```
###############################################################
###############################################################
###

Pig 1 Program Logs + Run Command:

[root@ip-10-0-0-212 Pig]# pig Pig1.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.
2018-04-27 08:56:43,801 [main] INFO  org.apache.pig.Main - Apache Pig
version 0.12.0-cdh5.14.0 (rUnversioned directory) compiled Jan 06 2018,
13:37:28
2018-04-27 08:56:43,801 [main] INFO  org.apache.pig.Main - Logging error
messages to: /root/Pig/pig_1524819403778.log
2018-04-27 08:56:44,593 [main] INFO  org.apache.pig.impl.util.Utils -
Default bootup file /root/.pigbootup not found
2018-04-27 08:56:44,696 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is
deprecated. Instead, use mapreduce.jobtracker.address
2018-04-27 08:56:44,696 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 08:56:44,696 [main] INFO
org.apache.pig.backend.hadoop.executionengine.HExecutionEngine -
Connecting to hadoop file system at: hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020
2018-04-27 08:56:45,178 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 08:56:45,216 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 08:56:45,248 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 08:56:45,284 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 08:56:45,331 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 08:56:45,362 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 08:56:45,390 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 08:56:45,418 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 08:56:46,114 [main] INFO
org.apache.pig.tools.pigstats.ScriptState - Pig features used in the
script: FILTER
2018-04-27 08:56:46,149 [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]}
2018-04-27 08:56:46,175 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation -
mapred.textoutputformat.separator is deprecated. Instead, use
mapreduce.output.textoutputformat.separator
2018-04-27 08:56:46,260 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler -
File concatenation threshold: 100 optimistic? false
2018-04-27 08:56:46,281 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOp
timizer - MR plan size before optimization: 1
2018-04-27 08:56:46,281 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOp
timizer - MR plan size after optimization: 1
2018-04-27 08:56:46,362 [main] INFO
org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at
ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:8032
2018-04-27 08:56:46,505 [main] INFO
org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added
to the job
2018-04-27 08:56:46,572 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use
mapreduce.reduce.markreset.buffer.percent
2018-04-27 08:56:46,572 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - mapred.job.reduce.markreset.buffer.percent is not set, set to
default 0.3
2018-04-27 08:56:46,572 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress
is deprecated. Instead, use mapreduce.output.fileoutputformat.compress
2018-04-27 08:56:48,337 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - creating jar file Job5405864530388369829.jar
2018-04-27 08:56:52,043 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - jar file Job5405864530388369829.jar created
2018-04-27 08:56:52,043 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.jar is
deprecated. Instead, use mapreduce.job.jar
2018-04-27 08:56:52,060 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Setting up single store job
2018-04-27 08:56:52,066 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false,
will not generate code.
2018-04-27 08:56:52,066 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Starting process to move
generated code to distributed cache
2018-04-27 08:56:52,066 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []
2018-04-27 08:56:52,119 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 1 map-reduce job(s) waiting for submission.
```

```
2018-04-27 08:56:52,121 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.tracker.http.address is deprecated. Instead, use
mapreduce.jobtracker.http.address
2018-04-27 08:56:52,127 [JobControl] INFO
org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at
ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:8032
2018-04-27 08:56:52,168 [JobControl] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 08:56:53,024 [JobControl] INFO
org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths
to process : 6
2018-04-27 08:56:53,024 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total
input paths to process : 6
2018-04-27 08:56:53,066 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total
input paths (combined) to process : 36
2018-04-27 08:56:53,516 [JobControl] INFO
org.apache.hadoop.mapreduce.JobSubmitter - number of splits:36
2018-04-27 08:56:54,025 [JobControl] INFO
org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job:
job_1524817480909_0002
2018-04-27 08:56:54,625 [JobControl] INFO
org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted
application application_1524817480909_0002
2018-04-27 08:56:54,658 [JobControl] INFO
org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-
0-0-212.ap-south-
1.compute.internal:8088/proxy/application_1524817480909_0002/
2018-04-27 08:56:54,658 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - HadoopJobId: job_1524817480909_0002
2018-04-27 08:56:54,658 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - Processing aliases data,filtered
2018-04-27 08:56:54,658 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - detailed locations: M: data[1,7],filtered[4,11],data[-1,-
1],null[-1,-1] C:  R:
2018-04-27 08:56:54,696 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 0% complete
2018-04-27 08:57:31,626 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 4% complete
2018-04-27 08:58:01,594 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 8% complete
2018-04-27 08:58:26,634 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 12% complete
2018-04-27 08:58:52,698 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 16% complete
2018-04-27 08:59:17,335 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 20% complete
```

```
2018-04-27 08:59:39,747 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 25% complete
2018-04-27 09:00:03,336 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 29% complete
2018-04-27 09:00:19,151 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 33% complete
2018-04-27 09:00:35,879 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 37% complete
2018-04-27 09:00:51,022 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 41% complete
2018-04-27 09:01:07,737 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 45% complete
2018-04-27 09:01:20,469 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 50% complete
2018-04-27 09:01:25,369 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is
deprecated. Instead, use mapreduce.job.reduces
2018-04-27 09:01:25,410 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 100% complete
2018-04-27 09:01:25,412 [main] INFO
org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion   PigVersion        UserId  StartedAt        FinishedAt
Features
2.6.0-cdh5.14.0 0.12.0-cdh5.14.0        root    2018-04-27 08:56:46
2018-04-27 09:01:25     FILTER

Success!

Job Stats (time in seconds):
JobId   Maps        Reduces MaxMapTime        MinMapTIme        AvgMapTime
MedianMapTime   MaxReduceTime   MinReduceTime AvgReduceTime
MedianReducetime        Alias   Feature Outputs
job_1524817480909_0002  36        0         29        10        20        20
n/a     n/a     n/a     n/a     data,filtered MAP_ONLY hdfs://ip-10-0-0-
212.ap-south-1.compute.internal:8020/user/root/pigoutput/PigOutput1.out,

Input(s):
Successfully read 54518004 records (4810074395 bytes) from: "hdfs://ip-
10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_*"

Output(s):
Successfully stored 1 records (94 bytes) in: "hdfs://ip-10-0-0-212.ap-
south-1.compute.internal:8020/user/root/pigoutput/PigOutput1.out"

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_1524817480909_0002

2018-04-27 09:01:25,561 [main] WARN
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 6
time(s).
2018-04-27 09:01:25,561 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - Success!

################################   End of Logs for
Pig1.pig Script   ##

Pig 2 Program Logs + Run Command:

[root@ip-10-0-0-212 Pig]# pig Pig2.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.
2018-04-27 10:41:26,290 [main] INFO org.apache.pig.Main - Apache Pig
version 0.12.0-cdh5.14.0 (rUnversion compiled Jan 06 2018, 13:37:28
2018-04-27 10:41:26,290 [main] INFO org.apache.pig.Main - Logging error
messages to: /root/Pig/pig_152482
2018-04-27 10:41:27,106 [main] INFO org.apache.pig.impl.util.Utils -
Default bootup file /root/.pigbootup
2018-04-27 10:41:27,213 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.job.trackerd.
Instead, use mapreduce.jobtracker.address
2018-04-27 10:41:27,213 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
isInstead, use fs.defaultFS
2018-04-27 10:41:27,213 [main] INFO
org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connoop
file system at: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020
2018-04-27 10:41:27,658 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
isInstead, use fs.defaultFS
2018-04-27 10:41:27,693 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
isInstead, use fs.defaultFS
2018-04-27 10:41:27,724 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
isInstead, use fs.defaultFS
2018-04-27 10:41:27,758 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
isInstead, use fs.defaultFS
2018-04-27 10:41:27,799 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
isInstead, use fs.defaultFS
2018-04-27 10:41:27,832 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
isInstead, use fs.defaultFS

```
2018-04-27 10:41:27,862 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
isInstead, use fs.defaultFS
2018-04-27 10:41:27,893 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
isInstead, use fs.defaultFS
2018-04-27 10:41:28,682 [main] INFO
org.apache.pig.tools.pigstats.ScriptState - Pig features used in the R
2018-04-27 10:41:28,713 [main] INFO
org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer -
{RULEdForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite,
GroupByConstParallelSetter, ImplicitSplitInserimizer,
LoadTypeCastInserter, MergeFilter, MergeForEach,
NewPartitionFilterOptimizer, PushDownForEachFlattter, SplitFilter,
StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier,
PartitionFilte
2018-04-27 10:41:28,737 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputfor
is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2018-04-27 10:41:28,822 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompilcate
nation threshold: 100 optimistic? false
2018-04-27 10:41:28,843 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQue MR
plan size before optimization: 1
2018-04-27 10:41:28,843 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQue MR
plan size after optimization: 1
2018-04-27 10:41:28,946 [main] INFO
org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager-
212.ap-south-1.compute.internal/10.0.0.212:8032
2018-04-27 10:41:29,139 [main] INFO
org.apache.pig.tools.pigstats.ScriptState - Pig script settings are aob
2018-04-27 10:41:29,193 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.reduce.fer.percent is deprecated. Instead, use
mapreduce.reduce.markreset.buffer.percent
2018-04-27 10:41:29,193 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobContrmapr
ed.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2018-04-27 10:41:29,193 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation -
mapred.output.compcated. Instead, use
mapreduce.output.fileoutputformat.compress
2018-04-27 10:41:30,133 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobContrcrea
ting jar file Job5868716944557052265.jar
2018-04-27 10:41:33,870 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobContrjar
file Job5868716944557052265.jar created
2018-04-27 10:41:33,870 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.jar is deprad,
use mapreduce.job.jar
2018-04-27 10:41:33,891 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobContrSett
ing up single store job
2018-04-27 10:41:33,900 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is fa
generate code.
```

2018-04-27 10:41:33,900 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Starting process to move geto
distributed cache
2018-04-27 10:41:33,900 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuplth
classes to deserialize []
2018-04-27 10:41:33,959 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc
map-reduce job(s) waiting for submission.
2018-04-27 10:41:33,961 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is
deprecated. Instead, use mapreduce.jobtracker.http.address
2018-04-27 10:41:33,968 [JobControl] INFO
org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceM10-0-0-
212.ap-south-1.compute.internal/10.0.0.212:8032
2018-04-27 10:41:34,007 [JobControl] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.nated.
Instead, use fs.defaultFS
2018-04-27 10:41:34,404 [JobControl] INFO
org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total i process :
6
2018-04-27 10:41:34,404 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil -paths to
process : 6
2018-04-27 10:41:34,442 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil -paths
(combined) to process : 36
2018-04-27 10:41:34,900 [JobControl] INFO
org.apache.hadoop.mapreduce.JobSubmitter - number of splits:36
2018-04-27 10:41:35,415 [JobControl] INFO
org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens
fo24825160576_0001
2018-04-27 10:41:36,016 [JobControl] INFO
org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submittn
application_1524825160576_0001
2018-04-27 10:41:36,049 [JobControl] INFO
org.apache.hadoop.mapreduce.Job - The url to track the job: htt-212.ap-
south-1.compute.internal:8088/proxy/application_1524825160576_0001/
2018-04-27 10:41:36,050 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReducadoo
pJobId: job_1524825160576_0001
2018-04-27 10:41:36,050 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReducroce
ssing aliases data,filtered
2018-04-27 10:41:36,050 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReducetai
led locations: M: data[1,7],data[-1,-1],filtered[4,11] C: R:
2018-04-27 10:41:36,083 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc%
complete
2018-04-27 10:42:11,666 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc%
complete
2018-04-27 10:42:36,179 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc%
complete
2018-04-27 10:43:01,964 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc2%
complete

2018-04-27 10:43:27,168 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc6%
complete
2018-04-27 10:43:53,249 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc0%
complete
2018-04-27 10:44:16,909 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc5%
complete
2018-04-27 10:44:40,938 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc9%
complete
2018-04-27 10:45:01,038 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc3%
complete
2018-04-27 10:45:22,101 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc7%
complete
2018-04-27 10:45:42,285 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc1%
complete
2018-04-27 10:46:00,040 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc5%
complete
2018-04-27 10:46:19,408 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc9%
complete
2018-04-27 10:46:26,891 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasked.
Instead, use mapreduce.job.reduces
2018-04-27 10:46:26,943 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduc00%
complete
2018-04-27 10:46:26,945 [main] INFO
org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion   PigVersion   UserId StartedAt   FinishedAt
Features
2.6.0-cdh5.14.0 0.12.0-cdh5.14.0   root   2018-04-27 10:41:29
2018-04-27 10:46:26   FILTER

Success!

Job Stats (time in seconds):
JobId   Maps   Reduces MaxMapTime   MinMapTIme   AvgMapTime
MedianMapTime   MaxReduceTime   MivgReduceTime MedianReducetime
Alias   Feature Outputs
job_1524825160576_0001 36   0   25   17   21   21
n/a   n/a   n/a   n/a   daAP_ONLY hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/pigoutput/PigOutput2.out,

Input(s):
Successfully read 54518004 records (4810074395 bytes) from: "hdfs://ip-
10-0-0-212.ap-south-
1.compute.inter/root/spark_assignment/input_dataset/yellow_tripdata_*"

Output(s):
Successfully stored 31029 records (2948843 bytes) in: "hdfs://ip-10-0-0-
212.ap-south-1.compute.internal:80pigoutput/PigOutput2.out"

```
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_1524825160576_0001


2018-04-27 10:46:27,111 [main] WARN
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReducncou
ntered Warning ACCESSING_NON_EXISTENT_FIELD 96 time(s).
2018-04-27 10:46:27,111 [main] WARN
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReducncou
ntered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 84 time(s).
2018-04-27 10:46:27,111 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReducucce
ss!

#######################################                End of Logs
for Pig2.pig Script
#########################################################

Pig 3 Program Logs + Run Command:

[root@ip-10-0-0-212 Pig]# pig Pig3.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.
2018-04-27 14:48:38,599 [main] INFO  org.apache.pig.Main - Apache Pig
version 0.12.0-cdh5.14.0 (rUnversioned directory) compiled Jan 06 2018,
13:37:28
2018-04-27 14:48:38,599 [main] INFO  org.apache.pig.Main - Logging error
messages to: /root/Pig/pig_1524840518575.log
2018-04-27 14:48:39,372 [main] INFO  org.apache.pig.impl.util.Utils -
Default bootup file /root/.pigbootup not found
2018-04-27 14:48:39,480 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is
deprecated. Instead, use mapreduce.jobtracker.address
2018-04-27 14:48:39,481 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 14:48:39,481 [main] INFO
org.apache.pig.backend.hadoop.executionengine.HExecutionEngine -
Connecting to hadoop file system at: hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020
2018-04-27 14:48:39,997 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 14:48:40,035 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 14:48:40,069 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
```

```
2018-04-27 14:48:40,099 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 14:48:40,143 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 14:48:40,174 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 14:48:40,204 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 14:48:40,234 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 14:48:40,987 [main] WARN  org.apache.pig.PigServer -
Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).
2018-04-27 14:48:41,006 [main] INFO
org.apache.pig.tools.pigstats.ScriptState - Pig features used in the
script: GROUP_BY,ORDER_BY,FILTER
2018-04-27 14:48:41,037 [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]}
2018-04-27 14:48:41,063 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation -
mapred.textoutputformat.separator is deprecated. Instead, use
mapreduce.output.textoutputformat.separator
2018-04-27 14:48:41,139 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler -
File concatenation threshold: 100 optimistic? false
2018-04-27 14:48:41,175 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.CombinerOpti
mizer - Choosing to move algebraic foreach to combiner
2018-04-27 14:48:41,194 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOp
timizer - MR plan size before optimization: 3
2018-04-27 14:48:41,194 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOp
timizer - MR plan size after optimization: 3
2018-04-27 14:48:41,278 [main] INFO
org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at
ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:8032
2018-04-27 14:48:41,433 [main] INFO
org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added
to the job
2018-04-27 14:48:41,504 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use
mapreduce.reduce.markreset.buffer.percent
2018-04-27 14:48:41,504 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - mapred.job.reduce.markreset.buffer.percent is not set, set to
default 0.3
```

```
2018-04-27 14:48:41,504 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress
is deprecated. Instead, use mapreduce.output.fileoutputformat.compress
2018-04-27 14:48:41,506 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Reduce phase detected, estimating # of required reducers.
2018-04-27 14:48:41,507 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Using reducer estimator:
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeRed
ucerEstimator
2018-04-27 14:48:41,511 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeRed
ucerEstimator - BytesPerReducer=1000000000 maxReducers=999
totalInputFileSize=4807895414
2018-04-27 14:48:41,512 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Setting Parallelism to 5
2018-04-27 14:48:41,512 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is
deprecated. Instead, use mapreduce.job.reduces
2018-04-27 14:48:42,848 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - creating jar file Job6463657566417871112.jar
2018-04-27 14:48:46,591 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - jar file Job6463657566417871112.jar created
2018-04-27 14:48:46,591 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation - mapred.jar is
deprecated. Instead, use mapreduce.job.jar
2018-04-27 14:48:46,610 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Setting up single store job
2018-04-27 14:48:46,616 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false,
will not generate code.
2018-04-27 14:48:46,616 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Starting process to move
generated code to distributed cache
2018-04-27 14:48:46,616 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []
2018-04-27 14:48:46,692 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 1 map-reduce job(s) waiting for submission.
2018-04-27 14:48:46,692 [main] INFO
org.apache.hadoop.conf.Configuration.deprecation -
mapred.job.tracker.http.address is deprecated. Instead, use
mapreduce.jobtracker.http.address
2018-04-27 14:48:46,697 [JobControl] INFO
org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at
ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:8032
2018-04-27 14:48:46,711 [JobControl] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 14:48:47,987 [JobControl] INFO
org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths
to process : 6
```

2018-04-27 14:48:47,987 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total
input paths to process : 6
2018-04-27 14:48:48,025 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total
input paths (combined) to process : 36
2018-04-27 14:48:48,893 [JobControl] INFO
org.apache.hadoop.mapreduce.JobSubmitter - number of splits:36
2018-04-27 14:48:49,486 [JobControl] INFO
org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job:
job_1524837966563_0014
2018-04-27 14:48:49,667 [JobControl] INFO
org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted
application application_1524837966563_0014
2018-04-27 14:48:49,701 [JobControl] INFO
org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-
0-0-212.ap-south-
1.compute.internal:8088/proxy/application_1524837966563_0014/
2018-04-27 14:48:49,701 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - HadoopJobId: job_1524837966563_0014
2018-04-27 14:48:49,701 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - Processing aliases data,data_not_null,group_count,grouped_data
2018-04-27 14:48:49,701 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - detailed locations: M:
data[1,7],data_not_null[20,16],group_count[22,14],grouped_data[21,15] C:
group_count[22,14],grouped_data[21,15] R: group_count[22,14]
2018-04-27 14:48:49,748 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 0% complete
2018-04-27 14:50:17,527 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 4% complete
2018-04-27 14:51:19,373 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 8% complete
2018-04-27 14:52:18,757 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 12% complete
2018-04-27 14:53:07,418 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 16% complete
2018-04-27 14:53:23,784 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 21% complete
2018-04-27 14:53:30,175 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 26% complete
2018-04-27 14:53:32,804 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 33% complete
2018-04-27 14:53:35,706 [main] INFO
org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added
to the job
2018-04-27 14:53:35,714 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - mapred.job.reduce.markreset.buffer.percent is not set, set to
default 0.3

```
2018-04-27 14:53:35,715 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Reduce phase detected, estimating # of required reducers.
2018-04-27 14:53:35,715 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Using reducer estimator:
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeRed
ucerEstimator
2018-04-27 14:53:35,723 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeRed
ucerEstimator - BytesPerReducer=1000000000 maxReducers=999
totalInputFileSize=76
2018-04-27 14:53:35,724 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Setting Parallelism to 1
2018-04-27 14:53:37,499 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - creating jar file Job6735876048159497102.jar
2018-04-27 14:53:41,214 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - jar file Job6735876048159497102.jar created
2018-04-27 14:53:41,225 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Setting up single store job
2018-04-27 14:53:41,225 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false,
will not generate code.
2018-04-27 14:53:41,225 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Starting process to move
generated code to distributed cache
2018-04-27 14:53:41,225 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []
2018-04-27 14:53:41,243 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 1 map-reduce job(s) waiting for submission.
2018-04-27 14:53:41,246 [JobControl] INFO
org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at
ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:8032
2018-04-27 14:53:41,250 [JobControl] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
2018-04-27 14:53:41,718 [JobControl] INFO
org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths
to process : 5
2018-04-27 14:53:41,718 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total
input paths to process : 5
2018-04-27 14:53:41,719 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total
input paths (combined) to process : 1
2018-04-27 14:53:42,538 [JobControl] INFO
org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2018-04-27 14:53:42,971 [JobControl] INFO
org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job:
job_1524837966563_0015
2018-04-27 14:53:42,994 [JobControl] INFO
org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted
application application_1524837966563_0015
```

```
2018-04-27 14:53:42,996 [JobControl] INFO
org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-
0-0-212.ap-south-
1.compute.internal:8088/proxy/application_1524837966563_0015/
2018-04-27 14:53:42,996 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - HadoopJobId: job_1524837966563_0015
2018-04-27 14:53:42,996 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - Processing aliases sorted_count
2018-04-27 14:53:42,996 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - detailed locations: M: sorted_count[23,15] C:  R:
2018-04-27 14:53:53,336 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 50% complete
2018-04-27 14:53:59,596 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 66% complete
2018-04-27 14:54:03,334 [main] INFO
org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added
to the job
2018-04-27 14:54:03,340 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - mapred.job.reduce.markreset.buffer.percent is not set, set to
default 0.3
2018-04-27 14:54:03,340 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Reduce phase detected, estimating # of required reducers.
2018-04-27 14:54:03,340 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Setting Parallelism to 1
2018-04-27 14:54:05,072 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - creating jar file Job7596507323199534623.jar
2018-04-27 14:54:08,694 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - jar file Job7596507323199534623.jar created
2018-04-27 14:54:08,704 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCo
mpiler - Setting up single store job
2018-04-27 14:54:08,705 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false,
will not generate code.
2018-04-27 14:54:08,705 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Starting process to move
generated code to distributed cache
2018-04-27 14:54:08,705 [main] INFO
org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []
2018-04-27 14:54:08,721 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 1 map-reduce job(s) waiting for submission.
2018-04-27 14:54:08,722 [JobControl] INFO
org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at
ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:8032
2018-04-27 14:54:08,727 [JobControl] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is
deprecated. Instead, use fs.defaultFS
```

```
2018-04-27 14:54:08,785 [JobControl] INFO
org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths
to process : 5
2018-04-27 14:54:08,785 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total
input paths to process : 5
2018-04-27 14:54:08,785 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total
input paths (combined) to process : 1
2018-04-27 14:54:08,805 [JobControl] INFO
org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2018-04-27 14:54:08,819 [JobControl] INFO
org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job:
job_1524837966563_0016
2018-04-27 14:54:08,852 [JobControl] INFO
org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted
application application_1524837966563_0016
2018-04-27 14:54:08,854 [JobControl] INFO
org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-
0-0-212.ap-south-
1.compute.internal:8088/proxy/application_1524837966563_0016/
2018-04-27 14:54:09,223 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - HadoopJobId: job_1524837966563_0016
2018-04-27 14:54:09,223 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - Processing aliases sorted_count
2018-04-27 14:54:09,223 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - detailed locations: M: sorted_count[23,15] C:  R:
2018-04-27 14:54:19,570 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 83% complete
2018-04-27 14:54:29,079 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - 100% complete
2018-04-27 14:54:29,100 [main] INFO
org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt
Features
2.6.0-cdh5.14.0 0.12.0-cdh5.14.0        root    2018-04-27 14:48:41
2018-04-27 14:54:29      GROUP_BY,ORDER_BY,FILTER

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTIme      AvgMapTime
MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime
MedianReducetime        Alias   Feature Outputs
job_1524837966563_0014  36      5       28      14      19      18
37      512     5       data,data_not_null,group_count,grouped_data
GROUP_BY,COMBINER
job_1524837966563_0015  1       1       2       2       2       2       3
33      3       sorted_count    SAMPLER
job_1524837966563_0016  1       1       3       3       3       3       4
44      4       sorted_count    ORDER_BY        hdfs://ip-10-0-0-212.ap-
south-1.compute.internal:8020/user/root/pigoutput/PigOutput3.out,

Input(s):
```

Successfully read 54518004 records (4810074395 bytes) from: "hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_*"

Output(s):
Successfully stored 4 records (39 bytes) in: "hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/pigoutput/PigOutput3.out"

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_1524837966563_0014 ->   job_1524837966563_0015,
job_1524837966563_0015 ->   job_1524837966563_0016,
job_1524837966563_0016

2018-04-27 14:54:30,103 [main] INFO org.apache.hadoop.ipc.Client -
Retrying connect to server: ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:45113. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-04-27 14:54:31,103 [main] INFO org.apache.hadoop.ipc.Client -
Retrying connect to server: ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:45113. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-04-27 14:54:32,104 [main] INFO org.apache.hadoop.ipc.Client -
Retrying connect to server: ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:45113. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-04-27 14:54:32,210 [main] INFO
org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2018-04-27 14:54:33,429 [main] INFO org.apache.hadoop.ipc.Client -
Retrying connect to server: ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:39681. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-04-27 14:54:34,429 [main] INFO org.apache.hadoop.ipc.Client -
Retrying connect to server: ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:39681. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-04-27 14:54:35,430 [main] INFO org.apache.hadoop.ipc.Client -
Retrying connect to server: ip-10-0-0-212.ap-south-1.compute.internal/10.0.0.212:39681. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)
2018-04-27 14:54:35,533 [main] INFO
org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-04-27 14:54:35,732 [main] INFO
org.apache.hadoop.mapred.ClientServiceDelegate - Application state is
completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history
server
2018-04-27 14:54:35,797 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLau
ncher - Success!

############################   End of Logs for
Pig3.pig Script
###

Spark 1 Program Logs + Submit Command

[root@ip-10-0-0-212 ~]# spark-submit --class
spark.assignment1.AssignmentProblem1 --master yarn --deploy-mode client -
-name sparkquestion1 --conf "spark.app.id=sparkquestion1" /home/ec2-
user/assignment1-0.0.1-SNAPSHOT-jar-with-dependencies.jar
/user/root/spark_assignment/input_dataset/yellow_tripdata_*
/user/root/spark_assignment/sparkoutput/que1_1out
Conf has been set
18/05/02 06:50:13 INFO spark.SparkContext: Running Spark version 1.6.0
18/05/02 06:50:14 INFO spark.SecurityManager: Changing view acls to: root
18/05/02 06:50:14 INFO spark.SecurityManager: Changing modify acls to:
root
18/05/02 06:50:14 INFO spark.SecurityManager: SecurityManager:
authentication disabled; ui acls disabled; users with view permissions:
Set(root); users with modify permissions: Set(root)
18/05/02 06:50:14 INFO util.Utils: Successfully started service
'sparkDriver' on port 39481.
18/05/02 06:50:15 INFO slf4j.Slf4jLogger: Slf4jLogger started
18/05/02 06:50:15 INFO Remoting: Starting remoting
18/05/02 06:50:15 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://sparkDriverActorSystem@10.0.0.212:35074]
18/05/02 06:50:15 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://sparkDriverActorSystem@10.0.0.212:35074]
18/05/02 06:50:15 INFO util.Utils: Successfully started service
'sparkDriverActorSystem' on port 35074.
18/05/02 06:50:15 INFO spark.SparkEnv: Registering MapOutputTracker
18/05/02 06:50:15 INFO spark.SparkEnv: Registering BlockManagerMaster
18/05/02 06:50:15 INFO storage.DiskBlockManager: Created local directory
at /tmp/blockmgr-4d5102c6-b9db-4575-b5d4-5fa52f0cd663
18/05/02 06:50:15 INFO storage.MemoryStore: MemoryStore started with
capacity 530.0 MB
18/05/02 06:50:15 INFO spark.SparkEnv: Registering
OutputCommitCoordinator
18/05/02 06:50:15 INFO util.Utils: Successfully started service 'SparkUI'
on port 4040.
18/05/02 06:50:15 INFO ui.SparkUI: Started SparkUI at
http://10.0.0.212:4040
18/05/02 06:50:15 INFO spark.SparkContext: Added JAR file:/home/ec2-
user/assignment1-0.0.1-SNAPSHOT-jar-with-dependencies.jar at
spark://10.0.0.212:39481/jars/assignment1-0.0.1-SNAPSHOT-jar-with-
dependencies.jar with timestamp 1525243815684
18/05/02 06:50:15 INFO executor.Executor: Starting executor ID driver on
host localhost
18/05/02 06:50:15 INFO util.Utils: Successfully started service
'org.apache.spark.network.netty.NettyBlockTransferService' on port 46178.
18/05/02 06:50:15 INFO netty.NettyBlockTransferService: Server created on
46178

18/05/02 06:50:15 INFO storage.BlockManager: external shuffle service
port = 7337
18/05/02 06:50:15 INFO storage.BlockManagerMaster: Trying to register
BlockManager
18/05/02 06:50:15 INFO storage.BlockManagerMasterEndpoint: Registering
block manager localhost:46178 with 530.0 MB RAM, BlockManagerId(driver,
localhost, 46178)
18/05/02 06:50:15 INFO storage.BlockManagerMaster: Registered
BlockManager
18/05/02 06:50:17 INFO scheduler.EventLoggingListener: Logging events to
hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/spark/applicationHistory/local-1525243815733
18/05/02 06:50:17 INFO spark.SparkContext: Registered listener
com.cloudera.spark.lineage.ClouderaNavigatorListener
Context has been initialized
18/05/02 06:50:18 INFO storage.MemoryStore: Block broadcast_0 stored as
values in memory (estimated size 303.4 KB, free 529.7 MB)
18/05/02 06:50:18 INFO storage.MemoryStore: Block broadcast_0_piece0
stored as bytes in memory (estimated size 26.1 KB, free 529.7 MB)
18/05/02 06:50:18 INFO storage.BlockManagerInfo: Added broadcast_0_piece0
in memory on localhost:46178 (size: 26.1 KB, free: 530.0 MB)
18/05/02 06:50:18 INFO spark.SparkContext: Created broadcast 0 from
textFile at AssignmentProblem1.java:23
File has been read successfully
Writing the output now
18/05/02 06:50:18 INFO Configuration.deprecation: mapred.tip.id is
deprecated. Instead, use mapreduce.task.id
18/05/02 06:50:18 INFO Configuration.deprecation: mapred.task.id is
deprecated. Instead, use mapreduce.task.attempt.id
18/05/02 06:50:18 INFO Configuration.deprecation: mapred.task.is.map is
deprecated. Instead, use mapreduce.task.ismap
18/05/02 06:50:18 INFO Configuration.deprecation: mapred.task.partition
is deprecated. Instead, use mapreduce.task.partition
18/05/02 06:50:18 INFO Configuration.deprecation: mapred.job.id is
deprecated. Instead, use mapreduce.job.id
18/05/02 06:50:18 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:50:18 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:50:18 INFO mapred.FileInputFormat: Total input paths to
process : 6
18/05/02 06:50:18 INFO spark.SparkContext: Starting job: saveAsTextFile
at AssignmentProblem1.java:42
18/05/02 06:50:18 INFO scheduler.DAGScheduler: Got job 0 (saveAsTextFile
at AssignmentProblem1.java:42) with 39 output partitions
18/05/02 06:50:18 INFO scheduler.DAGScheduler: Final stage: ResultStage 0
(saveAsTextFile at AssignmentProblem1.java:42)
18/05/02 06:50:18 INFO scheduler.DAGScheduler: Parents of final stage:
List()
18/05/02 06:50:18 INFO scheduler.DAGScheduler: Missing parents: List()
18/05/02 06:50:18 INFO scheduler.DAGScheduler: Submitting ResultStage 0
(MapPartitionsRDD[3] at saveAsTextFile at AssignmentProblem1.java:42),
which has no missing parents
18/05/02 06:50:18 INFO storage.MemoryStore: Block broadcast_1 stored as
values in memory (estimated size 81.3 KB, free 529.6 MB)
18/05/02 06:50:18 INFO storage.MemoryStore: Block broadcast_1_piece0
stored as bytes in memory (estimated size 29.2 KB, free 529.6 MB)
18/05/02 06:50:18 INFO storage.BlockManagerInfo: Added broadcast_1_piece0
in memory on localhost:46178 (size: 29.2 KB, free: 530.0 MB)

18/05/02 06:50:18 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1004
18/05/02 06:50:18 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 0 (MapPartitionsRDD[3] at saveAsTextFile at AssignmentProblem1.java:42) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))
18/05/02 06:50:18 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 39 tasks
18/05/02 06:50:18 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, executor driver, partition 0, ANY, 2316 bytes)
18/05/02 06:50:18 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, localhost, executor driver, partition 1, ANY, 2316 bytes)
18/05/02 06:50:18 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, localhost, executor driver, partition 2, ANY, 2316 bytes)
18/05/02 06:50:18 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, localhost, executor driver, partition 3, ANY, 2316 bytes)
18/05/02 06:50:18 INFO executor.Executor: Running task 0.0 in stage 0.0 (TID 0)
18/05/02 06:50:18 INFO executor.Executor: Running task 2.0 in stage 0.0 (TID 2)
18/05/02 06:50:18 INFO executor.Executor: Running task 1.0 in stage 0.0 (TID 1)
18/05/02 06:50:18 INFO executor.Executor: Running task 3.0 in stage 0.0 (TID 3)
18/05/02 06:50:18 INFO executor.Executor: Fetching spark://10.0.0.212:39481/jars/assignment1-0.0.1-SNAPSHOT-jar-with-dependencies.jar with timestamp 1525243815684
18/05/02 06:50:18 INFO spark.ExecutorAllocationManager: New executor driver has registered (new total is 1)
18/05/02 06:50:18 INFO util.Utils: Fetching spark://10.0.0.212:39481/jars/assignment1-0.0.1-SNAPSHOT-jar-with-dependencies.jar to /tmp/spark-87d79c82-ec11-4328-b741-2cc777087a63/userFiles-2da75ac5-c190-4e1d-94d6-04392d94d813/fetchFileTemp4739834040168063530.tmp
18/05/02 06:50:19 INFO executor.Executor: Adding file:/tmp/spark-87d79c82-ec11-4328-b741-2cc777087a63/userFiles-2da75ac5-c190-4e1d-94d6-04392d94d813/assignment1-0.0.1-SNAPSHOT-jar-with-dependencies.jar to class loader
18/05/02 06:50:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:134217728+134217728
18/05/02 06:50:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:0+134217728
18/05/02 06:50:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:268435456+134217728
18/05/02 06:50:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:402653184+134217728
18/05/02 06:50:19 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/05/02 06:50:19 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 06:50:19 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 06:50:19 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 06:50:19 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:19 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 2 total executors!
18/05/02 06:50:20 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 06:50:20 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 06:50:20 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 06:50:20 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 06:50:20 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:20 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 3 total executors!
18/05/02 06:50:21 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:21 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 4 total executors!
18/05/02 06:50:22 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:22 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 5 total executors!
18/05/02 06:50:23 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:23 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 6 total executors!
18/05/02 06:50:24 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:24 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 7 total executors!
18/05/02 06:50:25 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:25 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 8 total executors!
18/05/02 06:50:26 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:26 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 9 total executors!
18/05/02 06:50:27 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:27 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 10 total executors!
18/05/02 06:50:28 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:28 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 11 total executors!
18/05/02 06:50:29 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

```
18/05/02 06:50:29 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 12 total executors!
18/05/02 06:50:30 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:30 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 13 total executors!
18/05/02 06:50:31 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:31 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 14 total executors!
18/05/02 06:50:32 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:32 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 15 total executors!
18/05/02 06:50:33 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:33 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 16 total executors!
18/05/02 06:50:34 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:34 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 17 total executors!
18/05/02 06:50:35 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:35 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 18 total executors!
18/05/02 06:50:36 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:36 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 19 total executors!
18/05/02 06:50:37 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:37 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 20 total executors!
18/05/02 06:50:38 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:38 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 21 total executors!
18/05/02 06:50:39 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:39 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 22 total executors!
18/05/02 06:50:40 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:40 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 23 total executors!
18/05/02 06:50:41 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:41 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 24 total executors!
18/05/02 06:50:42 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:42 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 25 total executors!
18/05/02 06:50:43 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000001_1' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000001
18/05/02 06:50:43 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000001_1: Committed
```

18/05/02 06:50:43 INFO executor.Executor: Finished task 1.0 in stage 0.0 (TID 1). 2295 bytes result sent to driver
18/05/02 06:50:43 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 0.0 (TID 4, localhost, executor driver, partition 4, ANY, 2316 bytes)
18/05/02 06:50:43 INFO executor.Executor: Running task 4.0 in stage 0.0 (TID 4)
18/05/02 06:50:43 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 0.0 (TID 1) in 24644 ms on localhost (executor driver) (1/39)
18/05/02 06:50:43 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:536870912+134217728
18/05/02 06:50:43 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020650_0000_m_000002_2' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/_temporary/0/task_201805020650_0000_m_000002
18/05/02 06:50:43 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020650_0000_m_000002_2: Committed
18/05/02 06:50:43 INFO executor.Executor: Finished task 2.0 in stage 0.0 (TID 2). 2360 bytes result sent to driver
18/05/02 06:50:43 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 0.0 (TID 5, localhost, executor driver, partition 5, ANY, 2316 bytes)
18/05/02 06:50:43 INFO executor.Executor: Running task 5.0 in stage 0.0 (TID 5)
18/05/02 06:50:43 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 24773 ms on localhost (executor driver) (2/39)
18/05/02 06:50:43 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:671088640+86047889
18/05/02 06:50:43 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 06:50:43 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 06:50:43 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:43 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 26 total executors!
18/05/02 06:50:44 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 06:50:44 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 06:50:44 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:50:44 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 27 total executors!
18/05/02 06:50:45 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020650_0000_m_000003_3' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/_temporary/0/task_201805020650_0000_m_000003
18/05/02 06:50:45 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020650_0000_m_000003_3: Committed
18/05/02 06:50:45 INFO executor.Executor: Finished task 3.0 in stage 0.0 (TID 3). 2360 bytes result sent to driver

```
18/05/02 06:50:45 INFO scheduler.TaskSetManager: Starting task 6.0 in
stage 0.0 (TID 6, localhost, executor driver, partition 6, ANY, 2316
bytes)
18/05/02 06:50:45 INFO executor.Executor: Running task 6.0 in stage 0.0
(TID 6)
18/05/02 06:50:45 INFO scheduler.TaskSetManager: Finished task 3.0 in
stage 0.0 (TID 3) in 26257 ms on localhost (executor driver) (3/39)
18/05/02 06:50:45 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:0+134217728
18/05/02 06:50:45 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:50:45 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:50:45 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:45 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 28 total executors!
18/05/02 06:50:45 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000000_0' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000000
18/05/02 06:50:45 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000000_0: Committed
18/05/02 06:50:45 INFO executor.Executor: Finished task 0.0 in stage 0.0
(TID 0). 2360 bytes result sent to driver
18/05/02 06:50:45 INFO scheduler.TaskSetManager: Starting task 7.0 in
stage 0.0 (TID 7, localhost, executor driver, partition 7, ANY, 2316
bytes)
18/05/02 06:50:45 INFO executor.Executor: Running task 7.0 in stage 0.0
(TID 7)
18/05/02 06:50:45 INFO scheduler.TaskSetManager: Finished task 0.0 in
stage 0.0 (TID 0) in 27003 ms on localhost (executor driver) (4/39)
18/05/02 06:50:45 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:134217728+134217728
18/05/02 06:50:46 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:46 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 29 total executors!
18/05/02 06:50:47 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:50:47 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:50:47 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:47 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 30 total executors!
18/05/02 06:50:48 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:48 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 31 total executors!
18/05/02 06:50:49 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:49 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 32 total executors!
```

```
18/05/02 06:50:50 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:50 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 33 total executors!
18/05/02 06:50:51 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:51 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 34 total executors!
18/05/02 06:50:52 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:50:52 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 35 total executors!
18/05/02 06:51:04 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000005_5' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000005
18/05/02 06:51:04 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000005_5: Committed
18/05/02 06:51:04 INFO executor.Executor: Finished task 5.0 in stage 0.0
(TID 5). 2360 bytes result sent to driver
18/05/02 06:51:04 INFO scheduler.TaskSetManager: Starting task 8.0 in
stage 0.0 (TID 8, localhost, executor driver, partition 8, ANY, 2316
bytes)
18/05/02 06:51:04 INFO executor.Executor: Running task 8.0 in stage 0.0
(TID 8)
18/05/02 06:51:04 INFO scheduler.TaskSetManager: Finished task 5.0 in
stage 0.0 (TID 5) in 21018 ms on localhost (executor driver) (5/39)
18/05/02 06:51:04 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:268435456+134217728
18/05/02 06:51:04 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:05 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:05 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:06 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000004_4' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000004
18/05/02 06:51:06 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000004_4: Committed
18/05/02 06:51:06 INFO executor.Executor: Finished task 4.0 in stage 0.0
(TID 4). 2360 bytes result sent to driver
18/05/02 06:51:06 INFO scheduler.TaskSetManager: Starting task 9.0 in
stage 0.0 (TID 9, localhost, executor driver, partition 9, ANY, 2316
bytes)
18/05/02 06:51:06 INFO executor.Executor: Running task 9.0 in stage 0.0
(TID 9)
18/05/02 06:51:06 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:402653184+134217728
18/05/02 06:51:06 INFO scheduler.TaskSetManager: Finished task 4.0 in
stage 0.0 (TID 4) in 23066 ms on localhost (executor driver) (6/39)
18/05/02 06:51:06 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
```

18/05/02 06:51:07 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:07 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:11 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000006_6' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000006
18/05/02 06:51:11 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000006_6: Committed
18/05/02 06:51:11 INFO executor.Executor: Finished task 6.0 in stage 0.0
(TID 6). 2360 bytes result sent to driver
18/05/02 06:51:11 INFO scheduler.TaskSetManager: Starting task 10.0 in
stage 0.0 (TID 10, localhost, executor driver, partition 10, ANY, 2316
bytes)
18/05/02 06:51:11 INFO executor.Executor: Running task 10.0 in stage 0.0
(TID 10)
18/05/02 06:51:11 INFO scheduler.TaskSetManager: Finished task 6.0 in
stage 0.0 (TID 6) in 26832 ms on localhost (executor driver) (7/39)
18/05/02 06:51:11 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:536870912+134217728
18/05/02 06:51:11 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:12 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000007_7' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000007
18/05/02 06:51:12 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000007_7: Committed
18/05/02 06:51:12 INFO executor.Executor: Finished task 7.0 in stage 0.0
(TID 7). 2360 bytes result sent to driver
18/05/02 06:51:12 INFO scheduler.TaskSetManager: Starting task 11.0 in
stage 0.0 (TID 11, localhost, executor driver, partition 11, ANY, 2316
bytes)
18/05/02 06:51:12 INFO executor.Executor: Running task 11.0 in stage 0.0
(TID 11)
18/05/02 06:51:12 INFO scheduler.TaskSetManager: Finished task 7.0 in
stage 0.0 (TID 7) in 26469 ms on localhost (executor driver) (8/39)
18/05/02 06:51:12 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:671088640+71329760
18/05/02 06:51:12 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:12 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:12 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:13 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:13 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:25 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000008_8' to hdfs://ip-10-0-0-212.ap-south-

1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000008
18/05/02 06:51:25 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000008_8: Committed
18/05/02 06:51:25 INFO executor.Executor: Finished task 8.0 in stage 0.0
(TID 8). 2360 bytes result sent to driver
18/05/02 06:51:25 INFO scheduler.TaskSetManager: Starting task 12.0 in
stage 0.0 (TID 12, localhost, executor driver, partition 12, ANY, 2316
bytes)
18/05/02 06:51:25 INFO executor.Executor: Running task 12.0 in stage 0.0
(TID 12)
18/05/02 06:51:25 INFO scheduler.TaskSetManager: Finished task 8.0 in
stage 0.0 (TID 8) in 20598 ms on localhost (executor driver) (9/39)
18/05/02 06:51:25 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:0+134217728
18/05/02 06:51:25 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:25 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:25 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:28 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000011_11' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000011
18/05/02 06:51:28 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000011_11: Committed
18/05/02 06:51:28 INFO executor.Executor: Finished task 11.0 in stage 0.0
(TID 11). 2360 bytes result sent to driver
18/05/02 06:51:28 INFO scheduler.TaskSetManager: Starting task 13.0 in
stage 0.0 (TID 13, localhost, executor driver, partition 13, ANY, 2316
bytes)
18/05/02 06:51:28 INFO executor.Executor: Running task 13.0 in stage 0.0
(TID 13)
18/05/02 06:51:28 INFO scheduler.TaskSetManager: Finished task 11.0 in
stage 0.0 (TID 11) in 16149 ms on localhost (executor driver) (10/39)
18/05/02 06:51:28 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:134217728+134217728
18/05/02 06:51:28 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:28 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000009_9' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000009
18/05/02 06:51:28 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000009_9: Committed
18/05/02 06:51:28 INFO executor.Executor: Finished task 9.0 in stage 0.0
(TID 9). 2360 bytes result sent to driver
18/05/02 06:51:28 INFO scheduler.TaskSetManager: Starting task 14.0 in
stage 0.0 (TID 14, localhost, executor driver, partition 14, ANY, 2316
bytes)
18/05/02 06:51:28 INFO scheduler.TaskSetManager: Finished task 9.0 in
stage 0.0 (TID 9) in 22227 ms on localhost (executor driver) (11/39)
18/05/02 06:51:28 INFO executor.Executor: Running task 14.0 in stage 0.0
(TID 14)

18/05/02 06:51:28 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:268435456+134217728
18/05/02 06:51:28 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:29 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:29 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:29 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:29 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:32 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000010_10' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000010
18/05/02 06:51:32 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000010_10: Committed
18/05/02 06:51:32 INFO executor.Executor: Finished task 10.0 in stage 0.0
(TID 10). 2360 bytes result sent to driver
18/05/02 06:51:32 INFO scheduler.TaskSetManager: Starting task 15.0 in
stage 0.0 (TID 15, localhost, executor driver, partition 15, ANY, 2316
bytes)
18/05/02 06:51:32 INFO executor.Executor: Running task 15.0 in stage 0.0
(TID 15)
18/05/02 06:51:32 INFO scheduler.TaskSetManager: Finished task 10.0 in
stage 0.0 (TID 10) in 20764 ms on localhost (executor driver) (12/39)
18/05/02 06:51:32 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:402653184+134217728
18/05/02 06:51:32 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:33 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:33 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:47 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000012_12' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000012
18/05/02 06:51:47 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000012_12: Committed
18/05/02 06:51:47 INFO executor.Executor: Finished task 12.0 in stage 0.0
(TID 12). 2360 bytes result sent to driver
18/05/02 06:51:47 INFO scheduler.TaskSetManager: Starting task 16.0 in
stage 0.0 (TID 16, localhost, executor driver, partition 16, ANY, 2316
bytes)
18/05/02 06:51:47 INFO executor.Executor: Running task 16.0 in stage 0.0
(TID 16)
18/05/02 06:51:47 INFO scheduler.TaskSetManager: Finished task 12.0 in
stage 0.0 (TID 12) in 22034 ms on localhost (executor driver) (13/39)
18/05/02 06:51:47 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-

1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:536870912+134217728
18/05/02 06:51:47 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:49 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:49 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:50 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000013_13' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000013
18/05/02 06:51:50 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000013_13: Committed
18/05/02 06:51:50 INFO executor.Executor: Finished task 13.0 in stage 0.0
(TID 13). 2360 bytes result sent to driver
18/05/02 06:51:50 INFO scheduler.TaskSetManager: Starting task 17.0 in
stage 0.0 (TID 17, localhost, executor driver, partition 17, ANY, 2316
bytes)
18/05/02 06:51:50 INFO executor.Executor: Running task 17.0 in stage 0.0
(TID 17)
18/05/02 06:51:50 INFO scheduler.TaskSetManager: Finished task 13.0 in
stage 0.0 (TID 13) in 22515 ms on localhost (executor driver) (14/39)
18/05/02 06:51:50 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:671088640+117983715
18/05/02 06:51:50 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:52 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:52 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:52 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000014_14' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000014
18/05/02 06:51:52 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000014_14: Committed
18/05/02 06:51:52 INFO executor.Executor: Finished task 14.0 in stage 0.0
(TID 14). 2360 bytes result sent to driver
18/05/02 06:51:52 INFO scheduler.TaskSetManager: Starting task 18.0 in
stage 0.0 (TID 18, localhost, executor driver, partition 18, ANY, 2316
bytes)
18/05/02 06:51:52 INFO executor.Executor: Running task 18.0 in stage 0.0
(TID 18)
18/05/02 06:51:52 INFO scheduler.TaskSetManager: Finished task 14.0 in
stage 0.0 (TID 14) in 23812 ms on localhost (executor driver) (15/39)
18/05/02 06:51:52 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:0+134217728
18/05/02 06:51:52 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:52 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1

18/05/02 06:51:52 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:51:57 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000015_15' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000015
18/05/02 06:51:57 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000015_15: Committed
18/05/02 06:51:57 INFO executor.Executor: Finished task 15.0 in stage 0.0
(TID 15). 2360 bytes result sent to driver
18/05/02 06:51:57 INFO scheduler.TaskSetManager: Starting task 19.0 in
stage 0.0 (TID 19, localhost, executor driver, partition 19, ANY, 2316
bytes)
18/05/02 06:51:57 INFO executor.Executor: Running task 19.0 in stage 0.0
(TID 19)
18/05/02 06:51:57 INFO scheduler.TaskSetManager: Finished task 15.0 in
stage 0.0 (TID 15) in 24855 ms on localhost (executor driver) (16/39)
18/05/02 06:51:57 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:134217728+134217728
18/05/02 06:51:57 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:51:58 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:51:58 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:09 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000016_16' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000016
18/05/02 06:52:09 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000016_16: Committed
18/05/02 06:52:09 INFO executor.Executor: Finished task 16.0 in stage 0.0
(TID 16). 2360 bytes result sent to driver
18/05/02 06:52:09 INFO scheduler.TaskSetManager: Starting task 20.0 in
stage 0.0 (TID 20, localhost, executor driver, partition 20, ANY, 2316
bytes)
18/05/02 06:52:09 INFO scheduler.TaskSetManager: Finished task 16.0 in
stage 0.0 (TID 16) in 22248 ms on localhost (executor driver) (17/39)
18/05/02 06:52:09 INFO executor.Executor: Running task 20.0 in stage 0.0
(TID 20)
18/05/02 06:52:09 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:268435456+134217728
18/05/02 06:52:09 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:10 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:10 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:10 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000017_17' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000017

18/05/02 06:52:10 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000017_17: Committed
18/05/02 06:52:10 INFO executor.Executor: Finished task 17.0 in stage 0.0
(TID 17). 2360 bytes result sent to driver
18/05/02 06:52:10 INFO scheduler.TaskSetManager: Starting task 21.0 in
stage 0.0 (TID 21, localhost, executor driver, partition 21, ANY, 2316
bytes)
18/05/02 06:52:10 INFO scheduler.TaskSetManager: Finished task 17.0 in
stage 0.0 (TID 17) in 19743 ms on localhost (executor driver) (18/39)
18/05/02 06:52:10 INFO executor.Executor: Running task 21.0 in stage 0.0
(TID 21)
18/05/02 06:52:10 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:10 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:402653184+134217728
18/05/02 06:52:11 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:11 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:17 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000018_18' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000018
18/05/02 06:52:17 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000018_18: Committed
18/05/02 06:52:17 INFO executor.Executor: Finished task 18.0 in stage 0.0
(TID 18). 2360 bytes result sent to driver
18/05/02 06:52:17 INFO scheduler.TaskSetManager: Starting task 22.0 in
stage 0.0 (TID 22, localhost, executor driver, partition 22, ANY, 2316
bytes)
18/05/02 06:52:17 INFO executor.Executor: Running task 22.0 in stage 0.0
(TID 22)
18/05/02 06:52:17 INFO scheduler.TaskSetManager: Finished task 18.0 in
stage 0.0 (TID 18) in 24703 ms on localhost (executor driver) (19/39)
18/05/02 06:52:17 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:536870912+134217728
18/05/02 06:52:17 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:19 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:19 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:20 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000019_19' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000019
18/05/02 06:52:20 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000019_19: Committed
18/05/02 06:52:20 INFO executor.Executor: Finished task 19.0 in stage 0.0
(TID 19). 2360 bytes result sent to driver
18/05/02 06:52:20 INFO scheduler.TaskSetManager: Starting task 23.0 in
stage 0.0 (TID 23, localhost, executor driver, partition 23, ANY, 2316
bytes)

18/05/02 06:52:20 INFO executor.Executor: Running task 23.0 in stage 0.0
(TID 23)
18/05/02 06:52:20 INFO scheduler.TaskSetManager: Finished task 19.0 in
stage 0.0 (TID 19) in 23385 ms on localhost (executor driver) (20/39)
18/05/02 06:52:20 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:671088640+134217728
18/05/02 06:52:20 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:22 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:22 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:32 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000020_20' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000020
18/05/02 06:52:32 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000020_20: Committed
18/05/02 06:52:32 INFO executor.Executor: Finished task 20.0 in stage 0.0
(TID 20). 2360 bytes result sent to driver
18/05/02 06:52:32 INFO scheduler.TaskSetManager: Starting task 24.0 in
stage 0.0 (TID 24, localhost, executor driver, partition 24, ANY, 2316
bytes)
18/05/02 06:52:32 INFO executor.Executor: Running task 24.0 in stage 0.0
(TID 24)
18/05/02 06:52:32 INFO scheduler.TaskSetManager: Finished task 20.0 in
stage 0.0 (TID 20) in 22712 ms on localhost (executor driver) (21/39)
18/05/02 06:52:32 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:805306368+56688482
18/05/02 06:52:32 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:33 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:33 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:34 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000021_21' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000021
18/05/02 06:52:34 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000021_21: Committed
18/05/02 06:52:34 INFO executor.Executor: Finished task 21.0 in stage 0.0
(TID 21). 2360 bytes result sent to driver
18/05/02 06:52:34 INFO scheduler.TaskSetManager: Starting task 25.0 in
stage 0.0 (TID 25, localhost, executor driver, partition 25, ANY, 2316
bytes)
18/05/02 06:52:34 INFO executor.Executor: Running task 25.0 in stage 0.0
(TID 25)
18/05/02 06:52:34 INFO scheduler.TaskSetManager: Finished task 21.0 in
stage 0.0 (TID 21) in 23983 ms on localhost (executor driver) (22/39)
18/05/02 06:52:34 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:0+134217728

```
18/05/02 06:52:34 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:34 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:34 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:39 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000022_22' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000022
18/05/02 06:52:39 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000022_22: Committed
18/05/02 06:52:39 INFO executor.Executor: Finished task 22.0 in stage 0.0
(TID 22). 2360 bytes result sent to driver
18/05/02 06:52:39 INFO scheduler.TaskSetManager: Starting task 26.0 in
stage 0.0 (TID 26, localhost, executor driver, partition 26, ANY, 2316
bytes)
18/05/02 06:52:39 INFO executor.Executor: Running task 26.0 in stage 0.0
(TID 26)
18/05/02 06:52:39 INFO scheduler.TaskSetManager: Finished task 22.0 in
stage 0.0 (TID 22) in 22717 ms on localhost (executor driver) (23/39)
18/05/02 06:52:39 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:134217728+134217728
18/05/02 06:52:39 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:40 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:40 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:40 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000024_24' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000024
18/05/02 06:52:40 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000024_24: Committed
18/05/02 06:52:40 INFO executor.Executor: Finished task 24.0 in stage 0.0
(TID 24). 2360 bytes result sent to driver
18/05/02 06:52:40 INFO scheduler.TaskSetManager: Starting task 27.0 in
stage 0.0 (TID 27, localhost, executor driver, partition 27, ANY, 2316
bytes)
18/05/02 06:52:40 INFO scheduler.TaskSetManager: Finished task 24.0 in
stage 0.0 (TID 24) in 8522 ms on localhost (executor driver) (24/39)
18/05/02 06:52:40 INFO executor.Executor: Running task 27.0 in stage 0.0
(TID 27)
18/05/02 06:52:40 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:268435456+134217728
18/05/02 06:52:40 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:40 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:40 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
```

```
18/05/02 06:52:41 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000023_23' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000023
18/05/02 06:52:41 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000023_23: Committed
18/05/02 06:52:41 INFO executor.Executor: Finished task 23.0 in stage 0.0
(TID 23). 2360 bytes result sent to driver
18/05/02 06:52:41 INFO scheduler.TaskSetManager: Starting task 28.0 in
stage 0.0 (TID 28, localhost, executor driver, partition 28, ANY, 2316
bytes)
18/05/02 06:52:41 INFO scheduler.TaskSetManager: Finished task 23.0 in
stage 0.0 (TID 23) in 20478 ms on localhost (executor driver) (25/39)
18/05/02 06:52:41 INFO executor.Executor: Running task 28.0 in stage 0.0
(TID 28)
18/05/02 06:52:41 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:402653184+134217728
18/05/02 06:52:41 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:41 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:41 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:47 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000025_25' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000025
18/05/02 06:52:47 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000025_25: Committed
18/05/02 06:52:47 INFO executor.Executor: Finished task 25.0 in stage 0.0
(TID 25). 2360 bytes result sent to driver
18/05/02 06:52:47 INFO scheduler.TaskSetManager: Starting task 29.0 in
stage 0.0 (TID 29, localhost, executor driver, partition 29, ANY, 2316
bytes)
18/05/02 06:52:47 INFO scheduler.TaskSetManager: Finished task 25.0 in
stage 0.0 (TID 25) in 12517 ms on localhost (executor driver) (26/39)
18/05/02 06:52:47 INFO executor.Executor: Running task 29.0 in stage 0.0
(TID 29)
18/05/02 06:52:47 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:536870912+134217728
18/05/02 06:52:47 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:47 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:47 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:49 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000026_26' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000026
18/05/02 06:52:49 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000026_26: Committed
18/05/02 06:52:49 INFO executor.Executor: Finished task 26.0 in stage 0.0
(TID 26). 2360 bytes result sent to driver
```

```
18/05/02 06:52:49 INFO scheduler.TaskSetManager: Starting task 30.0 in
stage 0.0 (TID 30, localhost, executor driver, partition 30, ANY, 2316
bytes)
18/05/02 06:52:49 INFO executor.Executor: Running task 30.0 in stage 0.0
(TID 30)
18/05/02 06:52:49 INFO scheduler.TaskSetManager: Finished task 26.0 in
stage 0.0 (TID 26) in 10030 ms on localhost (executor driver) (27/39)
18/05/02 06:52:49 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:671088640+134217728
18/05/02 06:52:50 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:50 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:50 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:50 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000027_27' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000027
18/05/02 06:52:50 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000027_27: Committed
18/05/02 06:52:50 INFO executor.Executor: Finished task 27.0 in stage 0.0
(TID 27). 2360 bytes result sent to driver
18/05/02 06:52:50 INFO scheduler.TaskSetManager: Starting task 31.0 in
stage 0.0 (TID 31, localhost, executor driver, partition 31, ANY, 2316
bytes)
18/05/02 06:52:50 INFO scheduler.TaskSetManager: Finished task 27.0 in
stage 0.0 (TID 27) in 10002 ms on localhost (executor driver) (28/39)
18/05/02 06:52:50 INFO executor.Executor: Running task 31.0 in stage 0.0
(TID 31)
18/05/02 06:52:50 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:805306368+13877504
18/05/02 06:52:50 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:51 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:51 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:51 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000028_28' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000028
18/05/02 06:52:51 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000028_28: Committed
18/05/02 06:52:51 INFO executor.Executor: Finished task 28.0 in stage 0.0
(TID 28). 2360 bytes result sent to driver
18/05/02 06:52:51 INFO scheduler.TaskSetManager: Starting task 32.0 in
stage 0.0 (TID 32, localhost, executor driver, partition 32, ANY, 2316
bytes)
18/05/02 06:52:51 INFO scheduler.TaskSetManager: Finished task 28.0 in
stage 0.0 (TID 28) in 10383 ms on localhost (executor driver) (29/39)
18/05/02 06:52:51 INFO executor.Executor: Running task 32.0 in stage 0.0
(TID 32)
```

```
18/05/02 06:52:51 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:0+134217728
18/05/02 06:52:51 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:51 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:51 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:52 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000031_31' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000031
18/05/02 06:52:52 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000031_31: Committed
18/05/02 06:52:52 INFO executor.Executor: Finished task 31.0 in stage 0.0
(TID 31). 2360 bytes result sent to driver
18/05/02 06:52:52 INFO scheduler.TaskSetManager: Starting task 33.0 in
stage 0.0 (TID 33, localhost, executor driver, partition 33, ANY, 2316
bytes)
18/05/02 06:52:52 INFO scheduler.TaskSetManager: Finished task 31.0 in
stage 0.0 (TID 31) in 2001 ms on localhost (executor driver) (30/39)
18/05/02 06:52:52 INFO executor.Executor: Running task 33.0 in stage 0.0
(TID 33)
18/05/02 06:52:52 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:134217728+134217728
18/05/02 06:52:52 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:53 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:53 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:57 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000029_29' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000029
18/05/02 06:52:57 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000029_29: Committed
18/05/02 06:52:57 INFO executor.Executor: Finished task 29.0 in stage 0.0
(TID 29). 2360 bytes result sent to driver
18/05/02 06:52:57 INFO scheduler.TaskSetManager: Starting task 34.0 in
stage 0.0 (TID 34, localhost, executor driver, partition 34, ANY, 2316
bytes)
18/05/02 06:52:57 INFO executor.Executor: Running task 34.0 in stage 0.0
(TID 34)
18/05/02 06:52:57 INFO scheduler.TaskSetManager: Finished task 29.0 in
stage 0.0 (TID 29) in 9912 ms on localhost (executor driver) (31/39)
18/05/02 06:52:57 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:268435456+134217728
18/05/02 06:52:57 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:57 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
```

18/05/02 06:52:57 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:52:58 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000030_30' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000030
18/05/02 06:52:58 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000030_30: Committed
18/05/02 06:52:58 INFO executor.Executor: Finished task 30.0 in stage 0.0
(TID 30). 2360 bytes result sent to driver
18/05/02 06:52:58 INFO scheduler.TaskSetManager: Starting task 35.0 in
stage 0.0 (TID 35, localhost, executor driver, partition 35, ANY, 2316
bytes)
18/05/02 06:52:58 INFO scheduler.TaskSetManager: Finished task 30.0 in
stage 0.0 (TID 30) in 9001 ms on localhost (executor driver) (32/39)
18/05/02 06:52:58 INFO executor.Executor: Running task 35.0 in stage 0.0
(TID 35)
18/05/02 06:52:58 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:402653184+134217728
18/05/02 06:52:59 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:52:59 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:52:59 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:53:01 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000032_32' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000032
18/05/02 06:53:01 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000032_32: Committed
18/05/02 06:53:01 INFO executor.Executor: Finished task 32.0 in stage 0.0
(TID 32). 2360 bytes result sent to driver
18/05/02 06:53:01 INFO scheduler.TaskSetManager: Starting task 36.0 in
stage 0.0 (TID 36, localhost, executor driver, partition 36, ANY, 2316
bytes)
18/05/02 06:53:01 INFO executor.Executor: Running task 36.0 in stage 0.0
(TID 36)
18/05/02 06:53:01 INFO scheduler.TaskSetManager: Finished task 32.0 in
stage 0.0 (TID 32) in 10038 ms on localhost (executor driver) (33/39)
18/05/02 06:53:01 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:536870912+134217728
18/05/02 06:53:01 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:53:02 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:53:02 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:53:02 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000033_33' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000033

18/05/02 06:53:02 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000033_33: Committed
18/05/02 06:53:02 INFO executor.Executor: Finished task 33.0 in stage 0.0
(TID 33). 2360 bytes result sent to driver
18/05/02 06:53:02 INFO scheduler.TaskSetManager: Starting task 37.0 in
stage 0.0 (TID 37, localhost, executor driver, partition 37, ANY, 2316
bytes)
18/05/02 06:53:02 INFO executor.Executor: Running task 37.0 in stage 0.0
(TID 37)
18/05/02 06:53:02 INFO scheduler.TaskSetManager: Finished task 33.0 in
stage 0.0 (TID 33) in 10168 ms on localhost (executor driver) (34/39)
18/05/02 06:53:02 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:671088640+134217728
18/05/02 06:53:02 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:53:03 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:53:03 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:53:06 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000034_34' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000034
18/05/02 06:53:06 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000034_34: Committed
18/05/02 06:53:06 INFO executor.Executor: Finished task 34.0 in stage 0.0
(TID 34). 2360 bytes result sent to driver
18/05/02 06:53:06 INFO scheduler.TaskSetManager: Starting task 38.0 in
stage 0.0 (TID 38, localhost, executor driver, partition 38, ANY, 2316
bytes)
18/05/02 06:53:06 INFO scheduler.TaskSetManager: Finished task 34.0 in
stage 0.0 (TID 34) in 9499 ms on localhost (executor driver) (35/39)
18/05/02 06:53:06 INFO executor.Executor: Running task 38.0 in stage 0.0
(TID 38)
18/05/02 06:53:06 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:805306368+32783040
18/05/02 06:53:06 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:53:07 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:53:07 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:53:08 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020650_0000_m_000035_35' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/
_temporary/0/task_201805020650_0000_m_000035
18/05/02 06:53:08 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020650_0000_m_000035_35: Committed
18/05/02 06:53:08 INFO executor.Executor: Finished task 35.0 in stage 0.0
(TID 35). 2360 bytes result sent to driver
18/05/02 06:53:08 INFO scheduler.TaskSetManager: Finished task 35.0 in
stage 0.0 (TID 35) in 9540 ms on localhost (executor driver) (36/39)
18/05/02 06:53:08 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode

18/05/02 06:53:10 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020650_0000_m_000038_38' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/_temporary/0/task_201805020650_0000_m_000038
18/05/02 06:53:10 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020650_0000_m_000038_38: Committed
18/05/02 06:53:10 INFO executor.Executor: Finished task 38.0 in stage 0.0 (TID 38). 2360 bytes result sent to driver
18/05/02 06:53:10 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 0.0 (TID 38) in 3881 ms on localhost (executor driver) (37/39)
18/05/02 06:53:10 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:53:10 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020650_0000_m_000036_36' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/_temporary/0/task_201805020650_0000_m_000036
18/05/02 06:53:10 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020650_0000_m_000036_36: Committed
18/05/02 06:53:10 INFO executor.Executor: Finished task 36.0 in stage 0.0 (TID 36). 2360 bytes result sent to driver
18/05/02 06:53:10 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 0.0 (TID 36) in 8822 ms on localhost (executor driver) (38/39)
18/05/02 06:53:10 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:53:10 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020650_0000_m_000037_37' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que1_1out/_temporary/0/task_201805020650_0000_m_000037
18/05/02 06:53:10 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020650_0000_m_000037_37: Committed
18/05/02 06:53:10 INFO executor.Executor: Finished task 37.0 in stage 0.0 (TID 37). 2360 bytes result sent to driver
18/05/02 06:53:10 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 0.0 (TID 37) in 8023 ms on localhost (executor driver) (39/39)
18/05/02 06:53:10 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:53:10 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
18/05/02 06:53:10 INFO scheduler.DAGScheduler: ResultStage 0 (saveAsTextFile at AssignmentProblem1.java:42) finished in 172.078 s
18/05/02 06:53:10 INFO scheduler.DAGScheduler: Job 0 finished: saveAsTextFile at AssignmentProblem1.java:42, took 172.222560 s
File has been written with output
18/05/02 06:53:11 INFO spark.SparkContext: Invoking stop() from shutdown hook
18/05/02 06:53:11 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.212:4040
18/05/02 06:53:11 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/05/02 06:53:11 INFO storage.MemoryStore: MemoryStore cleared
18/05/02 06:53:11 INFO storage.BlockManager: BlockManager stopped
18/05/02 06:53:11 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
18/05/02 06:53:11 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/05/02 06:53:11 INFO remote.RemoteActorRefProvider$RemotingTerminator: Shutting down remote daemon.
18/05/02 06:53:11 INFO remote.RemoteActorRefProvider$RemotingTerminator: Remote daemon shut down; proceeding with flushing remote transports.

18/05/02 06:53:11 INFO spark.SparkContext: Successfully stopped
SparkContext
18/05/02 06:53:11 INFO util.ShutdownHookManager: Shutdown hook called
18/05/02 06:53:11 INFO util.ShutdownHookManager: Deleting directory
/tmp/spark-87d79c82-ec11-4328-b741-2cc777087a63
18/05/02 06:53:11 INFO Remoting: Remoting shut down
18/05/02 06:53:11 INFO remote.RemoteActorRefProvider$RemotingTerminator:
Remoting shut down.

######################################   End of Logs for
Spark - Assigment1 jar
###

Spark 2 Program Logs + Submit Command:

[root@ip-10-0-0-212 ~]# spark-submit --class
spark.assignment2.AssignmentProblem2 --master yarn --deploy-mode client -
-name sparkquestion2 --conf "spark.app.id=sparkquestion2" /home/ec2-
user/assignment2-0.0.1-SNAPSHOT-jar-with-dependencies.jar
/user/root/spark_assignment/input_dataset/yellow_tripdata_*
/user/root/spark_assignment/sparkoutput/que_2out
Conf has been set
18/05/02 05:17:15 INFO spark.SparkContext: Running Spark version 1.6.0
18/05/02 05:17:16 INFO spark.SecurityManager: Changing view acls to: root
18/05/02 05:17:16 INFO spark.SecurityManager: Changing modify acls to:
root
18/05/02 05:17:16 INFO spark.SecurityManager: SecurityManager:
authentication disabled; ui acls disabled; users with view permissions:
Set(root); users with modify permissions: Set(root)
18/05/02 05:17:17 INFO util.Utils: Successfully started service
'sparkDriver' on port 36063.
18/05/02 05:17:18 INFO slf4j.Slf4jLogger: Slf4jLogger started
18/05/02 05:17:18 INFO Remoting: Starting remoting
18/05/02 05:17:19 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://sparkDriverActorSystem@10.0.0.212:38801]
18/05/02 05:17:19 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://sparkDriverActorSystem@10.0.0.212:38801]
18/05/02 05:17:19 INFO util.Utils: Successfully started service
'sparkDriverActorSystem' on port 38801.
18/05/02 05:17:19 INFO spark.SparkEnv: Registering MapOutputTracker
18/05/02 05:17:19 INFO spark.SparkEnv: Registering BlockManagerMaster
18/05/02 05:17:19 INFO storage.DiskBlockManager: Created local directory
at /tmp/blockmgr-f1b725c6-52b4-4e3c-81f1-389dff702a24
18/05/02 05:17:19 INFO storage.MemoryStore: MemoryStore started with
capacity 530.0 MB
18/05/02 05:17:19 INFO spark.SparkEnv: Registering
OutputCommitCoordinator
18/05/02 05:17:20 INFO util.Utils: Successfully started service 'SparkUI'
on port 4040.
18/05/02 05:17:20 INFO ui.SparkUI: Started SparkUI at
http://10.0.0.212:4040
18/05/02 05:17:20 INFO spark.SparkContext: Added JAR file:/home/ec2-
user/assignment2-0.0.1-SNAPSHOT-jar-with-dependencies.jar at
spark://10.0.0.212:36063/jars/assignment2-0.0.1-SNAPSHOT-jar-with-
dependencies.jar with timestamp 1525238240607
18/05/02 05:17:20 INFO executor.Executor: Starting executor ID driver on
host localhost

```
18/05/02 05:17:20 INFO util.Utils: Successfully started service
'org.apache.spark.network.netty.NettyBlockTransferService' on port 36451.
18/05/02 05:17:20 INFO netty.NettyBlockTransferService: Server created on
36451
18/05/02 05:17:20 INFO storage.BlockManager: external shuffle service
port = 7337
18/05/02 05:17:20 INFO storage.BlockManagerMaster: Trying to register
BlockManager
18/05/02 05:17:20 INFO storage.BlockManagerMasterEndpoint: Registering
block manager localhost:36451 with 530.0 MB RAM, BlockManagerId(driver,
localhost, 36451)
18/05/02 05:17:20 INFO storage.BlockManagerMaster: Registered
BlockManager
18/05/02 05:17:23 INFO scheduler.EventLoggingListener: Logging events to
hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/spark/applicationHistory/local-1525238240684
18/05/02 05:17:23 INFO spark.SparkContext: Registered listener
com.cloudera.spark.lineage.ClouderaNavigatorListener
Context has been initialized
18/05/02 05:17:24 INFO storage.MemoryStore: Block broadcast_0 stored as
values in memory (estimated size 303.4 KB, free 529.7 MB)
18/05/02 05:17:25 INFO storage.MemoryStore: Block broadcast_0_piece0
stored as bytes in memory (estimated size 26.1 KB, free 529.7 MB)
18/05/02 05:17:25 INFO storage.BlockManagerInfo: Added broadcast_0_piece0
in memory on localhost:36451 (size: 26.1 KB, free: 530.0 MB)
18/05/02 05:17:25 INFO spark.SparkContext: Created broadcast 0 from
textFile at AssignmentProblem2.java:19
File has been read successfully
Writing the output now
18/05/02 05:17:25 INFO Configuration.deprecation: mapred.tip.id is
deprecated. Instead, use mapreduce.task.id
18/05/02 05:17:25 INFO Configuration.deprecation: mapred.task.id is
deprecated. Instead, use mapreduce.task.attempt.id
18/05/02 05:17:25 INFO Configuration.deprecation: mapred.task.is.map is
deprecated. Instead, use mapreduce.task.ismap
18/05/02 05:17:25 INFO Configuration.deprecation: mapred.task.partition
is deprecated. Instead, use mapreduce.task.partition
18/05/02 05:17:25 INFO Configuration.deprecation: mapred.job.id is
deprecated. Instead, use mapreduce.job.id
18/05/02 05:17:25 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:17:25 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:17:25 INFO mapred.FileInputFormat: Total input paths to
process : 6
18/05/02 05:17:25 INFO spark.SparkContext: Starting job: saveAsTextFile
at AssignmentProblem2.java:40
18/05/02 05:17:25 INFO scheduler.DAGScheduler: Got job 0 (saveAsTextFile
at AssignmentProblem2.java:40) with 39 output partitions
18/05/02 05:17:26 INFO scheduler.DAGScheduler: Final stage: ResultStage 0
(saveAsTextFile at AssignmentProblem2.java:40)
18/05/02 05:17:26 INFO scheduler.DAGScheduler: Parents of final stage:
List()
18/05/02 05:17:26 INFO scheduler.DAGScheduler: Missing parents: List()
18/05/02 05:17:26 INFO scheduler.DAGScheduler: Submitting ResultStage 0
(MapPartitionsRDD[3] at saveAsTextFile at AssignmentProblem2.java:40),
which has no missing parents
18/05/02 05:17:26 INFO storage.MemoryStore: Block broadcast_1 stored as
values in memory (estimated size 81.3 KB, free 529.6 MB)
```

18/05/02 05:17:26 INFO storage.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 29.2 KB, free 529.6 MB)
18/05/02 05:17:26 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on localhost:36451 (size: 29.2 KB, free: 530.0 MB)
18/05/02 05:17:26 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1004
18/05/02 05:17:26 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 0 (MapPartitionsRDD[3] at saveAsTextFile at AssignmentProblem2.java:40) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))
18/05/02 05:17:26 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 39 tasks
18/05/02 05:17:26 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, executor driver, partition 0, ANY, 2316 bytes)
18/05/02 05:17:26 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, localhost, executor driver, partition 1, ANY, 2316 bytes)
18/05/02 05:17:26 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, localhost, executor driver, partition 2, ANY, 2316 bytes)
18/05/02 05:17:26 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, localhost, executor driver, partition 3, ANY, 2316 bytes)
18/05/02 05:17:26 INFO executor.Executor: Running task 0.0 in stage 0.0 (TID 0)
18/05/02 05:17:26 INFO executor.Executor: Running task 2.0 in stage 0.0 (TID 2)
18/05/02 05:17:26 INFO executor.Executor: Running task 1.0 in stage 0.0 (TID 1)
18/05/02 05:17:26 INFO executor.Executor: Running task 3.0 in stage 0.0 (TID 3)
18/05/02 05:17:26 INFO executor.Executor: Fetching spark://10.0.0.212:36063/jars/assignment2-0.0.1-SNAPSHOT-jar-with-dependencies.jar with timestamp 1525238240607
18/05/02 05:17:26 INFO spark.ExecutorAllocationManager: New executor driver has registered (new total is 1)
18/05/02 05:17:26 INFO util.Utils: Fetching spark://10.0.0.212:36063/jars/assignment2-0.0.1-SNAPSHOT-jar-with-dependencies.jar to /tmp/spark-7a1a282d-ed32-474b-b74c-73e9d9498c89/userFiles-0f10e2e5-8c2a-426c-ad35-011a17fdcf6d/fetchFileTemp8073316898962413458.tmp
18/05/02 05:17:26 INFO executor.Executor: Adding file:/tmp/spark-7a1a282d-ed32-474b-b74c-73e9d9498c89/userFiles-0f10e2e5-8c2a-426c-ad35-011a17fdcf6d/assignment2-0.0.1-SNAPSHOT-jar-with-dependencies.jar to class loader
18/05/02 05:17:26 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:134217728+134217728
18/05/02 05:17:26 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:402653184+134217728
18/05/02 05:17:26 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:268435456+134217728
18/05/02 05:17:26 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-

1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-07.csv:0+134217728
18/05/02 05:17:26 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:17:26 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:17:27 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:17:27 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:17:27 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:27 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 2 total executors!
18/05/02 05:17:27 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:17:27 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:17:27 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:17:27 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:17:28 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:28 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 3 total executors!
18/05/02 05:17:29 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:29 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 4 total executors!
18/05/02 05:17:30 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:30 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 5 total executors!
18/05/02 05:17:31 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:31 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 6 total executors!
18/05/02 05:17:32 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:32 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 7 total executors!
18/05/02 05:17:33 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:33 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 8 total executors!
18/05/02 05:17:34 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:34 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 9 total executors!
18/05/02 05:17:35 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:35 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 10 total executors!
18/05/02 05:17:36 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode

```
18/05/02 05:17:36 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 11 total executors!
18/05/02 05:17:37 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:37 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 12 total executors!
18/05/02 05:17:38 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:38 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 13 total executors!
18/05/02 05:17:39 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:39 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 14 total executors!
18/05/02 05:17:40 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:40 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 15 total executors!
18/05/02 05:17:41 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:41 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 16 total executors!
18/05/02 05:17:42 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:42 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 17 total executors!
18/05/02 05:17:43 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:43 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 18 total executors!
18/05/02 05:17:44 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:44 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 19 total executors!
18/05/02 05:17:45 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:45 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 20 total executors!
18/05/02 05:17:46 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:46 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 21 total executors!
18/05/02 05:17:47 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:47 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 22 total executors!
18/05/02 05:17:48 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:48 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 23 total executors!
18/05/02 05:17:49 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:49 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 24 total executors!
18/05/02 05:17:50 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:50 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 25 total executors!
18/05/02 05:17:51 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
```

18/05/02 05:17:51 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 26 total executors!
18/05/02 05:17:51 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000002_2' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000002
18/05/02 05:17:51 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000002_2: Committed
18/05/02 05:17:51 INFO executor.Executor: Finished task 2.0 in stage 0.0
(TID 2). 2295 bytes result sent to driver
18/05/02 05:17:51 INFO scheduler.TaskSetManager: Starting task 4.0 in
stage 0.0 (TID 4, localhost, executor driver, partition 4, ANY, 2316
bytes)
18/05/02 05:17:51 INFO scheduler.TaskSetManager: Finished task 2.0 in
stage 0.0 (TID 2) in 25651 ms on localhost (executor driver) (1/39)
18/05/02 05:17:51 INFO executor.Executor: Running task 4.0 in stage 0.0
(TID 4)
18/05/02 05:17:51 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-07.csv:536870912+134217728
18/05/02 05:17:51 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:17:51 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:17:52 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000001_1' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000001
18/05/02 05:17:52 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000001_1: Committed
18/05/02 05:17:52 INFO executor.Executor: Finished task 1.0 in stage 0.0
(TID 1). 2360 bytes result sent to driver
18/05/02 05:17:52 INFO scheduler.TaskSetManager: Starting task 5.0 in
stage 0.0 (TID 5, localhost, executor driver, partition 5, ANY, 2316
bytes)
18/05/02 05:17:52 INFO executor.Executor: Running task 5.0 in stage 0.0
(TID 5)
18/05/02 05:17:52 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-07.csv:671088640+86047889
18/05/02 05:17:52 INFO scheduler.TaskSetManager: Finished task 1.0 in
stage 0.0 (TID 1) in 25915 ms on localhost (executor driver) (2/39)
18/05/02 05:17:52 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:17:52 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 27 total executors!
18/05/02 05:17:52 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000003_3' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000003
18/05/02 05:17:52 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000003_3: Committed
18/05/02 05:17:52 INFO executor.Executor: Finished task 3.0 in stage 0.0
(TID 3). 2360 bytes result sent to driver
18/05/02 05:17:52 INFO scheduler.TaskSetManager: Starting task 6.0 in
stage 0.0 (TID 6, localhost, executor driver, partition 6, ANY, 2316
bytes)

18/05/02 05:17:52 INFO executor.Executor: Running task 6.0 in stage 0.0 (TID 6)
18/05/02 05:17:52 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020517_0000_m_000000_0' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_temporary/0/task_201805020517_0000_m_000000
18/05/02 05:17:52 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020517_0000_m_000000_0: Committed
18/05/02 05:17:52 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 0.0 (TID 3) in 26167 ms on localhost (executor driver) (3/39)
18/05/02 05:17:52 INFO executor.Executor: Finished task 0.0 in stage 0.0 (TID 0). 2360 bytes result sent to driver
18/05/02 05:17:52 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:0+134217728
18/05/02 05:17:52 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 05:17:52 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 05:17:52 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, localhost, executor driver, partition 7, ANY, 2316 bytes)
18/05/02 05:17:52 INFO executor.Executor: Running task 7.0 in stage 0.0 (TID 7)
18/05/02 05:17:52 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 26395 ms on localhost (executor driver) (4/39)
18/05/02 05:17:52 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 05:17:52 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 05:17:52 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:134217728+134217728
18/05/02 05:17:53 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:17:53 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 28 total executors!
18/05/02 05:17:53 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 05:17:53 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 05:17:54 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:17:54 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 29 total executors!
18/05/02 05:17:55 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:17:55 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 30 total executors!
18/05/02 05:17:56 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:17:56 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 31 total executors!
18/05/02 05:17:57 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/05/02 05:17:57 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 32 total executors!
18/05/02 05:17:58 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:17:58 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 33 total executors!
18/05/02 05:17:59 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:17:59 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 34 total executors!
18/05/02 05:18:00 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:18:00 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 35 total executors!
18/05/02 05:18:11 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020517_0000_m_000005_5' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_temporary/0/task_201805020517_0000_m_000005
18/05/02 05:18:11 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020517_0000_m_000005_5: Committed
18/05/02 05:18:11 INFO executor.Executor: Finished task 5.0 in stage 0.0 (TID 5). 2360 bytes result sent to driver
18/05/02 05:18:11 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, localhost, executor driver, partition 8, ANY, 2316 bytes)
18/05/02 05:18:11 INFO executor.Executor: Running task 8.0 in stage 0.0 (TID 8)
18/05/02 05:18:11 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 0.0 (TID 5) in 18930 ms on localhost (executor driver) (5/39)
18/05/02 05:18:11 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:268435456+134217728
18/05/02 05:18:11 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:18:12 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 05:18:12 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 05:18:15 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020517_0000_m_000004_4' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_temporary/0/task_201805020517_0000_m_000004
18/05/02 05:18:15 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020517_0000_m_000004_4: Committed
18/05/02 05:18:15 INFO executor.Executor: Finished task 4.0 in stage 0.0 (TID 4). 2360 bytes result sent to driver
18/05/02 05:18:15 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, localhost, executor driver, partition 9, ANY, 2316 bytes)
18/05/02 05:18:15 INFO executor.Executor: Running task 9.0 in stage 0.0 (TID 9)
18/05/02 05:18:15 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 23717 ms on localhost (executor driver) (6/39)
18/05/02 05:18:15 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:402653184+134217728

18/05/02 05:18:15 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:18:16 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:18:16 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:18:19 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000007_7' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000007
18/05/02 05:18:19 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000007_7: Committed
18/05/02 05:18:19 INFO executor.Executor: Finished task 7.0 in stage 0.0
(TID 7). 2360 bytes result sent to driver
18/05/02 05:18:19 INFO scheduler.TaskSetManager: Starting task 10.0 in
stage 0.0 (TID 10, localhost, executor driver, partition 10, ANY, 2316
bytes)
18/05/02 05:18:19 INFO executor.Executor: Running task 10.0 in stage 0.0
(TID 10)
18/05/02 05:18:19 INFO scheduler.TaskSetManager: Finished task 7.0 in
stage 0.0 (TID 7) in 27022 ms on localhost (executor driver) (7/39)
18/05/02 05:18:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:536870912+134217728
18/05/02 05:18:19 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:18:19 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000006_6' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000006
18/05/02 05:18:19 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000006_6: Committed
18/05/02 05:18:19 INFO executor.Executor: Finished task 6.0 in stage 0.0
(TID 6). 2360 bytes result sent to driver
18/05/02 05:18:19 INFO scheduler.TaskSetManager: Starting task 11.0 in
stage 0.0 (TID 11, localhost, executor driver, partition 11, ANY, 2316
bytes)
18/05/02 05:18:19 INFO executor.Executor: Running task 11.0 in stage 0.0
(TID 11)
18/05/02 05:18:19 INFO scheduler.TaskSetManager: Finished task 6.0 in
stage 0.0 (TID 6) in 27492 ms on localhost (executor driver) (8/39)
18/05/02 05:18:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:671088640+71329760
18/05/02 05:18:19 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:18:20 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:18:20 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:18:21 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:18:21 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false

```
18/05/02 05:18:33 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000008_8' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000008
18/05/02 05:18:33 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000008_8: Committed
18/05/02 05:18:33 INFO executor.Executor: Finished task 8.0 in stage 0.0
(TID 8). 2360 bytes result sent to driver
18/05/02 05:18:33 INFO scheduler.TaskSetManager: Starting task 12.0 in
stage 0.0 (TID 12, localhost, executor driver, partition 12, ANY, 2316
bytes)
18/05/02 05:18:33 INFO scheduler.TaskSetManager: Finished task 8.0 in
stage 0.0 (TID 8) in 22689 ms on localhost (executor driver) (9/39)
18/05/02 05:18:33 INFO executor.Executor: Running task 12.0 in stage 0.0
(TID 12)
18/05/02 05:18:33 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:18:33 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:0+134217728
18/05/02 05:18:33 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:18:33 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:18:36 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000011_11' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000011
18/05/02 05:18:36 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000011_11: Committed
18/05/02 05:18:36 INFO executor.Executor: Finished task 11.0 in stage 0.0
(TID 11). 2360 bytes result sent to driver
18/05/02 05:18:36 INFO scheduler.TaskSetManager: Starting task 13.0 in
stage 0.0 (TID 13, localhost, executor driver, partition 13, ANY, 2316
bytes)
18/05/02 05:18:36 INFO executor.Executor: Running task 13.0 in stage 0.0
(TID 13)
18/05/02 05:18:36 INFO scheduler.TaskSetManager: Finished task 11.0 in
stage 0.0 (TID 11) in 17096 ms on localhost (executor driver) (10/39)
18/05/02 05:18:36 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:134217728+134217728
18/05/02 05:18:36 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:18:38 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:18:38 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:18:38 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000009_9' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000009
18/05/02 05:18:38 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000009_9: Committed
18/05/02 05:18:38 INFO executor.Executor: Finished task 9.0 in stage 0.0
(TID 9). 2360 bytes result sent to driver
```

18/05/02 05:18:38 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, localhost, executor driver, partition 14, ANY, 2316 bytes)
18/05/02 05:18:38 INFO executor.Executor: Running task 14.0 in stage 0.0 (TID 14)
18/05/02 05:18:38 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 23148 ms on localhost (executor driver) (11/39)
18/05/02 05:18:38 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:268435456+134217728
18/05/02 05:18:38 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:18:39 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 05:18:39 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 05:18:42 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020517_0000_m_000010_10' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_temporary/0/task_201805020517_0000_m_000010
18/05/02 05:18:42 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020517_0000_m_000010_10: Committed
18/05/02 05:18:42 INFO executor.Executor: Finished task 10.0 in stage 0.0 (TID 10). 2360 bytes result sent to driver
18/05/02 05:18:42 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, localhost, executor driver, partition 15, ANY, 2316 bytes)
18/05/02 05:18:42 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 22517 ms on localhost (executor driver) (12/39)
18/05/02 05:18:42 INFO executor.Executor: Running task 15.0 in stage 0.0 (TID 15)
18/05/02 05:18:42 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:402653184+134217728
18/05/02 05:18:42 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:18:43 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 05:18:43 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 05:18:57 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020517_0000_m_000012_12' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_temporary/0/task_201805020517_0000_m_000012
18/05/02 05:18:57 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020517_0000_m_000012_12: Committed
18/05/02 05:18:57 INFO executor.Executor: Finished task 12.0 in stage 0.0 (TID 12). 2360 bytes result sent to driver
18/05/02 05:18:57 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, localhost, executor driver, partition 16, ANY, 2316 bytes)
18/05/02 05:18:57 INFO executor.Executor: Running task 16.0 in stage 0.0 (TID 16)
18/05/02 05:18:57 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 23529 ms on localhost (executor driver) (13/39)

18/05/02 05:18:57 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:536870912+134217728
18/05/02 05:18:57 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:18:58 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:18:58 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:00 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000013_13' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000013
18/05/02 05:19:00 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000013_13: Committed
18/05/02 05:19:00 INFO executor.Executor: Finished task 13.0 in stage 0.0
(TID 13). 2360 bytes result sent to driver
18/05/02 05:19:00 INFO scheduler.TaskSetManager: Starting task 17.0 in
stage 0.0 (TID 17, localhost, executor driver, partition 17, ANY, 2316
bytes)
18/05/02 05:19:00 INFO scheduler.TaskSetManager: Finished task 13.0 in
stage 0.0 (TID 13) in 23404 ms on localhost (executor driver) (14/39)
18/05/02 05:19:00 INFO executor.Executor: Running task 17.0 in stage 0.0
(TID 17)
18/05/02 05:19:00 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:671088640+117983715
18/05/02 05:19:00 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:01 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:01 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:03 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000014_14' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000014
18/05/02 05:19:03 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000014_14: Committed
18/05/02 05:19:03 INFO executor.Executor: Finished task 14.0 in stage 0.0
(TID 14). 2360 bytes result sent to driver
18/05/02 05:19:03 INFO scheduler.TaskSetManager: Starting task 18.0 in
stage 0.0 (TID 18, localhost, executor driver, partition 18, ANY, 2316
bytes)
18/05/02 05:19:03 INFO executor.Executor: Running task 18.0 in stage 0.0
(TID 18)
18/05/02 05:19:03 INFO scheduler.TaskSetManager: Finished task 14.0 in
stage 0.0 (TID 14) in 24566 ms on localhost (executor driver) (15/39)
18/05/02 05:19:03 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:0+134217728
18/05/02 05:19:03 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1

18/05/02 05:19:03 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 05:19:03 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:19:04 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020517_0000_m_000015_15' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_temporary/0/task_201805020517_0000_m_000015
18/05/02 05:19:04 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020517_0000_m_000015_15: Committed
18/05/02 05:19:04 INFO executor.Executor: Finished task 15.0 in stage 0.0 (TID 15). 2360 bytes result sent to driver
18/05/02 05:19:04 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 0.0 (TID 19, localhost, executor driver, partition 19, ANY, 2316 bytes)
18/05/02 05:19:04 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 22103 ms on localhost (executor driver) (16/39)
18/05/02 05:19:04 INFO executor.Executor: Running task 19.0 in stage 0.0 (TID 19)
18/05/02 05:19:04 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:134217728+134217728
18/05/02 05:19:04 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:19:05 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 05:19:05 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 05:19:19 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020517_0000_m_000016_16' to hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_temporary/0/task_201805020517_0000_m_000016
18/05/02 05:19:19 INFO mapred.SparkHadoopMapRedUtil: attempt_201805020517_0000_m_000016_16: Committed
18/05/02 05:19:19 INFO executor.Executor: Finished task 16.0 in stage 0.0 (TID 16). 2360 bytes result sent to driver
18/05/02 05:19:19 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 0.0 (TID 20, localhost, executor driver, partition 20, ANY, 2316 bytes)
18/05/02 05:19:19 INFO executor.Executor: Running task 20.0 in stage 0.0 (TID 20)
18/05/02 05:19:19 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 22357 ms on localhost (executor driver) (17/39)
18/05/02 05:19:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:268435456+134217728
18/05/02 05:19:19 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 05:19:20 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/05/02 05:19:20 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
18/05/02 05:19:20 INFO output.FileOutputCommitter: Saved output of task 'attempt_201805020517_0000_m_000017_17' to hdfs://ip-10-0-0-212.ap-south-

1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000017
18/05/02 05:19:20 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000017_17: Committed
18/05/02 05:19:20 INFO executor.Executor: Finished task 17.0 in stage 0.0
(TID 17). 2360 bytes result sent to driver
18/05/02 05:19:20 INFO scheduler.TaskSetManager: Starting task 21.0 in
stage 0.0 (TID 21, localhost, executor driver, partition 21, ANY, 2316
bytes)
18/05/02 05:19:20 INFO executor.Executor: Running task 21.0 in stage 0.0
(TID 21)
18/05/02 05:19:20 INFO scheduler.TaskSetManager: Finished task 17.0 in
stage 0.0 (TID 17) in 20437 ms on localhost (executor driver) (18/39)
18/05/02 05:19:20 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:402653184+134217728
18/05/02 05:19:20 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:21 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:21 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:27 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000019_19' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000019
18/05/02 05:19:27 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000019_19: Committed
18/05/02 05:19:27 INFO executor.Executor: Finished task 19.0 in stage 0.0
(TID 19). 2360 bytes result sent to driver
18/05/02 05:19:27 INFO scheduler.TaskSetManager: Starting task 22.0 in
stage 0.0 (TID 22, localhost, executor driver, partition 22, ANY, 2316
bytes)
18/05/02 05:19:27 INFO executor.Executor: Running task 22.0 in stage 0.0
(TID 22)
18/05/02 05:19:27 INFO scheduler.TaskSetManager: Finished task 19.0 in
stage 0.0 (TID 19) in 23813 ms on localhost (executor driver) (19/39)
18/05/02 05:19:27 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:27 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:536870912+134217728
18/05/02 05:19:28 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000018_18' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000018
18/05/02 05:19:28 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000018_18: Committed
18/05/02 05:19:28 INFO executor.Executor: Finished task 18.0 in stage 0.0
(TID 18). 2360 bytes result sent to driver
18/05/02 05:19:28 INFO scheduler.TaskSetManager: Starting task 23.0 in
stage 0.0 (TID 23, localhost, executor driver, partition 23, ANY, 2316
bytes)
18/05/02 05:19:28 INFO executor.Executor: Running task 23.0 in stage 0.0
(TID 23)
18/05/02 05:19:28 INFO scheduler.TaskSetManager: Finished task 18.0 in
stage 0.0 (TID 18) in 25336 ms on localhost (executor driver) (20/39)

18/05/02 05:19:28 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:671088640+134217728
18/05/02 05:19:28 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:28 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:28 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:29 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:29 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:42 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000020_20' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000020
18/05/02 05:19:42 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000020_20: Committed
18/05/02 05:19:42 INFO executor.Executor: Finished task 20.0 in stage 0.0
(TID 20). 2360 bytes result sent to driver
18/05/02 05:19:42 INFO scheduler.TaskSetManager: Starting task 24.0 in
stage 0.0 (TID 24, localhost, executor driver, partition 24, ANY, 2316
bytes)
18/05/02 05:19:42 INFO executor.Executor: Running task 24.0 in stage 0.0
(TID 24)
18/05/02 05:19:42 INFO scheduler.TaskSetManager: Finished task 20.0 in
stage 0.0 (TID 20) in 22741 ms on localhost (executor driver) (21/39)
18/05/02 05:19:42 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:805306368+56688482
18/05/02 05:19:42 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:43 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:43 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:43 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000021_21' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000021
18/05/02 05:19:43 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000021_21: Committed
18/05/02 05:19:43 INFO executor.Executor: Finished task 21.0 in stage 0.0
(TID 21). 2360 bytes result sent to driver
18/05/02 05:19:43 INFO scheduler.TaskSetManager: Starting task 25.0 in
stage 0.0 (TID 25, localhost, executor driver, partition 25, ANY, 2316
bytes)
18/05/02 05:19:43 INFO executor.Executor: Running task 25.0 in stage 0.0
(TID 25)
18/05/02 05:19:43 INFO scheduler.TaskSetManager: Finished task 21.0 in
stage 0.0 (TID 21) in 23186 ms on localhost (executor driver) (22/39)
18/05/02 05:19:43 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-

1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:0+134217728
18/05/02 05:19:43 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:43 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:44 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:48 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000022_22' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000022
18/05/02 05:19:48 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000022_22: Committed
18/05/02 05:19:48 INFO executor.Executor: Finished task 22.0 in stage 0.0
(TID 22). 2360 bytes result sent to driver
18/05/02 05:19:48 INFO scheduler.TaskSetManager: Starting task 26.0 in
stage 0.0 (TID 26, localhost, executor driver, partition 26, ANY, 2316
bytes)
18/05/02 05:19:48 INFO scheduler.TaskSetManager: Finished task 22.0 in
stage 0.0 (TID 22) in 20997 ms on localhost (executor driver) (23/39)
18/05/02 05:19:48 INFO executor.Executor: Running task 26.0 in stage 0.0
(TID 26)
18/05/02 05:19:48 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:134217728+134217728
18/05/02 05:19:49 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:49 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:49 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:50 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000023_23' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000023
18/05/02 05:19:50 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000023_23: Committed
18/05/02 05:19:50 INFO executor.Executor: Finished task 23.0 in stage 0.0
(TID 23). 2360 bytes result sent to driver
18/05/02 05:19:50 INFO scheduler.TaskSetManager: Starting task 27.0 in
stage 0.0 (TID 27, localhost, executor driver, partition 27, ANY, 2316
bytes)
18/05/02 05:19:50 INFO scheduler.TaskSetManager: Finished task 23.0 in
stage 0.0 (TID 23) in 21939 ms on localhost (executor driver) (24/39)
18/05/02 05:19:50 INFO executor.Executor: Running task 27.0 in stage 0.0
(TID 27)
18/05/02 05:19:50 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:268435456+134217728
18/05/02 05:19:50 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:50 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1

18/05/02 05:19:50 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:50 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000024_24' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000024
18/05/02 05:19:50 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000024_24: Committed
18/05/02 05:19:50 INFO executor.Executor: Finished task 24.0 in stage 0.0
(TID 24). 2360 bytes result sent to driver
18/05/02 05:19:50 INFO scheduler.TaskSetManager: Starting task 28.0 in
stage 0.0 (TID 28, localhost, executor driver, partition 28, ANY, 2316
bytes)
18/05/02 05:19:50 INFO executor.Executor: Running task 28.0 in stage 0.0
(TID 28)
18/05/02 05:19:50 INFO scheduler.TaskSetManager: Finished task 24.0 in
stage 0.0 (TID 24) in 8435 ms on localhost (executor driver) (25/39)
18/05/02 05:19:50 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:402653184+134217728
18/05/02 05:19:50 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:51 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:51 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:55 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000025_25' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000025
18/05/02 05:19:55 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000025_25: Committed
18/05/02 05:19:55 INFO executor.Executor: Finished task 25.0 in stage 0.0
(TID 25). 2360 bytes result sent to driver
18/05/02 05:19:55 INFO scheduler.TaskSetManager: Starting task 29.0 in
stage 0.0 (TID 29, localhost, executor driver, partition 29, ANY, 2316
bytes)
18/05/02 05:19:55 INFO scheduler.TaskSetManager: Finished task 25.0 in
stage 0.0 (TID 25) in 11089 ms on localhost (executor driver) (26/39)
18/05/02 05:19:55 INFO executor.Executor: Running task 29.0 in stage 0.0
(TID 29)
18/05/02 05:19:55 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:536870912+134217728
18/05/02 05:19:55 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:55 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:55 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:57 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000026_26' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000026

18/05/02 05:19:57 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000026_26: Committed
18/05/02 05:19:57 INFO executor.Executor: Finished task 26.0 in stage 0.0
(TID 26). 2360 bytes result sent to driver
18/05/02 05:19:57 INFO scheduler.TaskSetManager: Starting task 30.0 in
stage 0.0 (TID 30, localhost, executor driver, partition 30, ANY, 2316
bytes)
18/05/02 05:19:57 INFO executor.Executor: Running task 30.0 in stage 0.0
(TID 30)
18/05/02 05:19:57 INFO scheduler.TaskSetManager: Finished task 26.0 in
stage 0.0 (TID 26) in 9021 ms on localhost (executor driver) (27/39)
18/05/02 05:19:57 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:671088640+134217728
18/05/02 05:19:58 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:58 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:58 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:19:59 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000027_27' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000027
18/05/02 05:19:59 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000027_27: Committed
18/05/02 05:19:59 INFO executor.Executor: Finished task 27.0 in stage 0.0
(TID 27). 2360 bytes result sent to driver
18/05/02 05:19:59 INFO scheduler.TaskSetManager: Starting task 31.0 in
stage 0.0 (TID 31, localhost, executor driver, partition 31, ANY, 2316
bytes)
18/05/02 05:19:59 INFO executor.Executor: Running task 31.0 in stage 0.0
(TID 31)
18/05/02 05:19:59 INFO scheduler.TaskSetManager: Finished task 27.0 in
stage 0.0 (TID 27) in 8718 ms on localhost (executor driver) (28/39)
18/05/02 05:19:59 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:805306368+13877504
18/05/02 05:19:59 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:19:59 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:19:59 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:20:00 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000028_28' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000028
18/05/02 05:20:00 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000028_28: Committed
18/05/02 05:20:00 INFO executor.Executor: Finished task 28.0 in stage 0.0
(TID 28). 2360 bytes result sent to driver
18/05/02 05:20:00 INFO scheduler.TaskSetManager: Starting task 32.0 in
stage 0.0 (TID 32, localhost, executor driver, partition 32, ANY, 2316
bytes)

18/05/02 05:20:00 INFO scheduler.TaskSetManager: Finished task 28.0 in
stage 0.0 (TID 28) in 9503 ms on localhost (executor driver) (29/39)
18/05/02 05:20:00 INFO executor.Executor: Running task 32.0 in stage 0.0
(TID 32)
18/05/02 05:20:00 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:0+134217728
18/05/02 05:20:00 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:20:00 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:20:00 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:20:03 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000031_31' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000031
18/05/02 05:20:03 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000031_31: Committed
18/05/02 05:20:03 INFO executor.Executor: Finished task 31.0 in stage 0.0
(TID 31). 2360 bytes result sent to driver
18/05/02 05:20:03 INFO scheduler.TaskSetManager: Starting task 33.0 in
stage 0.0 (TID 33, localhost, executor driver, partition 33, ANY, 2316
bytes)
18/05/02 05:20:03 INFO executor.Executor: Running task 33.0 in stage 0.0
(TID 33)
18/05/02 05:20:03 INFO scheduler.TaskSetManager: Finished task 31.0 in
stage 0.0 (TID 31) in 4100 ms on localhost (executor driver) (30/39)
18/05/02 05:20:03 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:134217728+134217728
18/05/02 05:20:03 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:20:04 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:20:04 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:20:07 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000029_29' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000029
18/05/02 05:20:07 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000029_29: Committed
18/05/02 05:20:07 INFO executor.Executor: Finished task 29.0 in stage 0.0
(TID 29). 2360 bytes result sent to driver
18/05/02 05:20:07 INFO scheduler.TaskSetManager: Starting task 34.0 in
stage 0.0 (TID 34, localhost, executor driver, partition 34, ANY, 2316
bytes)
18/05/02 05:20:07 INFO scheduler.TaskSetManager: Finished task 29.0 in
stage 0.0 (TID 29) in 12409 ms on localhost (executor driver) (31/39)
18/05/02 05:20:07 INFO executor.Executor: Running task 34.0 in stage 0.0
(TID 34)
18/05/02 05:20:07 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:268435456+134217728

18/05/02 05:20:07 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:20:07 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:20:07 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:20:10 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000030_30' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000030
18/05/02 05:20:10 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000030_30: Committed
18/05/02 05:20:10 INFO executor.Executor: Finished task 30.0 in stage 0.0
(TID 30). 2360 bytes result sent to driver
18/05/02 05:20:10 INFO scheduler.TaskSetManager: Starting task 35.0 in
stage 0.0 (TID 35, localhost, executor driver, partition 35, ANY, 2316
bytes)
18/05/02 05:20:10 INFO executor.Executor: Running task 35.0 in stage 0.0
(TID 35)
18/05/02 05:20:10 INFO scheduler.TaskSetManager: Finished task 30.0 in
stage 0.0 (TID 30) in 12285 ms on localhost (executor driver) (32/39)
18/05/02 05:20:10 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:402653184+134217728
18/05/02 05:20:10 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:20:10 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:20:10 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:20:14 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000032_32' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000032
18/05/02 05:20:14 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000032_32: Committed
18/05/02 05:20:14 INFO executor.Executor: Finished task 32.0 in stage 0.0
(TID 32). 2360 bytes result sent to driver
18/05/02 05:20:14 INFO scheduler.TaskSetManager: Starting task 36.0 in
stage 0.0 (TID 36, localhost, executor driver, partition 36, ANY, 2316
bytes)
18/05/02 05:20:14 INFO executor.Executor: Running task 36.0 in stage 0.0
(TID 36)
18/05/02 05:20:14 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:536870912+134217728
18/05/02 05:20:14 INFO scheduler.TaskSetManager: Finished task 32.0 in
stage 0.0 (TID 32) in 14198 ms on localhost (executor driver) (33/39)
18/05/02 05:20:14 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:20:14 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:20:14 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false

```
18/05/02 05:20:15 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000033_33' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000033
18/05/02 05:20:15 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000033_33: Committed
18/05/02 05:20:15 INFO executor.Executor: Finished task 33.0 in stage 0.0
(TID 33). 2360 bytes result sent to driver
18/05/02 05:20:15 INFO scheduler.TaskSetManager: Starting task 37.0 in
stage 0.0 (TID 37, localhost, executor driver, partition 37, ANY, 2316
bytes)
18/05/02 05:20:15 INFO scheduler.TaskSetManager: Finished task 33.0 in
stage 0.0 (TID 33) in 12262 ms on localhost (executor driver) (34/39)
18/05/02 05:20:15 INFO executor.Executor: Running task 37.0 in stage 0.0
(TID 37)
18/05/02 05:20:15 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:671088640+134217728
18/05/02 05:20:15 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:20:16 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:20:16 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:20:16 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000034_34' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000034
18/05/02 05:20:16 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000034_34: Committed
18/05/02 05:20:16 INFO executor.Executor: Finished task 34.0 in stage 0.0
(TID 34). 2360 bytes result sent to driver
18/05/02 05:20:16 INFO scheduler.TaskSetManager: Starting task 38.0 in
stage 0.0 (TID 38, localhost, executor driver, partition 38, ANY, 2316
bytes)
18/05/02 05:20:16 INFO executor.Executor: Running task 38.0 in stage 0.0
(TID 38)
18/05/02 05:20:16 INFO scheduler.TaskSetManager: Finished task 34.0 in
stage 0.0 (TID 34) in 9355 ms on localhost (executor driver) (35/39)
18/05/02 05:20:16 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:805306368+32783040
18/05/02 05:20:16 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:20:17 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 05:20:17 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 05:20:19 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000035_35' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000035
18/05/02 05:20:19 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000035_35: Committed
18/05/02 05:20:19 INFO executor.Executor: Finished task 35.0 in stage 0.0
(TID 35). 2360 bytes result sent to driver
```

```
18/05/02 05:20:19 INFO scheduler.TaskSetManager: Finished task 35.0 in
stage 0.0 (TID 35) in 9416 ms on localhost (executor driver) (36/39)
18/05/02 05:20:19 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:20:21 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000038_38' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000038
18/05/02 05:20:21 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000038_38: Committed
18/05/02 05:20:21 INFO executor.Executor: Finished task 38.0 in stage 0.0
(TID 38). 2360 bytes result sent to driver
18/05/02 05:20:21 INFO scheduler.TaskSetManager: Finished task 38.0 in
stage 0.0 (TID 38) in 4230 ms on localhost (executor driver) (37/39)
18/05/02 05:20:21 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 05:20:22 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000036_36' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000036
18/05/02 05:20:22 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000036_36: Committed
18/05/02 05:20:22 INFO executor.Executor: Finished task 36.0 in stage 0.0
(TID 36). 2360 bytes result sent to driver
18/05/02 05:20:22 INFO scheduler.TaskSetManager: Finished task 36.0 in
stage 0.0 (TID 36) in 8074 ms on localhost (executor driver) (38/39)
18/05/02 05:20:22 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020517_0000_m_000037_37' to hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que_2out/_
temporary/0/task_201805020517_0000_m_000037
18/05/02 05:20:22 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020517_0000_m_000037_37: Committed
18/05/02 05:20:22 INFO executor.Executor: Finished task 37.0 in stage 0.0
(TID 37). 2360 bytes result sent to driver
18/05/02 05:20:22 INFO scheduler.TaskSetManager: Finished task 37.0 in
stage 0.0 (TID 37) in 6927 ms on localhost (executor driver) (39/39)
18/05/02 05:20:22 INFO scheduler.DAGScheduler: ResultStage 0
(saveAsTextFile at AssignmentProblem2.java:40) finished in 176.351 s
18/05/02 05:20:22 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0,
whose tasks have all completed, from pool
18/05/02 05:20:22 INFO scheduler.DAGScheduler: Job 0 finished:
saveAsTextFile at AssignmentProblem2.java:40, took 176.606710 s
18/05/02 05:20:22 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
File has been written with output
18/05/02 05:20:22 INFO spark.SparkContext: Invoking stop() from shutdown
hook
18/05/02 05:20:22 INFO ui.SparkUI: Stopped Spark web UI at
http://10.0.0.212:4040
18/05/02 05:20:22 INFO spark.MapOutputTrackerMasterEndpoint:
MapOutputTrackerMasterEndpoint stopped!
18/05/02 05:20:22 INFO storage.MemoryStore: MemoryStore cleared
18/05/02 05:20:22 INFO storage.BlockManager: BlockManager stopped
18/05/02 05:20:22 INFO storage.BlockManagerMaster: BlockManagerMaster
stopped
18/05/02 05:20:22 INFO
scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:
OutputCommitCoordinator stopped!
18/05/02 05:20:22 INFO remote.RemoteActorRefProvider$RemotingTerminator:
Shutting down remote daemon.
```

18/05/02 05:20:22 INFO spark.SparkContext: Successfully stopped
SparkContext
18/05/02 05:20:22 INFO util.ShutdownHookManager: Shutdown hook called
18/05/02 05:20:22 INFO util.ShutdownHookManager: Deleting directory
/tmp/spark-7a1a282d-ed32-474b-b74c-73e9d9498c89
18/05/02 05:20:22 INFO remote.RemoteActorRefProvider$RemotingTerminator:
Remote daemon shut down; proceeding with flushing remote transports.
18/05/02 05:20:22 INFO Remoting: Remoting shut down
18/05/02 05:20:22 INFO remote.RemoteActorRefProvider$RemotingTerminator:
Remoting shut down.

####################################   End of Logs for
Spark - Assigment2 jar
###

Spark 3 Program Logs + Submit Command:

[root@ip-10-0-0-212 ~]# spark-submit --class
spark.assignment3.AssignmentProblem3 --master yarn --deploy-mode client -
-name sparkquestion3 --conf "spark.app.id=sparkquestion3" /home/ec2-
user/assignment3-0.0.1-SNAPSHOT-jar-with-dependencies.jar
/user/root/spark_assignment/input_dataset/yellow_tripdata_*
/user/root/spark_assignment/sparkoutput/que3out
Spark Conf has been set successfully
18/05/02 06:03:04 INFO spark.SparkContext: Running Spark version 1.6.0
18/05/02 06:03:05 INFO spark.SecurityManager: Changing view acls to: root
18/05/02 06:03:05 INFO spark.SecurityManager: Changing modify acls to:
root
18/05/02 06:03:05 INFO spark.SecurityManager: SecurityManager:
authentication disabled; ui acls disabled; users with view permissions:
Set(root); users with modify permissions: Set(root)
18/05/02 06:03:06 INFO util.Utils: Successfully started service
'sparkDriver' on port 37576.
18/05/02 06:03:07 INFO slf4j.Slf4jLogger: Slf4jLogger started
18/05/02 06:03:07 INFO Remoting: Starting remoting
18/05/02 06:03:08 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://sparkDriverActorSystem@10.0.0.212:34811]
18/05/02 06:03:08 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://sparkDriverActorSystem@10.0.0.212:34811]
18/05/02 06:03:08 INFO util.Utils: Successfully started service
'sparkDriverActorSystem' on port 34811.
18/05/02 06:03:08 INFO spark.SparkEnv: Registering MapOutputTracker
18/05/02 06:03:08 INFO spark.SparkEnv: Registering BlockManagerMaster
18/05/02 06:03:08 INFO storage.DiskBlockManager: Created local directory
at /tmp/blockmgr-a9f15c53-8eda-4439-ad0b-5ec6cd2476fe
18/05/02 06:03:08 INFO storage.MemoryStore: MemoryStore started with
capacity 530.0 MB
18/05/02 06:03:08 INFO spark.SparkEnv: Registering
OutputCommitCoordinator
18/05/02 06:03:09 INFO util.Utils: Successfully started service 'SparkUI'
on port 4040.
18/05/02 06:03:09 INFO ui.SparkUI: Started SparkUI at
http://10.0.0.212:4040
18/05/02 06:03:09 INFO spark.SparkContext: Added JAR file:/home/ec2-
user/assignment3-0.0.1-SNAPSHOT-jar-with-dependencies.jar at
spark://10.0.0.212:37576/jars/assignment3-0.0.1-SNAPSHOT-jar-with-
dependencies.jar with timestamp 1525240989638
18/05/02 06:03:09 INFO executor.Executor: Starting executor ID driver on
host localhost

18/05/02 06:03:09 INFO util.Utils: Successfully started service
'org.apache.spark.network.netty.NettyBlockTransferService' on port 43712.
18/05/02 06:03:09 INFO netty.NettyBlockTransferService: Server created on
43712
18/05/02 06:03:09 INFO storage.BlockManager: external shuffle service
port = 7337
18/05/02 06:03:09 INFO storage.BlockManagerMaster: Trying to register
BlockManager
18/05/02 06:03:09 INFO storage.BlockManagerMasterEndpoint: Registering
block manager localhost:43712 with 530.0 MB RAM, BlockManagerId(driver,
localhost, 43712)
18/05/02 06:03:09 INFO storage.BlockManagerMaster: Registered
BlockManager
18/05/02 06:03:14 INFO scheduler.EventLoggingListener: Logging events to
hdfs://ip-10-0-0-212.ap-south-
1.compute.internal:8020/user/spark/applicationHistory/local-1525240989683
18/05/02 06:03:14 INFO spark.SparkContext: Registered listener
com.cloudera.spark.lineage.ClouderaNavigatorListener
Context has been initialized
18/05/02 06:03:16 INFO storage.MemoryStore: Block broadcast_0 stored as
values in memory (estimated size 303.4 KB, free 529.7 MB)
18/05/02 06:03:16 INFO storage.MemoryStore: Block broadcast_0_piece0
stored as bytes in memory (estimated size 26.1 KB, free 529.7 MB)
18/05/02 06:03:16 INFO storage.BlockManagerInfo: Added broadcast_0_piece0
in memory on localhost:43712 (size: 26.1 KB, free: 530.0 MB)
18/05/02 06:03:16 INFO spark.SparkContext: Created broadcast 0 from
textFile at AssignmentProblem3.java:22
File has been read successfully
18/05/02 06:03:17 INFO mapred.FileInputFormat: Total input paths to
process : 6
18/05/02 06:03:17 INFO spark.SparkContext: Starting job: sortByKey at
AssignmentProblem3.java:54
18/05/02 06:03:17 INFO scheduler.DAGScheduler: Registering RDD 3
(mapToPair at AssignmentProblem3.java:44)
18/05/02 06:03:17 INFO scheduler.DAGScheduler: Got job 0 (sortByKey at
AssignmentProblem3.java:54) with 39 output partitions
18/05/02 06:03:17 INFO scheduler.DAGScheduler: Final stage: ResultStage 1
(sortByKey at AssignmentProblem3.java:54)
18/05/02 06:03:17 INFO scheduler.DAGScheduler: Parents of final stage:
List(ShuffleMapStage 0)
18/05/02 06:03:17 INFO scheduler.DAGScheduler: Missing parents:
List(ShuffleMapStage 0)
18/05/02 06:03:17 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage
0 (MapPartitionsRDD[3] at mapToPair at AssignmentProblem3.java:44), which
has no missing parents
18/05/02 06:03:17 INFO storage.MemoryStore: Block broadcast_1 stored as
values in memory (estimated size 5.3 KB, free 529.7 MB)
18/05/02 06:03:17 INFO storage.MemoryStore: Block broadcast_1_piece0
stored as bytes in memory (estimated size 2.9 KB, free 529.7 MB)
18/05/02 06:03:17 INFO storage.BlockManagerInfo: Added broadcast_1_piece0
in memory on localhost:43712 (size: 2.9 KB, free: 530.0 MB)
18/05/02 06:03:17 INFO spark.SparkContext: Created broadcast 1 from
broadcast at DAGScheduler.scala:1004
18/05/02 06:03:17 INFO scheduler.DAGScheduler: Submitting 39 missing
tasks from ShuffleMapStage 0 (MapPartitionsRDD[3] at mapToPair at
AssignmentProblem3.java:44) (first 15 tasks are for partitions Vector(0,
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))
18/05/02 06:03:17 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0
with 39 tasks

18/05/02 06:03:17 INFO scheduler.TaskSetManager: Starting task 0.0 in
stage 0.0 (TID 0, localhost, executor driver, partition 0, ANY, 2305
bytes)
18/05/02 06:03:17 INFO scheduler.TaskSetManager: Starting task 1.0 in
stage 0.0 (TID 1, localhost, executor driver, partition 1, ANY, 2305
bytes)
18/05/02 06:03:17 INFO scheduler.TaskSetManager: Starting task 2.0 in
stage 0.0 (TID 2, localhost, executor driver, partition 2, ANY, 2305
bytes)
18/05/02 06:03:17 INFO scheduler.TaskSetManager: Starting task 3.0 in
stage 0.0 (TID 3, localhost, executor driver, partition 3, ANY, 2305
bytes)
18/05/02 06:03:17 INFO executor.Executor: Running task 3.0 in stage 0.0
(TID 3)
18/05/02 06:03:17 INFO executor.Executor: Running task 0.0 in stage 0.0
(TID 0)
18/05/02 06:03:17 INFO executor.Executor: Running task 1.0 in stage 0.0
(TID 1)
18/05/02 06:03:17 INFO executor.Executor: Fetching
spark://10.0.0.212:37576/jars/assignment3-0.0.1-SNAPSHOT-jar-with-
dependencies.jar with timestamp 1525240989638
18/05/02 06:03:17 INFO executor.Executor: Running task 2.0 in stage 0.0
(TID 2)
18/05/02 06:03:17 INFO spark.ExecutorAllocationManager: New executor
driver has registered (new total is 1)
18/05/02 06:03:18 INFO util.Utils: Fetching
spark://10.0.0.212:37576/jars/assignment3-0.0.1-SNAPSHOT-jar-with-
dependencies.jar to /tmp/spark-ce991521-bfda-41a8-8e62-
9c2e5fab6ba7/userFiles-b1c7bde3-1f47-4616-abbd-
61555d530504/fetchFileTemp7248503448150539829.tmp
18/05/02 06:03:18 INFO executor.Executor: Adding file:/tmp/spark-
ce991521-bfda-41a8-8e62-9c2e5fab6ba7/userFiles-b1c7bde3-1f47-4616-abbd-
61555d530504/assignment3-0.0.1-SNAPSHOT-jar-with-dependencies.jar to
class loader
18/05/02 06:03:18 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-07.csv:268435456+134217728
18/05/02 06:03:18 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:03:18 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 2 total executors!
18/05/02 06:03:18 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-07.csv:0+134217728
18/05/02 06:03:18 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-07.csv:134217728+134217728
18/05/02 06:03:18 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-07.csv:402653184+134217728
18/05/02 06:03:18 INFO Configuration.deprecation: mapred.tip.id is
deprecated. Instead, use mapreduce.task.id
18/05/02 06:03:18 INFO Configuration.deprecation: mapred.task.id is
deprecated. Instead, use mapreduce.task.attempt.id
18/05/02 06:03:18 INFO Configuration.deprecation: mapred.task.is.map is
deprecated. Instead, use mapreduce.task.ismap

18/05/02 06:03:18 INFO Configuration.deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.partition
18/05/02 06:03:18 INFO Configuration.deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id
18/05/02 06:03:19 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:19 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 3 total executors!
18/05/02 06:03:20 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:20 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 4 total executors!
18/05/02 06:03:21 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:21 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 5 total executors!
18/05/02 06:03:22 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:22 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 6 total executors!
18/05/02 06:03:23 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:23 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 7 total executors!
18/05/02 06:03:24 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:24 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 8 total executors!
18/05/02 06:03:25 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:25 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 9 total executors!
18/05/02 06:03:26 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:26 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 10 total executors!
18/05/02 06:03:27 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:27 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 11 total executors!
18/05/02 06:03:28 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:28 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 12 total executors!
18/05/02 06:03:29 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:29 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 13 total executors!
18/05/02 06:03:30 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:30 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 14 total executors!
18/05/02 06:03:31 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:31 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 15 total executors!
18/05/02 06:03:32 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:32 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 16 total executors!

18/05/02 06:03:33 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:33 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 17 total executors!
18/05/02 06:03:34 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:34 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 18 total executors!
18/05/02 06:03:35 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:35 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 19 total executors!
18/05/02 06:03:36 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:36 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 20 total executors!
18/05/02 06:03:37 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:37 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 21 total executors!
18/05/02 06:03:38 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:38 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 22 total executors!
18/05/02 06:03:39 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:39 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 23 total executors!
18/05/02 06:03:40 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:40 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 24 total executors!
18/05/02 06:03:41 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:41 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 25 total executors!
18/05/02 06:03:42 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:42 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 26 total executors!
18/05/02 06:03:43 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:43 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 27 total executors!
18/05/02 06:03:44 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:44 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 28 total executors!
18/05/02 06:03:45 INFO executor.Executor: Finished task 1.0 in stage 0.0 (TID 1). 2219 bytes result sent to driver
18/05/02 06:03:45 INFO executor.Executor: Finished task 2.0 in stage 0.0 (TID 2). 2219 bytes result sent to driver
18/05/02 06:03:45 INFO executor.Executor: Finished task 3.0 in stage 0.0 (TID 3). 2219 bytes result sent to driver
18/05/02 06:03:45 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 0.0 (TID 4, localhost, executor driver, partition 4, ANY, 2305 bytes)
18/05/02 06:03:45 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 0.0 (TID 5, localhost, executor driver, partition 5, ANY, 2305 bytes)

18/05/02 06:03:45 INFO executor.Executor: Running task 5.0 in stage 0.0
(TID 5)
18/05/02 06:03:45 INFO executor.Executor: Running task 4.0 in stage 0.0
(TID 4)
18/05/02 06:03:45 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-07.csv:536870912+134217728
18/05/02 06:03:45 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-07.csv:671088640+86047889
18/05/02 06:03:45 INFO scheduler.TaskSetManager: Starting task 6.0 in
stage 0.0 (TID 6, localhost, executor driver, partition 6, ANY, 2305
bytes)
18/05/02 06:03:45 INFO executor.Executor: Running task 6.0 in stage 0.0
(TID 6)
18/05/02 06:03:45 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:0+134217728
18/05/02 06:03:45 INFO scheduler.TaskSetManager: Finished task 2.0 in
stage 0.0 (TID 2) in 27384 ms on localhost (executor driver) (1/39)
18/05/02 06:03:45 INFO scheduler.TaskSetManager: Finished task 3.0 in
stage 0.0 (TID 3) in 27385 ms on localhost (executor driver) (2/39)
18/05/02 06:03:45 INFO scheduler.TaskSetManager: Finished task 1.0 in
stage 0.0 (TID 1) in 27391 ms on localhost (executor driver) (3/39)
18/05/02 06:03:45 INFO executor.Executor: Finished task 0.0 in stage 0.0
(TID 0). 2274 bytes result sent to driver
18/05/02 06:03:45 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:03:45 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 29 total executors!
18/05/02 06:03:45 INFO scheduler.TaskSetManager: Starting task 7.0 in
stage 0.0 (TID 7, localhost, executor driver, partition 7, ANY, 2305
bytes)
18/05/02 06:03:45 INFO executor.Executor: Running task 7.0 in stage 0.0
(TID 7)
18/05/02 06:03:45 INFO scheduler.TaskSetManager: Finished task 0.0 in
stage 0.0 (TID 0) in 28065 ms on localhost (executor driver) (4/39)
18/05/02 06:03:45 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:134217728+134217728
18/05/02 06:03:46 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:03:46 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 30 total executors!
18/05/02 06:03:47 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:03:47 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 31 total executors!
18/05/02 06:03:48 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:03:48 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 32 total executors!
18/05/02 06:03:49 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:03:49 WARN spark.ExecutorAllocationManager: Unable to reach
the cluster manager to request 33 total executors!

18/05/02 06:03:50 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:50 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 34 total executors!
18/05/02 06:03:51 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:03:51 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 35 total executors!
18/05/02 06:04:00 INFO executor.Executor: Finished task 5.0 in stage 0.0 (TID 5). 2274 bytes result sent to driver
18/05/02 06:04:00 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, localhost, executor driver, partition 8, ANY, 2305 bytes)
18/05/02 06:04:00 INFO executor.Executor: Running task 8.0 in stage 0.0 (TID 8)
18/05/02 06:04:00 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:268435456+134217728
18/05/02 06:04:00 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 0.0 (TID 5) in 15229 ms on localhost (executor driver) (5/39)
18/05/02 06:04:00 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:04:07 INFO executor.Executor: Finished task 4.0 in stage 0.0 (TID 4). 2274 bytes result sent to driver
18/05/02 06:04:07 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, localhost, executor driver, partition 9, ANY, 2305 bytes)
18/05/02 06:04:07 INFO executor.Executor: Running task 9.0 in stage 0.0 (TID 9)
18/05/02 06:04:07 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:402653184+134217728
18/05/02 06:04:07 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 22729 ms on localhost (executor driver) (6/39)
18/05/02 06:04:07 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:04:08 INFO executor.Executor: Finished task 6.0 in stage 0.0 (TID 6). 2274 bytes result sent to driver
18/05/02 06:04:08 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, localhost, executor driver, partition 10, ANY, 2305 bytes)
18/05/02 06:04:08 INFO executor.Executor: Running task 10.0 in stage 0.0 (TID 10)
18/05/02 06:04:08 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:536870912+134217728
18/05/02 06:04:08 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 23049 ms on localhost (executor driver) (7/39)
18/05/02 06:04:08 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:04:10 INFO executor.Executor: Finished task 7.0 in stage 0.0 (TID 7). 2274 bytes result sent to driver
18/05/02 06:04:10 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, localhost, executor driver, partition 11, ANY, 2305 bytes)
18/05/02 06:04:10 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 24920 ms on localhost (executor driver) (8/39)

18/05/02 06:04:10 INFO executor.Executor: Running task 11.0 in stage 0.0
(TID 11)
18/05/02 06:04:10 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-08.csv:671088640+71329760
18/05/02 06:04:10 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:04:23 INFO executor.Executor: Finished task 8.0 in stage 0.0
(TID 8). 2274 bytes result sent to driver
18/05/02 06:04:23 INFO scheduler.TaskSetManager: Starting task 12.0 in
stage 0.0 (TID 12, localhost, executor driver, partition 12, ANY, 2305
bytes)
18/05/02 06:04:23 INFO scheduler.TaskSetManager: Finished task 8.0 in
stage 0.0 (TID 8) in 23329 ms on localhost (executor driver) (9/39)
18/05/02 06:04:23 INFO executor.Executor: Running task 12.0 in stage 0.0
(TID 12)
18/05/02 06:04:23 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:0+134217728
18/05/02 06:04:23 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:04:25 INFO executor.Executor: Finished task 11.0 in stage 0.0
(TID 11). 2274 bytes result sent to driver
18/05/02 06:04:25 INFO scheduler.TaskSetManager: Starting task 13.0 in
stage 0.0 (TID 13, localhost, executor driver, partition 13, ANY, 2305
bytes)
18/05/02 06:04:25 INFO executor.Executor: Running task 13.0 in stage 0.0
(TID 13)
18/05/02 06:04:25 INFO scheduler.TaskSetManager: Finished task 11.0 in
stage 0.0 (TID 11) in 14859 ms on localhost (executor driver) (10/39)
18/05/02 06:04:25 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:134217728+134217728
18/05/02 06:04:25 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:04:29 INFO executor.Executor: Finished task 10.0 in stage 0.0
(TID 10). 2274 bytes result sent to driver
18/05/02 06:04:29 INFO scheduler.TaskSetManager: Starting task 14.0 in
stage 0.0 (TID 14, localhost, executor driver, partition 14, ANY, 2305
bytes)
18/05/02 06:04:29 INFO scheduler.TaskSetManager: Finished task 10.0 in
stage 0.0 (TID 10) in 21336 ms on localhost (executor driver) (11/39)
18/05/02 06:04:29 INFO executor.Executor: Running task 14.0 in stage 0.0
(TID 14)
18/05/02 06:04:29 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:268435456+134217728
18/05/02 06:04:29 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:04:30 INFO executor.Executor: Finished task 9.0 in stage 0.0
(TID 9). 2274 bytes result sent to driver
18/05/02 06:04:30 INFO scheduler.TaskSetManager: Starting task 15.0 in
stage 0.0 (TID 15, localhost, executor driver, partition 15, ANY, 2305
bytes)
18/05/02 06:04:30 INFO scheduler.TaskSetManager: Finished task 9.0 in
stage 0.0 (TID 9) in 22537 ms on localhost (executor driver) (12/39)

```
18/05/02 06:04:30 INFO executor.Executor: Running task 15.0 in stage 0.0
(TID 15)
18/05/02 06:04:30 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:402653184+134217728
18/05/02 06:04:30 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:04:47 INFO executor.Executor: Finished task 12.0 in stage 0.0
(TID 12). 2274 bytes result sent to driver
18/05/02 06:04:47 INFO scheduler.TaskSetManager: Starting task 16.0 in
stage 0.0 (TID 16, localhost, executor driver, partition 16, ANY, 2305
bytes)
18/05/02 06:04:47 INFO scheduler.TaskSetManager: Finished task 12.0 in
stage 0.0 (TID 12) in 23495 ms on localhost (executor driver) (13/39)
18/05/02 06:04:47 INFO executor.Executor: Running task 16.0 in stage 0.0
(TID 16)
18/05/02 06:04:47 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:536870912+134217728
18/05/02 06:04:47 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:04:49 INFO executor.Executor: Finished task 13.0 in stage 0.0
(TID 13). 2274 bytes result sent to driver
18/05/02 06:04:49 INFO scheduler.TaskSetManager: Starting task 17.0 in
stage 0.0 (TID 17, localhost, executor driver, partition 17, ANY, 2305
bytes)
18/05/02 06:04:49 INFO executor.Executor: Running task 17.0 in stage 0.0
(TID 17)
18/05/02 06:04:49 INFO scheduler.TaskSetManager: Finished task 13.0 in
stage 0.0 (TID 13) in 24078 ms on localhost (executor driver) (14/39)
18/05/02 06:04:49 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-09.csv:671088640+117983715
18/05/02 06:04:49 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:04:53 INFO executor.Executor: Finished task 14.0 in stage 0.0
(TID 14). 2274 bytes result sent to driver
18/05/02 06:04:53 INFO scheduler.TaskSetManager: Starting task 18.0 in
stage 0.0 (TID 18, localhost, executor driver, partition 18, ANY, 2305
bytes)
18/05/02 06:04:53 INFO executor.Executor: Running task 18.0 in stage 0.0
(TID 18)
18/05/02 06:04:53 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:0+134217728
18/05/02 06:04:53 INFO scheduler.TaskSetManager: Finished task 14.0 in
stage 0.0 (TID 14) in 24434 ms on localhost (executor driver) (15/39)
18/05/02 06:04:53 INFO executor.Executor: Finished task 15.0 in stage 0.0
(TID 15). 2274 bytes result sent to driver
18/05/02 06:04:53 INFO scheduler.TaskSetManager: Starting task 19.0 in
stage 0.0 (TID 19, localhost, executor driver, partition 19, ANY, 2305
bytes)
18/05/02 06:04:53 INFO executor.Executor: Running task 19.0 in stage 0.0
(TID 19)
18/05/02 06:04:53 INFO scheduler.TaskSetManager: Finished task 15.0 in
stage 0.0 (TID 15) in 23572 ms on localhost (executor driver) (16/39)
```

```
18/05/02 06:04:53 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:134217728+134217728
18/05/02 06:04:54 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:05:09 INFO executor.Executor: Finished task 17.0 in stage 0.0
(TID 17). 2274 bytes result sent to driver
18/05/02 06:05:09 INFO scheduler.TaskSetManager: Starting task 20.0 in
stage 0.0 (TID 20, localhost, executor driver, partition 20, ANY, 2305
bytes)
18/05/02 06:05:09 INFO scheduler.TaskSetManager: Finished task 17.0 in
stage 0.0 (TID 17) in 19816 ms on localhost (executor driver) (17/39)
18/05/02 06:05:09 INFO executor.Executor: Running task 20.0 in stage 0.0
(TID 20)
18/05/02 06:05:09 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:268435456+134217728
18/05/02 06:05:09 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:05:10 INFO executor.Executor: Finished task 16.0 in stage 0.0
(TID 16). 2274 bytes result sent to driver
18/05/02 06:05:10 INFO scheduler.TaskSetManager: Starting task 21.0 in
stage 0.0 (TID 21, localhost, executor driver, partition 21, ANY, 2305
bytes)
18/05/02 06:05:10 INFO scheduler.TaskSetManager: Finished task 16.0 in
stage 0.0 (TID 16) in 23272 ms on localhost (executor driver) (18/39)
18/05/02 06:05:10 INFO executor.Executor: Running task 21.0 in stage 0.0
(TID 21)
18/05/02 06:05:10 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:402653184+134217728
18/05/02 06:05:10 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:05:17 INFO executor.Executor: Finished task 18.0 in stage 0.0
(TID 18). 2274 bytes result sent to driver
18/05/02 06:05:17 INFO scheduler.TaskSetManager: Starting task 22.0 in
stage 0.0 (TID 22, localhost, executor driver, partition 22, ANY, 2305
bytes)
18/05/02 06:05:17 INFO scheduler.TaskSetManager: Finished task 18.0 in
stage 0.0 (TID 18) in 23675 ms on localhost (executor driver) (19/39)
18/05/02 06:05:17 INFO executor.Executor: Running task 22.0 in stage 0.0
(TID 22)
18/05/02 06:05:17 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-10.csv:536870912+134217728
18/05/02 06:05:17 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:05:19 INFO executor.Executor: Finished task 19.0 in stage 0.0
(TID 19). 2274 bytes result sent to driver
18/05/02 06:05:19 INFO scheduler.TaskSetManager: Starting task 23.0 in
stage 0.0 (TID 23, localhost, executor driver, partition 23, ANY, 2305
bytes)
18/05/02 06:05:19 INFO executor.Executor: Running task 23.0 in stage 0.0
(TID 23)
18/05/02 06:05:19 INFO scheduler.TaskSetManager: Finished task 19.0 in
stage 0.0 (TID 19) in 25171 ms on localhost (executor driver) (20/39)
```

18/05/02 06:05:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:671088640+134217728
18/05/02 06:05:19 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:05:32 INFO executor.Executor: Finished task 20.0 in stage 0.0 (TID 20). 2274 bytes result sent to driver
18/05/02 06:05:32 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 0.0 (TID 24, localhost, executor driver, partition 24, ANY, 2305 bytes)
18/05/02 06:05:32 INFO executor.Executor: Running task 24.0 in stage 0.0 (TID 24)
18/05/02 06:05:32 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:805306368+56688482
18/05/02 06:05:32 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 0.0 (TID 20) in 23252 ms on localhost (executor driver) (21/39)
18/05/02 06:05:32 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:05:33 INFO executor.Executor: Finished task 21.0 in stage 0.0 (TID 21). 2274 bytes result sent to driver
18/05/02 06:05:33 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 0.0 (TID 25, localhost, executor driver, partition 25, ANY, 2305 bytes)
18/05/02 06:05:33 INFO executor.Executor: Running task 25.0 in stage 0.0 (TID 25)
18/05/02 06:05:33 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 0.0 (TID 21) in 23197 ms on localhost (executor driver) (22/39)
18/05/02 06:05:33 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:0+134217728
18/05/02 06:05:33 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:05:39 INFO executor.Executor: Finished task 22.0 in stage 0.0 (TID 22). 2274 bytes result sent to driver
18/05/02 06:05:39 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 0.0 (TID 26, localhost, executor driver, partition 26, ANY, 2305 bytes)
18/05/02 06:05:39 INFO executor.Executor: Running task 26.0 in stage 0.0 (TID 26)
18/05/02 06:05:39 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 0.0 (TID 22) in 21908 ms on localhost (executor driver) (23/39)
18/05/02 06:05:39 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:134217728+134217728
18/05/02 06:05:39 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:05:41 INFO executor.Executor: Finished task 23.0 in stage 0.0 (TID 23). 2274 bytes result sent to driver
18/05/02 06:05:41 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 0.0 (TID 27, localhost, executor driver, partition 27, ANY, 2305 bytes)
18/05/02 06:05:41 INFO executor.Executor: Running task 27.0 in stage 0.0 (TID 27)
18/05/02 06:05:41 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 0.0 (TID 23) in 22806 ms on localhost (executor driver) (24/39)

18/05/02 06:05:41 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:268435456+134217728
18/05/02 06:05:41 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:05:42 INFO executor.Executor: Finished task 24.0 in stage 0.0
(TID 24). 2274 bytes result sent to driver
18/05/02 06:05:42 INFO scheduler.TaskSetManager: Starting task 28.0 in
stage 0.0 (TID 28, localhost, executor driver, partition 28, ANY, 2305
bytes)
18/05/02 06:05:42 INFO scheduler.TaskSetManager: Finished task 24.0 in
stage 0.0 (TID 24) in 9920 ms on localhost (executor driver) (25/39)
18/05/02 06:05:42 INFO executor.Executor: Running task 28.0 in stage 0.0
(TID 28)
18/05/02 06:05:42 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:402653184+134217728
18/05/02 06:05:42 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:05:48 INFO executor.Executor: Finished task 25.0 in stage 0.0
(TID 25). 2274 bytes result sent to driver
18/05/02 06:05:48 INFO scheduler.TaskSetManager: Starting task 29.0 in
stage 0.0 (TID 29, localhost, executor driver, partition 29, ANY, 2305
bytes)
18/05/02 06:05:48 INFO executor.Executor: Running task 29.0 in stage 0.0
(TID 29)
18/05/02 06:05:48 INFO scheduler.TaskSetManager: Finished task 25.0 in
stage 0.0 (TID 25) in 15254 ms on localhost (executor driver) (26/39)
18/05/02 06:05:48 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:536870912+134217728
18/05/02 06:05:48 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:05:51 INFO executor.Executor: Finished task 26.0 in stage 0.0
(TID 26). 2274 bytes result sent to driver
18/05/02 06:05:51 INFO scheduler.TaskSetManager: Starting task 30.0 in
stage 0.0 (TID 30, localhost, executor driver, partition 30, ANY, 2305
bytes)
18/05/02 06:05:51 INFO scheduler.TaskSetManager: Finished task 26.0 in
stage 0.0 (TID 26) in 11786 ms on localhost (executor driver) (27/39)
18/05/02 06:05:51 INFO executor.Executor: Running task 30.0 in stage 0.0
(TID 30)
18/05/02 06:05:51 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-11.csv:671088640+134217728
18/05/02 06:05:51 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:05:51 INFO executor.Executor: Finished task 27.0 in stage 0.0
(TID 27). 2274 bytes result sent to driver
18/05/02 06:05:51 INFO scheduler.TaskSetManager: Starting task 31.0 in
stage 0.0 (TID 31, localhost, executor driver, partition 31, ANY, 2305
bytes)
18/05/02 06:05:51 INFO scheduler.TaskSetManager: Finished task 27.0 in
stage 0.0 (TID 27) in 9906 ms on localhost (executor driver) (28/39)
18/05/02 06:05:51 INFO executor.Executor: Running task 31.0 in stage 0.0
(TID 31)

18/05/02 06:05:51 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:805306368+13877504
18/05/02 06:05:51 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:05:53 INFO executor.Executor: Finished task 28.0 in stage 0.0 (TID 28). 2274 bytes result sent to driver
18/05/02 06:05:53 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 0.0 (TID 32, localhost, executor driver, partition 32, ANY, 2305 bytes)
18/05/02 06:05:53 INFO executor.Executor: Running task 32.0 in stage 0.0 (TID 32)
18/05/02 06:05:53 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 0.0 (TID 28) in 10595 ms on localhost (executor driver) (29/39)
18/05/02 06:05:53 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:0+134217728
18/05/02 06:05:53 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:05:54 INFO executor.Executor: Finished task 31.0 in stage 0.0 (TID 31). 2274 bytes result sent to driver
18/05/02 06:05:54 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 0.0 (TID 33, localhost, executor driver, partition 33, ANY, 2305 bytes)
18/05/02 06:05:54 INFO executor.Executor: Running task 33.0 in stage 0.0 (TID 33)
18/05/02 06:05:54 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:134217728+134217728
18/05/02 06:05:54 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 0.0 (TID 31) in 2190 ms on localhost (executor driver) (30/39)
18/05/02 06:05:54 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:05:59 INFO executor.Executor: Finished task 29.0 in stage 0.0 (TID 29). 2274 bytes result sent to driver
18/05/02 06:05:59 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 0.0 (TID 34, localhost, executor driver, partition 34, ANY, 2305 bytes)
18/05/02 06:05:59 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 0.0 (TID 29) in 10314 ms on localhost (executor driver) (31/39)
18/05/02 06:05:59 INFO executor.Executor: Running task 34.0 in stage 0.0 (TID 34)
18/05/02 06:05:59 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-212.ap-south-1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:268435456+134217728
18/05/02 06:05:59 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:06:02 INFO executor.Executor: Finished task 30.0 in stage 0.0 (TID 30). 2274 bytes result sent to driver
18/05/02 06:06:02 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 0.0 (TID 35, localhost, executor driver, partition 35, ANY, 2305 bytes)
18/05/02 06:06:02 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 0.0 (TID 30) in 10940 ms on localhost (executor driver) (32/39)
18/05/02 06:06:02 INFO executor.Executor: Running task 35.0 in stage 0.0 (TID 35)

```
18/05/02 06:06:02 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:402653184+134217728
18/05/02 06:06:02 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:06:04 INFO executor.Executor: Finished task 32.0 in stage 0.0
(TID 32). 2274 bytes result sent to driver
18/05/02 06:06:04 INFO scheduler.TaskSetManager: Starting task 36.0 in
stage 0.0 (TID 36, localhost, executor driver, partition 36, ANY, 2305
bytes)
18/05/02 06:06:04 INFO scheduler.TaskSetManager: Finished task 32.0 in
stage 0.0 (TID 32) in 11702 ms on localhost (executor driver) (33/39)
18/05/02 06:06:04 INFO executor.Executor: Running task 36.0 in stage 0.0
(TID 36)
18/05/02 06:06:04 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:536870912+134217728
18/05/02 06:06:04 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:06:06 INFO executor.Executor: Finished task 33.0 in stage 0.0
(TID 33). 2274 bytes result sent to driver
18/05/02 06:06:06 INFO scheduler.TaskSetManager: Starting task 37.0 in
stage 0.0 (TID 37, localhost, executor driver, partition 37, ANY, 2305
bytes)
18/05/02 06:06:06 INFO scheduler.TaskSetManager: Finished task 33.0 in
stage 0.0 (TID 33) in 12013 ms on localhost (executor driver) (34/39)
18/05/02 06:06:06 INFO executor.Executor: Running task 37.0 in stage 0.0
(TID 37)
18/05/02 06:06:06 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:671088640+134217728
18/05/02 06:06:06 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:06:09 INFO executor.Executor: Finished task 34.0 in stage 0.0
(TID 34). 2274 bytes result sent to driver
18/05/02 06:06:09 INFO scheduler.TaskSetManager: Starting task 38.0 in
stage 0.0 (TID 38, localhost, executor driver, partition 38, ANY, 2305
bytes)
18/05/02 06:06:09 INFO scheduler.TaskSetManager: Finished task 34.0 in
stage 0.0 (TID 34) in 10240 ms on localhost (executor driver) (35/39)
18/05/02 06:06:09 INFO executor.Executor: Running task 38.0 in stage 0.0
(TID 38)
18/05/02 06:06:09 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-
212.ap-south-
1.compute.internal:8020/user/root/spark_assignment/input_dataset/yellow_t
ripdata_2017-12.csv:805306368+32783040
18/05/02 06:06:09 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:06:13 INFO executor.Executor: Finished task 35.0 in stage 0.0
(TID 35). 2274 bytes result sent to driver
18/05/02 06:06:13 INFO scheduler.TaskSetManager: Finished task 35.0 in
stage 0.0 (TID 35) in 10942 ms on localhost (executor driver) (36/39)
18/05/02 06:06:13 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
18/05/02 06:06:14 INFO executor.Executor: Finished task 38.0 in stage 0.0
(TID 38). 2274 bytes result sent to driver
```

18/05/02 06:06:14 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 0.0 (TID 38) in 4726 ms on localhost (executor driver) (37/39)
18/05/02 06:06:14 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:06:14 INFO executor.Executor: Finished task 36.0 in stage 0.0 (TID 36). 2274 bytes result sent to driver
18/05/02 06:06:14 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 0.0 (TID 36) in 10049 ms on localhost (executor driver) (38/39)
18/05/02 06:06:14 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode
18/05/02 06:06:15 INFO executor.Executor: Finished task 37.0 in stage 0.0 (TID 37). 2274 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 0.0 (TID 37) in 9519 ms on localhost (executor driver) (39/39)
18/05/02 06:06:15 INFO scheduler.DAGScheduler: ShuffleMapStage 0 (mapToPair at AssignmentProblem3.java:44) finished in 177.881 s
18/05/02 06:06:15 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
18/05/02 06:06:15 INFO scheduler.DAGScheduler: looking for newly runnable stages
18/05/02 06:06:15 INFO scheduler.DAGScheduler: running: Set()
18/05/02 06:06:15 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 1)
18/05/02 06:06:15 INFO scheduler.DAGScheduler: failed: Set()
18/05/02 06:06:15 INFO scheduler.DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[7] at sortByKey at AssignmentProblem3.java:54), which has no missing parents
18/05/02 06:06:15 INFO storage.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 4.7 KB, free 529.7 MB)
18/05/02 06:06:15 INFO storage.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 2.5 KB, free 529.7 MB)
18/05/02 06:06:15 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on localhost:43712 (size: 2.5 KB, free: 530.0 MB)
18/05/02 06:06:15 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1004
18/05/02 06:06:15 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 1 (MapPartitionsRDD[7] at sortByKey at AssignmentProblem3.java:54) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))
18/05/02 06:06:15 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 39 tasks
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 1.0 (TID 39, localhost, executor driver, partition 10, NODE_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 1.0 (TID 40, localhost, executor driver, partition 11, NODE_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 1.0 (TID 41, localhost, executor driver, partition 12, NODE_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 1.0 (TID 42, localhost, executor driver, partition 13, NODE_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 12.0 in stage 1.0 (TID 41)
18/05/02 06:06:15 INFO executor.Executor: Running task 10.0 in stage 1.0 (TID 39)
18/05/02 06:06:15 INFO executor.Executor: Running task 11.0 in stage 1.0 (TID 40)

18/05/02 06:06:15 INFO executor.Executor: Running task 13.0 in stage 1.0 (TID 42)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 2 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 5 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 13.0 in stage 1.0 (TID 42). 1157 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Finished task 11.0 in stage 1.0 (TID 40). 1158 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Finished task 10.0 in stage 1.0 (TID 39). 1158 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Finished task 12.0 in stage 1.0 (TID 41). 1157 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 43, localhost, executor driver, partition 0, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 44, localhost, executor driver, partition 1, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 0.0 in stage 1.0 (TID 43)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 45, localhost, executor driver, partition 2, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 1.0 in stage 1.0 (TID 44)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 46, localhost, executor driver, partition 3, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 1.0 (TID 41) in 83 ms on localhost (executor driver) (1/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 1.0 in stage 1.0 (TID 44). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Finished task 0.0 in stage 1.0 (TID 43). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Running task 3.0 in stage 1.0 (TID 46)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 3.0 in stage 1.0 (TID 46). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 1.0 (TID 42) in 90 ms on localhost (executor driver) (2/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 1.0 (TID 47, localhost, executor driver, partition 4, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 2.0 in stage 1.0 (TID 45)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 1.0 (TID 48, localhost, executor driver, partition 5, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 1.0 (TID 49, localhost, executor driver, partition 6, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 1.0 (TID 39) in 101 ms on localhost (executor driver) (3/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 1.0 (TID 44) in 23 ms on localhost (executor driver) (4/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 43) in 26 ms on localhost (executor driver) (5/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 2.0 in stage 1.0 (TID 45). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 1.0 (TID 50, localhost, executor driver, partition 7, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 7.0 in stage 1.0 (TID 50)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 1.0 (TID 46) in 26 ms on localhost (executor driver) (6/39)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 1.0 (TID 45) in 28 ms on localhost (executor driver) (7/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 1.0 (TID 40) in 108 ms on localhost (executor driver) (8/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 7.0 in stage 1.0 (TID 50). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 1.0 (TID 51, localhost, executor driver, partition 8, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 5.0 in stage 1.0 (TID 48)
18/05/02 06:06:15 INFO executor.Executor: Running task 4.0 in stage 1.0 (TID 47)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 5.0 in stage 1.0
(TID 48). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Finished task 4.0 in stage 1.0
(TID 47). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 9.0 in
stage 1.0 (TID 52, localhost, executor driver, partition 9,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 6.0 in stage 1.0
(TID 49)
18/05/02 06:06:15 INFO executor.Executor: Running task 8.0 in stage 1.0
(TID 51)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 6.0 in stage 1.0
(TID 49). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Running task 9.0 in stage 1.0
(TID 52)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 1 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 9.0 in stage 1.0
(TID 52). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Finished task 8.0 in stage 1.0
(TID 51). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 14.0 in
stage 1.0 (TID 53, localhost, executor driver, partition 14,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 14.0 in stage 1.0
(TID 53)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 15.0 in
stage 1.0 (TID 54, localhost, executor driver, partition 15,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 16.0 in
stage 1.0 (TID 55, localhost, executor driver, partition 16,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 17.0 in
stage 1.0 (TID 56, localhost, executor driver, partition 17,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Running task 15.0 in stage 1.0
(TID 54)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 6.0 in
stage 1.0 (TID 49) in 47 ms on localhost (executor driver) (9/39)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms

18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 1.0 (TID 52) in 25 ms on localhost (executor driver) (10/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 14.0 in stage 1.0 (TID 53). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 1.0 (TID 51) in 37 ms on localhost (executor driver) (11/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 15.0 in stage 1.0 (TID 54). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 1.0 (TID 57, localhost, executor driver, partition 18, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 18.0 in stage 1.0 (TID 57)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 1.0 (TID 58, localhost, executor driver, partition 19, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 17.0 in stage 1.0 (TID 56)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 1.0 (TID 48) in 54 ms on localhost (executor driver) (12/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 1.0 (TID 53) in 31 ms on localhost (executor driver) (13/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 1.0 (TID 54) in 14 ms on localhost (executor driver) (14/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 18.0 in stage 1.0 (TID 57). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Running task 16.0 in stage 1.0 (TID 55)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 1.0 (TID 59, localhost, executor driver, partition 20, PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 1.0 (TID 47) in 62 ms on localhost (executor driver) (15/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 1.0 (TID 57) in 11 ms on localhost (executor driver) (16/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 1.0 (TID 50) in 57 ms on localhost (executor driver) (17/39)
18/05/02 06:06:15 INFO executor.Executor: Running task 19.0 in stage 1.0 (TID 58)
18/05/02 06:06:15 INFO executor.Executor: Finished task 16.0 in stage 1.0 (TID 55). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 1.0 (TID 60, localhost, executor driver, partition 21, PROCESS_LOCAL, 1986 bytes)

18/05/02 06:06:15 INFO executor.Executor: Finished task 17.0 in stage 1.0
(TID 56). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Running task 21.0 in stage 1.0
(TID 60)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 19.0 in stage 1.0
(TID 58). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 22.0 in
stage 1.0 (TID 61, localhost, executor driver, partition 22,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 23.0 in
stage 1.0 (TID 62, localhost, executor driver, partition 23,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 20.0 in stage 1.0
(TID 59)
18/05/02 06:06:15 INFO executor.Executor: Finished task 21.0 in stage 1.0
(TID 60). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Running task 23.0 in stage 1.0
(TID 62)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 16.0 in
stage 1.0 (TID 55) in 30 ms on localhost (executor driver) (18/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 17.0 in
stage 1.0 (TID 56) in 32 ms on localhost (executor driver) (19/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 20.0 in stage 1.0
(TID 59). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Running task 22.0 in stage 1.0
(TID 61)
18/05/02 06:06:15 INFO executor.Executor: Finished task 23.0 in stage 1.0
(TID 62). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 24.0 in
stage 1.0 (TID 63, localhost, executor driver, partition 24,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 25.0 in
stage 1.0 (TID 64, localhost, executor driver, partition 25,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 26.0 in
stage 1.0 (TID 65, localhost, executor driver, partition 26,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 24.0 in stage 1.0
(TID 63)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 21.0 in
stage 1.0 (TID 60) in 15 ms on localhost (executor driver) (20/39)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks

18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 20.0 in
stage 1.0 (TID 59) in 23 ms on localhost (executor driver) (21/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 22.0 in stage 1.0
(TID 61). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 23.0 in
stage 1.0 (TID 62) in 14 ms on localhost (executor driver) (22/39)
18/05/02 06:06:15 INFO executor.Executor: Running task 25.0 in stage 1.0
(TID 64)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 19.0 in
stage 1.0 (TID 58) in 34 ms on localhost (executor driver) (23/39)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Running task 26.0 in stage 1.0
(TID 65)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 24.0 in stage 1.0
(TID 63). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 27.0 in
stage 1.0 (TID 66, localhost, executor driver, partition 27,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Finished task 25.0 in stage 1.0
(TID 64). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 28.0 in
stage 1.0 (TID 67, localhost, executor driver, partition 28,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 28.0 in stage 1.0
(TID 67)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 29.0 in
stage 1.0 (TID 68, localhost, executor driver, partition 29,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 26.0 in stage 1.0
(TID 65). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 22.0 in
stage 1.0 (TID 61) in 23 ms on localhost (executor driver) (24/39)
18/05/02 06:06:15 INFO executor.Executor: Running task 27.0 in stage 1.0
(TID 66)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 24.0 in
stage 1.0 (TID 63) in 17 ms on localhost (executor driver) (25/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 25.0 in
stage 1.0 (TID 64) in 18 ms on localhost (executor driver) (26/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 28.0 in stage 1.0
(TID 67). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Running task 29.0 in stage 1.0
(TID 68)

18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 27.0 in stage 1.0
(TID 66). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 26.0 in
stage 1.0 (TID 65) in 20 ms on localhost (executor driver) (27/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 30.0 in
stage 1.0 (TID 69, localhost, executor driver, partition 30,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 31.0 in
stage 1.0 (TID 70, localhost, executor driver, partition 31,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 32.0 in
stage 1.0 (TID 71, localhost, executor driver, partition 32,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 28.0 in
stage 1.0 (TID 67) in 13 ms on localhost (executor driver) (28/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 29.0 in stage 1.0
(TID 68). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Running task 32.0 in stage 1.0
(TID 71)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 27.0 in
stage 1.0 (TID 66) in 14 ms on localhost (executor driver) (29/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 33.0 in
stage 1.0 (TID 72, localhost, executor driver, partition 33,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Running task 33.0 in stage 1.0
(TID 72)
18/05/02 06:06:15 INFO executor.Executor: Running task 30.0 in stage 1.0
(TID 69)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Running task 31.0 in stage 1.0
(TID 70)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 29.0 in
stage 1.0 (TID 68) in 20 ms on localhost (executor driver) (30/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 31.0 in stage 1.0
(TID 70). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 34.0 in
stage 1.0 (TID 73, localhost, executor driver, partition 34,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Finished task 30.0 in stage 1.0
(TID 69). 1153 bytes result sent to driver

```
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 31.0 in
stage 1.0 (TID 70) in 12 ms on localhost (executor driver) (31/39)
18/05/02 06:06:15 INFO executor.Executor: Running task 34.0 in stage 1.0
(TID 73)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 35.0 in
stage 1.0 (TID 74, localhost, executor driver, partition 35,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 30.0 in
stage 1.0 (TID 69) in 15 ms on localhost (executor driver) (32/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 34.0 in stage 1.0
(TID 73). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Finished task 33.0 in stage 1.0
(TID 72). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 36.0 in
stage 1.0 (TID 75, localhost, executor driver, partition 36,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 37.0 in
stage 1.0 (TID 76, localhost, executor driver, partition 37,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 36.0 in stage 1.0
(TID 75)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 32.0 in stage 1.0
(TID 71). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 38.0 in
stage 1.0 (TID 77, localhost, executor driver, partition 38,
PROCESS_LOCAL, 1986 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 38.0 in stage 1.0
(TID 77)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 36.0 in stage 1.0
(TID 75). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Running task 37.0 in stage 1.0
(TID 76)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 34.0 in
stage 1.0 (TID 73) in 19 ms on localhost (executor driver) (33/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 38.0 in stage 1.0
(TID 77). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO executor.Executor: Running task 35.0 in stage 1.0
(TID 74)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 33.0 in
stage 1.0 (TID 72) in 28 ms on localhost (executor driver) (34/39)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
```

18/05/02 06:06:15 INFO executor.Executor: Finished task 37.0 in stage 1.0
(TID 76). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 32.0 in
stage 1.0 (TID 71) in 33 ms on localhost (executor driver) (35/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 36.0 in
stage 1.0 (TID 75) in 14 ms on localhost (executor driver) (36/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 38.0 in
stage 1.0 (TID 77) in 11 ms on localhost (executor driver) (37/39)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 37.0 in
stage 1.0 (TID 76) in 16 ms on localhost (executor driver) (38/39)
18/05/02 06:06:15 INFO executor.Executor: Finished task 35.0 in stage 1.0
(TID 74). 1153 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 35.0 in
stage 1.0 (TID 74) in 24 ms on localhost (executor driver) (39/39)
18/05/02 06:06:15 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0,
whose tasks have all completed, from pool
18/05/02 06:06:15 INFO scheduler.DAGScheduler: ResultStage 1 (sortByKey
at AssignmentProblem3.java:54) finished in 0.242 s
18/05/02 06:06:15 INFO scheduler.DAGScheduler: Job 0 finished: sortByKey
at AssignmentProblem3.java:54, took 178.446261 s
18/05/02 06:06:15 WARN spark.SparkContext: Requesting executors is only
supported in coarse-grained mode
Writing the output now
18/05/02 06:06:15 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:06:15 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:06:15 INFO spark.SparkContext: Starting job: saveAsTextFile
at AssignmentProblem3.java:57
18/05/02 06:06:15 INFO spark.MapOutputTrackerMaster: Size of output
statuses for shuffle 0 is 195 bytes
18/05/02 06:06:15 INFO scheduler.DAGScheduler: Registering RDD 5
(mapToPair at AssignmentProblem3.java:54)
18/05/02 06:06:15 INFO scheduler.DAGScheduler: Got job 1 (saveAsTextFile
at AssignmentProblem3.java:57) with 5 output partitions
18/05/02 06:06:15 INFO scheduler.DAGScheduler: Final stage: ResultStage 4
(saveAsTextFile at AssignmentProblem3.java:57)
18/05/02 06:06:15 INFO scheduler.DAGScheduler: Parents of final stage:
List(ShuffleMapStage 3)
18/05/02 06:06:15 INFO scheduler.DAGScheduler: Missing parents:
List(ShuffleMapStage 3)
18/05/02 06:06:15 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage
3 (MapPartitionsRDD[5] at mapToPair at AssignmentProblem3.java:54), which
has no missing parents
18/05/02 06:06:15 INFO storage.MemoryStore: Block broadcast_3 stored as
values in memory (estimated size 4.8 KB, free 529.7 MB)
18/05/02 06:06:15 INFO storage.MemoryStore: Block broadcast_3_piece0
stored as bytes in memory (estimated size 2.6 KB, free 529.7 MB)
18/05/02 06:06:15 INFO storage.BlockManagerInfo: Added broadcast_3_piece0
in memory on localhost:43712 (size: 2.6 KB, free: 530.0 MB)
18/05/02 06:06:15 INFO spark.SparkContext: Created broadcast 3 from
broadcast at DAGScheduler.scala:1004
18/05/02 06:06:15 INFO scheduler.DAGScheduler: Submitting 39 missing
tasks from ShuffleMapStage 3 (MapPartitionsRDD[5] at mapToPair at
AssignmentProblem3.java:54) (first 15 tasks are for partitions Vector(0,
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))
18/05/02 06:06:15 INFO scheduler.TaskSchedulerImpl: Adding task set 3.0
with 39 tasks

18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 3.0 (TID 78, localhost, executor driver, partition 10, NODE_LOCAL, 1975 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 3.0 (TID 79, localhost, executor driver, partition 11, NODE_LOCAL, 1975 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 3.0 (TID 80, localhost, executor driver, partition 12, NODE_LOCAL, 1975 bytes)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 3.0 (TID 81, localhost, executor driver, partition 13, NODE_LOCAL, 1975 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 10.0 in stage 3.0 (TID 78)
18/05/02 06:06:15 INFO executor.Executor: Running task 11.0 in stage 3.0 (TID 79)
18/05/02 06:06:15 INFO executor.Executor: Running task 13.0 in stage 3.0 (TID 81)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Running task 12.0 in stage 3.0 (TID 80)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 11.0 in stage 3.0 (TID 79). 1306 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 3.0 (TID 82, localhost, executor driver, partition 0, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:15 INFO executor.Executor: Finished task 13.0 in stage 3.0 (TID 81). 1306 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 3.0 (TID 83, localhost, executor driver, partition 1, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:15 INFO executor.Executor: Finished task 12.0 in stage 3.0 (TID 80). 1306 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 3.0 (TID 84, localhost, executor driver, partition 2, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:15 INFO executor.Executor: Running task 1.0 in stage 3.0 (TID 83)
18/05/02 06:06:15 INFO executor.Executor: Running task 2.0 in stage 3.0 (TID 84)
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 3.0 (TID 81) in 72 ms on localhost (executor driver) (1/39)
18/05/02 06:06:15 INFO executor.Executor: Running task 0.0 in stage 3.0 (TID 82)

18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 3.0 (TID 80) in 74 ms on localhost (executor driver) (2/39)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 3.0 (TID 79) in 78 ms on localhost (executor driver) (3/39)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO executor.Executor: Finished task 2.0 in stage 3.0 (TID 84). 1306 bytes result sent to driver
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 3.0 (TID 85, localhost, executor driver, partition 3, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:15 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:15 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 3.0 (TID 84) in 14 ms on localhost (executor driver) (4/39)
18/05/02 06:06:16 INFO executor.Executor: Finished task 0.0 in stage 3.0 (TID 82). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 3.0 (TID 86, localhost, executor driver, partition 4, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Finished task 1.0 in stage 3.0 (TID 83). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 3.0 (TID 82) in 34 ms on localhost (executor driver) (5/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 4.0 in stage 3.0 (TID 86)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 3.0 (TID 87, localhost, executor driver, partition 5, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 3.0 (TID 83) in 23 ms on localhost (executor driver) (6/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 5.0 in stage 3.0 (TID 87)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 4.0 in stage 3.0 (TID 86). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO executor.Executor: Finished task 5.0 in stage 3.0 (TID 87). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 3.0 (TID 88, localhost, executor driver, partition 6, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Running task 6.0 in stage 3.0 (TID 88)

18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 3.0 (TID 89, localhost, executor driver, partition 7, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Running task 7.0 in stage 3.0 (TID 89)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 3.0 (TID 87) in 14 ms on localhost (executor driver) (7/39)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 3.0 (TID 86) in 19 ms on localhost (executor driver) (8/39)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Running task 3.0 in stage 3.0 (TID 85)
18/05/02 06:06:16 INFO executor.Executor: Finished task 6.0 in stage 3.0 (TID 88). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 3.0 (TID 90, localhost, executor driver, partition 8, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Finished task 3.0 in stage 3.0 (TID 85). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 3.0 (TID 88) in 18 ms on localhost (executor driver) (9/39)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 3.0 (TID 91, localhost, executor driver, partition 9, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Running task 9.0 in stage 3.0 (TID 91)
18/05/02 06:06:16 INFO executor.Executor: Finished task 7.0 in stage 3.0 (TID 89). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 3.0 (TID 92, localhost, executor driver, partition 14, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 3.0 (TID 85) in 40 ms on localhost (executor driver) (10/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 14.0 in stage 3.0 (TID 92)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 10.0 in stage 3.0 (TID 78). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 3.0 (TID 93, localhost, executor driver, partition 15, PROCESS_LOCAL, 1975 bytes)

18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 7.0 in
stage 3.0 (TID 89) in 32 ms on localhost (executor driver) (11/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 8.0 in stage 3.0
(TID 90)
18/05/02 06:06:16 INFO executor.Executor: Finished task 9.0 in stage 3.0
(TID 91). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 16.0 in
stage 3.0 (TID 94, localhost, executor driver, partition 16,
PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Running task 16.0 in stage 3.0
(TID 94)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Running task 15.0 in stage 3.0
(TID 93)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 8.0 in stage 3.0
(TID 90). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 17.0 in
stage 3.0 (TID 95, localhost, executor driver, partition 17,
PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 8.0 in
stage 3.0 (TID 90) in 29 ms on localhost (executor driver) (12/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 17.0 in stage 3.0
(TID 95)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 16.0 in stage 3.0
(TID 94). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 9.0 in
stage 3.0 (TID 91) in 30 ms on localhost (executor driver) (13/39)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 18.0 in
stage 3.0 (TID 96, localhost, executor driver, partition 18,
PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 15.0 in stage 3.0
(TID 93). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 16.0 in
stage 3.0 (TID 94) in 15 ms on localhost (executor driver) (14/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 18.0 in stage 3.0
(TID 96)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 19.0 in
stage 3.0 (TID 97, localhost, executor driver, partition 19,
PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 15.0 in
stage 3.0 (TID 93) in 27 ms on localhost (executor driver) (15/39)

18/05/02 06:06:16 INFO executor.Executor: Finished task 18.0 in stage 3.0 (TID 96). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 3.0 (TID 98, localhost, executor driver, partition 20, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Running task 19.0 in stage 3.0 (TID 97)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 3.0 (TID 96) in 13 ms on localhost (executor driver) (16/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 20.0 in stage 3.0 (TID 98)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 19.0 in stage 3.0 (TID 97). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO executor.Executor: Finished task 20.0 in stage 3.0 (TID 98). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 3.0 (TID 99, localhost, executor driver, partition 21, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 3.0 (TID 100, localhost, executor driver, partition 22, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 3.0 (TID 97) in 19 ms on localhost (executor driver) (17/39)
18/05/02 06:06:16 INFO executor.Executor: Finished task 14.0 in stage 3.0 (TID 92). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 3.0 (TID 98) in 16 ms on localhost (executor driver) (18/39)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 3.0 (TID 101, localhost, executor driver, partition 23, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Running task 23.0 in stage 3.0 (TID 101)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 3.0 (TID 92) in 59 ms on localhost (executor driver) (19/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 22.0 in stage 3.0 (TID 100)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 23.0 in stage 3.0 (TID 101). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 3.0 (TID 102, localhost, executor driver, partition 24, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Finished task 22.0 in stage 3.0 (TID 100). 1306 bytes result sent to driver

18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 23.0 in
stage 3.0 (TID 101) in 12 ms on localhost (executor driver) (20/39)
18/05/02 06:06:16 INFO executor.Executor: Finished task 17.0 in stage 3.0
(TID 95). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 10.0 in
stage 3.0 (TID 78) in 189 ms on localhost (executor driver) (21/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 21.0 in stage 3.0
(TID 99)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 21.0 in stage 3.0
(TID 99). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 25.0 in
stage 3.0 (TID 103, localhost, executor driver, partition 25,
PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 26.0 in
stage 3.0 (TID 104, localhost, executor driver, partition 26,
PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 27.0 in
stage 3.0 (TID 105, localhost, executor driver, partition 27,
PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 21.0 in
stage 3.0 (TID 99) in 39 ms on localhost (executor driver) (22/39)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 17.0 in
stage 3.0 (TID 95) in 70 ms on localhost (executor driver) (23/39)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 22.0 in
stage 3.0 (TID 100) in 43 ms on localhost (executor driver) (24/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 27.0 in stage 3.0
(TID 105)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 1 ms
18/05/02 06:06:16 INFO executor.Executor: Running task 26.0 in stage 3.0
(TID 104)
18/05/02 06:06:16 INFO executor.Executor: Finished task 27.0 in stage 3.0
(TID 105). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 28.0 in
stage 3.0 (TID 106, localhost, executor driver, partition 28,
PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 27.0 in
stage 3.0 (TID 105) in 19 ms on localhost (executor driver) (25/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 24.0 in stage 3.0
(TID 102)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 14 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 26.0 in stage 3.0
(TID 104). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO executor.Executor: Running task 28.0 in stage 3.0
(TID 106)

18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 3.0 (TID 107, localhost, executor driver, partition 29, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Finished task 24.0 in stage 3.0 (TID 102). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 3.0 (TID 104) in 40 ms on localhost (executor driver) (26/39)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Running task 25.0 in stage 3.0 (TID 103)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 3.0 (TID 108, localhost, executor driver, partition 30, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 3.0 (TID 102) in 74 ms on localhost (executor driver) (27/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 29.0 in stage 3.0 (TID 107)
18/05/02 06:06:16 INFO executor.Executor: Finished task 25.0 in stage 3.0 (TID 103). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 3.0 (TID 109, localhost, executor driver, partition 31, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Running task 31.0 in stage 3.0 (TID 109)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 29.0 in stage 3.0 (TID 107). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO executor.Executor: Finished task 31.0 in stage 3.0 (TID 109). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 3.0 (TID 103) in 64 ms on localhost (executor driver) (28/39)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 3.0 (TID 110, localhost, executor driver, partition 32, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 3.0 (TID 111, localhost, executor driver, partition 33, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 3.0 (TID 107) in 30 ms on localhost (executor driver) (29/39)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 3.0 (TID 109) in 14 ms on localhost (executor driver) (30/39)
18/05/02 06:06:16 INFO executor.Executor: Finished task 28.0 in stage 3.0 (TID 106). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 3.0 (TID 112, localhost, executor driver, partition 34, PROCESS_LOCAL, 1975 bytes)

18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 3.0 (TID 106) in 53 ms on localhost (executor driver) (31/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 34.0 in stage 3.0 (TID 112)
18/05/02 06:06:16 INFO executor.Executor: Running task 33.0 in stage 3.0 (TID 111)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 34.0 in stage 3.0 (TID 112). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 3.0 (TID 113, localhost, executor driver, partition 35, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 3.0 (TID 112) in 14 ms on localhost (executor driver) (32/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 30.0 in stage 3.0 (TID 108)
18/05/02 06:06:16 INFO executor.Executor: Finished task 33.0 in stage 3.0 (TID 111). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 3.0 (TID 114, localhost, executor driver, partition 36, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Running task 32.0 in stage 3.0 (TID 110)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 3.0 (TID 111) in 29 ms on localhost (executor driver) (33/39)
18/05/02 06:06:16 INFO executor.Executor: Finished task 30.0 in stage 3.0 (TID 108). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO executor.Executor: Running task 35.0 in stage 3.0 (TID 113)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 3.0 (TID 115, localhost, executor driver, partition 37, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 3.0 (TID 108) in 57 ms on localhost (executor driver) (34/39)
18/05/02 06:06:16 INFO executor.Executor: Finished task 35.0 in stage 3.0 (TID 113). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 3.0 (TID 116, localhost, executor driver, partition 38, PROCESS_LOCAL, 1975 bytes)
18/05/02 06:06:16 INFO executor.Executor: Running task 37.0 in stage 3.0 (TID 115)

18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 3.0 (TID 113) in 25 ms on localhost (executor driver) (35/39)
18/05/02 06:06:16 INFO executor.Executor: Finished task 32.0 in stage 3.0 (TID 110). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 3.0 (TID 110) in 48 ms on localhost (executor driver) (36/39)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Running task 36.0 in stage 3.0 (TID 114)
18/05/02 06:06:16 INFO executor.Executor: Finished task 37.0 in stage 3.0 (TID 115). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 36.0 in stage 3.0 (TID 114). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 3.0 (TID 115) in 28 ms on localhost (executor driver) (37/39)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 3.0 (TID 114) in 39 ms on localhost (executor driver) (38/39)
18/05/02 06:06:16 INFO executor.Executor: Running task 38.0 in stage 3.0 (TID 116)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/05/02 06:06:16 INFO executor.Executor: Finished task 38.0 in stage 3.0 (TID 116). 1306 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 3.0 (TID 116) in 42 ms on localhost (executor driver) (39/39)
18/05/02 06:06:16 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
18/05/02 06:06:16 INFO scheduler.DAGScheduler: ShuffleMapStage 3 (mapToPair at AssignmentProblem3.java:54) finished in 0.355 s
18/05/02 06:06:16 INFO scheduler.DAGScheduler: looking for newly runnable stages
18/05/02 06:06:16 INFO scheduler.DAGScheduler: running: Set()
18/05/02 06:06:16 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 4)
18/05/02 06:06:16 INFO scheduler.DAGScheduler: failed: Set()
18/05/02 06:06:16 INFO scheduler.DAGScheduler: Submitting ResultStage 4 (MapPartitionsRDD[10] at saveAsTextFile at AssignmentProblem3.java:57), which has no missing parents
18/05/02 06:06:16 INFO storage.MemoryStore: Block broadcast_4 stored as values in memory (estimated size 82.2 KB, free 529.6 MB)
18/05/02 06:06:16 INFO storage.MemoryStore: Block broadcast_4_piece0 stored as bytes in memory (estimated size 29.8 KB, free 529.6 MB)
18/05/02 06:06:16 INFO storage.BlockManagerInfo: Added broadcast_4_piece0 in memory on localhost:43712 (size: 29.8 KB, free: 529.9 MB)
18/05/02 06:06:16 INFO spark.SparkContext: Created broadcast 4 from broadcast at DAGScheduler.scala:1004
18/05/02 06:06:16 INFO scheduler.DAGScheduler: Submitting 5 missing tasks from ResultStage 4 (MapPartitionsRDD[10] at saveAsTextFile at AssignmentProblem3.java:57) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4))

18/05/02 06:06:16 INFO scheduler.TaskSchedulerImpl: Adding task set 4.0
with 5 tasks
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 0.0 in
stage 4.0 (TID 117, localhost, executor driver, partition 0, NODE_LOCAL,
1986 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 1.0 in
stage 4.0 (TID 118, localhost, executor driver, partition 1, NODE_LOCAL,
1986 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 2.0 in
stage 4.0 (TID 119, localhost, executor driver, partition 2, NODE_LOCAL,
1986 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 3.0 in
stage 4.0 (TID 120, localhost, executor driver, partition 3, NODE_LOCAL,
1986 bytes)
18/05/02 06:06:16 INFO executor.Executor: Running task 3.0 in stage 4.0
(TID 120)
18/05/02 06:06:16 INFO executor.Executor: Running task 2.0 in stage 4.0
(TID 119)
18/05/02 06:06:16 INFO executor.Executor: Running task 1.0 in stage 4.0
(TID 118)
18/05/02 06:06:16 INFO executor.Executor: Running task 0.0 in stage 4.0
(TID 117)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 4
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 4
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 4
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 1 ms
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 4
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 0 ms
18/05/02 06:06:16 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:06:16 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:06:16 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:06:16 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:06:16 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:06:16 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:06:16 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:06:16 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:06:16 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020606_0004_m_000002_119' to hdfs://ip-10-0-0-212.ap-

south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que3out/_t
emporary/0/task_201805020606_0004_m_000002
18/05/02 06:06:16 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020606_0004_m_000002_119: Committed
18/05/02 06:06:16 INFO executor.Executor: Finished task 2.0 in stage 4.0
(TID 119). 2038 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Starting task 4.0 in
stage 4.0 (TID 121, localhost, executor driver, partition 4, NODE_LOCAL,
1986 bytes)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 2.0 in
stage 4.0 (TID 119) in 182 ms on localhost (executor driver) (1/5)
18/05/02 06:06:16 INFO executor.Executor: Running task 4.0 in stage 4.0
(TID 121)
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Getting 4
non-empty blocks out of 39 blocks
18/05/02 06:06:16 INFO storage.ShuffleBlockFetcherIterator: Started 0
remote fetches in 1 ms
18/05/02 06:06:16 INFO output.FileOutputCommitter: File Output Committer
Algorithm version is 1
18/05/02 06:06:16 INFO output.FileOutputCommitter: FileOutputCommitter
skip cleanup _temporary folders under output directory:false, ignore
cleanup failures: false
18/05/02 06:06:16 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020606_0004_m_000004_121' to hdfs://ip-10-0-0-212.ap-
south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que3out/_t
emporary/0/task_201805020606_0004_m_000004
18/05/02 06:06:16 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020606_0004_m_000004_121: Committed
18/05/02 06:06:16 INFO executor.Executor: Finished task 4.0 in stage 4.0
(TID 121). 2052 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 4.0 in
stage 4.0 (TID 121) in 41 ms on localhost (executor driver) (2/5)
18/05/02 06:06:16 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020606_0004_m_000001_118' to hdfs://ip-10-0-0-212.ap-
south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que3out/_t
emporary/0/task_201805020606_0004_m_000001
18/05/02 06:06:16 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020606_0004_m_000001_118: Committed
18/05/02 06:06:16 INFO executor.Executor: Finished task 1.0 in stage 4.0
(TID 118). 2052 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 1.0 in
stage 4.0 (TID 118) in 570 ms on localhost (executor driver) (3/5)
18/05/02 06:06:16 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020606_0004_m_000003_120' to hdfs://ip-10-0-0-212.ap-
south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que3out/_t
emporary/0/task_201805020606_0004_m_000003
18/05/02 06:06:16 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020606_0004_m_000003_120: Committed
18/05/02 06:06:16 INFO output.FileOutputCommitter: Saved output of task
'attempt_201805020606_0004_m_000000_117' to hdfs://ip-10-0-0-212.ap-
south-
1.compute.internal:8020/user/root/spark_assignment/sparkoutput/que3out/_t
emporary/0/task_201805020606_0004_m_000000
18/05/02 06:06:16 INFO mapred.SparkHadoopMapRedUtil:
attempt_201805020606_0004_m_000000_117: Committed

18/05/02 06:06:16 INFO executor.Executor: Finished task 3.0 in stage 4.0 (TID 120). 2052 bytes result sent to driver
18/05/02 06:06:16 INFO executor.Executor: Finished task 0.0 in stage 4.0 (TID 117). 2052 bytes result sent to driver
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 4.0 (TID 117) in 582 ms on localhost (executor driver) (4/5)
18/05/02 06:06:16 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 4.0 (TID 120) in 582 ms on localhost (executor driver) (5/5)
18/05/02 06:06:16 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool
18/05/02 06:06:16 INFO scheduler.DAGScheduler: ResultStage 4 (saveAsTextFile at AssignmentProblem3.java:57) finished in 0.583 s
18/05/02 06:06:16 INFO scheduler.DAGScheduler: Job 1 finished: saveAsTextFile at AssignmentProblem3.java:57, took 0.986941 s
18/05/02 06:06:17 INFO storage.BlockManagerInfo: Removed broadcast_2_piece0 on localhost:43712 in memory (size: 2.5 KB, free: 530.0 MB)
File has been written with output
18/05/02 06:06:17 INFO spark.SparkContext: Invoking stop() from shutdown hook
18/05/02 06:06:17 INFO storage.BlockManagerInfo: Removed broadcast_4_piece0 on localhost:43712 in memory (size: 29.8 KB, free: 530.0 MB)
18/05/02 06:06:17 INFO spark.ContextCleaner: Cleaned accumulator 4
18/05/02 06:06:17 INFO storage.BlockManagerInfo: Removed broadcast_3_piece0 on localhost:43712 in memory (size: 2.6 KB, free: 530.0 MB)
18/05/02 06:06:17 INFO spark.ContextCleaner: Cleaned accumulator 3
18/05/02 06:06:17 INFO spark.ContextCleaner: Cleaned accumulator 2
18/05/02 06:06:17 INFO storage.BlockManagerInfo: Removed broadcast_1_piece0 on localhost:43712 in memory (size: 2.9 KB, free: 530.0 MB)
18/05/02 06:06:17 INFO spark.ContextCleaner: Cleaned accumulator 1
18/05/02 06:06:17 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.212:4040
18/05/02 06:06:17 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/05/02 06:06:17 INFO storage.MemoryStore: MemoryStore cleared
18/05/02 06:06:17 INFO storage.BlockManager: BlockManager stopped
18/05/02 06:06:17 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
18/05/02 06:06:17 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/05/02 06:06:17 INFO remote.RemoteActorRefProvider$RemotingTerminator: Shutting down remote daemon.
18/05/02 06:06:17 INFO remote.RemoteActorRefProvider$RemotingTerminator: Remote daemon shut down; proceeding with flushing remote transports.
18/05/02 06:06:17 INFO spark.SparkContext: Successfully stopped SparkContext
18/05/02 06:06:17 INFO util.ShutdownHookManager: Shutdown hook called
18/05/02 06:06:17 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-ce991521-bfda-41a8-8e62-9c2e5fab6ba7
18/05/02 06:06:17 INFO Remoting: Remoting shut down
18/05/02 06:06:17 INFO remote.RemoteActorRefProvider$RemotingTerminator: Remoting shut down.

##################################   End of Logs for
Spark - Assigment3 jar
##