backlogged (new desired total will be 5)

18/09/08 14:30:31 INFO spark.ExecutorAllocationManager: Requesting 2 new executors because tasks are backlogged (new desired total will be 7)

18/09/08 14:30:32 INFO spark.ExecutorAllocationManager: Requesting 4 new executors because tasks are backlogged (new desired total will be 11)

18/09/08 14:30:33 INFO spark.ExecutorAllocationManager: Requesting 8 new executors because tasks are backlogged (new desired total will be 19)

18/09/08 14:30:34 INFO spark.ExecutorAllocationManager: Requesting 16 new executors because tasks are backlogged (new desired total will be 35)

18/09/08 14:30:35 INFO spark.ExecutorAllocationManager: Requesting 4 new executors because tasks are backlogged (new desired total will be 39)

18/09/08 14:30:47 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 42, ip-172-31-84-37.ec2.internal, executor 2, partition 3, NODE_LOCAL, 7963 bytes)

18/09/08 14:30:47 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 1.0 (TID 41) in 18637 ms on ip-172-31-84-37.ec2.internal (executor 2) (1/39)

18/09/08 14:30:48 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 1.0 (TID 43, ip-172-31-84-37.ec2.internal, executor 1, partition 4, NODE_LOCAL, 7963 bytes)

18/09/08 14:30:48 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 39) in 19838 ms on ip-172-31-84-37.ec2.internal (executor 1) (2/39)

18/09/08 14:30:51 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 1.0 (TID 44, ip-172-31-84-37.ec2.internal, executor 3, partition 5, NODE_LOCAL, 7963 bytes)

18/09/08 14:30:51 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 1.0 (TID 40) in 23059 ms on ip-172-31-84-37.ec2.internal (executor 3) (3/39)

18/09/08 14:31:01 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 1.0 (TID 45, ip-172-31-84-37.ec2.internal, executor 3, partition 6, NODE_LOCAL, 7963 bytes)

18/09/08 14:31:01 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 1.0 (TID 44) in 10300 ms on ip-172-31-84-37.ec2.internal (executor 3) (4/39)

18/09/08 14:31:12 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 1.0 (TID 46, ip-172-31-84-37.ec2.internal, executor 1, partition 7, NODE_LOCAL, 7963 bytes)

18/09/08 14:31:12 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 1.0 (TID 43) in 23576 ms on ip-172-31-84-37.ec2.internal (executor 1) (5/39)

18/09/08 14:31:12 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 1.0 (TID 47, ip-172-31-84-37.ec2.internal, executor 2, partition 8, NODE_LOCAL, 7963 bytes)

18/09/08 14:31:12 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 1.0 (TID 42) in 24843 ms on ip-172-31-84-37.ec2.internal (executor 2) (6/39)

18/09/08 14:31:19 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 1.0 (TID 48, ip-172-31-84-37.ec2.internal, executor 3, partition 9, NODE_LOCAL, 7963 bytes)

18/09/08 14:31:19 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 1.0 (TID 45) in 17778 ms on ip-172-31-84-37.ec2.internal (executor 3) (7/39)

18/09/08 14:31:32 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 1.0 (TID 49, ip-172-31-84-37.ec2.internal, executor 2, partition 10, NODE_LOCAL, 7963 bytes)

18/09/08 14:31:32 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 1.0 (TID 47) in 20875 ms on ip-172-31-84-37.ec2.internal (executor 2) (8/39)

18/09/08 14:31:35 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 1.0 (TID 50, ip-172-31-84-37.ec2.internal, executor 1, partition 11, NODE_LOCAL, 7963 bytes)

18/09/08 14:31:35 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 1.0 (TID 46) in 23542 ms on ip-172-31-84-37.ec2.internal (executor 1) (9/39)

18/09/08 14:31:35 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 1.0 (TID 51, ip-172-31-84-37.ec2.internal, executor 3, partition 12, NODE_LOCAL, 7963 bytes)

18/09/08 14:31:35 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 1.0 (TID 48) in 15996 ms on ip-172-31-84-37.ec2.internal (executor 3) (10/39)

18/09/08 14:31:47 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 1.0 (TID 52, ip-172-31-84-37.ec2.internal, executor 1, partition 13, NODE_LOCAL, 7963 bytes)

18/09/08 14:31:47 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 1.0 (TID 50) in 11977 ms on ip-172-31-84-37.ec2.internal (executor 1) (11/39)

18/09/08 14:31:48 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 1.0 (TID 53, ip-172-31-84-37.ec2.internal, executor 2, partition 14, NODE_LOCAL, 7963 bytes)

18/09/08 14:31:48 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 1.0 (TID 49) in 15850 ms on ip-172-31-84-37.ec2.internal (executor 2) (12/39)

18/09/08 14:32:00 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 1.0 (TID 54, ip-172-31-84-37.ec2.internal, executor 3, partition 15, NODE_LOCAL, 7963 bytes)

18/09/08 14:32:00 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 1.0 (TID 51) in 24756 ms on ip-172-31-84-37.ec2.internal (executor 3) (13/39)

18/09/08 14:32:05 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 1.0 (TID 55, ip-172-31-84-37.ec2.internal, executor 2, partition 16, NODE_LOCAL, 7963 bytes)

18/09/08 14:32:05 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 1.0 (TID 53) in 16739 ms on ip-172-31-84-37.ec2.internal (executor 2) (14/39)

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 11

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 21

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 3

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 7

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 5

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 24

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 22

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 6

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 17

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 15

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 14

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 9

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 10

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 19

18/09/08 14:32:09 INFO spark.ContextCleaner: Cleaned accumulator 13

18/09/08 14:32:09 INFO storage.BlockManagerInfo: Removed broadcast_1_piece0 on ip-172-31-84-37.ec2.internal:33027 in memory (size: 2.4 KB, free: 366.2 MB)

18/09/08 14:32:09 INFO storage.BlockManagerInfo: Removed broadcast_1_piece0 on ip-172-31-84-37.ec2.internal:35508 in memory (size: 2.4 KB, free: 366.2 MB)

18/09/08 14:32:09 INFO storage.BlockManagerInfo: Removed broadcast_1_piece0 on ip-172-31-84-37.ec2.internal:42678 in memory (size: 2.4 KB, free: 366.2 MB)

18/09/08 14:32:10 INFO storage.BlockManagerInfo: Removed broadcast_1_piece0 on ip-172-31-84-37.ec2.internal:40642 in memory (size: 2.4 KB, free: 366.2 MB)

18/09/08 14:32:10 INFO spark.ContextCleaner: Cleaned accumulator 18

18/09/08 14:32:10 INFO spark.ContextCleaner: Cleaned accumulator 23

18/09/08 14:32:10 INFO spark.ContextCleaner: Cleaned accumulator 20

18/09/08 14:32:10 INFO spark.ContextCleaner: Cleaned accumulator 4

18/09/08 14:32:10 INFO spark.ContextCleaner: Cleaned accumulator 1

18/09/08 14:32:10 INFO spark.ContextCleaner: Cleaned accumulator 0

18/09/08 14:32:10 INFO spark.ContextCleaner: Cleaned accumulator 8

18/09/08 14:32:10 INFO spark.ContextCleaner: Cleaned accumulator 2

18/09/08 14:32:10 INFO spark.ContextCleaner: Cleaned accumulator 12

18/09/08 14:32:10 INFO spark.ContextCleaner: Cleaned accumulator 16

18/09/08 14:32:11 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 1.0 (TID 56, ip-172-31-84-37.ec2.internal, executor 1, partition 17, NODE_LOCAL, 7963 bytes)

18/09/08 14:32:11 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 1.0 (TID 52) in 23906 ms on ip-172-31-84-37.ec2.internal (executor 1) (15/39)

18/09/08 14:32:21 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 1.0 (TID 57, ip-172-31-84-37.ec2.internal, executor 3, partition 18, NODE_LOCAL, 7963 bytes)

18/09/08 14:32:21 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 1.0 (TID 54) in 20775 ms on ip-172-31-84-37.ec2.internal (executor 3) (16/39)

18/09/08 14:32:27 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 1.0 (TID 58, ip-172-31-84-37.ec2.internal, executor 2, partition 19, NODE_LOCAL, 7963 bytes)

18/09/08 14:32:27 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 1.0 (TID 55) in 21868 ms on ip-172-31-84-37.ec2.internal (executor 2) (17/39)

18/09/08 14:32:29 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 1.0 (TID 59, ip-172-31-84-37.ec2.internal, executor 1, partition 20, NODE_LOCAL, 7963 bytes)

18/09/08 14:32:29 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 1.0 (TID 56) in 17841 ms on ip-172-31-84-37.ec2.internal (executor 1) (18/39)

18/09/08 14:32:42 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 1.0 (TID 60, ip-172-31-84-37.ec2.internal, executor 3, partition 21, NODE_LOCAL, 7963 bytes)

18/09/08 14:32:42 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 1.0 (TID 57) in 21381 ms on ip-172-31-84-37.ec2.internal (executor 3) (19/39)

18/09/08 14:32:43 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 1.0 (TID 61, ip-172-31-84-37.ec2.internal, executor 2, partition 22, NODE_LOCAL, 7963 bytes)

18/09/08 14:32:43 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 1.0 (TID 58) in 16079 ms on ip-172-31-84-37.ec2.internal (executor 2) (20/39)

18/09/08 14:32:49 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 1.0 (TID 62, ip-172-31-84-37.ec2.internal, executor 1, partition 23, NODE_LOCAL, 7963 bytes)

18/09/08 14:32:49 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 1.0 (TID 59) in 20113 ms on ip-172-31-84-37.ec2.internal (executor 1) (21/39)

18/09/08 14:33:01 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 1.0 (TID 63, ip-172-31-84-37.ec2.internal, executor 3, partition 24, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:01 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 1.0 (TID 60) in 18723 ms on ip-172-31-84-37.ec2.internal (executor 3) (22/39)

18/09/08 14:33:06 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 1.0 (TID 64, ip-172-31-84-37.ec2.internal, executor 2, partition 25, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:06 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 1.0 (TID 61) in 23482 ms on ip-172-31-84-37.ec2.internal (executor 2) (23/39)

18/09/08 14:33:08 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 1.0 (TID 65, ip-172-31-84-37.ec2.internal, executor 3, partition 26, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:08 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 1.0 (TID 63) in 7005 ms on ip-172-31-84-37.ec2.internal (executor 3) (24/39)

18/09/08 14:33:10 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 1.0 (TID 66, ip-172-31-84-37.ec2.internal, executor 1, partition 27, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:10 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 1.0 (TID 62) in 20767 ms on ip-172-31-84-37.ec2.internal (executor 1) (25/39)

18/09/08 14:33:30 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 1.0 (TID 67, ip-172-31-84-37.ec2.internal, executor 2, partition 28, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:30 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 1.0 (TID 64) in 23745 ms on ip-172-31-84-37.ec2.internal (executor 2) (26/39)

18/09/08 14:33:31 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 1.0 (TID 68, ip-172-31-84-37.ec2.internal, executor 3, partition 29, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:31 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 1.0 (TID 65) in 22821 ms on ip-172-31-84-37.ec2.internal (executor 3) (27/39)

18/09/08 14:33:33 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 1.0 (TID 69, ip-172-31-84-37.ec2.internal, executor 1, partition 30, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:33 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 1.0 (TID 66) in 23849 ms on ip-172-31-84-37.ec2.internal (executor 1) (28/39)

18/09/08 14:33:44 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 1.0 (TID 70, ip-172-31-84-37.ec2.internal, executor 2, partition 31, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:44 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 1.0 (TID 67) in 14274 ms on ip-172-31-84-37.ec2.internal (executor 2) (29/39)

18/09/08 14:33:48 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 1.0 (TID 71, ip-172-31-84-37.ec2.internal, executor 2, partition 32, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:48 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 1.0 (TID 70) in 3329 ms on ip-172-31-84-37.ec2.internal (executor 2) (30/39)

18/09/08 14:33:52 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 1.0 (TID 72, ip-172-31-84-37.ec2.internal, executor 3, partition 33, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:52 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 1.0 (TID 68) in 21305 ms on ip-172-31-84-37.ec2.internal (executor 3) (31/39)

18/09/08 14:33:54 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 1.0 (TID 73, ip-172-31-84-37.ec2.internal, executor 1, partition 34, NODE_LOCAL, 7963 bytes)

18/09/08 14:33:54 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 1.0 (TID 69) in 20061 ms on ip-172-31-84-37.ec2.internal (executor 1) (32/39)

18/09/08 14:34:10 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 1.0 (TID 74, ip-172-31-84-37.ec2.internal, executor 2, partition 35, NODE_LOCAL, 7963 bytes)

18/09/08 14:34:10 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 1.0 (TID 71) in 22335 ms on ip-172-31-84-37.ec2.internal (executor 2) (33/39)

18/09/08 14:34:17 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 1.0 (TID 75, ip-172-31-84-37.ec2.internal, executor 3, partition 36, NODE_LOCAL, 7963 bytes)

18/09/08 14:34:17 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 1.0 (TID 72) in 24788 ms on ip-172-31-84-37.ec2.internal (executor 3) (34/39)

18/09/08 14:34:17 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 1.0 (TID 76, ip-172-31-84-37.ec2.internal, executor 1, partition 37, NODE_LOCAL, 7963 bytes)

18/09/08 14:34:17 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 1.0 (TID 73) in 23608 ms on ip-172-31-84-37.ec2.internal (executor 1) (35/39)

18/09/08 14:34:32 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 1.0 (TID 77, ip-172-31-84-37.ec2.internal, executor 3, partition 38, NODE_LOCAL, 7963 bytes)

18/09/08 14:34:32 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 1.0 (TID 75) in 14988 ms on ip-172-31-84-37.ec2.internal (executor 3) (36/39)

18/09/08 14:34:32 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 1.0 (TID 74) in 22203 ms on ip-172-31-84-37.ec2.internal (executor 2) (37/39)

18/09/08 14:34:36 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 1.0 (TID 77) in 3843 ms on ip-172-31-84-37.ec2.internal (executor 3) (38/39)

18/09/08 14:34:38 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 1.0 (TID 76) in 20579 ms on ip-172-31-84-37.ec2.internal (executor 1) (39/39)

18/09/08 14:34:38 INFO cluster.YarnScheduler: Removed TaskSet 1.0, whose tasks have all completed, from pool

18/09/08 14:34:38 INFO scheduler.DAGScheduler: ResultStage 1 (runJob at SparkHadoopWriter.scala:78) finished in 249.696 s

18/09/08 14:34:38 INFO scheduler.DAGScheduler: Job 1 finished: runJob at SparkHadoopWriter.scala:78, took 249.711678 s

18/09/08 14:34:40 INFO io.SparkHadoopWriter: Job job_20180908143028_0004 committed.

2,2017-10-01 00:15:30,2017-10-01 00:25:11,1,2.17,1,N,141,142,1,9,0.5,0.5,2.06,0,0.3,12.36

18/09/08 14:34:40 INFO spark.SparkContext: Invoking stop() from shutdown hook

18/09/08 14:34:40 INFO server.AbstractConnector: Stopped Spark@22ee2d0{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}

18/09/08 14:34:40 INFO ui.SparkUI: Stopped Spark web UI at http://ip-172-31-84-37.ec2.internal:4040

18/09/08 14:34:40 INFO cluster.YarnClientSchedulerBackend: Interrupting monitor thread

18/09/08 14:34:40 INFO cluster.YarnClientSchedulerBackend: Shutting down all executors

18/09/08 14:34:40 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Asking each executor to shut down

18/09/08 14:34:40 INFO cluster.SchedulerExtensionServices: Stopping SchedulerExtensionServices

(serviceOption=None,

services=List(),

started=false)

18/09/08 14:34:40 INFO cluster.YarnClientSchedulerBackend: Stopped

18/09/08 14:34:40 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

18/09/08 14:34:41 INFO memory.MemoryStore: MemoryStore cleared

18/09/08 14:34:41 INFO storage.BlockManager: BlockManager stopped

18/09/08 14:34:41 INFO storage.BlockManagerMaster: BlockManagerMaster stopped

18/09/08 14:34:41 INFO scheduler.OutputCommitCoordinator\$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

18/09/08 14:34:41 INFO spark.SparkContext: Successfully stopped SparkContext

18/09/08 14:34:41 INFO util.ShutdownHookManager: Shutdown hook called

18/09/08 14:34:41 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-33cd9f9e-ocd3-4b90-b6dd-9of834a4adf3

18/09/08 14:34:41 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-57bd6d85-7780-4f09-8fod-obc43eed2530

```
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/pig1_single_row_lookup/part* | wc -l
```

```
1
```

```
[ec2-user@ip-172-31-84-37 ~]$
```

SPARK LOG2: FILTER

```
ec2-user@ip-172-31-84-37:~$ cat /var/log/hadoop-hdfs/hadoop-ec2-user@ip-172-31-84-37-16-1.log
18/09/08 14:04:06 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 1.0 (TID 73, ip-172-31-84-37.ec2.internal, executor 2, partition 34, MODE_LOCAL, 7963 bytes)
18/09/08 14:04:06 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 1.0 (TID 73) in 11023 ms on ip-172-31-84-37.ec2.internal (executor 2) (32/39)
18/09/08 14:04:14 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 1.0 (TID 74, ip-172-31-84-37.ec2.internal, executor 1, partition 35, MODE_LOCAL, 7963 bytes)
18/09/08 14:04:14 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 1.0 (TID 74) in 8434 ms on ip-172-31-84-37.ec2.internal (executor 1) (33/39)
18/09/08 14:04:16 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 1.0 (TID 75, ip-172-31-84-37.ec2.internal, executor 2, partition 36, MODE_LOCAL, 7963 bytes)
18/09/08 14:04:16 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 1.0 (TID 75) in 8143 ms on ip-172-31-84-37.ec2.internal (executor 2) (34/39)
18/09/08 14:04:16 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 1.0 (TID 76, ip-172-31-84-37.ec2.internal, executor 3, partition 37, MODE_LOCAL, 7963 bytes)
18/09/08 14:04:16 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 1.0 (TID 76) in 8292 ms on ip-172-31-84-37.ec2.internal (executor 3) (35/39)
18/09/08 14:04:21 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 1.0 (TID 77, ip-172-31-84-37.ec2.internal, executor 1, partition 38, MODE_LOCAL, 7963 bytes)
18/09/08 14:04:21 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 1.0 (TID 77) in 7290 ms on ip-172-31-84-37.ec2.internal (executor 1) (36/39)
18/09/08 14:04:23 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 1.0 (TID 75) in 8894 ms on ip-172-31-84-37.ec2.internal (executor 2) (37/39)
18/09/08 14:04:24 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 1.0 (TID 77) in 2544 ms on ip-172-31-84-37.ec2.internal (executor 1) (38/39)
18/09/08 14:04:24 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 1.0 (TID 76) in 7963 ms on ip-172-31-84-37.ec2.internal (executor 3) (39/39)
18/09/08 14:04:24 INFO cluster.YarnScheduler: Removed TaskSet 1.0, whose tasks have all completed, from pool
18/09/08 14:04:24 INFO scheduler.DAGScheduler: ResultStage 1 (runJob at SparkHadoopWriter.scala:70) finished in 97.062 s
18/09/08 14:04:24 INFO io.SparkHadoopWriter: Job job_20180908140217_0004 committed.
18/09/08 14:04:26 INFO server.AbstractConnector: Accepted Spark2WebSocketClient (http://172.31.84.37:4040)
18/09/08 14:04:26 INFO ui.SparkUI: Stopped Spark web UI at http://ip-172-31-84-37.ec2.internal:4040
18/09/08 14:04:26 INFO cluster.YarnClientSchedulerBackend: Interrupting monitor thread
18/09/08 14:04:27 INFO cluster.YarnClientSchedulerBackend: Shutting down all executors
18/09/08 14:04:27 INFO cluster.YarnClientSchedulerBackend: Asking each executor to shut down
18/09/08 14:04:27 INFO cluster.SchedulerExtensionService: Stopping SchedulerExtensionServices
(services=List(),
services=List()),
started=List())
18/09/08 14:04:27 INFO cluster.YarnClientSchedulerBackend: Stopped
18/09/08 14:04:27 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/09/08 14:04:27 INFO memory.MemoryStore: MemoryStore cleared
18/09/08 14:04:27 INFO storage.BlockManager: BlockManager stopped
18/09/08 14:04:27 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
18/09/08 14:04:27 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/09/08 14:04:27 INFO spark.SparkContext: Successfully stopped SparkContext
18/09/08 14:04:27 INFO util.ShutdownHookManager: Shutdown hook called
18/09/08 14:04:27 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-af08ed1-614a-4c10-84db-6d561e1fa889
18/09/08 14:04:27 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-54c5b91-567c-493a-83f1-b7b1a17eeb2
ec2-user@ip-172-31-84-37 ~$ hadoop fs -cat /user/ec2-user/spark_assignment/output/spark2_filter/part* | wc -l
31024
ec2-user@ip-172-31-84-37 ~$
```

```
[ec2-user@ip-172-31-84-37 ~]$ spark2-submit --class sparkAssignment.Filter.Filter \
```

```
> --master yarn --deploy-mode client \
```

```
> --name program2 --conf "spark.app.id=program2_id" \
```

```
> Filter-o.o.1-SNAPSHOT-jar-with-dependencies.jar /user/ec2-
```

```
user/spark_assignment/input/yellow_tripdata* /user/ec2-user/spark_assignment/output/spark2_filter
```

```
18/09/08 14:00:04 INFO spark.SparkContext: Running Spark version 2.3.0.cloudera2
```

```
18/09/08 14:00:04 INFO spark.SparkContext: Submitted application: Filter
```

```
18/09/08 14:00:04 INFO spark.SecurityManager: Changing view acls to: ec2-user
```

```
18/09/08 14:00:04 INFO spark.SecurityManager: Changing modify acls to: ec2-user
```

```
18/09/08 14:00:04 INFO spark.SecurityManager: Changing view acls groups to:
```

```
18/09/08 14:00:04 INFO spark.SecurityManager: Changing modify acls groups to:
```

```
18/09/08 14:00:04 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(ec2-user); groups with view permissions: Set(); users with modify permissions: Set(ec2-user); groups with modify permissions: Set()
```

18/09/08 14:00:05 INFO util.Utils: Successfully started service 'sparkDriver' on port 42432.

18/09/08 14:00:05 INFO spark.SparkEnv: Registering MapOutputTracker

18/09/08 14:00:05 INFO spark.SparkEnv: Registering BlockManagerMaster

18/09/08 14:00:05 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information

18/09/08 14:00:05 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up

18/09/08 14:00:05 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-e300d5ee-03a1-4ca4-b32e-8c8ado82c95e

18/09/08 14:00:05 INFO memory.MemoryStore: MemoryStore started with capacity 366.3 MB

18/09/08 14:00:06 INFO spark.SparkEnv: Registering OutputCommitCoordinator

18/09/08 14:00:06 INFO util.log: Logging initialized @9085ms

18/09/08 14:00:07 INFO server.Server: jetty-9.3.z-SNAPSHOT

18/09/08 14:00:07 INFO server.Server: Started @9550ms

18/09/08 14:00:07 INFO server.AbstractConnector: Started ServerConnector@22ee2do{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}

18/09/08 14:00:07 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.

18/09/08 14:00:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5cad8b7d{/jobs,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@10567255{/jobs/json,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1c4ee95c{/jobs/job,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6548bb7d{/jobs/job/json,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@54336c81{/stages,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@35e52059{/stages/json,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@49bd54f7{/stages/stage,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@79ab3a71{/stages/stage/json,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3d829787{/stages/pool,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@51bde877{/stages/pool/json,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@492fc69e{/storage,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@2fb68ec6{/storage/json,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3add81c4{/storage/rdd,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@1c65121{/storage/rdd/json,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@57dc9128{/environment,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@17aeg8d7{/environment/json,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@6ac4944a{/executors,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@39fc6b2c{/executors/json,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3ee39dao{/executors/threadDump,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@7cc9ce8{/executors/threadDump/json,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@cob41d6{/static,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@75201592{/,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@aa5455e{/api,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3b4ef7{/jobs/job/kill,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5987e932{/stages/stage/kill,null,AVAILABLE,@Spark}

18/09/08 14:00:07 INFO ui.SparkUI: Bound SparkUI to o.o.o.o, and started at http://ip-172-31-84-37.ec2.internal:4040

18/09/08 14:00:07 INFO spark.SparkContext: Added JAR file:/home/ec2-user/Filter-o.o.1-SNAPSHOT-jar-with-dependencies.jar at spark://ip-172-31-84-37.ec2.internal:42432/jars/Filter-o.o.1-SNAPSHOT-jar-with-dependencies.jar with timestamp 1536415207621

18/09/08 14:00:07 INFO util.Utils: Using initial executors = 0, max of
spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and
spark.executor.instances

18/09/08 14:00:11 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

18/09/08 14:00:12 INFO yarn.Client: Requesting a new application from cluster with 1 NodeManagers

18/09/08 14:00:12 INFO yarn.Client: Verifying our application has not requested more than the
maximum memory capability of the cluster (8192 MB per container)

18/09/08 14:00:12 INFO yarn.Client: Will allocate AM container, with 896 MB memory including 384 MB
overhead

18/09/08 14:00:12 INFO yarn.Client: Setting up container launch context for our AM

18/09/08 14:00:12 INFO yarn.Client: Setting up the launch environment for our AM container

18/09/08 14:00:12 INFO yarn.Client: Preparing resources for our AM container

18/09/08 14:00:13 INFO yarn.Client: Uploading resource file:/tmp/spark-afdo86d1-614a-4c10-8dcd-bd05e3e8a889/___spark_conf__5625926052993618680.zip -> hdfs://ip-172-31-84-37.ec2.internal:8020/user/ec2-user/.sparkStaging/application_1536410193861_0144/___spark_conf__.zip

18/09/08 14:00:14 INFO spark.SecurityManager: Changing view acls to: ec2-user

18/09/08 14:00:14 INFO spark.SecurityManager: Changing modify acls to: ec2-user

18/09/08 14:00:14 INFO spark.SecurityManager: Changing view acls groups to:

18/09/08 14:00:14 INFO spark.SecurityManager: Changing modify acls groups to:

18/09/08 14:00:14 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(ec2-user); groups with view permissions: Set(); users with modify permissions: Set(ec2-user); groups with modify permissions: Set()

18/09/08 14:00:18 INFO yarn.Client: Submitting application application_1536410193861_0144 to ResourceManager

18/09/08 14:00:18 INFO impl.YarnClientImpl: Submitted application application_1536410193861_0144

18/09/08 14:00:18 INFO cluster.SchedulerExtensionServices: Starting Yarn extension services with app application_1536410193861_0144 and attemptId None

18/09/08 14:00:19 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:19 INFO yarn.Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: N/A

ApplicationMaster RPC port: -1

queue: root.users.ec2-user

start time: 1536415218600

final status: UNDEFINED

tracking URL: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0144/

user: ec2-user

18/09/08 14:00:20 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:21 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:22 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:23 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:24 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:25 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:26 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:27 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:28 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:29 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:30 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:31 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:32 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:33 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:35 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:36 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:37 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:38 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:38 INFO cluster.YarnClientSchedulerBackend: Add WebUI Filter.
org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter, Map(PROXY_HOSTS -> ip-172-31-84-37.ec2.internal, PROXY_URI_BASES -> http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0144),
/proxy/application_1536410193861_0144

18/09/08 14:00:38 INFO ui.JettyUtils: Adding filter:
org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter

18/09/08 14:00:39 INFO yarn.Client: Application report for application_1536410193861_0144 (state: ACCEPTED)

18/09/08 14:00:39 INFO cluster.YarnSchedulerBackend\$YarnSchedulerEndpoint: ApplicationMaster registered as NettyRpcEndpointRef(spark-client://YarnAM)

18/09/08 14:00:40 INFO yarn.Client: Application report for application_1536410193861_0144 (state: RUNNING)

18/09/08 14:00:40 INFO yarn.Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: 172.31.84.37

ApplicationMaster RPC port: 0

queue: root.users.ec2-user

start time: 1536415218600

final status: UNDEFINED

tracking URL: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0144/

user: ec2-user

18/09/08 14:00:40 INFO cluster.YarnClientSchedulerBackend: Application application_1536410193861_0144 has started running.

18/09/08 14:00:40 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 37977.

18/09/08 14:00:40 INFO netty.NettyBlockTransferService: Server created on ip-172-31-84-37.ec2.internal:37977

18/09/08 14:00:40 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy

18/09/08 14:00:40 INFO storage.BlockManagerMaster: Registering BlockManager BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 37977, None)

18/09/08 14:00:40 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:37977 with 366.3 MB RAM, BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 37977, None)

18/09/08 14:00:40 INFO storage.BlockManagerMaster: Registered BlockManager BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 37977, None)

18/09/08 14:00:40 INFO storage.BlockManager: external shuffle service port = 7337

18/09/08 14:00:40 INFO storage.BlockManager: Initialized BlockManager: BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 37977, None)

18/09/08 14:00:40 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@4aaecabd{/metrics/json,null,AVAILABLE,@Spark}

18/09/08 14:00:41 INFO scheduler.EventLoggingListener: Logging events to hdfs://ip-172-31-84-37.ec2.internal:8020/user/spark/spark2ApplicationHistory/application_1536410193861_0144

18/09/08 14:00:41 INFO util.Utils: Using initial executors = 0, max of
spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and
spark.executor.instances

18/09/08 14:00:41 INFO cluster.YarnClientSchedulerBackend: SchedulerBackend is ready for scheduling
beginning after reached minRegisteredResourcesRatio: 0.8

18/09/08 14:00:42 INFO memory.MemoryStore: Block broadcast_0 stored as values in memory
(estimated size 303.3 KB, free 366.0 MB)

18/09/08 14:00:43 INFO memory.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory
(estimated size 27.1 KB, free 366.0 MB)

18/09/08 14:00:43 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on ip-172-31-84-37.ec2.internal:37977 (size: 27.1 KB, free: 366.3 MB)

18/09/08 14:00:43 INFO spark.SparkContext: Created broadcast 0 from textFile at Filter.java:15

18/09/08 14:00:44 INFO mapred.FileInputFormat: Total input paths to process : 6

18/09/08 14:00:44 INFO spark.SparkContext: Starting job: collect at Filter.java:36

18/09/08 14:00:44 INFO scheduler.DAGScheduler: Got job 0 (collect at Filter.java:36) with 39 output
partitions

18/09/08 14:00:44 INFO scheduler.DAGScheduler: Final stage: ResultStage 0 (collect at Filter.java:36)

18/09/08 14:00:44 INFO scheduler.DAGScheduler: Parents of final stage: List()

18/09/08 14:00:44 INFO scheduler.DAGScheduler: Missing parents: List()

18/09/08 14:00:44 INFO scheduler.DAGScheduler: Submitting ResultStage 0 (MapPartitionsRDD[3] at
filter at Filter.java:21), which has no missing parents

18/09/08 14:00:44 INFO memory.MemoryStore: Block broadcast_1 stored as values in memory
(estimated size 4.3 KB, free 366.0 MB)

18/09/08 14:00:44 INFO memory.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 2.4 KB, free 366.0 MB)

18/09/08 14:00:44 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-84-37.ec2.internal:37977 (size: 2.4 KB, free: 366.3 MB)

18/09/08 14:00:44 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1039

18/09/08 14:00:44 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 0 (MapPartitionsRDD[3] at filter at Filter.java:21) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/09/08 14:00:44 INFO cluster.YarnScheduler: Adding task set 0.0 with 39 tasks

18/09/08 14:00:45 INFO spark.ExecutorAllocationManager: Requesting 1 new executor because tasks are backlogged (new desired total will be 1)

18/09/08 14:00:46 INFO spark.ExecutorAllocationManager: Requesting 2 new executors because tasks are backlogged (new desired total will be 3)

18/09/08 14:00:47 INFO spark.ExecutorAllocationManager: Requesting 4 new executors because tasks are backlogged (new desired total will be 7)

18/09/08 14:00:48 INFO spark.ExecutorAllocationManager: Requesting 8 new executors because tasks are backlogged (new desired total will be 15)

18/09/08 14:00:49 INFO spark.ExecutorAllocationManager: Requesting 16 new executors because tasks are backlogged (new desired total will be 31)

18/09/08 14:00:50 INFO spark.ExecutorAllocationManager: Requesting 8 new executors because tasks are backlogged (new desired total will be 39)

18/09/08 14:00:58 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (172.31.84.37:43606) with ID 2

18/09/08 14:00:58 INFO spark.ExecutorAllocationManager: New executor 2 has registered (new total is 1)

18/09/08 14:00:58 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, ip-172-31-84-37.ec2.internal, executor 2, partition 0, NODE_LOCAL, 7963 bytes)

18/09/08 14:00:58 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:45181 with 366.3 MB RAM, BlockManagerId(2, ip-172-31-84-37.ec2.internal, 45181, None)

18/09/08 14:00:59 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (172.31.84.37:43608) with ID 1

18/09/08 14:00:59 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, ip-172-31-84-37.ec2.internal, executor 1, partition 1, NODE_LOCAL, 7963 bytes)

18/09/08 14:00:59 INFO spark.ExecutorAllocationManager: New executor 1 has registered (new total is 2)

18/09/08 14:00:59 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:36030 with 366.3 MB RAM, BlockManagerId(1, ip-172-31-84-37.ec2.internal, 36030, None)

18/09/08 14:00:59 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (172.31.84.37:43614) with ID 3

18/09/08 14:00:59 INFO spark.ExecutorAllocationManager: New executor 3 has registered (new total is 3)

18/09/08 14:00:59 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, ip-172-31-84-37.ec2.internal, executor 3, partition 2, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:00 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:41502 with 366.3 MB RAM, BlockManagerId(3, ip-172-31-84-37.ec2.internal, 41502, None)

18/09/08 14:01:03 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-84-37.ec2.internal:41502 (size: 2.4 KB, free: 366.3 MB)

18/09/08 14:01:03 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-84-37.ec2.internal:36030 (size: 2.4 KB, free: 366.3 MB)

18/09/08 14:01:03 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-84-37.ec2.internal:45181 (size: 2.4 KB, free: 366.3 MB)

18/09/08 14:01:04 INFO storage.BlockManagerInfo: Added broadcast_o_piece0 in memory on ip-172-31-84-37.ec2.internal:41502 (size: 27.1 KB, free: 366.3 MB)

18/09/08 14:01:04 INFO storage.BlockManagerInfo: Added broadcast_o_piece0 in memory on ip-172-31-84-37.ec2.internal:45181 (size: 27.1 KB, free: 366.3 MB)

18/09/08 14:01:05 INFO storage.BlockManagerInfo: Added broadcast_o_piece0 in memory on ip-172-31-84-37.ec2.internal:36030 (size: 27.1 KB, free: 366.3 MB)

18/09/08 14:01:19 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, ip-172-31-84-37.ec2.internal, executor 3, partition 3, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:20 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 20251 ms on ip-172-31-84-37.ec2.internal (executor 3) (1/39)

18/09/08 14:01:21 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 0.0 (TID 4, ip-172-31-84-37.ec2.internal, executor 2, partition 4, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:21 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 23287 ms on ip-172-31-84-37.ec2.internal (executor 2) (2/39)

18/09/08 14:01:22 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 0.0 (TID 5, ip-172-31-84-37.ec2.internal, executor 1, partition 5, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:22 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 0.0 (TID 1) in 22860 ms on ip-172-31-84-37.ec2.internal (executor 1) (3/39)

18/09/08 14:01:26 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 0.0 (TID 6, ip-172-31-84-37.ec2.internal, executor 1, partition 6, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:26 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 0.0 (TID 5) in 4234 ms on ip-172-31-84-37.ec2.internal (executor 1) (4/39)

18/09/08 14:01:27 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, ip-172-31-84-37.ec2.internal, executor 3, partition 7, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:27 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 0.0 (TID 3) in 7054 ms on ip-172-31-84-37.ec2.internal (executor 3) (5/39)

18/09/08 14:01:31 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, ip-172-31-84-37.ec2.internal, executor 2, partition 8, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:31 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 10148 ms on ip-172-31-84-37.ec2.internal (executor 2) (6/39)

18/09/08 14:01:31 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, ip-172-31-84-37.ec2.internal, executor 3, partition 9, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:31 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 4884 ms on ip-172-31-84-37.ec2.internal (executor 3) (7/39)

18/09/08 14:01:35 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, ip-172-31-84-37.ec2.internal, executor 1, partition 10, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:35 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 9462 ms on ip-172-31-84-37.ec2.internal (executor 1) (8/39)

18/09/08 14:01:39 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, ip-172-31-84-37.ec2.internal, executor 3, partition 11, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:39 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 7745 ms on ip-172-31-84-37.ec2.internal (executor 3) (9/39)

18/09/08 14:01:40 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 0.0 (TID 12, ip-172-31-84-37.ec2.internal, executor 2, partition 12, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:40 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 8293 ms on ip-172-31-84-37.ec2.internal (executor 2) (10/39)

18/09/08 14:01:43 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13, ip-172-31-84-37.ec2.internal, executor 1, partition 13, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:43 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 7862 ms on ip-172-31-84-37.ec2.internal (executor 1) (11/39)

18/09/08 14:01:43 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, ip-172-31-84-37.ec2.internal, executor 3, partition 14, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:43 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 0.0 (TID 11) in 4017 ms on ip-172-31-84-37.ec2.internal (executor 3) (12/39)

18/09/08 14:01:48 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, ip-172-31-84-37.ec2.internal, executor 1, partition 15, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:48 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 0.0 (TID 13) in 4856 ms on ip-172-31-84-37.ec2.internal (executor 1) (13/39)

18/09/08 14:01:49 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, ip-172-31-84-37.ec2.internal, executor 2, partition 16, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:49 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 9166 ms on ip-172-31-84-37.ec2.internal (executor 2) (14/39)

18/09/08 14:01:51 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 0.0 (TID 17, ip-172-31-84-37.ec2.internal, executor 3, partition 17, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:51 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 0.0 (TID 14) in 8280 ms on ip-172-31-84-37.ec2.internal (executor 3) (15/39)

18/09/08 14:01:55 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 0.0 (TID 18, ip-172-31-84-37.ec2.internal, executor 1, partition 18, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:55 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 7346 ms on ip-172-31-84-37.ec2.internal (executor 1) (16/39)

18/09/08 14:01:59 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 0.0 (TID 19, ip-172-31-84-37.ec2.internal, executor 2, partition 19, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:59 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 9909 ms on ip-172-31-84-37.ec2.internal (executor 2) (17/39)

18/09/08 14:01:59 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 0.0 (TID 20, ip-172-31-84-37.ec2.internal, executor 3, partition 20, NODE_LOCAL, 7963 bytes)

18/09/08 14:01:59 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 0.0 (TID 17) in 7863 ms on ip-172-31-84-37.ec2.internal (executor 3) (18/39)

18/09/08 14:02:03 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 0.0 (TID 21, ip-172-31-84-37.ec2.internal, executor 1, partition 21, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:03 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 0.0 (TID 18) in 7960 ms on ip-172-31-84-37.ec2.internal (executor 1) (19/39)

18/09/08 14:02:07 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 0.0 (TID 22, ip-172-31-84-37.ec2.internal, executor 2, partition 22, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:07 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 0.0 (TID 19) in 8037 ms on ip-172-31-84-37.ec2.internal (executor 2) (20/39)

18/09/08 14:02:08 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 0.0 (TID 23, ip-172-31-84-37.ec2.internal, executor 3, partition 23, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:08 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 0.0 (TID 20) in 8518 ms on ip-172-31-84-37.ec2.internal (executor 3) (21/39)

18/09/08 14:02:13 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 0.0 (TID 24, ip-172-31-84-37.ec2.internal, executor 1, partition 24, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:13 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 0.0 (TID 21) in 10155 ms on ip-172-31-84-37.ec2.internal (executor 1) (22/39)

18/09/08 14:02:16 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 0.0 (TID 25, ip-172-31-84-37.ec2.internal, executor 2, partition 25, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:16 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 0.0 (TID 22) in 9503 ms on ip-172-31-84-37.ec2.internal (executor 2) (23/39)

18/09/08 14:02:17 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 0.0 (TID 26, ip-172-31-84-37.ec2.internal, executor 1, partition 26, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:17 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 0.0 (TID 24) in 3112 ms on ip-172-31-84-37.ec2.internal (executor 1) (24/39)

18/09/08 14:02:18 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 0.0 (TID 27, ip-172-31-84-37.ec2.internal, executor 3, partition 27, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:18 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 0.0 (TID 23) in 10433 ms on ip-172-31-84-37.ec2.internal (executor 3) (25/39)

18/09/08 14:02:23 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 0.0 (TID 28, ip-172-31-84-37.ec2.internal, executor 2, partition 28, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:23 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 0.0 (TID 25) in 6524 ms on ip-172-31-84-37.ec2.internal (executor 2) (26/39)

18/09/08 14:02:23 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 0.0 (TID 29, ip-172-31-84-37.ec2.internal, executor 1, partition 29, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:23 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 0.0 (TID 26) in 6278 ms on ip-172-31-84-37.ec2.internal (executor 1) (27/39)

18/09/08 14:02:26 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 0.0 (TID 30, ip-172-31-84-37.ec2.internal, executor 3, partition 30, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:26 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 0.0 (TID 27) in 8009 ms on ip-172-31-84-37.ec2.internal (executor 3) (28/39)

18/09/08 14:02:29 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 0.0 (TID 31, ip-172-31-84-37.ec2.internal, executor 2, partition 31, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:29 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 0.0 (TID 28) in 6325 ms on ip-172-31-84-37.ec2.internal (executor 2) (29/39)

18/09/08 14:02:30 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 0.0 (TID 32, ip-172-31-84-37.ec2.internal, executor 2, partition 32, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:30 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 0.0 (TID 31) in 883 ms on ip-172-31-84-37.ec2.internal (executor 2) (30/39)

18/09/08 14:02:31 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 0.0 (TID 33, ip-172-31-84-37.ec2.internal, executor 1, partition 33, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:31 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 0.0 (TID 29) in 8569 ms on ip-172-31-84-37.ec2.internal (executor 1) (31/39)

18/09/08 14:02:32 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 0.0 (TID 34, ip-172-31-84-37.ec2.internal, executor 3, partition 34, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:32 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 0.0 (TID 30) in 5875 ms on ip-172-31-84-37.ec2.internal (executor 3) (32/39)

18/09/08 14:02:35 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 0.0 (TID 35, ip-172-31-84-37.ec2.internal, executor 2, partition 35, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:35 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 0.0 (TID 32) in 5144 ms on ip-172-31-84-37.ec2.internal (executor 2) (33/39)

18/09/08 14:02:39 INFO scheduler.TaskSetManager: Starting task 36.o in stage 0.o (TID 36, ip-172-31-84-37.ec2.internal, executor 1, partition 36, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:39 INFO scheduler.TaskSetManager: Finished task 33.o in stage 0.o (TID 33) in 7155 ms on ip-172-31-84-37.ec2.internal (executor 1) (34/39)

18/09/08 14:02:40 INFO scheduler.TaskSetManager: Starting task 37.o in stage 0.o (TID 37, ip-172-31-84-37.ec2.internal, executor 3, partition 37, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:40 INFO scheduler.TaskSetManager: Finished task 34.o in stage 0.o (TID 34) in 8261 ms on ip-172-31-84-37.ec2.internal (executor 3) (35/39)

18/09/08 14:02:42 INFO scheduler.TaskSetManager: Starting task 38.o in stage 0.o (TID 38, ip-172-31-84-37.ec2.internal, executor 2, partition 38, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:42 INFO scheduler.TaskSetManager: Finished task 35.o in stage 0.o (TID 35) in 7269 ms on ip-172-31-84-37.ec2.internal (executor 2) (36/39)

18/09/08 14:02:44 INFO scheduler.TaskSetManager: Finished task 38.o in stage 0.o (TID 38) in 1449 ms on ip-172-31-84-37.ec2.internal (executor 2) (37/39)

18/09/08 14:02:46 INFO scheduler.TaskSetManager: Finished task 37.o in stage 0.o (TID 37) in 5743 ms on ip-172-31-84-37.ec2.internal (executor 3) (38/39)

18/09/08 14:02:46 INFO scheduler.TaskSetManager: Finished task 36.o in stage 0.o (TID 36) in 7868 ms on ip-172-31-84-37.ec2.internal (executor 1) (39/39)

18/09/08 14:02:46 INFO cluster.YarnScheduler: Removed TaskSet 0.o, whose tasks have all completed, from pool

18/09/08 14:02:47 INFO scheduler.DAGScheduler: ResultStage 0 (collect at Filter.java:36) finished in 122.409 s

18/09/08 14:02:47 INFO scheduler.DAGScheduler: Job 0 finished: collect at Filter.java:36, took 122.956739 s

18/09/08 14:02:47 INFO Configuration.deprecation: mapred.output.dir is deprecated. Instead, use mapreduce.output.fileoutputformat.outputdir

18/09/08 14:02:47 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/09/08 14:02:47 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/09/08 14:02:47 INFO spark.SparkContext: Starting job: runJob at SparkHadoopWriter.scala:78

18/09/08 14:02:47 INFO scheduler.DAGScheduler: Got job 1 (runJob at SparkHadoopWriter.scala:78) with 39 output partitions

18/09/08 14:02:47 INFO scheduler.DAGScheduler: Final stage: ResultStage 1 (runJob at SparkHadoopWriter.scala:78)

18/09/08 14:02:47 INFO scheduler.DAGScheduler: Parents of final stage: List()

18/09/08 14:02:47 INFO scheduler.DAGScheduler: Missing parents: List()

18/09/08 14:02:47 INFO scheduler.DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[4] at saveAsTextFile at Filter.java:39), which has no missing parents

18/09/08 14:02:47 INFO memory.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 81.5 KB, free 365.9 MB)

18/09/08 14:02:47 INFO memory.MemoryStore: Block broadcast_2_pieceo stored as bytes in memory (estimated size 30.3 KB, free 365.9 MB)

18/09/08 14:02:47 INFO storage.BlockManagerInfo: Added broadcast_2_pieceo in memory on ip-172-31-84-37.ec2.internal:37977 (size: 30.3 KB, free: 366.2 MB)

18/09/08 14:02:47 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1039

18/09/08 14:02:47 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 1 (MapPartitionsRDD[4] at saveAsTextFile at Filter.java:39) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/09/08 14:02:47 INFO cluster.YarnScheduler: Adding task set 1.0 with 39 tasks

18/09/08 14:02:47 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 39, ip-172-31-84-37.ec2.internal, executor 1, partition 0, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:47 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 40, ip-172-31-84-37.ec2.internal, executor 3, partition 1, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:47 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 41, ip-172-31-84-37.ec2.internal, executor 2, partition 2, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:47 INFO storage.BlockManagerInfo: Added broadcast_2_pieceo in memory on ip-172-31-84-37.ec2.internal:41502 (size: 30.3 KB, free: 366.2 MB)

18/09/08 14:02:47 INFO storage.BlockManagerInfo: Added broadcast_2_pieceo in memory on ip-172-31-84-37.ec2.internal:45181 (size: 30.3 KB, free: 366.2 MB)

18/09/08 14:02:47 INFO storage.BlockManagerInfo: Added broadcast_2_pieceo in memory on ip-172-31-84-37.ec2.internal:36030 (size: 30.3 KB, free: 366.2 MB)

18/09/08 14:02:48 INFO spark.ExecutorAllocationManager: Requesting 4 new executors because tasks are backlogged (new desired total will be 4)

18/09/08 14:02:49 INFO spark.ExecutorAllocationManager: Requesting 1 new executor because tasks are backlogged (new desired total will be 5)

18/09/08 14:02:50 INFO spark.ExecutorAllocationManager: Requesting 2 new executors because tasks are backlogged (new desired total will be 7)

18/09/08 14:02:51 INFO spark.ExecutorAllocationManager: Requesting 4 new executors because tasks are backlogged (new desired total will be 11)

18/09/08 14:02:52 INFO spark.ExecutorAllocationManager: Requesting 8 new executors because tasks are backlogged (new desired total will be 19)

18/09/08 14:02:53 INFO spark.ExecutorAllocationManager: Requesting 16 new executors because tasks are backlogged (new desired total will be 35)

18/09/08 14:02:54 INFO spark.ExecutorAllocationManager: Requesting 4 new executors because tasks are backlogged (new desired total will be 39)

18/09/08 14:02:55 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 42, ip-172-31-84-37.ec2.internal, executor 1, partition 3, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:55 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 39) in 7893 ms on ip-172-31-84-37.ec2.internal (executor 1) (1/39)

18/09/08 14:02:56 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 1.0 (TID 43, ip-172-31-84-37.ec2.internal, executor 2, partition 4, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:56 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 1.0 (TID 41) in 8751 ms on ip-172-31-84-37.ec2.internal (executor 2) (2/39)

18/09/08 14:02:56 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 1.0 (TID 44, ip-172-31-84-37.ec2.internal, executor 3, partition 5, NODE_LOCAL, 7963 bytes)

18/09/08 14:02:56 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 1.0 (TID 40) in 9430 ms on ip-172-31-84-37.ec2.internal (executor 3) (3/39)

18/09/08 14:03:01 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 1.0 (TID 45, ip-172-31-84-37.ec2.internal, executor 3, partition 6, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:01 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 1.0 (TID 44) in 4569 ms on ip-172-31-84-37.ec2.internal (executor 3) (4/39)

18/09/08 14:03:01 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 1.0 (TID 46, ip-172-31-84-37.ec2.internal, executor 1, partition 7, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:01 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 1.0 (TID 42) in 6574 ms on ip-172-31-84-37.ec2.internal (executor 1) (5/39)

18/09/08 14:03:04 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 1.0 (TID 47, ip-172-31-84-37.ec2.internal, executor 2, partition 8, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:04 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 1.0 (TID 43) in 7903 ms on ip-172-31-84-37.ec2.internal (executor 2) (6/39)

18/09/08 14:03:08 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 1.0 (TID 48, ip-172-31-84-37.ec2.internal, executor 3, partition 9, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:08 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 1.0 (TID 45) in 6788 ms on ip-172-31-84-37.ec2.internal (executor 3) (7/39)

18/09/08 14:03:10 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 1.0 (TID 49, ip-172-31-84-37.ec2.internal, executor 1, partition 10, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:10 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 1.0 (TID 46) in 8770 ms on ip-172-31-84-37.ec2.internal (executor 1) (8/39)

18/09/08 14:03:12 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 1.0 (TID 50, ip-172-31-84-37.ec2.internal, executor 2, partition 11, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:12 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 1.0 (TID 47) in 8423 ms on ip-172-31-84-37.ec2.internal (executor 2) (9/39)

18/09/08 14:03:15 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 1.0 (TID 51, ip-172-31-84-37.ec2.internal, executor 3, partition 12, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:15 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 1.0 (TID 48) in 7654 ms on ip-172-31-84-37.ec2.internal (executor 3) (10/39)

18/09/08 14:03:16 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 1.0 (TID 52, ip-172-31-84-37.ec2.internal, executor 1, partition 13, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:16 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 1.0 (TID 49) in 6140 ms on ip-172-31-84-37.ec2.internal (executor 1) (11/39)

18/09/08 14:03:17 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 1.0 (TID 53, ip-172-31-84-37.ec2.internal, executor 2, partition 14, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:17 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 1.0 (TID 50) in 4782 ms on ip-172-31-84-37.ec2.internal (executor 2) (12/39)

18/09/08 14:03:23 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 1.0 (TID 54, ip-172-31-84-37.ec2.internal, executor 3, partition 15, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:23 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 1.0 (TID 51) in 7313 ms on ip-172-31-84-37.ec2.internal (executor 3) (13/39)

18/09/08 14:03:24 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 1.0 (TID 55, ip-172-31-84-37.ec2.internal, executor 1, partition 16, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:24 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 1.0 (TID 52) in 7456 ms on ip-172-31-84-37.ec2.internal (executor 1) (14/39)

18/09/08 14:03:24 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 1.0 (TID 56, ip-172-31-84-37.ec2.internal, executor 2, partition 17, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:24 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 1.0 (TID 53) in 7169 ms on ip-172-31-84-37.ec2.internal (executor 2) (15/39)

18/09/08 14:03:30 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 1.0 (TID 57, ip-172-31-84-37.ec2.internal, executor 3, partition 18, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:30 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 1.0 (TID 54) in 6859 ms on ip-172-31-84-37.ec2.internal (executor 3) (16/39)

18/09/08 14:03:30 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 1.0 (TID 58, ip-172-31-84-37.ec2.internal, executor 1, partition 19, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:30 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 1.0 (TID 55) in 6083 ms on ip-172-31-84-37.ec2.internal (executor 1) (17/39)

18/09/08 14:03:32 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 1.0 (TID 59, ip-172-31-84-37.ec2.internal, executor 2, partition 20, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:32 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 1.0 (TID 56) in 7836 ms on ip-172-31-84-37.ec2.internal (executor 2) (18/39)

18/09/08 14:03:39 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 1.0 (TID 60, ip-172-31-84-37.ec2.internal, executor 3, partition 21, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:39 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 1.0 (TID 57) in 9336 ms on ip-172-31-84-37.ec2.internal (executor 3) (19/39)

18/09/08 14:03:41 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 1.0 (TID 61, ip-172-31-84-37.ec2.internal, executor 1, partition 22, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:41 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 1.0 (TID 58) in 10924 ms on ip-172-31-84-37.ec2.internal (executor 1) (20/39)

18/09/08 14:03:43 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 1.0 (TID 62, ip-172-31-84-37.ec2.internal, executor 2, partition 23, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:43 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 1.0 (TID 59) in 11497 ms on ip-172-31-84-37.ec2.internal (executor 2) (21/39)

18/09/08 14:03:48 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 1.0 (TID 63, ip-172-31-84-37.ec2.internal, executor 3, partition 24, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:48 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 1.0 (TID 60) in 9131 ms on ip-172-31-84-37.ec2.internal (executor 3) (22/39)

18/09/08 14:03:48 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 1.0 (TID 64, ip-172-31-84-37.ec2.internal, executor 2, partition 25, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:48 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 1.0 (TID 62) in 4800 ms on ip-172-31-84-37.ec2.internal (executor 2) (23/39)

18/09/08 14:03:49 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 1.0 (TID 65, ip-172-31-84-37.ec2.internal, executor 1, partition 26, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:49 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 1.0 (TID 61) in 7909 ms on ip-172-31-84-37.ec2.internal (executor 1) (24/39)

18/09/08 14:03:52 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 1.0 (TID 66, ip-172-31-84-37.ec2.internal, executor 3, partition 27, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:52 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 1.0 (TID 63) in 3924 ms on ip-172-31-84-37.ec2.internal (executor 3) (25/39)

18/09/08 14:03:57 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 1.0 (TID 67, ip-172-31-84-37.ec2.internal, executor 2, partition 28, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:57 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 1.0 (TID 64) in 8716 ms on ip-172-31-84-37.ec2.internal (executor 2) (26/39)

18/09/08 14:03:57 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 1.0 (TID 68, ip-172-31-84-37.ec2.internal, executor 1, partition 29, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:57 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 1.0 (TID 65) in 8493 ms on ip-172-31-84-37.ec2.internal (executor 1) (27/39)

18/09/08 14:03:58 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 1.0 (TID 69, ip-172-31-84-37.ec2.internal, executor 3, partition 30, NODE_LOCAL, 7963 bytes)

18/09/08 14:03:58 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 1.0 (TID 66) in 6167 ms on ip-172-31-84-37.ec2.internal (executor 3) (28/39)

18/09/08 14:04:06 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 1.0 (TID 70, ip-172-31-84-37.ec2.internal, executor 1, partition 31, NODE_LOCAL, 7963 bytes)

18/09/08 14:04:06 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 1.0 (TID 68) in 9198 ms on ip-172-31-84-37.ec2.internal (executor 1) (29/39)

18/09/08 14:04:07 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 1.0 (TID 71, ip-172-31-84-37.ec2.internal, executor 1, partition 32, NODE_LOCAL, 7963 bytes)

18/09/08 14:04:07 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 1.0 (TID 70) in 1118 ms on ip-172-31-84-37.ec2.internal (executor 1) (30/39)

18/09/08 14:04:08 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 1.0 (TID 72, ip-172-31-84-37.ec2.internal, executor 3, partition 33, NODE_LOCAL, 7963 bytes)

18/09/08 14:04:08 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 1.0 (TID 69) in 9606 ms on ip-172-31-84-37.ec2.internal (executor 3) (31/39)

18/09/08 14:04:08 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 1.0 (TID 73, ip-172-31-84-37.ec2.internal, executor 2, partition 34, NODE_LOCAL, 7963 bytes)

18/09/08 14:04:08 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 1.0 (TID 67) in 11023 ms on ip-172-31-84-37.ec2.internal (executor 2) (32/39)

18/09/08 14:04:14 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 1.0 (TID 74, ip-172-31-84-37.ec2.internal, executor 1, partition 35, NODE_LOCAL, 7963 bytes)

18/09/08 14:04:14 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 1.0 (TID 71) in 6434 ms on ip-172-31-84-37.ec2.internal (executor 1) (33/39)

18/09/08 14:04:16 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 1.0 (TID 75, ip-172-31-84-37.ec2.internal, executor 2, partition 36, NODE_LOCAL, 7963 bytes)

18/09/08 14:04:16 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 1.0 (TID 73) in 8142 ms on ip-172-31-84-37.ec2.internal (executor 2) (34/39)

18/09/08 14:04:16 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 1.0 (TID 76, ip-172-31-84-37.ec2.internal, executor 3, partition 37, NODE_LOCAL, 7963 bytes)

18/09/08 14:04:16 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 1.0 (TID 72) in 8292 ms on ip-172-31-84-37.ec2.internal (executor 3) (35/39)

18/09/08 14:04:21 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 1.0 (TID 77, ip-172-31-84-37.ec2.internal, executor 1, partition 38, NODE_LOCAL, 7963 bytes)

18/09/08 14:04:21 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 1.0 (TID 74) in 7290 ms on ip-172-31-84-37.ec2.internal (executor 1) (36/39)

18/09/08 14:04:23 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 1.0 (TID 75) in 6894 ms on ip-172-31-84-37.ec2.internal (executor 2) (37/39)

18/09/08 14:04:24 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 1.0 (TID 77) in 2546 ms on ip-172-31-84-37.ec2.internal (executor 1) (38/39)

18/09/08 14:04:24 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 1.0 (TID 76) in 7961 ms on ip-172-31-84-37.ec2.internal (executor 3) (39/39)

18/09/08 14:04:24 INFO cluster.YarnScheduler: Removed TaskSet 1.0, whose tasks have all completed, from pool

18/09/08 14:04:24 INFO scheduler.DAGScheduler: ResultStage 1 (runJob at SparkHadoopWriter.scala:78) finished in 97.062 s

18/09/08 14:04:24 INFO scheduler.DAGScheduler: Job 1 finished: runJob at SparkHadoopWriter.scala:78, took 97.072818 s

18/09/08 14:04:26 INFO io.SparkHadoopWriter: Job job_20180908140247_0004 committed.

18/09/08 14:04:26 INFO server.AbstractConnector: Stopped Spark@22ee2d0{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}

18/09/08 14:04:26 INFO ui.SparkUI: Stopped Spark web UI at http://ip-172-31-84-37.ec2.internal:4040

18/09/08 14:04:26 INFO cluster.YarnClientSchedulerBackend: Interrupting monitor thread

18/09/08 14:04:27 INFO cluster.YarnClientSchedulerBackend: Shutting down all executors

18/09/08 14:04:27 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Asking each executor to shut down

18/09/08 14:04:27 INFO cluster.SchedulerExtensionServices: Stopping SchedulerExtensionServices

(serviceOption=None,

services=List(),

started=false)

18/09/08 14:04:27 INFO cluster.YarnClientSchedulerBackend: Stopped

18/09/08 14:04:27 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

18/09/08 14:04:27 INFO memory.MemoryStore: MemoryStore cleared

18/09/08 14:04:27 INFO storage.BlockManager: BlockManager stopped

18/09/08 14:04:27 INFO storage.BlockManagerMaster: BlockManagerMaster stopped

18/09/08 14:04:27 INFO scheduler.OutputCommitCoordinator\$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

18/09/08 14:04:27 INFO spark.SparkContext: Successfully stopped SparkContext

18/09/08 14:04:27 INFO util.ShutdownHookManager: Shutdown hook called

18/09/08 14:04:27 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-afdo86d1-614a-4c10-8dcd-bd05e3e8a889

18/09/08 14:04:27 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-58c5b991-5c7c-493e-83f6-b77da37bccb2

```
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/spark2_filter/part* | wc -l
```

```
31029
```

```
[ec2-user@ip-172-31-84-37 ~]$
```


SPARK LOG₃: GROUPBY



```
ec2-user@ip-172-31-84-37~  
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/spark3_groupby_orderby/part* | wc -l  
1  
[ec2-user@ip-172-31-84-37 ~]$
```

```
[ec2-user@ip-172-31-84-37 ~]$ spark2-submit --class sparkAssignment.GroupBy.GroupBy \
```

```
> --master yarn --deploy-mode client \
```

```
> --name program3 --conf "spark.app.id=program3_id" \
```

```
> GroupBy-0.0.1-SNAPSHOT-jar-with-dependencies.jar /user/ec2-  
user/spark_assignment/input/yellow_tripdata* /user/ec2-  
user/spark_assignment/output/spark3_groupby_orderby
```

```
18/09/08 12:50:36 INFO spark.SparkContext: Running Spark version 2.3.0.cloudera2
```

```
18/09/08 12:50:36 INFO spark.SparkContext: Submitted application: count
```

```
18/09/08 12:50:36 INFO spark.SecurityManager: Changing view acls to: ec2-user
```

```
18/09/08 12:50:36 INFO spark.SecurityManager: Changing modify acls to: ec2-user
```

```
18/09/08 12:50:36 INFO spark.SecurityManager: Changing view acls groups to:
```

```
18/09/08 12:50:36 INFO spark.SecurityManager: Changing modify acls groups to:
```

```
18/09/08 12:50:36 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls  
disabled; users with view permissions: Set(ec2-user); groups with view permissions: Set(); users with  
modify permissions: Set(ec2-user); groups with modify permissions: Set()
```

```
18/09/08 12:50:36 INFO util.Utils: Successfully started service 'sparkDriver' on port 46178.
```

```
18/09/08 12:50:36 INFO spark.SparkEnv: Registering MapOutputTracker
```

```
18/09/08 12:50:36 INFO spark.SparkEnv: Registering BlockManagerMaster
```

```
18/09/08 12:50:36 INFO storage.BlockManagerMasterEndpoint: Using  
org.apache.spark.storage.DefaultTopologyMapper for getting topology information
```

18/09/08 12:50:36 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up

18/09/08 12:50:36 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-53a7dd8d-224a-49cc-8103-97079cb6dbd3

18/09/08 12:50:36 INFO memory.MemoryStore: MemoryStore started with capacity 366.3 MB

18/09/08 12:50:36 INFO spark.SparkEnv: Registering OutputCommitCoordinator

18/09/08 12:50:37 INFO util.log: Logging initialized @2771ms

18/09/08 12:50:37 INFO server.Server: jetty-9.3.z-SNAPSHOT

18/09/08 12:50:37 INFO server.Server: Started @2870ms

18/09/08 12:50:37 INFO server.AbstractConnector: Started
ServerConnector@3e792ce3{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}

18/09/08 12:50:37 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@25243bc1{/jobs,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@1c4ee95c{/jobs/json,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5aa360ea{/jobs/job,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@54336c81{/jobs/job/json,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@35e52059{/stages,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@49bd54f7{/stages/json,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@772485dd{/stages/stage,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3d829787{/stages/stage/json,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@51bde877{/stages/pool,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@492fc69e{/stages/pool/json,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@2fb68ec6{/storage,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3add81c4{/storage/json,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@1c65121{/storage/rdd,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@57dc9128{/storage/rdd/json,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@17aeg8d7{/environment,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@6ac4944a{/environment/json,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@39fc6b2c{/executors,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3ee39dao{/executors/json,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@7cc9ce8{/executors/threadDump,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@cob41d6{/executors/threadDump/json,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@2bffa76d{/static,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@aa5455e{/null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5dda14do{/api,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5987e932{/jobs/job/kill,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5bbbdd4b{/stages/stage/kill,null,AVAILABLE,@Spark}

18/09/08 12:50:37 INFO ui.SparkUI: Bound SparkUI to o.o.o.o, and started at http://ip-172-31-84-37.ec2.internal:4040

18/09/08 12:50:37 INFO spark.SparkContext: Added JAR file:/home/ec2-user/GroupBy-o.o.1-SNAPSHOT-jar-with-dependencies.jar at spark://ip-172-31-84-37.ec2.internal:46178/jars/GroupBy-o.o.1-SNAPSHOT-jar-with-dependencies.jar with timestamp 1536411037270

18/09/08 12:50:37 INFO util.Utils: Using initial executors = 0, max of spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and spark.executor.instances

18/09/08 12:50:38 INFO client.RMPProxy: Connecting to ResourceManager at ip-172-31-84-37.ec2.internal/172.31.84.37:8032

18/09/08 12:50:38 INFO yarn.Client: Requesting a new application from cluster with 1 NodeManagers

18/09/08 12:50:38 INFO yarn.Client: Verifying our application has not requested more than the maximum memory capability of the cluster (8192 MB per container)

18/09/08 12:50:38 INFO yarn.Client: Will allocate AM container, with 896 MB memory including 384 MB overhead

18/09/08 12:50:38 INFO yarn.Client: Setting up container launch context for our AM

18/09/08 12:50:38 INFO yarn.Client: Setting up the launch environment for our AM container

18/09/08 12:50:38 INFO yarn.Client: Preparing resources for our AM container

18/09/08 12:50:39 INFO yarn.Client: Uploading resource file:/tmp/spark-4359450e-d7c6-4177-874b-ed3615db796f/___spark_conf___1651506697540072221.zip -> hdfs://ip-172-31-84-37.ec2.internal:8020/user/ec2-user/.sparkStaging/application_1536410193861_0001/___spark_conf___.zip

18/09/08 12:50:39 INFO spark.SecurityManager: Changing view acls to: ec2-user

18/09/08 12:50:39 INFO spark.SecurityManager: Changing modify acls to: ec2-user

18/09/08 12:50:39 INFO spark.SecurityManager: Changing view acls groups to:

18/09/08 12:50:39 INFO spark.SecurityManager: Changing modify acls groups to:

18/09/08 12:50:39 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(ec2-user); groups with view permissions: Set(); users with modify permissions: Set(ec2-user); groups with modify permissions: Set()

18/09/08 12:50:41 INFO yarn.Client: Submitting application application_1536410193861_0001 to ResourceManager

18/09/08 12:50:41 INFO impl.YarnClientImpl: Submitted application application_1536410193861_0001

18/09/08 12:50:41 INFO cluster.SchedulerExtensionServices: Starting Yarn extension services with app application_1536410193861_0001 and attemptId None

18/09/08 12:50:42 INFO yarn.Client: Application report for application_1536410193861_0001 (state: ACCEPTED)

18/09/08 12:50:42 INFO yarn.Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: N/A

ApplicationMaster RPC port: -1

queue: root.users.ec2-user

start time: 1536411041345

final status: UNDEFINED

tracking URL: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0001/

user: ec2-user

18/09/08 12:50:43 INFO yarn.Client: Application report for application_1536410193861_0001 (state: ACCEPTED)

18/09/08 12:50:44 INFO yarn.Client: Application report for application_1536410193861_0001 (state: ACCEPTED)

18/09/08 12:50:45 INFO yarn.Client: Application report for application_1536410193861_0001 (state: ACCEPTED)

18/09/08 12:50:46 INFO yarn.Client: Application report for application_1536410193861_0001 (state: ACCEPTED)

18/09/08 12:50:47 INFO yarn.Client: Application report for application_1536410193861_0001 (state: ACCEPTED)

18/09/08 12:50:48 INFO cluster.YarnClientSchedulerBackend: Add WebUI Filter.
org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter, Map(PROXY_HOSTS -> ip-172-31-84-37.ec2.internal, PROXY_URI_BASES -> http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0001),
/proxy/application_1536410193861_0001

18/09/08 12:50:48 INFO ui.JettyUtils: Adding filter:
org.apache.hadoop.yarn.server.webproxy.amfilter.AmIpFilter

18/09/08 12:50:48 INFO yarn.Client: Application report for application_1536410193861_0001 (state: ACCEPTED)

18/09/08 12:50:48 INFO cluster.YarnSchedulerBackend\$YarnSchedulerEndpoint: ApplicationMaster registered as NettyRpcEndpointRef(spark-client://YarnAM)

18/09/08 12:50:49 INFO yarn.Client: Application report for application_1536410193861_0001 (state: RUNNING)

18/09/08 12:50:49 INFO yarn.Client:

client token: N/A

diagnostics: N/A

ApplicationMaster host: 172.31.84.37

ApplicationMaster RPC port: 0

queue: root.users.ec2-user

start time: 1536411041345

final status: UNDEFINED

tracking URL: http://ip-172-31-84-37.ec2.internal:8088/proxy/application_1536410193861_0001/

user: ec2-user

18/09/08 12:50:49 INFO cluster.YarnClientSchedulerBackend: Application application_1536410193861_0001 has started running.

18/09/08 12:50:49 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 45790.

18/09/08 12:50:49 INFO netty.NettyBlockTransferService: Server created on ip-172-31-84-37.ec2.internal:45790

18/09/08 12:50:49 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy

18/09/08 12:50:49 INFO storage.BlockManagerMaster: Registering BlockManager BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 45790, None)

18/09/08 12:50:49 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:45790 with 366.3 MB RAM, BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 45790, None)

18/09/08 12:50:49 INFO storage.BlockManagerMaster: Registered BlockManager BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 45790, None)

18/09/08 12:50:49 INFO storage.BlockManager: external shuffle service port = 7337

18/09/08 12:50:49 INFO storage.BlockManager: Initialized BlockManager: BlockManagerId(driver, ip-172-31-84-37.ec2.internal, 45790, None)

18/09/08 12:50:49 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@51888019{/metrics/json,null,AVAILABLE,@Spark}

18/09/08 12:50:50 INFO scheduler.EventLoggingListener: Logging events to hdfs://ip-172-31-84-37.ec2.internal:8020/user/spark/spark2ApplicationHistory/application_1536410193861_0001

18/09/08 12:50:50 INFO util.Utils: Using initial executors = 0, max of
spark.dynamicAllocation.initialExecutors, spark.dynamicAllocation.minExecutors and
spark.executor.instances

18/09/08 12:50:50 INFO cluster.YarnClientSchedulerBackend: SchedulerBackend is ready for scheduling
beginning after reached minRegisteredResourcesRatio: 0.8

18/09/08 12:50:50 INFO memory.MemoryStore: Block broadcast_0 stored as values in memory
(estimated size 303.3 KB, free 366.0 MB)

18/09/08 12:50:50 INFO memory.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory
(estimated size 27.1 KB, free 366.0 MB)

18/09/08 12:50:50 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on ip-172-31-84-37.ec2.internal:45790 (size: 27.1 KB, free: 366.3 MB)

18/09/08 12:50:50 INFO spark.SparkContext: Created broadcast 0 from textFile at GroupBy.java:17

18/09/08 12:50:51 INFO mapred.FileInputFormat: Total input paths to process : 6

18/09/08 12:50:51 INFO spark.SparkContext: Starting job: count at GroupBy.java:55

18/09/08 12:50:51 INFO scheduler.DAGScheduler: Registering RDD 3 (mapToPair at GroupBy.java:24)

18/09/08 12:50:51 INFO scheduler.DAGScheduler: Got job 0 (count at GroupBy.java:55) with 39 output
partitions

18/09/08 12:50:51 INFO scheduler.DAGScheduler: Final stage: ResultStage 1 (count at GroupBy.java:55)

18/09/08 12:50:51 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 0)

18/09/08 12:50:51 INFO scheduler.DAGScheduler: Missing parents: List(ShuffleMapStage 0)

18/09/08 12:50:51 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 0
(MapPartitionsRDD[3] at mapToPair at GroupBy.java:24), which has no missing parents

18/09/08 12:50:51 INFO memory.MemoryStore: Block broadcast_1 stored as values in memory
(estimated size 5.9 KB, free 366.0 MB)

18/09/08 12:50:51 INFO memory.MemoryStore: Block broadcast_1_pieceo stored as bytes in memory (estimated size 3.3 KB, free 366.0 MB)

18/09/08 12:50:51 INFO storage.BlockManagerInfo: Added broadcast_1_pieceo in memory on ip-172-31-84-37.ec2.internal:45790 (size: 3.3 KB, free: 366.3 MB)

18/09/08 12:50:51 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1039

18/09/08 12:50:51 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ShuffleMapStage 0 (MapPartitionsRDD[3] at mapToPair at GroupBy.java:24) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/09/08 12:50:51 INFO cluster.YarnScheduler: Adding task set 0.0 with 39 tasks

18/09/08 12:50:52 INFO spark.ExecutorAllocationManager: Requesting 1 new executor because tasks are backlogged (new desired total will be 1)

18/09/08 12:50:53 INFO spark.ExecutorAllocationManager: Requesting 2 new executors because tasks are backlogged (new desired total will be 3)

18/09/08 12:50:54 INFO spark.ExecutorAllocationManager: Requesting 4 new executors because tasks are backlogged (new desired total will be 7)

18/09/08 12:50:55 INFO spark.ExecutorAllocationManager: Requesting 8 new executors because tasks are backlogged (new desired total will be 15)

18/09/08 12:50:56 INFO spark.ExecutorAllocationManager: Requesting 16 new executors because tasks are backlogged (new desired total will be 31)

18/09/08 12:50:57 INFO spark.ExecutorAllocationManager: Requesting 8 new executors because tasks are backlogged (new desired total will be 39)

18/09/08 12:50:58 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (172.31.84.37:35530) with ID 1

18/09/08 12:50:58 INFO spark.ExecutorAllocationManager: New executor 1 has registered (new total is 1)

18/09/08 12:50:58 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, ip-172-31-84-37.ec2.internal, executor 1, partition 0, NODE_LOCAL, 7952 bytes)

18/09/08 12:50:58 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:40608 with 366.3 MB RAM, BlockManagerId(1, ip-172-31-84-37.ec2.internal, 40608, None)

18/09/08 12:50:59 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (172.31.84.37:35538) with ID 2

18/09/08 12:50:59 INFO spark.ExecutorAllocationManager: New executor 2 has registered (new total is 2)

18/09/08 12:50:59 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, ip-172-31-84-37.ec2.internal, executor 2, partition 1, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:00 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Registered executor NettyRpcEndpointRef(spark-client://Executor) (172.31.84.37:35542) with ID 3

18/09/08 12:51:00 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, ip-172-31-84-37.ec2.internal, executor 3, partition 2, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:00 INFO spark.ExecutorAllocationManager: New executor 3 has registered (new total is 3)

18/09/08 12:51:00 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:39965 with 366.3 MB RAM, BlockManagerId(2, ip-172-31-84-37.ec2.internal, 39965, None)

18/09/08 12:51:00 INFO storage.BlockManagerMasterEndpoint: Registering block manager ip-172-31-84-37.ec2.internal:42352 with 366.3 MB RAM, BlockManagerId(3, ip-172-31-84-37.ec2.internal, 42352, None)

18/09/08 12:51:00 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-84-37.ec2.internal:40608 (size: 3.3 KB, free: 366.3 MB)

18/09/08 12:51:01 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on ip-172-31-84-37.ec2.internal:40608 (size: 27.1 KB, free: 366.3 MB)

18/09/08 12:51:01 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-84-37.ec2.internal:39965 (size: 3.3 KB, free: 366.3 MB)

18/09/08 12:51:02 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on ip-172-31-84-37.ec2.internal:42352 (size: 3.3 KB, free: 366.3 MB)

18/09/08 12:51:02 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on ip-172-31-84-37.ec2.internal:39965 (size: 27.1 KB, free: 366.3 MB)

18/09/08 12:51:02 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on ip-172-31-84-37.ec2.internal:42352 (size: 27.1 KB, free: 366.3 MB)

18/09/08 12:51:14 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, ip-172-31-84-37.ec2.internal, executor 1, partition 3, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:14 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 16469 ms on ip-172-31-84-37.ec2.internal (executor 1) (1/39)

18/09/08 12:51:15 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 0.0 (TID 4, ip-172-31-84-37.ec2.internal, executor 2, partition 4, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:15 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 0.0 (TID 1) in 15266 ms on ip-172-31-84-37.ec2.internal (executor 2) (2/39)

18/09/08 12:51:15 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 0.0 (TID 5, ip-172-31-84-37.ec2.internal, executor 3, partition 5, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:15 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 15754 ms on ip-172-31-84-37.ec2.internal (executor 3) (3/39)

18/09/08 12:51:21 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 0.0 (TID 6, ip-172-31-84-37.ec2.internal, executor 3, partition 6, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:21 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 0.0 (TID 5) in 6149 ms on ip-172-31-84-37.ec2.internal (executor 3) (4/39)

18/09/08 12:51:23 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, ip-172-31-84-37.ec2.internal, executor 1, partition 7, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:23 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 0.0 (TID 3) in 8584 ms on ip-172-31-84-37.ec2.internal (executor 1) (5/39)

18/09/08 12:51:23 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, ip-172-31-84-37.ec2.internal, executor 2, partition 8, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:23 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 8676 ms on ip-172-31-84-37.ec2.internal (executor 2) (6/39)

18/09/08 12:51:29 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, ip-172-31-84-37.ec2.internal, executor 3, partition 9, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:29 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 7924 ms on ip-172-31-84-37.ec2.internal (executor 3) (7/39)

18/09/08 12:51:29 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, ip-172-31-84-37.ec2.internal, executor 1, partition 10, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:29 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 6772 ms on ip-172-31-84-37.ec2.internal (executor 1) (8/39)

18/09/08 12:51:30 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, ip-172-31-84-37.ec2.internal, executor 2, partition 11, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:30 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 6193 ms on ip-172-31-84-37.ec2.internal (executor 2) (9/39)

18/09/08 12:51:37 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 0.0 (TID 12, ip-172-31-84-37.ec2.internal, executor 2, partition 12, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:37 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 0.0 (TID 11) in 7098 ms on ip-172-31-84-37.ec2.internal (executor 2) (10/39)

18/09/08 12:51:39 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13, ip-172-31-84-37.ec2.internal, executor 3, partition 13, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:39 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 9405 ms on ip-172-31-84-37.ec2.internal (executor 3) (11/39)

18/09/08 12:51:39 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, ip-172-31-84-37.ec2.internal, executor 1, partition 14, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:39 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 9814 ms on ip-172-31-84-37.ec2.internal (executor 1) (12/39)

18/09/08 12:51:41 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, ip-172-31-84-37.ec2.internal, executor 2, partition 15, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:41 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 4850 ms on ip-172-31-84-37.ec2.internal (executor 2) (13/39)

18/09/08 12:51:44 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, ip-172-31-84-37.ec2.internal, executor 3, partition 16, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:44 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 0.0 (TID 13) in 4877 ms on ip-172-31-84-37.ec2.internal (executor 3) (14/39)

18/09/08 12:51:44 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 0.0 (TID 17, ip-172-31-84-37.ec2.internal, executor 1, partition 17, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:44 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 0.0 (TID 14) in 4485 ms on ip-172-31-84-37.ec2.internal (executor 1) (15/39)

18/09/08 12:51:46 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 0.0 (TID 18, ip-172-31-84-37.ec2.internal, executor 2, partition 18, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:46 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 4523 ms on ip-172-31-84-37.ec2.internal (executor 2) (16/39)

18/09/08 12:51:49 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 0.0 (TID 19, ip-172-31-84-37.ec2.internal, executor 1, partition 19, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:49 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 0.0 (TID 17) in 5111 ms on ip-172-31-84-37.ec2.internal (executor 1) (17/39)

18/09/08 12:51:49 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 0.0 (TID 20, ip-172-31-84-37.ec2.internal, executor 3, partition 20, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:49 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 5373 ms on ip-172-31-84-37.ec2.internal (executor 3) (18/39)

18/09/08 12:51:52 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 0.0 (TID 21, ip-172-31-84-37.ec2.internal, executor 2, partition 21, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:52 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 0.0 (TID 18) in 5737 ms on ip-172-31-84-37.ec2.internal (executor 2) (19/39)

18/09/08 12:51:53 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 0.0 (TID 22, ip-172-31-84-37.ec2.internal, executor 1, partition 22, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:53 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 0.0 (TID 19) in 4425 ms on ip-172-31-84-37.ec2.internal (executor 1) (20/39)

18/09/08 12:51:53 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 0.0 (TID 23, ip-172-31-84-37.ec2.internal, executor 3, partition 23, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:53 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 0.0 (TID 20) in 4414 ms on ip-172-31-84-37.ec2.internal (executor 3) (21/39)

18/09/08 12:51:57 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 0.0 (TID 24, ip-172-31-84-37.ec2.internal, executor 2, partition 24, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:57 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 0.0 (TID 21) in 5374 ms on ip-172-31-84-37.ec2.internal (executor 2) (22/39)

18/09/08 12:51:58 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 0.0 (TID 25, ip-172-31-84-37.ec2.internal, executor 3, partition 25, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:58 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 0.0 (TID 23) in 4482 ms on ip-172-31-84-37.ec2.internal (executor 3) (23/39)

18/09/08 12:51:58 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 0.0 (TID 26, ip-172-31-84-37.ec2.internal, executor 1, partition 26, NODE_LOCAL, 7952 bytes)

18/09/08 12:51:58 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 0.0 (TID 22) in 4582 ms on ip-172-31-84-37.ec2.internal (executor 1) (24/39)

18/09/08 12:52:00 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 0.0 (TID 27, ip-172-31-84-37.ec2.internal, executor 2, partition 27, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:00 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 0.0 (TID 24) in 2689 ms on ip-172-31-84-37.ec2.internal (executor 2) (25/39)

18/09/08 12:52:04 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 0.0 (TID 28, ip-172-31-84-37.ec2.internal, executor 1, partition 28, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:04 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 0.0 (TID 26) in 5716 ms on ip-172-31-84-37.ec2.internal (executor 1) (26/39)

18/09/08 12:52:04 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 0.0 (TID 29, ip-172-31-84-37.ec2.internal, executor 3, partition 29, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:04 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 0.0 (TID 25) in 6030 ms on ip-172-31-84-37.ec2.internal (executor 3) (27/39)

18/09/08 12:52:06 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 0.0 (TID 30, ip-172-31-84-37.ec2.internal, executor 2, partition 30, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:06 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 0.0 (TID 27) in 6350 ms on ip-172-31-84-37.ec2.internal (executor 2) (28/39)

18/09/08 12:52:11 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 0.0 (TID 31, ip-172-31-84-37.ec2.internal, executor 1, partition 31, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:11 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 0.0 (TID 28) in 7722 ms on ip-172-31-84-37.ec2.internal (executor 1) (29/39)

18/09/08 12:52:12 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 0.0 (TID 32, ip-172-31-84-37.ec2.internal, executor 3, partition 32, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:12 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 0.0 (TID 29) in 7800 ms on ip-172-31-84-37.ec2.internal (executor 3) (30/39)

18/09/08 12:52:12 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 0.0 (TID 33, ip-172-31-84-37.ec2.internal, executor 1, partition 33, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:12 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 0.0 (TID 31) in 622 ms on ip-172-31-84-37.ec2.internal (executor 1) (31/39)

18/09/08 12:52:13 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 0.0 (TID 34, ip-172-31-84-37.ec2.internal, executor 2, partition 34, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:13 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 0.0 (TID 30) in 6555 ms on ip-172-31-84-37.ec2.internal (executor 2) (32/39)

18/09/08 12:52:16 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 0.0 (TID 35, ip-172-31-84-37.ec2.internal, executor 3, partition 35, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:16 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 0.0 (TID 32) in 4395 ms on ip-172-31-84-37.ec2.internal (executor 3) (33/39)

18/09/08 12:52:16 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 0.0 (TID 36, ip-172-31-84-37.ec2.internal, executor 1, partition 36, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:16 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 0.0 (TID 33) in 4227 ms on ip-172-31-84-37.ec2.internal (executor 1) (34/39)

18/09/08 12:52:17 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 0.0 (TID 37, ip-172-31-84-37.ec2.internal, executor 2, partition 37, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:17 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 0.0 (TID 34) in 4669 ms on ip-172-31-84-37.ec2.internal (executor 2) (35/39)

18/09/08 12:52:20 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 0.0 (TID 38, ip-172-31-84-37.ec2.internal, executor 1, partition 38, NODE_LOCAL, 7952 bytes)

18/09/08 12:52:20 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 0.0 (TID 36) in 4099 ms on ip-172-31-84-37.ec2.internal (executor 1) (36/39)

18/09/08 12:52:20 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 0.0 (TID 35) in 4350 ms on ip-172-31-84-37.ec2.internal (executor 3) (37/39)

18/09/08 12:52:21 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 0.0 (TID 38) in 786 ms on ip-172-31-84-37.ec2.internal (executor 1) (38/39)

18/09/08 12:52:21 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 0.0 (TID 37) in 4072 ms on ip-172-31-84-37.ec2.internal (executor 2) (39/39)

18/09/08 12:52:21 INFO cluster.YarnScheduler: Removed TaskSet 0.0, whose tasks have all completed, from pool

18/09/08 12:52:21 INFO scheduler.DAGScheduler: ShuffleMapStage 0 (mapToPair at GroupBy.java:24) finished in 90.744 s

18/09/08 12:52:21 INFO scheduler.DAGScheduler: looking for newly runnable stages

18/09/08 12:52:21 INFO scheduler.DAGScheduler: running: Set()

18/09/08 12:52:21 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 1)

18/09/08 12:52:21 INFO scheduler.DAGScheduler: failed: Set()

18/09/08 12:52:21 INFO scheduler.DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[5] at filter at GroupBy.java:43), which has no missing parents

18/09/08 12:52:21 INFO memory.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 4.5 KB, free 366.0 MB)

18/09/08 12:52:21 INFO memory.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 2.6 KB, free 366.0 MB)

18/09/08 12:52:21 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-84-37.ec2.internal:45790 (size: 2.6 KB, free: 366.3 MB)

18/09/08 12:52:21 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1039

18/09/08 12:52:21 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 1 (MapPartitionsRDD[5] at filter at GroupBy.java:43) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/09/08 12:52:21 INFO cluster.YarnScheduler: Adding task set 1.0 with 39 tasks

18/09/08 12:52:21 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 39, ip-172-31-84-37.ec2.internal, executor 3, partition 0, NODE_LOCAL, 7660 bytes)

18/09/08 12:52:21 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 40, ip-172-31-84-37.ec2.internal, executor 2, partition 1, NODE_LOCAL, 7660 bytes)

18/09/08 12:52:21 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 41, ip-172-31-84-37.ec2.internal, executor 1, partition 2, NODE_LOCAL, 7660 bytes)

18/09/08 12:52:21 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-84-37.ec2.internal:39965 (size: 2.6 KB, free: 366.3 MB)

18/09/08 12:52:22 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-84-37.ec2.internal:40608 (size: 2.6 KB, free: 366.3 MB)

18/09/08 12:52:22 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on ip-172-31-84-37.ec2.internal:42352 (size: 2.6 KB, free: 366.3 MB)

18/09/08 12:52:22 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 0 to 172.31.84.37:35538

18/09/08 12:52:22 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 0 to 172.31.84.37:35542

18/09/08 12:52:22 INFO spark.MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 0 to 172.31.84.37:35530

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 42, ip-172-31-84-37.ec2.internal, executor 3, partition 3, NODE_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 39) in 267 ms on ip-172-31-84-37.ec2.internal (executor 3) (1/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 1.0 (TID 43, ip-172-31-84-37.ec2.internal, executor 1, partition 4, NODE_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 1.0 (TID 41) in 317 ms on ip-172-31-84-37.ec2.internal (executor 1) (2/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 1.0 (TID 44, ip-172-31-84-37.ec2.internal, executor 3, partition 5, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 1.0 (TID 42) in 117 ms on ip-172-31-84-37.ec2.internal (executor 3) (3/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 1.0 (TID 45, ip-172-31-84-37.ec2.internal, executor 3, partition 6, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 1.0 (TID 44) in 51 ms on ip-172-31-84-37.ec2.internal (executor 3) (4/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 1.0 (TID 46, ip-172-31-84-37.ec2.internal, executor 2, partition 7, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 1.0 (TID 40) in 453 ms on ip-172-31-84-37.ec2.internal (executor 2) (5/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 1.0 (TID 47, ip-172-31-84-37.ec2.internal, executor 3, partition 8, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 1.0 (TID 45) in 63 ms on ip-172-31-84-37.ec2.internal (executor 3) (6/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 1.0 (TID 48, ip-172-31-84-37.ec2.internal, executor 1, partition 9, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 1.0 (TID 43) in 180 ms on ip-172-31-84-37.ec2.internal (executor 1) (7/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 1.0 (TID 49, ip-172-31-84-37.ec2.internal, executor 3, partition 10, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 1.0 (TID 47) in 35 ms on ip-172-31-84-37.ec2.internal (executor 3) (8/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 1.0 (TID 50, ip-172-31-84-37.ec2.internal, executor 1, partition 11, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 1.0 (TID 48) in 46 ms on ip-172-31-84-37.ec2.internal (executor 1) (9/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 1.0 (TID 51, ip-172-31-84-37.ec2.internal, executor 3, partition 12, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 1.0 (TID 49) in 24 ms on ip-172-31-84-37.ec2.internal (executor 3) (10/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 1.0 (TID 52, ip-172-31-84-37.ec2.internal, executor 2, partition 13, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 1.0 (TID 46) in 121 ms on ip-172-31-84-37.ec2.internal (executor 2) (11/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 1.0 (TID 53, ip-172-31-84-37.ec2.internal, executor 1, partition 14, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 1.0 (TID 54, ip-172-31-84-37.ec2.internal, executor 3, partition 15, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 1.0 (TID 55, ip-172-31-84-37.ec2.internal, executor 2, partition 16, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 1.0 (TID 50) in 78 ms on ip-172-31-84-37.ec2.internal (executor 1) (12/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 1.0 (TID 51) in 77 ms on ip-172-31-84-37.ec2.internal (executor 3) (13/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 1.0 (TID 52) in 54 ms on ip-172-31-84-37.ec2.internal (executor 2) (14/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 1.0 (TID 56, ip-172-31-84-37.ec2.internal, executor 3, partition 17, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 1.0 (TID 54) in 46 ms on ip-172-31-84-37.ec2.internal (executor 3) (15/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 1.0 (TID 57, ip-172-31-84-37.ec2.internal, executor 1, partition 18, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 1.0 (TID 53) in 65 ms on ip-172-31-84-37.ec2.internal (executor 1) (16/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 1.0 (TID 58, ip-172-31-84-37.ec2.internal, executor 2, partition 19, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 1.0 (TID 55) in 97 ms on ip-172-31-84-37.ec2.internal (executor 2) (17/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 1.0 (TID 59, ip-172-31-84-37.ec2.internal, executor 3, partition 20, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 1.0 (TID 56) in 82 ms on ip-172-31-84-37.ec2.internal (executor 3) (18/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 1.0 (TID 60, ip-172-31-84-37.ec2.internal, executor 1, partition 21, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 1.0 (TID 61, ip-172-31-84-37.ec2.internal, executor 2, partition 22, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 1.0 (TID 57) in 92 ms on ip-172-31-84-37.ec2.internal (executor 1) (19/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 1.0 (TID 58) in 58 ms on ip-172-31-84-37.ec2.internal (executor 2) (20/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 1.0 (TID 62, ip-172-31-84-37.ec2.internal, executor 1, partition 23, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 1.0 (TID 63, ip-172-31-84-37.ec2.internal, executor 3, partition 24, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 1.0 (TID 64, ip-172-31-84-37.ec2.internal, executor 2, partition 25, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 1.0 (TID 61) in 43 ms on ip-172-31-84-37.ec2.internal (executor 2) (21/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 1.0 (TID 60) in 47 ms on ip-172-31-84-37.ec2.internal (executor 1) (22/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 1.0 (TID 59) in 75 ms on ip-172-31-84-37.ec2.internal (executor 3) (23/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 1.0 (TID 65, ip-172-31-84-37.ec2.internal, executor 3, partition 26, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 1.0 (TID 63) in 36 ms on ip-172-31-84-37.ec2.internal (executor 3) (24/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 1.0 (TID 66, ip-172-31-84-37.ec2.internal, executor 1, partition 27, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 1.0 (TID 62) in 59 ms on ip-172-31-84-37.ec2.internal (executor 1) (25/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 1.0 (TID 67, ip-172-31-84-37.ec2.internal, executor 3, partition 28, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 1.0 (TID 65) in 45 ms on ip-172-31-84-37.ec2.internal (executor 3) (26/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 1.0 (TID 68, ip-172-31-84-37.ec2.internal, executor 1, partition 29, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 1.0 (TID 66) in 33 ms on ip-172-31-84-37.ec2.internal (executor 1) (27/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 1.0 (TID 69, ip-172-31-84-37.ec2.internal, executor 2, partition 30, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 1.0 (TID 70, ip-172-31-84-37.ec2.internal, executor 3, partition 31, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 1.0 (TID 64) in 114 ms on ip-172-31-84-37.ec2.internal (executor 2) (28/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 1.0 (TID 67) in 42 ms on ip-172-31-84-37.ec2.internal (executor 3) (29/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 1.0 (TID 71, ip-172-31-84-37.ec2.internal, executor 1, partition 32, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 1.0 (TID 68) in 65 ms on ip-172-31-84-37.ec2.internal (executor 1) (30/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 1.0 (TID 72, ip-172-31-84-37.ec2.internal, executor 2, partition 33, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 1.0 (TID 69) in 45 ms on ip-172-31-84-37.ec2.internal (executor 2) (31/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 1.0 (TID 73, ip-172-31-84-37.ec2.internal, executor 3, partition 34, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 1.0 (TID 70) in 59 ms on ip-172-31-84-37.ec2.internal (executor 3) (32/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 1.0 (TID 74, ip-172-31-84-37.ec2.internal, executor 3, partition 35, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 1.0 (TID 73) in 34 ms on ip-172-31-84-37.ec2.internal (executor 3) (33/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 1.0 (TID 72) in 48 ms on ip-172-31-84-37.ec2.internal (executor 2) (34/39)

18/09/08 12:52:22 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 1.0 (TID 75, ip-172-31-84-37.ec2.internal, executor 2, partition 36, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO spark.ExecutorAllocationManager: Requesting 3 new executors because tasks are backlogged (new desired total will be 4)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 1.0 (TID 76, ip-172-31-84-37.ec2.internal, executor 3, partition 37, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 1.0 (TID 74) in 32 ms on ip-172-31-84-37.ec2.internal (executor 3) (35/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 1.0 (TID 77, ip-172-31-84-37.ec2.internal, executor 1, partition 38, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 1.0 (TID 71) in 124 ms on ip-172-31-84-37.ec2.internal (executor 1) (36/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 1.0 (TID 76) in 54 ms on ip-172-31-84-37.ec2.internal (executor 3) (37/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 1.0 (TID 77) in 31 ms on ip-172-31-84-37.ec2.internal (executor 1) (38/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 1.0 (TID 75) in 94 ms on ip-172-31-84-37.ec2.internal (executor 2) (39/39)

18/09/08 12:52:23 INFO cluster.YarnScheduler: Removed TaskSet 1.0, whose tasks have all completed, from pool

18/09/08 12:52:23 INFO scheduler.DAGScheduler: ResultStage 1 (count at GroupBy.java:55) finished in 1.123 s

18/09/08 12:52:23 INFO scheduler.DAGScheduler: Job 0 finished: count at GroupBy.java:55, took 92.009687 s

18/09/08 12:52:23 INFO spark.SparkContext: Starting job: top at GroupBy.java:55

18/09/08 12:52:23 INFO scheduler.DAGScheduler: Got job 1 (top at GroupBy.java:55) with 39 output partitions

18/09/08 12:52:23 INFO scheduler.DAGScheduler: Final stage: ResultStage 3 (top at GroupBy.java:55)

18/09/08 12:52:23 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 2)

18/09/08 12:52:23 INFO scheduler.DAGScheduler: Missing parents: List()

18/09/08 12:52:23 INFO scheduler.DAGScheduler: Submitting ResultStage 3 (MapPartitionsRDD[6] at top at GroupBy.java:55), which has no missing parents

18/09/08 12:52:23 INFO memory.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 5.3 KB, free 366.0 MB)

18/09/08 12:52:23 INFO memory.MemoryStore: Block broadcast_3_pieceo stored as bytes in memory (estimated size 3.0 KB, free 366.0 MB)

18/09/08 12:52:23 INFO storage.BlockManagerInfo: Added broadcast_3_pieceo in memory on ip-172-31-84-37.ec2.internal:45790 (size: 3.0 KB, free: 366.3 MB)

18/09/08 12:52:23 INFO spark.SparkContext: Created broadcast 3 from broadcast at DAGScheduler.scala:1039

18/09/08 12:52:23 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 3 (MapPartitionsRDD[6] at top at GroupBy.java:55) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/09/08 12:52:23 INFO cluster.YarnScheduler: Adding task set 3.0 with 39 tasks

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 3.0 (TID 78, ip-172-31-84-37.ec2.internal, executor 3, partition 0, NODE_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 3.0 (TID 79, ip-172-31-84-37.ec2.internal, executor 1, partition 1, NODE_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 3.0 (TID 80, ip-172-31-84-37.ec2.internal, executor 2, partition 2, NODE_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO storage.BlockManagerInfo: Added broadcast_3_pieceo in memory on ip-172-31-84-37.ec2.internal:39965 (size: 3.0 KB, free: 366.3 MB)

18/09/08 12:52:23 INFO storage.BlockManagerInfo: Added broadcast_3_pieceo in memory on ip-172-31-84-37.ec2.internal:40608 (size: 3.0 KB, free: 366.3 MB)

18/09/08 12:52:23 INFO storage.BlockManagerInfo: Added broadcast_3_pieceo in memory on ip-172-31-84-37.ec2.internal:42352 (size: 3.0 KB, free: 366.3 MB)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 3.0 (TID 81, ip-172-31-84-37.ec2.internal, executor 2, partition 3, NODE_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 3.0 (TID 82, ip-172-31-84-37.ec2.internal, executor 1, partition 4, NODE_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 3.0 (TID 79) in 362 ms on ip-172-31-84-37.ec2.internal (executor 1) (1/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 3.0 (TID 80) in 379 ms on ip-172-31-84-37.ec2.internal (executor 2) (2/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 3.0 (TID 83, ip-172-31-84-37.ec2.internal, executor 3, partition 5, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 3.0 (TID 78) in 453 ms on ip-172-31-84-37.ec2.internal (executor 3) (3/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 3.0 (TID 84, ip-172-31-84-37.ec2.internal, executor 3, partition 6, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 3.0 (TID 83) in 28 ms on ip-172-31-84-37.ec2.internal (executor 3) (4/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 3.0 (TID 85, ip-172-31-84-37.ec2.internal, executor 2, partition 7, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 3.0 (TID 81) in 152 ms on ip-172-31-84-37.ec2.internal (executor 2) (5/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 3.0 (TID 86, ip-172-31-84-37.ec2.internal, executor 3, partition 8, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 3.0 (TID 84) in 32 ms on ip-172-31-84-37.ec2.internal (executor 3) (6/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 3.0 (TID 87, ip-172-31-84-37.ec2.internal, executor 3, partition 9, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 3.0 (TID 86) in 17 ms on ip-172-31-84-37.ec2.internal (executor 3) (7/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 3.0 (TID 88, ip-172-31-84-37.ec2.internal, executor 1, partition 10, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 3.0 (TID 82) in 219 ms on ip-172-31-84-37.ec2.internal (executor 1) (8/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 3.0 (TID 89, ip-172-31-84-37.ec2.internal, executor 2, partition 11, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 3.0 (TID 85) in 99 ms on ip-172-31-84-37.ec2.internal (executor 2) (9/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 3.0 (TID 90, ip-172-31-84-37.ec2.internal, executor 3, partition 12, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 3.0 (TID 87) in 107 ms on ip-172-31-84-37.ec2.internal (executor 3) (10/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 3.0 (TID 91, ip-172-31-84-37.ec2.internal, executor 3, partition 13, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 3.0 (TID 90) in 17 ms on ip-172-31-84-37.ec2.internal (executor 3) (11/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 3.0 (TID 92, ip-172-31-84-37.ec2.internal, executor 3, partition 14, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 3.0 (TID 91) in 24 ms on ip-172-31-84-37.ec2.internal (executor 3) (12/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 3.0 (TID 93, ip-172-31-84-37.ec2.internal, executor 1, partition 15, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 3.0 (TID 94, ip-172-31-84-37.ec2.internal, executor 3, partition 16, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 3.0 (TID 88) in 126 ms on ip-172-31-84-37.ec2.internal (executor 1) (13/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 3.0 (TID 92) in 27 ms on ip-172-31-84-37.ec2.internal (executor 3) (14/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 3.0 (TID 95, ip-172-31-84-37.ec2.internal, executor 1, partition 17, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 3.0 (TID 96, ip-172-31-84-37.ec2.internal, executor 3, partition 18, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 3.0 (TID 93) in 45 ms on ip-172-31-84-37.ec2.internal (executor 1) (15/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 3.0 (TID 94) in 45 ms on ip-172-31-84-37.ec2.internal (executor 3) (16/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 3.0 (TID 97, ip-172-31-84-37.ec2.internal, executor 2, partition 19, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 3.0 (TID 89) in 153 ms on ip-172-31-84-37.ec2.internal (executor 2) (17/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 3.0 (TID 98, ip-172-31-84-37.ec2.internal, executor 2, partition 20, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 3.0 (TID 99, ip-172-31-84-37.ec2.internal, executor 3, partition 21, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 3.0 (TID 100, ip-172-31-84-37.ec2.internal, executor 1, partition 22, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 3.0 (TID 97) in 49 ms on ip-172-31-84-37.ec2.internal (executor 2) (18/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 3.0 (TID 96) in 57 ms on ip-172-31-84-37.ec2.internal (executor 3) (19/39)

18/09/08 12:52:23 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 3.0 (TID 95) in 57 ms on ip-172-31-84-37.ec2.internal (executor 1) (20/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 3.0 (TID 101, ip-172-31-84-37.ec2.internal, executor 3, partition 23, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 3.0 (TID 99) in 63 ms on ip-172-31-84-37.ec2.internal (executor 3) (21/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 3.0 (TID 102, ip-172-31-84-37.ec2.internal, executor 3, partition 24, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 3.0 (TID 101) in 17 ms on ip-172-31-84-37.ec2.internal (executor 3) (22/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 3.0 (TID 103, ip-172-31-84-37.ec2.internal, executor 1, partition 25, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 3.0 (TID 100) in 109 ms on ip-172-31-84-37.ec2.internal (executor 1) (23/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 3.0 (TID 104, ip-172-31-84-37.ec2.internal, executor 2, partition 26, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 3.0 (TID 98) in 141 ms on ip-172-31-84-37.ec2.internal (executor 2) (24/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 3.0 (TID 105, ip-172-31-84-37.ec2.internal, executor 2, partition 27, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 3.0 (TID 104) in 45 ms on ip-172-31-84-37.ec2.internal (executor 2) (25/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 3.0 (TID 106, ip-172-31-84-37.ec2.internal, executor 1, partition 28, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 3.0 (TID 103) in 98 ms on ip-172-31-84-37.ec2.internal (executor 1) (26/39)

18/09/08 12:52:24 INFO spark.ExecutorAllocationManager: Requesting 4 new executors because tasks are backlogged (new desired total will be 4)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 3.0 (TID 107, ip-172-31-84-37.ec2.internal, executor 3, partition 29, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 3.0 (TID 102) in 145 ms on ip-172-31-84-37.ec2.internal (executor 3) (27/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 3.0 (TID 108, ip-172-31-84-37.ec2.internal, executor 1, partition 30, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 3.0 (TID 106) in 28 ms on ip-172-31-84-37.ec2.internal (executor 1) (28/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 3.0 (TID 109, ip-172-31-84-37.ec2.internal, executor 2, partition 31, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 3.0 (TID 105) in 57 ms on ip-172-31-84-37.ec2.internal (executor 2) (29/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 3.0 (TID 110, ip-172-31-84-37.ec2.internal, executor 3, partition 32, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 3.0 (TID 107) in 39 ms on ip-172-31-84-37.ec2.internal (executor 3) (30/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 3.0 (TID 111, ip-172-31-84-37.ec2.internal, executor 3, partition 33, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 3.0 (TID 112, ip-172-31-84-37.ec2.internal, executor 2, partition 34, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 3.0 (TID 110) in 39 ms on ip-172-31-84-37.ec2.internal (executor 3) (31/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 3.0 (TID 109) in 60 ms on ip-172-31-84-37.ec2.internal (executor 2) (32/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 3.0 (TID 113, ip-172-31-84-37.ec2.internal, executor 3, partition 35, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 3.0 (TID 111) in 62 ms on ip-172-31-84-37.ec2.internal (executor 3) (33/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 3.0 (TID 114, ip-172-31-84-37.ec2.internal, executor 1, partition 36, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 3.0 (TID 108) in 125 ms on ip-172-31-84-37.ec2.internal (executor 1) (34/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 3.0 (TID 115, ip-172-31-84-37.ec2.internal, executor 1, partition 37, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 3.0 (TID 114) in 22 ms on ip-172-31-84-37.ec2.internal (executor 1) (35/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 3.0 (TID 116, ip-172-31-84-37.ec2.internal, executor 2, partition 38, PROCESS_LOCAL, 7660 bytes)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 3.0 (TID 112) in 112 ms on ip-172-31-84-37.ec2.internal (executor 2) (36/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 3.0 (TID 113) in 55 ms on ip-172-31-84-37.ec2.internal (executor 3) (37/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 3.0 (TID 115) in 61 ms on ip-172-31-84-37.ec2.internal (executor 1) (38/39)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 3.0 (TID 116) in 42 ms on ip-172-31-84-37.ec2.internal (executor 2) (39/39)

18/09/08 12:52:24 INFO cluster.YarnScheduler: Removed TaskSet 3.0, whose tasks have all completed, from pool

18/09/08 12:52:24 INFO scheduler.DAGScheduler: ResultStage 3 (top at GroupBy.java:55) finished in 1.280 s

18/09/08 12:52:24 INFO scheduler.DAGScheduler: Job 1 finished: top at GroupBy.java:55, took 1.301255 s

18/09/08 12:52:24 INFO spark.SparkContext: Starting job: collect at GroupBy.java:59

18/09/08 12:52:24 INFO scheduler.DAGScheduler: Got job 2 (collect at GroupBy.java:59) with 1 output partitions

18/09/08 12:52:24 INFO scheduler.DAGScheduler: Final stage: ResultStage 4 (collect at GroupBy.java:59)

18/09/08 12:52:24 INFO scheduler.DAGScheduler: Parents of final stage: List()

18/09/08 12:52:24 INFO scheduler.DAGScheduler: Missing parents: List()

18/09/08 12:52:24 INFO scheduler.DAGScheduler: Submitting ResultStage 4 (ParallelCollectionRDD[7] at parallelize at GroupBy.java:57), which has no missing parents

18/09/08 12:52:24 INFO memory.MemoryStore: Block broadcast_4 stored as values in memory (estimated size 1616.0 B, free 366.0 MB)

18/09/08 12:52:24 INFO memory.MemoryStore: Block broadcast_4_piece0 stored as bytes in memory (estimated size 1033.0 B, free 366.0 MB)

18/09/08 12:52:24 INFO storage.BlockManagerInfo: Added broadcast_4_piece0 in memory on ip-172-31-84-37.ec2.internal:45790 (size: 1033.0 B, free: 366.3 MB)

18/09/08 12:52:24 INFO spark.SparkContext: Created broadcast 4 from broadcast at DAGScheduler.scala:1039

18/09/08 12:52:24 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 4 (ParallelCollectionRDD[7] at parallelize at GroupBy.java:57) (first 15 tasks are for partitions Vector(0))

18/09/08 12:52:24 INFO cluster.YarnScheduler: Adding task set 4.0 with 1 tasks

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 4.0 (TID 117, ip-172-31-84-37.ec2.internal, executor 2, partition 0, PROCESS_LOCAL, 7833 bytes)

18/09/08 12:52:24 INFO storage.BlockManagerInfo: Added broadcast_4_piece0 in memory on ip-172-31-84-37.ec2.internal:39965 (size: 1033.0 B, free: 366.3 MB)

18/09/08 12:52:24 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 4.0 (TID 117) in 195 ms on ip-172-31-84-37.ec2.internal (executor 2) (1/1)

18/09/08 12:52:24 INFO cluster.YarnScheduler: Removed TaskSet 4.0, whose tasks have all completed, from pool

18/09/08 12:52:24 INFO scheduler.DAGScheduler: ResultStage 4 (collect at GroupBy.java:59) finished in 0.245 s

18/09/08 12:52:24 INFO scheduler.DAGScheduler: Job 2 finished: collect at GroupBy.java:59, took 0.257083 s

18/09/08 12:52:24 INFO Configuration.deprecation: mapred.output.dir is deprecated. Instead, use mapreduce.output.fileoutputformat.outputdir

18/09/08 12:52:24 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/09/08 12:52:24 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/09/08 12:52:25 INFO spark.SparkContext: Starting job: runJob at SparkHadoopWriter.scala:78

18/09/08 12:52:25 INFO scheduler.DAGScheduler: Got job 3 (runJob at SparkHadoopWriter.scala:78) with 1 output partitions

18/09/08 12:52:25 INFO scheduler.DAGScheduler: Final stage: ResultStage 5 (runJob at SparkHadoopWriter.scala:78)

18/09/08 12:52:25 INFO scheduler.DAGScheduler: Parents of final stage: List()

18/09/08 12:52:25 INFO scheduler.DAGScheduler: Missing parents: List()

18/09/08 12:52:25 INFO scheduler.DAGScheduler: Submitting ResultStage 5 (MapPartitionsRDD[8] at saveAsTextFile at GroupBy.java:62), which has no missing parents

18/09/08 12:52:25 INFO memory.MemoryStore: Block broadcast_5 stored as values in memory (estimated size 79.1 KB, free 365.9 MB)

18/09/08 12:52:25 INFO memory.MemoryStore: Block broadcast_5_piece0 stored as bytes in memory (estimated size 29.1 KB, free 365.8 MB)

18/09/08 12:52:25 INFO storage.BlockManagerInfo: Added broadcast_5_piece0 in memory on ip-172-31-84-37.ec2.internal:45790 (size: 29.1 KB, free: 366.2 MB)

18/09/08 12:52:25 INFO spark.SparkContext: Created broadcast 5 from broadcast at DAGScheduler.scala:1039

18/09/08 12:52:25 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 5 (MapPartitionsRDD[8] at saveAsTextFile at GroupBy.java:62) (first 15 tasks are for partitions Vector(0))

18/09/08 12:52:25 INFO cluster.YarnScheduler: Adding task set 5.0 with 1 tasks

18/09/08 12:52:25 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 5.0 (TID 118, ip-172-31-84-37.ec2.internal, executor 2, partition 0, PROCESS_LOCAL, 7833 bytes)

18/09/08 12:52:25 INFO storage.BlockManagerInfo: Added broadcast_5_piece0 in memory on ip-172-31-84-37.ec2.internal:39965 (size: 29.1 KB, free: 366.2 MB)

18/09/08 12:52:25 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 5.0 (TID 118) in 520 ms on ip-172-31-84-37.ec2.internal (executor 2) (1/1)

18/09/08 12:52:25 INFO cluster.YarnScheduler: Removed TaskSet 5.0, whose tasks have all completed, from pool

18/09/08 12:52:25 INFO scheduler.DAGScheduler: ResultStage 5 (runJob at SparkHadoopWriter.scala:78) finished in 0.576 s

18/09/08 12:52:25 INFO scheduler.DAGScheduler: Job 3 finished: runJob at SparkHadoopWriter.scala:78, took 0.585924 s

18/09/08 12:52:25 INFO io.SparkHadoopWriter: Job job_20180908125224_0008 committed.

(4,85287)

(3,292090)

(2,17648209)

(1,36492406)

18/09/08 12:52:25 INFO server.AbstractConnector: Stopped
Spark@3e792ce3{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}

18/09/08 12:52:25 INFO ui.SparkUI: Stopped Spark web UI at http://ip-172-31-84-37.ec2.internal:4040

18/09/08 12:52:26 INFO cluster.YarnClientSchedulerBackend: Interrupting monitor thread

18/09/08 12:52:26 INFO cluster.YarnClientSchedulerBackend: Shutting down all executors

18/09/08 12:52:26 INFO cluster.YarnSchedulerBackend\$YarnDriverEndpoint: Asking each executor to shut down

18/09/08 12:52:26 INFO cluster.SchedulerExtensionServices: Stopping SchedulerExtensionServices

(serviceOption=None,

services=List(),

started=false)

18/09/08 12:52:26 INFO cluster.YarnClientSchedulerBackend: Stopped

18/09/08 12:52:26 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

18/09/08 12:52:26 INFO memory.MemoryStore: MemoryStore cleared

18/09/08 12:52:26 INFO storage.BlockManager: BlockManager stopped

18/09/08 12:52:26 INFO storage.BlockManagerMaster: BlockManagerMaster stopped

18/09/08 12:52:26 INFO scheduler.OutputCommitCoordinator\$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

18/09/08 12:52:26 INFO spark.SparkContext: Successfully stopped SparkContext

18/09/08 12:52:26 INFO util.ShutdownHookManager: Shutdown hook called

18/09/08 12:52:26 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-49267164-gd6e-47ea-8665-17bf9079b329

18/09/08 12:52:26 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-4359450e-d7c6-4177-874b-ed3615db796f

```
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-  
user/spark_assignment/output/spark3_groupby_orderby/part* | wc -l
```

```
4
```

```
[ec2-user@ip-172-31-84-37 ~]$
```