

Logs for spark Assignment

LookUpRDD:

[root@ip-10-0-0-161 spark Assignment]# spark-submit --class spark.question1.lookUpRDD \

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

> --name aws_spark --conf "spark.app.id=spark_aws_run" \

> lookUpRDD.jar /user/root/spark_assignment/input_dataset/yellow_tripdata*
/user/root/spark_assignment/output/single_row_lookup_SparkRDD

18/04/21 06:19:39 INFO spark.SparkContext: Running Spark version 1.6.0

18/04/21 06:19:39 INFO spark.SecurityManager: Changing view acls to: root

18/04/21 06:19:39 INFO spark.SecurityManager: Changing modify acls to: root

18/04/21 06:19:39 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set(root)

18/04/21 06:19:39 INFO util.Utils: Successfully started service 'sparkDriver' on port 38383.

18/04/21 06:19:40 INFO slf4j.Slf4jLogger: Slf4jLogger started

18/04/21 06:19:40 INFO Remoting: Starting remoting

18/04/21 06:19:40 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://sparkDriverActorSystem@10.0.0.161:42128]

18/04/21 06:19:40 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://sparkDriverActorSystem@10.0.0.161:42128]

18/04/21 06:19:40 INFO util.Utils: Successfully started service 'sparkDriverActorSystem' on port 42128.

18/04/21 06:19:40 INFO spark.SparkEnv: Registering MapOutputTracker

18/04/21 06:19:40 INFO spark.SparkEnv: Registering BlockManagerMaster

18/04/21 06:19:40 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-6a29ccd0-f155-49e9-81cd-b4b022049594

18/04/21 06:19:40 INFO storage.MemoryStore: MemoryStore started with capacity 530.0 MB

18/04/21 06:19:40 INFO spark.SparkEnv: Registering OutputCommitCoordinator

18/04/21 06:19:40 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.

18/04/21 06:19:40 INFO ui.SparkUI: Started SparkUI at http://10.0.0.161:4040

18/04/21 06:19:40 INFO spark.SparkContext: Added JAR file:/root/spark_Assignment/lookUpRDD.jar at spark://10.0.0.161:38383/jars/lookUpRDD.jar with timestamp 1524291580807

18/04/21 06:19:40 INFO executor.Executor: Starting executor ID driver on host localhost

18/04/21 06:19:40 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 39186.

18/04/21 06:19:40 INFO netty.NettyBlockTransferService: Server created on 39186

18/04/21 06:19:40 INFO storage.BlockManager: external shuffle service port = 7337

18/04/21 06:19:40 INFO storage.BlockManagerMaster: Trying to register BlockManager

18/04/21 06:19:40 INFO storage.BlockManagerMasterEndpoint: Registering block manager localhost:39186 with 530.0 MB RAM, BlockManagerId(driver, localhost, 39186)

18/04/21 06:19:40 INFO storage.BlockManagerMaster: Registered BlockManager

18/04/21 06:19:42 INFO scheduler.EventLoggingListener: Logging events to hdfs://ip-10-0-0-161.ec2.internal:8020/user/spark/applicationHistory/local-1524291580855

18/04/21 06:19:42 INFO spark.SparkContext: Registered listener com.cloudera.spark.lineage.ClouderaNavigatorListener

18/04/21 06:19:42 INFO storage.MemoryStore: Block broadcast_0 stored as values in memory (estimated size 302.5 KB, free 529.7 MB)

18/04/21 06:19:42 INFO storage.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 26.0 KB, free 529.7 MB)

18/04/21 06:19:42 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on localhost:39186 (size: 26.0 KB, free: 530.0 MB)

18/04/21 06:19:42 INFO spark.SparkContext: Created broadcast 0 from textFile at lookUpRDD.java:32

18/04/21 06:19:42 INFO Configuration.deprecation: mapred.tip.id is deprecated. Instead, use mapreduce.task.id

18/04/21 06:19:42 INFO Configuration.deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id

18/04/21 06:19:42 INFO Configuration.deprecation: mapred.task.is.map is deprecated. Instead, use mapreduce.task.ismap

18/04/21 06:19:42 INFO Configuration.deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.partition

18/04/21 06:19:42 INFO Configuration.deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id

18/04/21 06:19:42 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:42 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:42 INFO mapred.FileInputFormat: Total input paths to process : 6

18/04/21 06:19:43 INFO spark.SparkContext: Starting job: saveAsTextFile at lookUpRDD.java:70

18/04/21 06:19:43 INFO scheduler.DAGScheduler: Got job 0 (saveAsTextFile at lookUpRDD.java:70) with 39 output partitions

18/04/21 06:19:43 INFO scheduler.DAGScheduler: Final stage: ResultStage 0 (saveAsTextFile at lookUpRDD.java:70)

18/04/21 06:19:43 INFO scheduler.DAGScheduler: Parents of final stage: List()

18/04/21 06:19:43 INFO scheduler.DAGScheduler: Missing parents: List()

18/04/21 06:19:43 INFO scheduler.DAGScheduler: Submitting ResultStage 0 (MapPartitionsRDD[8] at saveAsTextFile at lookUpRDD.java:70), which has no missing parents

18/04/21 06:19:43 INFO storage.MemoryStore: Block broadcast_1 stored as values in memory (estimated size 82.4 KB, free 529.6 MB)

18/04/21 06:19:43 INFO storage.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 29.7 KB, free 529.6 MB)

18/04/21 06:19:43 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on localhost:39186 (size: 29.7 KB, free: 530.0 MB)

18/04/21 06:19:43 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1004

18/04/21 06:19:43 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 0 (MapPartitionsRDD[8] at saveAsTextFile at lookUpRDD.java:70) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/04/21 06:19:43 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 39 tasks

18/04/21 06:19:43 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, executor driver, partition 0, ANY, 2262 bytes)

18/04/21 06:19:43 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, localhost, executor driver, partition 1, ANY, 2262 bytes)

18/04/21 06:19:43 INFO executor.Executor: Running task 1.0 in stage 0.0 (TID 1)

18/04/21 06:19:43 INFO executor.Executor: Running task 0.0 in stage 0.0 (TID 0)

18/04/21 06:19:43 INFO executor.Executor: Fetching spark://10.0.0.161:38383/jars/lookUpRDD.jar with timestamp 1524291580807

18/04/21 06:19:43 INFO spark.ExecutorAllocationManager: New executor driver has registered (new total is 1)

18/04/21 06:19:43 INFO util.Utils: Fetching spark://10.0.0.161:38383/jars/lookUpRDD.jar to /tmp/spark-aac54069-939f-4817-a34e-14d7ac19d4e7/userFiles-bb86d0ec-9c54-4271-bb2d-d4bfd7e465f2/fetchFileTemp7694889265009811309.tmp

18/04/21 06:19:43 INFO executor.Executor: Adding file:/tmp/spark-aac54069-939f-4817-a34e-14d7ac19d4e7/userFiles-bb86d0ec-9c54-4271-bb2d-d4bfd7e465f2/lookUpRDD.jar to class loader

18/04/21 06:19:43 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:134217728+134217728

18/04/21 06:19:43 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:0+134217728

18/04/21 06:19:43 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:43 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:43 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:43 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:44 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:44 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 2 total executors!

18/04/21 06:19:45 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:45 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 3 total executors!

18/04/21 06:19:46 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:46 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 4 total executors!

18/04/21 06:19:46 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000001 1' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task_201804210619 0000 m 000001

18/04/21 06:19:46 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000001 1: Committed

18/04/21 06:19:47 INFO executor.Executor: Finished task 1.0 in stage 0.0 (TID 1). 2295 bytes result sent to driver

18/04/21 06:19:47 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, localhost, executor driver, partition 2, ANY, 2262 bytes)

18/04/21 06:19:47 INFO executor.Executor: Running task 2.0 in stage 0.0 (TID 2)

18/04/21 06:19:47 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 0.0 (TID 1) in 3822 ms on localhost (executor driver) (1/39)

18/04/21 06:19:47 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:268435456+134217728

18/04/21 06:19:47 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:47 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:47 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000000 0' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task 201804210619 0000 m 000000

18/04/21 06:19:47 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000000 0: Committed

18/04/21 06:19:47 INFO executor.Executor: Finished task 0.0 in stage 0.0 (TID 0). 2360 bytes result sent to driver

18/04/21 06:19:47 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, localhost, executor driver, partition 3, ANY, 2262 bytes)

18/04/21 06:19:47 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 3957 ms on localhost (executor driver) (2/39)

18/04/21 06:19:47 INFO executor.Executor: Running task 3.0 in stage 0.0 (TID 3)

18/04/21 06:19:47 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:402653184+134217728

18/04/21 06:19:47 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:47 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:47 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:47 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 5 total executors!

18/04/21 06:19:48 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:48 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 6 total executors!

18/04/21 06:19:49 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:49 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 7 total executors!

18/04/21 06:19:50 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000003 3' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/_temporary/0/task_201804210619_0000_m_000003

18/04/21 06:19:50 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000003 3: Committed

18/04/21 06:19:50 INFO executor.Executor: Finished task 3.0 in stage 0.0 (TID 3). 2360 bytes result sent to driver

18/04/21 06:19:50 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 0.0 (TID 4, localhost, executor driver, partition 4, ANY, 2262 bytes)

18/04/21 06:19:50 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 0.0 (TID 3) in 2880 ms on localhost (executor driver) (3/39)

18/04/21 06:19:50 INFO executor.Executor: Running task 4.0 in stage 0.0 (TID 4)

18/04/21 06:19:50 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:536870912+134217728

18/04/21 06:19:50 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:50 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:50 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000002 2' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/_temporary/0/task_201804210619_0000_m_000002

18/04/21 06:19:50 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000002 2: Committed

18/04/21 06:19:50 INFO executor.Executor: Finished task 2.0 in stage 0.0 (TID 2). 2360 bytes result sent to driver

18/04/21 06:19:50 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 0.0 (TID 5, localhost, executor driver, partition 5, ANY, 2262 bytes)

18/04/21 06:19:50 INFO executor.Executor: Running task 5.0 in stage 0.0 (TID 5)

18/04/21 06:19:50 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 3074 ms on localhost (executor driver) (4/39)

18/04/21 06:19:50 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:671088640+86047889

18/04/21 06:19:50 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:50 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:50 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:50 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 8 total executors!

18/04/21 06:19:51 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:51 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 9 total executors!

18/04/21 06:19:51 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210619_0000_m_000005_5' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task_201804210619_0000_m_000005

18/04/21 06:19:51 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000005_5: Committed

18/04/21 06:19:51 INFO executor.Executor: Finished task 5.0 in stage 0.0 (TID 5). 2360 bytes result sent to driver

18/04/21 06:19:51 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 0.0 (TID 6, localhost, executor driver, partition 6, ANY, 2262 bytes)

18/04/21 06:19:51 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 0.0 (TID 5) in 1875 ms on localhost (executor driver) (5/39)

18/04/21 06:19:51 INFO executor.Executor: Running task 6.0 in stage 0.0 (TID 6)

18/04/21 06:19:51 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:0+134217728

18/04/21 06:19:51 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:51 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:52 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:52 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 10 total executors!

18/04/21 06:19:52 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210619_0000_m_000004_4' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/single row lookup SparkRDD/ temporary/0/task 201804210619_0000_m_000004

18/04/21 06:19:52 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000004_4: Committed

18/04/21 06:19:52 INFO executor.Executor: Finished task 4.0 in stage 0.0 (TID 4). 2360 bytes result sent to driver

18/04/21 06:19:52 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, localhost, executor driver, partition 7, ANY, 2262 bytes)

18/04/21 06:19:52 INFO executor.Executor: Running task 7.0 in stage 0.0 (TID 7)

18/04/21 06:19:52 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 2911 ms on localhost (executor driver) (6/39)

18/04/21 06:19:52 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-08.csv:134217728+134217728

18/04/21 06:19:52 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:52 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:53 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:53 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 11 total executors!

18/04/21 06:19:54 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:54 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 12 total executors!

18/04/21 06:19:54 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210619_0000_m_000006_6' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/single row lookup SparkRDD/ temporary/0/task 201804210619_0000_m_000006

18/04/21 06:19:54 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000006_6: Committed

18/04/21 06:19:54 INFO executor.Executor: Finished task 6.0 in stage 0.0 (TID 6). 2360 bytes result sent to driver

18/04/21 06:19:54 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, localhost, executor driver, partition 8, ANY, 2262 bytes)

18/04/21 06:19:54 INFO executor.Executor: Running task 8.0 in stage 0.0 (TID 8)

18/04/21 06:19:54 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:268435456+134217728

18/04/21 06:19:54 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 2797 ms on localhost (executor driver) (7/39)

18/04/21 06:19:54 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:54 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:55 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:55 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 13 total executors!

18/04/21 06:19:55 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210619_0000_m_000007_7' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task_201804210619_0000_m_000007

18/04/21 06:19:55 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000007_7: Committed

18/04/21 06:19:55 INFO executor.Executor: Finished task 7.0 in stage 0.0 (TID 7). 2360 bytes result sent to driver

18/04/21 06:19:55 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, localhost, executor driver, partition 9, ANY, 2262 bytes)

18/04/21 06:19:55 INFO executor.Executor: Running task 9.0 in stage 0.0 (TID 9)

18/04/21 06:19:55 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:402653184+134217728

18/04/21 06:19:55 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 2791 ms on localhost (executor driver) (8/39)

18/04/21 06:19:55 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:55 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:56 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:56 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 14 total executors!

18/04/21 06:19:57 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:57 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 15 total executors!

18/04/21 06:19:57 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000008 8' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task 201804210619 0000 m 000008

18/04/21 06:19:57 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000008 8: Committed

18/04/21 06:19:57 INFO executor.Executor: Finished task 8.0 in stage 0.0 (TID 8). 2360 bytes result sent to driver

18/04/21 06:19:57 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, localhost, executor driver, partition 10, ANY, 2262 bytes)

18/04/21 06:19:57 INFO executor.Executor: Running task 10.0 in stage 0.0 (TID 10)

18/04/21 06:19:57 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:536870912+134217728

18/04/21 06:19:57 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 3190 ms on localhost (executor driver) (9/39)

18/04/21 06:19:57 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:57 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:58 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:58 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 16 total executors!

18/04/21 06:19:59 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000009 9' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task 201804210619 0000 m 000009

18/04/21 06:19:59 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000009 9: Committed

18/04/21 06:19:59 INFO executor.Executor: Finished task 9.0 in stage 0.0 (TID 9). 2360 bytes result sent to driver

18/04/21 06:19:59 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, localhost, executor driver, partition 11, ANY, 2262 bytes)

18/04/21 06:19:59 INFO executor.Executor: Running task 11.0 in stage 0.0 (TID 11)

18/04/21 06:19:59 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 3406 ms on localhost (executor driver) (10/39)

18/04/21 06:19:59 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:671088640+71329760

18/04/21 06:19:59 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:19:59 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:19:59 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:19:59 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 17 total executors!

18/04/21 06:20:00 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:00 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 18 total executors!

18/04/21 06:20:00 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210619_0000_m_000011_11' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/_temporary/0/task_201804210619_0000_m_000011

18/04/21 06:20:00 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000011_11: Committed

18/04/21 06:20:00 INFO executor.Executor: Finished task 11.0 in stage 0.0 (TID 11). 2360 bytes result sent to driver

18/04/21 06:20:00 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 0.0 (TID 12, localhost, executor driver, partition 12, ANY, 2262 bytes)

18/04/21 06:20:00 INFO executor.Executor: Running task 12.0 in stage 0.0 (TID 12)

18/04/21 06:20:00 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 0.0 (TID 11) in 1525 ms on localhost (executor driver) (11/39)

18/04/21 06:20:00 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:0+134217728

18/04/21 06:20:00 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:00 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:00 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210619_0000_m_000010_10' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/temporary/0/task_201804210619_0000_m_000010

18/04/21 06:20:00 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000010_10: Committed

18/04/21 06:20:00 INFO executor.Executor: Finished task 10.0 in stage 0.0 (TID 10). 2360 bytes result sent to driver

18/04/21 06:20:00 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13, localhost, executor driver, partition 13, ANY, 2262 bytes)

18/04/21 06:20:00 INFO executor.Executor: Running task 13.0 in stage 0.0 (TID 13)

18/04/21 06:20:00 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:134217728+134217728

18/04/21 06:20:00 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 2894 ms on localhost (executor driver) (12/39)

18/04/21 06:20:00 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:00 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:01 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:01 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 19 total executors!

18/04/21 06:20:02 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:02 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 20 total executors!

18/04/21 06:20:03 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:03 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 21 total executors!

18/04/21 06:20:03 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210619_0000_m_000012_12' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/temporary/0/task_201804210619_0000_m_000012

18/04/21 06:20:03 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000012_12: Committed

18/04/21 06:20:03 INFO executor.Executor: Finished task 12.0 in stage 0.0 (TID 12). 2360 bytes result sent to driver

18/04/21 06:20:03 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, localhost, executor driver, partition 14, ANY, 2262 bytes)

18/04/21 06:20:03 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 2911 ms on localhost (executor driver) (13/39)

18/04/21 06:20:03 INFO executor.Executor: Running task 14.0 in stage 0.0 (TID 14)

18/04/21 06:20:03 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:268435456+134217728

18/04/21 06:20:03 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:03 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:03 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210619_0000_m_000013_13' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/_temporary/0/task_201804210619_0000_m_000013

18/04/21 06:20:03 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000013_13: Committed

18/04/21 06:20:03 INFO executor.Executor: Finished task 13.0 in stage 0.0 (TID 13). 2360 bytes result sent to driver

18/04/21 06:20:03 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, localhost, executor driver, partition 15, ANY, 2262 bytes)

18/04/21 06:20:03 INFO executor.Executor: Running task 15.0 in stage 0.0 (TID 15)

18/04/21 06:20:03 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 0.0 (TID 13) in 2900 ms on localhost (executor driver) (14/39)

18/04/21 06:20:03 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:402653184+134217728

18/04/21 06:20:03 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:03 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:04 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:04 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 22 total executors!

18/04/21 06:20:05 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:05 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 23 total executors!

18/04/21 06:20:06 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:06 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 24 total executors!

18/04/21 06:20:06 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000014 14' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/single row lookup SparkRDD/ temporary/0/task 201804210619 0000 m 000014

18/04/21 06:20:06 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000014 14: Committed

18/04/21 06:20:06 INFO executor.Executor: Finished task 14.0 in stage 0.0 (TID 14). 2360 bytes result sent to driver

18/04/21 06:20:06 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, localhost, executor driver, partition 16, ANY, 2262 bytes)

18/04/21 06:20:06 INFO executor.Executor: Running task 16.0 in stage 0.0 (TID 16)

18/04/21 06:20:06 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 0.0 (TID 14) in 2772 ms on localhost (executor driver) (15/39)

18/04/21 06:20:06 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow tripdata 2017-09.csv:536870912+134217728

18/04/21 06:20:06 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:06 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:06 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000015 15' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/single row lookup SparkRDD/ temporary/0/task 201804210619 0000 m 000015

18/04/21 06:20:06 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000015 15: Committed

18/04/21 06:20:06 INFO executor.Executor: Finished task 15.0 in stage 0.0 (TID 15). 2360 bytes result sent to driver

18/04/21 06:20:06 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 0.0 (TID 17, localhost, executor driver, partition 17, ANY, 2262 bytes)

18/04/21 06:20:06 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 2928 ms on localhost (executor driver) (16/39)

18/04/21 06:20:06 INFO executor.Executor: Running task 17.0 in stage 0.0 (TID 17)

18/04/21 06:20:06 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:671088640+117983715

18/04/21 06:20:06 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:06 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:06 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:09 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210619_0000_m_000017_17' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task_201804210619_0000_m_000017

18/04/21 06:20:09 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000017_17: Committed

18/04/21 06:20:09 INFO executor.Executor: Finished task 17.0 in stage 0.0 (TID 17). 2360 bytes result sent to driver

18/04/21 06:20:09 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 0.0 (TID 18, localhost, executor driver, partition 18, ANY, 2262 bytes)

18/04/21 06:20:09 INFO executor.Executor: Running task 18.0 in stage 0.0 (TID 18)

18/04/21 06:20:09 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 0.0 (TID 17) in 2645 ms on localhost (executor driver) (17/39)

18/04/21 06:20:09 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:0+134217728

18/04/21 06:20:09 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:09 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:09 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210619_0000_m_000016_16' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task_201804210619_0000_m_000016

18/04/21 06:20:09 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000016_16: Committed

18/04/21 06:20:09 INFO executor.Executor: Finished task 16.0 in stage 0.0 (TID 16). 2360 bytes result sent to driver

18/04/21 06:20:09 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 0.0 (TID 19, localhost, executor driver, partition 19, ANY, 2262 bytes)

18/04/21 06:20:09 INFO executor.Executor: Running task 19.0 in stage 0.0 (TID 19)

18/04/21 06:20:09 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 2997 ms on localhost (executor driver) (18/39)

18/04/21 06:20:09 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:134217728+134217728

18/04/21 06:20:09 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:09 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:09 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:11 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000019 19' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/_temporary/0/task_201804210619_0000_m_000019

18/04/21 06:20:11 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000019 19: Committed

18/04/21 06:20:11 INFO executor.Executor: Finished task 19.0 in stage 0.0 (TID 19). 2360 bytes result sent to driver

18/04/21 06:20:11 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 0.0 (TID 20, localhost, executor driver, partition 20, ANY, 2262 bytes)

18/04/21 06:20:11 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 0.0 (TID 19) in 2613 ms on localhost (executor driver) (19/39)

18/04/21 06:20:11 INFO executor.Executor: Running task 20.0 in stage 0.0 (TID 20)

18/04/21 06:20:11 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:268435456+134217728

18/04/21 06:20:11 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:11 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:11 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:12 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000018 18' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/_temporary/0/task_201804210619_0000_m_000018

18/04/21 06:20:12 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000018 18: Committed

18/04/21 06:20:12 INFO executor.Executor: Finished task 18.0 in stage 0.0 (TID 18). 2360 bytes result sent to driver

18/04/21 06:20:12 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 0.0 (TID 21, localhost, executor driver, partition 21, ANY, 2262 bytes)

18/04/21 06:20:12 INFO executor.Executor: Running task 21.0 in stage 0.0 (TID 21)

18/04/21 06:20:12 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:402653184+134217728

18/04/21 06:20:12 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 0.0 (TID 18) in 3059 ms on localhost (executor driver) (20/39)

18/04/21 06:20:12 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:12 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:12 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:14 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000020 20' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task 201804210619_0000_m_000020

18/04/21 06:20:14 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000020 20: Committed

18/04/21 06:20:14 INFO executor.Executor: Finished task 20.0 in stage 0.0 (TID 20). 2360 bytes result sent to driver

18/04/21 06:20:14 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 0.0 (TID 22, localhost, executor driver, partition 22, ANY, 2262 bytes)

18/04/21 06:20:14 INFO executor.Executor: Running task 22.0 in stage 0.0 (TID 22)

18/04/21 06:20:14 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 0.0 (TID 20) in 3071 ms on localhost (executor driver) (21/39)

18/04/21 06:20:14 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:536870912+134217728

18/04/21 06:20:14 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:14 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:15 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:15 INFO output.FileOutputCommitter: Saved output of task

'attempt_201804210619_0000_m_000021_21' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark assignment/output/single row lookup SparkRDD/ temporary/0/task 20

1804210619_0000_m_000021

18/04/21 06:20:15 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000021_21:

Committed

18/04/21 06:20:15 INFO executor.Executor: Finished task 21.0 in stage 0.0 (TID 21). 2360 bytes result sent to driver

18/04/21 06:20:15 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 0.0 (TID 23, localhost, executor driver, partition 23, ANY, 2262 bytes)

18/04/21 06:20:15 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 0.0 (TID 21) in 3240 ms on localhost (executor driver) (22/39)

18/04/21 06:20:15 INFO executor.Executor: Running task 23.0 in stage 0.0 (TID 23)

18/04/21 06:20:15 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-

10.csv:671088640+134217728

18/04/21 06:20:15 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:15 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:15 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:17 INFO output.FileOutputCommitter: Saved output of task

'attempt_201804210619_0000_m_000022_22' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark assignment/output/single row lookup SparkRDD/ temporary/0/task 20

1804210619_0000_m_000022

18/04/21 06:20:17 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000022_22:

Committed

18/04/21 06:20:17 INFO executor.Executor: Finished task 22.0 in stage 0.0 (TID 22). 2360 bytes result sent to driver

18/04/21 06:20:17 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 0.0 (TID 24, localhost, executor driver, partition 24, ANY, 2262 bytes)

18/04/21 06:20:17 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 0.0 (TID 22) in 2818 ms on localhost (executor driver) (23/39)

18/04/21 06:20:17 INFO executor.Executor: Running task 24.0 in stage 0.0 (TID 24)

18/04/21 06:20:17 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:805306368+56688482

18/04/21 06:20:17 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:17 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:17 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:18 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000023 23' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task_201804210619 0000 m 000023

18/04/21 06:20:18 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000023 23: Committed

18/04/21 06:20:18 INFO executor.Executor: Finished task 23.0 in stage 0.0 (TID 23). 2360 bytes result sent to driver

18/04/21 06:20:18 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 0.0 (TID 25, localhost, executor driver, partition 25, ANY, 2262 bytes)

18/04/21 06:20:18 INFO executor.Executor: Running task 25.0 in stage 0.0 (TID 25)

18/04/21 06:20:18 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 0.0 (TID 23) in 2949 ms on localhost (executor driver) (24/39)

18/04/21 06:20:18 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:0+134217728

18/04/21 06:20:18 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:18 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:18 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:19 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000024 24' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task_201804210619 0000 m 000024

18/04/21 06:20:19 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000024 24: Committed

18/04/21 06:20:19 INFO executor.Executor: Finished task 24.0 in stage 0.0 (TID 24). 2360 bytes result sent to driver

18/04/21 06:20:19 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 0.0 (TID 26, localhost, executor driver, partition 26, ANY, 2262 bytes)

18/04/21 06:20:19 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 0.0 (TID 24) in 1289 ms on localhost (executor driver) (25/39)

18/04/21 06:20:19 INFO executor.Executor: Running task 26.0 in stage 0.0 (TID 26)

18/04/21 06:20:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:134217728+134217728

18/04/21 06:20:19 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:19 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:19 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:21 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000025 25' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/_temporary/0/task_201804210619_0000_m_000025

18/04/21 06:20:21 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000025 25: Committed

18/04/21 06:20:21 INFO executor.Executor: Finished task 25.0 in stage 0.0 (TID 25). 2360 bytes result sent to driver

18/04/21 06:20:21 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 0.0 (TID 27, localhost, executor driver, partition 27, ANY, 2262 bytes)

18/04/21 06:20:21 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 0.0 (TID 25) in 2759 ms on localhost (executor driver) (26/39)

18/04/21 06:20:21 INFO executor.Executor: Running task 27.0 in stage 0.0 (TID 27)

18/04/21 06:20:21 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:268435456+134217728

18/04/21 06:20:21 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:21 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:21 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:21 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000026 26' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/_temporary/0/task_201804210619_0000_m_000026

18/04/21 06:20:21 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000026 26: Committed

18/04/21 06:20:21 INFO executor.Executor: Finished task 26.0 in stage 0.0 (TID 26). 2360 bytes result sent to driver

18/04/21 06:20:21 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 0.0 (TID 28, localhost, executor driver, partition 28, ANY, 2262 bytes)

18/04/21 06:20:21 INFO executor.Executor: Running task 28.0 in stage 0.0 (TID 28)

18/04/21 06:20:21 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 0.0 (TID 26) in 2809 ms on localhost (executor driver) (27/39)

18/04/21 06:20:21 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:402653184+134217728

18/04/21 06:20:21 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:21 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:21 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:24 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000027 27' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task 201804210619_0000_m_000027

18/04/21 06:20:24 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000027 27: Committed

18/04/21 06:20:24 INFO executor.Executor: Finished task 27.0 in stage 0.0 (TID 27). 2360 bytes result sent to driver

18/04/21 06:20:24 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 0.0 (TID 29, localhost, executor driver, partition 29, ANY, 2262 bytes)

18/04/21 06:20:24 INFO executor.Executor: Running task 29.0 in stage 0.0 (TID 29)

18/04/21 06:20:24 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 0.0 (TID 27) in 3036 ms on localhost (executor driver) (28/39)

18/04/21 06:20:24 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:536870912+134217728

18/04/21 06:20:24 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:24 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:24 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:24 INFO output.FileOutputCommitter: Saved output of task

'attempt_201804210619_0000_m_000028_28' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark assignment/output/single row lookup SparkRDD/ temporary/0/task 20

1804210619_0000_m_000028

18/04/21 06:20:24 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000028_28:

Committed

18/04/21 06:20:24 INFO executor.Executor: Finished task 28.0 in stage 0.0 (TID 28). 2360 bytes result sent to driver

18/04/21 06:20:24 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 0.0 (TID 30, localhost, executor

driver, partition 30, ANY, 2262 bytes)

18/04/21 06:20:24 INFO executor.Executor: Running task 30.0 in stage 0.0 (TID 30)

18/04/21 06:20:24 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 0.0 (TID 28) in 2857 ms on

localhost (executor driver) (29/39)

18/04/21 06:20:24 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-

11.csv:671088640+134217728

18/04/21 06:20:24 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:24 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders

under output directory:false, ignore cleanup failures: false

18/04/21 06:20:24 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:27 INFO output.FileOutputCommitter: Saved output of task

'attempt_201804210619_0000_m_000029_29' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark assignment/output/single row lookup SparkRDD/ temporary/0/task 20

1804210619_0000_m_000029

18/04/21 06:20:27 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000029_29:

Committed

18/04/21 06:20:27 INFO executor.Executor: Finished task 29.0 in stage 0.0 (TID 29). 2360 bytes result sent to driver

18/04/21 06:20:27 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 0.0 (TID 31, localhost, executor

driver, partition 31, ANY, 2262 bytes)

18/04/21 06:20:27 INFO executor.Executor: Running task 31.0 in stage 0.0 (TID 31)

18/04/21 06:20:27 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 0.0 (TID 29) in 2766 ms on

localhost (executor driver) (30/39)

18/04/21 06:20:27 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:805306368+13877504

18/04/21 06:20:27 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:27 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:27 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:27 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000031 31' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task_201804210619 0000 m 000031

18/04/21 06:20:27 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000031 31: Committed

18/04/21 06:20:27 INFO executor.Executor: Finished task 31.0 in stage 0.0 (TID 31). 2360 bytes result sent to driver

18/04/21 06:20:27 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 0.0 (TID 32, localhost, executor driver, partition 32, ANY, 2262 bytes)

18/04/21 06:20:27 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 0.0 (TID 31) in 307 ms on localhost (executor driver) (31/39)

18/04/21 06:20:27 INFO executor.Executor: Running task 32.0 in stage 0.0 (TID 32)

18/04/21 06:20:27 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:0+134217728

18/04/21 06:20:27 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:27 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:27 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:27 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000030 30' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task_201804210619 0000 m 000030

18/04/21 06:20:27 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000030 30: Committed

18/04/21 06:20:27 INFO executor.Executor: Finished task 30.0 in stage 0.0 (TID 30). 2360 bytes result sent to driver

18/04/21 06:20:27 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 0.0 (TID 33, localhost, executor driver, partition 33, ANY, 2262 bytes)

18/04/21 06:20:27 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 0.0 (TID 30) in 2714 ms on localhost (executor driver) (32/39)

18/04/21 06:20:27 INFO executor.Executor: Running task 33.0 in stage 0.0 (TID 33)

18/04/21 06:20:27 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:134217728+134217728

18/04/21 06:20:27 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:27 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:27 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:30 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000033 33' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/_temporary/0/task_201804210619_0000_m_000033

18/04/21 06:20:30 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000033 33: Committed

18/04/21 06:20:30 INFO executor.Executor: Finished task 33.0 in stage 0.0 (TID 33). 2360 bytes result sent to driver

18/04/21 06:20:30 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 0.0 (TID 34, localhost, executor driver, partition 34, ANY, 2262 bytes)

18/04/21 06:20:30 INFO executor.Executor: Running task 34.0 in stage 0.0 (TID 34)

18/04/21 06:20:30 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 0.0 (TID 33) in 2694 ms on localhost (executor driver) (33/39)

18/04/21 06:20:30 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:268435456+134217728

18/04/21 06:20:30 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:30 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:30 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:30 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000032 32' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/_temporary/0/task_201804210619_0000_m_000032

18/04/21 06:20:30 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000032 32: Committed

18/04/21 06:20:30 INFO executor.Executor: Finished task 32.0 in stage 0.0 (TID 32). 2360 bytes result sent to driver

18/04/21 06:20:30 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 0.0 (TID 35, localhost, executor driver, partition 35, ANY, 2262 bytes)

18/04/21 06:20:30 INFO executor.Executor: Running task 35.0 in stage 0.0 (TID 35)

18/04/21 06:20:30 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 0.0 (TID 32) in 2900 ms on localhost (executor driver) (34/39)

18/04/21 06:20:30 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:402653184+134217728

18/04/21 06:20:30 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:30 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:30 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:32 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000034 34' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task 201804210619_0000_m_000034

18/04/21 06:20:32 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000034 34: Committed

18/04/21 06:20:32 INFO executor.Executor: Finished task 34.0 in stage 0.0 (TID 34). 2360 bytes result sent to driver

18/04/21 06:20:32 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 0.0 (TID 36, localhost, executor driver, partition 36, ANY, 2262 bytes)

18/04/21 06:20:32 INFO executor.Executor: Running task 36.0 in stage 0.0 (TID 36)

18/04/21 06:20:32 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:536870912+134217728

18/04/21 06:20:32 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 0.0 (TID 34) in 2864 ms on localhost (executor driver) (35/39)

18/04/21 06:20:33 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:33 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:33 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:33 INFO output.FileOutputCommitter: Saved output of task

'attempt_201804210619_0000_m_000035_35' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark assignment/output/single row lookup SparkRDD/ temporary/0/task 20

1804210619_0000_m_000035

18/04/21 06:20:33 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000035_35:

Committed

18/04/21 06:20:33 INFO executor.Executor: Finished task 35.0 in stage 0.0 (TID 35). 2360 bytes result sent to driver

18/04/21 06:20:33 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 0.0 (TID 37, localhost, executor

driver, partition 37, ANY, 2262 bytes)

18/04/21 06:20:33 INFO executor.Executor: Running task 37.0 in stage 0.0 (TID 37)

18/04/21 06:20:33 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 0.0 (TID 35) in 2944 ms on

localhost (executor driver) (36/39)

18/04/21 06:20:33 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-

12.csv:671088640+134217728

18/04/21 06:20:33 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:33 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders

under output directory:false, ignore cleanup failures: false

18/04/21 06:20:33 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:35 INFO output.FileOutputCommitter: Saved output of task

'attempt_201804210619_0000_m_000036_36' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark assignment/output/single row lookup SparkRDD/ temporary/0/task 20

1804210619_0000_m_000036

18/04/21 06:20:35 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210619_0000_m_000036_36:

Committed

18/04/21 06:20:35 INFO executor.Executor: Finished task 36.0 in stage 0.0 (TID 36). 2360 bytes result sent to driver

18/04/21 06:20:35 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 0.0 (TID 38, localhost, executor

driver, partition 38, ANY, 2262 bytes)

18/04/21 06:20:35 INFO executor.Executor: Running task 38.0 in stage 0.0 (TID 38)

18/04/21 06:20:35 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 0.0 (TID 36) in 2975 ms on

localhost (executor driver) (37/39)

18/04/21 06:20:35 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:805306368+32783040

18/04/21 06:20:35 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:20:35 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:20:36 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:36 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000037 37' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task 201804210619 0000 m 000037

18/04/21 06:20:36 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000037 37: Committed

18/04/21 06:20:36 INFO executor.Executor: Finished task 37.0 in stage 0.0 (TID 37). 2360 bytes result sent to driver

18/04/21 06:20:36 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 0.0 (TID 37) in 2986 ms on localhost (executor driver) (38/39)

18/04/21 06:20:36 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:36 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210619 0000 m 000038 38' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/single_row_lookup_SparkRDD/ temporary/0/task 201804210619 0000 m 000038

18/04/21 06:20:36 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210619 0000 m 000038 38: Committed

18/04/21 06:20:36 INFO executor.Executor: Finished task 38.0 in stage 0.0 (TID 38). 2360 bytes result sent to driver

18/04/21 06:20:36 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 0.0 (TID 38) in 709 ms on localhost (executor driver) (39/39)

18/04/21 06:20:36 INFO scheduler.DAGScheduler: ResultStage 0 (saveAsTextFile at lookUpRDD.java:70) finished in 53.485 s

18/04/21 06:20:36 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

18/04/21 06:20:36 INFO scheduler.DAGScheduler: Job 0 finished: saveAsTextFile at lookUpRDD.java:70, took 53.612822 s

18/04/21 06:20:36 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:20:37 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.161:4040

18/04/21 06:20:37 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

18/04/21 06:20:37 INFO storage.MemoryStore: MemoryStore cleared

18/04/21 06:20:37 INFO storage.BlockManager: BlockManager stopped

18/04/21 06:20:37 INFO storage.BlockManagerMaster: BlockManagerMaster stopped

18/04/21 06:20:37 INFO scheduler.OutputCommitCoordinator\$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

18/04/21 06:20:37 INFO remote.RemoteActorRefProvider\$RemotingTerminator: Shutting down remote daemon.

18/04/21 06:20:37 INFO remote.RemoteActorRefProvider\$RemotingTerminator: Remote daemon shut down; proceeding with flushing remote transports.

18/04/21 06:20:37 INFO spark.SparkContext: Successfully stopped SparkContext

18/04/21 06:20:37 INFO util.ShutdownHookManager: Shutdown hook called

18/04/21 06:20:37 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-aac54069-939f-4817-a34e-14d7ac19d4e7

18/04/21 06:20:37 INFO Remoting: Remoting shut down

18/04/21 06:20:37 INFO remote.RemoteActorRefProvider\$RemotingTerminator: Remoting shut down.

2. FilterByRateCodeIdRDD

[root@ip-10-0-0-161 spark Assignment]# spark-submit --class spark.question2.FilterByRateCodeIdRDD \

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

> --name aws_spark --conf "spark.app.id=spark_aws_run" \

> FilterByRateCodeIdRDD.jar /user/root/spark_assignment/input_dataset/yellow_tripdata*
/user/root/spark_assignment/output/Filter By RateCodeId SparkRDD

18/04/21 06:22:52 INFO spark.SparkContext: Running Spark version 1.6.0

18/04/21 06:22:52 INFO spark.SecurityManager: Changing view acls to: root

18/04/21 06:22:52 INFO spark.SecurityManager: Changing modify acls to: root

18/04/21 06:22:52 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set(root)

18/04/21 06:22:53 INFO util.Utils: Successfully started service 'sparkDriver' on port 40773.

18/04/21 06:22:53 INFO slf4j.Slf4jLogger: Slf4jLogger started

18/04/21 06:22:53 INFO Remoting: Starting remoting

18/04/21 06:22:53 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://sparkDriverActorSystem@10.0.0.161:33364]

18/04/21 06:22:53 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://sparkDriverActorSystem@10.0.0.161:33364]

18/04/21 06:22:53 INFO util.Utils: Successfully started service 'sparkDriverActorSystem' on port 33364.

18/04/21 06:22:53 INFO spark.SparkEnv: Registering MapOutputTracker

18/04/21 06:22:53 INFO spark.SparkEnv: Registering BlockManagerMaster

18/04/21 06:22:53 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-82f78905-9bcb-4d78-8fcf-03dd3ec4e76e

18/04/21 06:22:53 INFO storage.MemoryStore: MemoryStore started with capacity 530.0 MB

18/04/21 06:22:53 INFO spark.SparkEnv: Registering OutputCommitCoordinator

18/04/21 06:22:53 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.

18/04/21 06:22:53 INFO ui.SparkUI: Started SparkUI at http://10.0.0.161:4040

18/04/21 06:22:53 INFO spark.SparkContext: Added JAR file:/root/spark_Assignment/FilterByRateCodeIdRDD.jar
at spark://10.0.0.161:40773/jars/FilterByRateCodeIdRDD.jar with timestamp 1524291773912

18/04/21 06:22:53 INFO executor.Executor: Starting executor ID driver on host localhost

18/04/21 06:22:54 INFO util.Utils: Successfully started service
'org.apache.spark.network.netty.NettyBlockTransferService' on port 37255.

18/04/21 06:22:54 INFO netty.NettyBlockTransferService: Server created on 37255

18/04/21 06:22:54 INFO storage.BlockManager: external shuffle service port = 7337

18/04/21 06:22:54 INFO storage.BlockManagerMaster: Trying to register BlockManager

18/04/21 06:22:54 INFO storage.BlockManagerMasterEndpoint: Registering block manager localhost:37255 with
530.0 MB RAM, BlockManagerId(driver, localhost, 37255)

18/04/21 06:22:54 INFO storage.BlockManagerMaster: Registered BlockManager

18/04/21 06:22:55 INFO scheduler.EventLoggingListener: Logging events to hdfs://ip-10-0-0-
161.ec2.internal:8020/user/spark/applicationHistory/local-1524291773953

18/04/21 06:22:55 INFO spark.SparkContext: Registered listener
com.cloudera.spark.lineage.ClouderaNavigatorListener

18/04/21 06:22:55 INFO storage.MemoryStore: Block broadcast_0 stored as values in memory (estimated size
302.5 KB, free 529.7 MB)

18/04/21 06:22:56 INFO storage.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 26.0 KB, free 529.7 MB)

18/04/21 06:22:56 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on localhost:37255 (size: 26.0 KB, free: 530.0 MB)

18/04/21 06:22:56 INFO spark.SparkContext: Created broadcast 0 from textFile at FilterByRateCodeIdRDD.java:26

18/04/21 06:22:56 INFO Configuration.deprecation: mapred.tip.id is deprecated. Instead, use mapreduce.task.id

18/04/21 06:22:56 INFO Configuration.deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id

18/04/21 06:22:56 INFO Configuration.deprecation: mapred.task.is.map is deprecated. Instead, use mapreduce.task.ismap

18/04/21 06:22:56 INFO Configuration.deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.partition

18/04/21 06:22:56 INFO Configuration.deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id

18/04/21 06:22:56 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:22:56 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:22:56 INFO mapred.FileInputFormat: Total input paths to process : 6

18/04/21 06:22:56 INFO spark.SparkContext: Starting job: saveAsTextFile at FilterByRateCodeIdRDD.java:42

18/04/21 06:22:56 INFO scheduler.DAGScheduler: Got job 0 (saveAsTextFile at FilterByRateCodeIdRDD.java:42) with 39 output partitions

18/04/21 06:22:56 INFO scheduler.DAGScheduler: Final stage: ResultStage 0 (saveAsTextFile at FilterByRateCodeIdRDD.java:42)

18/04/21 06:22:56 INFO scheduler.DAGScheduler: Parents of final stage: List()

18/04/21 06:22:56 INFO scheduler.DAGScheduler: Missing parents: List()

18/04/21 06:22:56 INFO scheduler.DAGScheduler: Submitting ResultStage 0 (MapPartitionsRDD[4] at saveAsTextFile at FilterByRateCodeIdRDD.java:42), which has no missing parents

18/04/21 06:22:56 INFO storage.MemoryStore: Block broadcast_1 stored as values in memory (estimated size 81.8 KB, free 529.6 MB)

18/04/21 06:22:56 INFO storage.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 29.6 KB, free 529.6 MB)

18/04/21 06:22:56 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on localhost:37255 (size: 29.6 KB, free: 530.0 MB)

18/04/21 06:22:56 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1004

18/04/21 06:22:56 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 0 (MapPartitionsRDD[4] at saveAsTextFile at FilterByRateCodeIdRDD.java:42) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/04/21 06:22:56 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 39 tasks

18/04/21 06:22:56 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, executor driver, partition 0, ANY, 2274 bytes)

18/04/21 06:22:56 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, localhost, executor driver, partition 1, ANY, 2274 bytes)

18/04/21 06:22:56 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, localhost, executor driver, partition 2, ANY, 2274 bytes)

18/04/21 06:22:56 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, localhost, executor driver, partition 3, ANY, 2274 bytes)

18/04/21 06:22:56 INFO executor.Executor: Running task 0.0 in stage 0.0 (TID 0)

18/04/21 06:22:56 INFO executor.Executor: Running task 3.0 in stage 0.0 (TID 3)

18/04/21 06:22:56 INFO executor.Executor: Running task 1.0 in stage 0.0 (TID 1)

18/04/21 06:22:56 INFO executor.Executor: Running task 2.0 in stage 0.0 (TID 2)

18/04/21 06:22:56 INFO executor.Executor: Fetching spark://10.0.0.161:40773/jars/FilterByRateCodeIdRDD.jar with timestamp 1524291773912

18/04/21 06:22:56 INFO spark.ExecutorAllocationManager: New executor driver has registered (new total is 1)

18/04/21 06:22:56 INFO util.Utils: Fetching spark://10.0.0.161:40773/jars/FilterByRateCodeIdRDD.jar to /tmp/spark-8c8ebda4-a11c-440d-841d-3edab7d3450a/userFiles-1c704fde-5563-4621-9203-6d4e101e0741/fetchFileTemp6335295350163998163.tmp

18/04/21 06:22:56 INFO executor.Executor: Adding file:/tmp/spark-8c8ebda4-a11c-440d-841d-3edab7d3450a/userFiles-1c704fde-5563-4621-9203-6d4e101e0741/FilterByRateCodeIdRDD.jar to class loader

18/04/21 06:22:57 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow tripdata 2017-07.csv:268435456+134217728

18/04/21 06:22:57 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow tripdata 2017-07.csv:134217728+134217728

18/04/21 06:22:57 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow tripdata 2017-07.csv:0+134217728

18/04/21 06:22:57 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:402653184+134217728

18/04/21 06:22:57 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:22:57 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:22:57 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:22:57 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:22:57 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:22:57 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:22:57 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:22:57 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:22:57 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:22:57 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 2 total executors!

18/04/21 06:22:58 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:22:58 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 3 total executors!

18/04/21 06:22:59 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:22:59 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 4 total executors!

18/04/21 06:23:00 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:00 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 5 total executors!

18/04/21 06:23:01 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:01 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 6 total executors!

18/04/21 06:23:01 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000003_3' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task 201804210622 0000 m 000003

18/04/21 06:23:01 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622 0000 m 000003 3: Committed

18/04/21 06:23:01 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210622 0000 m 000002 2' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task 201804210622 0000 m 000002

18/04/21 06:23:01 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622 0000 m 000002 2: Committed

18/04/21 06:23:01 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210622 0000 m 000001 1' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task 201804210622 0000 m 000001

18/04/21 06:23:01 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622 0000 m 000001 1: Committed

18/04/21 06:23:01 INFO executor.Executor: Finished task 1.0 in stage 0.0 (TID 1). 2295 bytes result sent to driver

18/04/21 06:23:01 INFO executor.Executor: Finished task 3.0 in stage 0.0 (TID 3). 2295 bytes result sent to driver

18/04/21 06:23:01 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 0.0 (TID 4, localhost, executor driver, partition 4, ANY, 2274 bytes)

18/04/21 06:23:01 INFO executor.Executor: Running task 4.0 in stage 0.0 (TID 4)

18/04/21 06:23:01 INFO executor.Executor: Finished task 2.0 in stage 0.0 (TID 2). 2295 bytes result sent to driver

18/04/21 06:23:01 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 0.0 (TID 5, localhost, executor driver, partition 5, ANY, 2274 bytes)

18/04/21 06:23:01 INFO executor.Executor: Running task 5.0 in stage 0.0 (TID 5)

18/04/21 06:23:01 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow tripdata 2017-07.csv:536870912+134217728

18/04/21 06:23:01 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 0.0 (TID 6, localhost, executor driver, partition 6, ANY, 2274 bytes)

18/04/21 06:23:01 INFO executor.Executor: Running task 6.0 in stage 0.0 (TID 6)

18/04/21 06:23:01 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow tripdata 2017-07.csv:671088640+86047889

18/04/21 06:23:01 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:0+134217728

18/04/21 06:23:01 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:01 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:01 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:01 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:01 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 5238 ms on localhost (executor driver) (1/39)

18/04/21 06:23:01 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 0.0 (TID 3) in 5239 ms on localhost (executor driver) (2/39)

18/04/21 06:23:01 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 0.0 (TID 1) in 5240 ms on localhost (executor driver) (3/39)

18/04/21 06:23:01 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:01 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:01 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000000_0' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/temporary/0/task_201804210622_0000_m_000000

18/04/21 06:23:01 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000000_0: Committed

18/04/21 06:23:01 INFO executor.Executor: Finished task 0.0 in stage 0.0 (TID 0). 2360 bytes result sent to driver

18/04/21 06:23:01 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, localhost, executor driver, partition 7, ANY, 2274 bytes)

18/04/21 06:23:01 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 5542 ms on localhost (executor driver) (4/39)

18/04/21 06:23:01 INFO executor.Executor: Running task 7.0 in stage 0.0 (TID 7)

18/04/21 06:23:02 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:134217728+134217728

18/04/21 06:23:02 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:02 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:02 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:02 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 7 total executors!

18/04/21 06:23:03 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:03 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 8 total executors!

18/04/21 06:23:04 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000005_5' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000005

18/04/21 06:23:04 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000005_5: Committed

18/04/21 06:23:04 INFO executor.Executor: Finished task 5.0 in stage 0.0 (TID 5). 2360 bytes result sent to driver

18/04/21 06:23:04 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, localhost, executor driver, partition 8, ANY, 2274 bytes)

18/04/21 06:23:04 INFO executor.Executor: Running task 8.0 in stage 0.0 (TID 8)

18/04/21 06:23:04 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 0.0 (TID 5) in 2755 ms on localhost (executor driver) (5/39)

18/04/21 06:23:04 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-08.csv:268435456+134217728

18/04/21 06:23:04 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:04 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:04 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:04 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 9 total executors!

18/04/21 06:23:05 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:05 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 10 total executors!

18/04/21 06:23:05 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000004_4' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000004

18/04/21 06:23:05 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000004_4: Committed

18/04/21 06:23:05 INFO executor.Executor: Finished task 4.0 in stage 0.0 (TID 4). 2360 bytes result sent to driver

18/04/21 06:23:05 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, localhost, executor driver, partition 9, ANY, 2274 bytes)

18/04/21 06:23:05 INFO executor.Executor: Running task 9.0 in stage 0.0 (TID 9)

18/04/21 06:23:05 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 3851 ms on localhost (executor driver) (6/39)

18/04/21 06:23:05 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-08.csv:402653184+134217728

18/04/21 06:23:05 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:05 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:05 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000006_6' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000006

18/04/21 06:23:05 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000006_6: Committed

18/04/21 06:23:05 INFO executor.Executor: Finished task 6.0 in stage 0.0 (TID 6). 2360 bytes result sent to driver

18/04/21 06:23:05 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, localhost, executor driver, partition 10, ANY, 2274 bytes)

18/04/21 06:23:05 INFO executor.Executor: Running task 10.0 in stage 0.0 (TID 10)

18/04/21 06:23:05 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 4028 ms on localhost (executor driver) (7/39)

18/04/21 06:23:05 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-08.csv:536870912+134217728

18/04/21 06:23:05 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:05 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:05 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000007_7' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622_0000_m_000007

18/04/21 06:23:05 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000007_7: Committed

18/04/21 06:23:05 INFO executor.Executor: Finished task 7.0 in stage 0.0 (TID 7). 2360 bytes result sent to driver

18/04/21 06:23:05 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, localhost, executor driver, partition 11, ANY, 2274 bytes)

18/04/21 06:23:05 INFO executor.Executor: Running task 11.0 in stage 0.0 (TID 11)

18/04/21 06:23:05 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:671088640+71329760

18/04/21 06:23:05 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 3931 ms on localhost (executor driver) (8/39)

18/04/21 06:23:05 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:05 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:06 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:06 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 11 total executors!

18/04/21 06:23:07 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:07 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 12 total executors!

18/04/21 06:23:07 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000011_11' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622_0000_m_000011

18/04/21 06:23:07 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000011_11: Committed

18/04/21 06:23:07 INFO executor.Executor: Finished task 11.0 in stage 0.0 (TID 11). 2360 bytes result sent to driver

18/04/21 06:23:07 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 0.0 (TID 12, localhost, executor driver, partition 12, ANY, 2274 bytes)

18/04/21 06:23:07 INFO executor.Executor: Running task 12.0 in stage 0.0 (TID 12)

18/04/21 06:23:07 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:0+134217728

18/04/21 06:23:07 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 0.0 (TID 11) in 1901 ms on localhost (executor driver) (9/39)

18/04/21 06:23:07 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:07 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:08 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000008_8' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/_temporary/0/task_201804210622_0000_m_000008

18/04/21 06:23:08 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000008_8: Committed

18/04/21 06:23:08 INFO executor.Executor: Finished task 8.0 in stage 0.0 (TID 8). 2360 bytes result sent to driver

18/04/21 06:23:08 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13, localhost, executor driver, partition 13, ANY, 2274 bytes)

18/04/21 06:23:08 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 3838 ms on localhost (executor driver) (10/39)

18/04/21 06:23:08 INFO executor.Executor: Running task 13.0 in stage 0.0 (TID 13)

18/04/21 06:23:08 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:134217728+134217728

18/04/21 06:23:08 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:08 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:08 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:08 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 13 total executors!

18/04/21 06:23:09 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000009_9' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark_assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000009

18/04/21 06:23:09 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000009_9: Committed

18/04/21 06:23:09 INFO executor.Executor: Finished task 9.0 in stage 0.0 (TID 9). 2360 bytes result sent to driver

18/04/21 06:23:09 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, localhost, executor driver, partition 14, ANY, 2274 bytes)

18/04/21 06:23:09 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 3641 ms on localhost (executor driver) (11/39)

18/04/21 06:23:09 INFO executor.Executor: Running task 14.0 in stage 0.0 (TID 14)

18/04/21 06:23:09 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:268435456+134217728

18/04/21 06:23:09 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:09 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:09 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:09 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 14 total executors!

18/04/21 06:23:09 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000010_10' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000010

18/04/21 06:23:09 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000010_10: Committed

18/04/21 06:23:09 INFO executor.Executor: Finished task 10.0 in stage 0.0 (TID 10). 2360 bytes result sent to driver

18/04/21 06:23:09 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, localhost, executor driver, partition 15, ANY, 2274 bytes)

18/04/21 06:23:09 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 3839 ms on localhost (executor driver) (12/39)

18/04/21 06:23:09 INFO executor.Executor: Running task 15.0 in stage 0.0 (TID 15)

18/04/21 06:23:09 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:402653184+134217728

18/04/21 06:23:09 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:09 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:10 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:10 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 15 total executors!

18/04/21 06:23:11 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:11 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 16 total executors!

18/04/21 06:23:11 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000012_12' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000012

18/04/21 06:23:11 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000012_12: Committed

18/04/21 06:23:11 INFO executor.Executor: Finished task 12.0 in stage 0.0 (TID 12). 2360 bytes result sent to driver

18/04/21 06:23:11 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 3696 ms on localhost (executor driver) (13/39)

18/04/21 06:23:11 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, localhost, executor driver, partition 16, ANY, 2274 bytes)

18/04/21 06:23:11 INFO executor.Executor: Running task 16.0 in stage 0.0 (TID 16)

18/04/21 06:23:11 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-09.csv:536870912+134217728

18/04/21 06:23:11 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:11 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:11 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000013_13' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000013

18/04/21 06:23:11 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000013_13: Committed

18/04/21 06:23:11 INFO executor.Executor: Finished task 13.0 in stage 0.0 (TID 13). 2360 bytes result sent to driver

18/04/21 06:23:11 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 0.0 (TID 17, localhost, executor driver, partition 17, ANY, 2274 bytes)

18/04/21 06:23:11 INFO executor.Executor: Running task 17.0 in stage 0.0 (TID 17)

18/04/21 06:23:11 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:671088640+117983715

18/04/21 06:23:11 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 0.0 (TID 13) in 3797 ms on localhost (executor driver) (14/39)

18/04/21 06:23:12 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:12 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:12 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:12 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 17 total executors!

18/04/21 06:23:13 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000014_14' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/_temporary/0/task_201804210622_0000_m_000014

18/04/21 06:23:13 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000014_14: Committed

18/04/21 06:23:13 INFO executor.Executor: Finished task 14.0 in stage 0.0 (TID 14). 2360 bytes result sent to driver

18/04/21 06:23:13 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 0.0 (TID 18, localhost, executor driver, partition 18, ANY, 2274 bytes)

18/04/21 06:23:13 INFO executor.Executor: Running task 18.0 in stage 0.0 (TID 18)

18/04/21 06:23:13 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 0.0 (TID 14) in 4103 ms on localhost (executor driver) (15/39)

18/04/21 06:23:13 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:0+134217728

18/04/21 06:23:13 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:13 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:13 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000015_15' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622_0000_m_000015

18/04/21 06:23:13 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000015_15: Committed

18/04/21 06:23:13 INFO executor.Executor: Finished task 15.0 in stage 0.0 (TID 15). 2360 bytes result sent to driver

18/04/21 06:23:13 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 0.0 (TID 19, localhost, executor driver, partition 19, ANY, 2274 bytes)

18/04/21 06:23:13 INFO executor.Executor: Running task 19.0 in stage 0.0 (TID 19)

18/04/21 06:23:13 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 3741 ms on localhost (executor driver) (16/39)

18/04/21 06:23:13 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:134217728+134217728

18/04/21 06:23:13 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:13 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup_temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:13 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:13 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 18 total executors!

18/04/21 06:23:14 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:14 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 19 total executors!

18/04/21 06:23:14 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000017_17' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622_0000_m_000017

18/04/21 06:23:14 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000017_17: Committed

18/04/21 06:23:14 INFO executor.Executor: Finished task 17.0 in stage 0.0 (TID 17). 2360 bytes result sent to driver

18/04/21 06:23:14 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 0.0 (TID 20, localhost, executor driver, partition 20, ANY, 2274 bytes)

18/04/21 06:23:14 INFO executor.Executor: Running task 20.0 in stage 0.0 (TID 20)

18/04/21 06:23:14 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:268435456+134217728

18/04/21 06:23:14 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 0.0 (TID 17) in 2814 ms on localhost (executor driver) (17/39)

18/04/21 06:23:14 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:14 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:15 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:15 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 20 total executors!

18/04/21 06:23:15 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000016_16' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000016

18/04/21 06:23:15 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000016_16: Committed

18/04/21 06:23:15 INFO executor.Executor: Finished task 16.0 in stage 0.0 (TID 16). 2360 bytes result sent to driver

18/04/21 06:23:15 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 0.0 (TID 21, localhost, executor driver, partition 21, ANY, 2274 bytes)

18/04/21 06:23:15 INFO executor.Executor: Running task 21.0 in stage 0.0 (TID 21)

18/04/21 06:23:15 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 3996 ms on localhost (executor driver) (18/39)

18/04/21 06:23:15 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:402653184+134217728

18/04/21 06:23:15 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:15 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:16 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:16 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 21 total executors!

18/04/21 06:23:16 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000018_18' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622_0000_m_000018

18/04/21 06:23:16 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000018_18: Committed

18/04/21 06:23:16 INFO executor.Executor: Finished task 18.0 in stage 0.0 (TID 18). 2360 bytes result sent to driver

18/04/21 06:23:16 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 0.0 (TID 22, localhost, executor driver, partition 22, ANY, 2274 bytes)

18/04/21 06:23:16 INFO executor.Executor: Running task 22.0 in stage 0.0 (TID 22)

18/04/21 06:23:16 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 0.0 (TID 18) in 3328 ms on localhost (executor driver) (19/39)

18/04/21 06:23:16 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:536870912+134217728

18/04/21 06:23:16 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:16 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:16 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:16 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000019_19' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622_0000_m_000019

18/04/21 06:23:16 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000019_19: Committed

18/04/21 06:23:16 INFO executor.Executor: Finished task 19.0 in stage 0.0 (TID 19). 2360 bytes result sent to driver

18/04/21 06:23:16 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 0.0 (TID 23, localhost, executor driver, partition 23, ANY, 2274 bytes)

18/04/21 06:23:16 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 0.0 (TID 19) in 3552 ms on localhost (executor driver) (20/39)

18/04/21 06:23:16 INFO executor.Executor: Running task 23.0 in stage 0.0 (TID 23)

18/04/21 06:23:16 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:671088640+134217728

18/04/21 06:23:16 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:16 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:16 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:18 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000020_20' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622_0000_m_000020

18/04/21 06:23:18 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000020_20: Committed

18/04/21 06:23:18 INFO executor.Executor: Finished task 20.0 in stage 0.0 (TID 20). 2360 bytes result sent to driver

18/04/21 06:23:18 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 0.0 (TID 24, localhost, executor driver, partition 24, ANY, 2274 bytes)

18/04/21 06:23:18 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 0.0 (TID 20) in 3343 ms on localhost (executor driver) (21/39)

18/04/21 06:23:18 INFO executor.Executor: Running task 24.0 in stage 0.0 (TID 24)

18/04/21 06:23:18 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:805306368+56688482

18/04/21 06:23:18 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:18 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:18 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:19 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000021_21' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622_0000_m_000021

18/04/21 06:23:19 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000021_21: Committed

18/04/21 06:23:19 INFO executor.Executor: Finished task 21.0 in stage 0.0 (TID 21). 2360 bytes result sent to driver

18/04/21 06:23:19 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 0.0 (TID 25, localhost, executor driver, partition 25, ANY, 2274 bytes)

18/04/21 06:23:19 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 0.0 (TID 21) in 4052 ms on localhost (executor driver) (22/39)

18/04/21 06:23:19 INFO executor.Executor: Running task 25.0 in stage 0.0 (TID 25)

18/04/21 06:23:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:0+134217728

18/04/21 06:23:19 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:19 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:19 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:19 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210622 0000 m 000022 22' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622 0000 m 000022

18/04/21 06:23:19 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622 0000 m 000022 22: Committed

18/04/21 06:23:19 INFO executor.Executor: Finished task 22.0 in stage 0.0 (TID 22). 2360 bytes result sent to driver

18/04/21 06:23:19 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 0.0 (TID 26, localhost, executor driver, partition 26, ANY, 2274 bytes)

18/04/21 06:23:19 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 0.0 (TID 22) in 3362 ms on localhost (executor driver) (23/39)

18/04/21 06:23:19 INFO executor.Executor: Running task 26.0 in stage 0.0 (TID 26)

18/04/21 06:23:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:134217728+134217728

18/04/21 06:23:19 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:19 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:19 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:20 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210622 0000 m 000024 24' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622 0000 m 000024

18/04/21 06:23:20 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622 0000 m 000024 24: Committed

18/04/21 06:23:20 INFO executor.Executor: Finished task 24.0 in stage 0.0 (TID 24). 2360 bytes result sent to driver

18/04/21 06:23:20 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 0.0 (TID 27, localhost, executor driver, partition 27, ANY, 2274 bytes)

18/04/21 06:23:20 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 0.0 (TID 24) in 2063 ms on localhost (executor driver) (24/39)

18/04/21 06:23:20 INFO executor.Executor: Running task 27.0 in stage 0.0 (TID 27)

18/04/21 06:23:20 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:268435456+134217728

18/04/21 06:23:20 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:20 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:20 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210622_0000_m_000023_23' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/_temporary/0/task_201804210622_0000_m_000023

18/04/21 06:23:20 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622_0000_m_000023_23: Committed

18/04/21 06:23:20 INFO executor.Executor: Finished task 23.0 in stage 0.0 (TID 23). 2360 bytes result sent to driver

18/04/21 06:23:20 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 0.0 (TID 28, localhost, executor driver, partition 28, ANY, 2274 bytes)

18/04/21 06:23:20 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 0.0 (TID 23) in 3474 ms on localhost (executor driver) (25/39)

18/04/21 06:23:20 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:20 INFO executor.Executor: Running task 28.0 in stage 0.0 (TID 28)

18/04/21 06:23:20 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:402653184+134217728

18/04/21 06:23:20 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:20 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:23 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210622_0000_m_000025_25' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/_temporary/0/task_201804210622_0000_m_000025

18/04/21 06:23:23 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622_0000_m_000025_25: Committed

18/04/21 06:23:23 INFO executor.Executor: Finished task 25.0 in stage 0.0 (TID 25). 2360 bytes result sent to driver

18/04/21 06:23:23 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 0.0 (TID 29, localhost, executor driver, partition 29, ANY, 2274 bytes)

18/04/21 06:23:23 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 0.0 (TID 25) in 3792 ms on localhost (executor driver) (26/39)

18/04/21 06:23:23 INFO executor.Executor: Running task 29.0 in stage 0.0 (TID 29)

18/04/21 06:23:23 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:536870912+134217728

18/04/21 06:23:23 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:23 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:23 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:23 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000027_27' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000027

18/04/21 06:23:23 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000027_27: Committed

18/04/21 06:23:23 INFO executor.Executor: Finished task 27.0 in stage 0.0 (TID 27). 2360 bytes result sent to driver

18/04/21 06:23:23 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 0.0 (TID 30, localhost, executor driver, partition 30, ANY, 2274 bytes)

18/04/21 06:23:23 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 0.0 (TID 27) in 3526 ms on localhost (executor driver) (27/39)

18/04/21 06:23:23 INFO executor.Executor: Running task 30.0 in stage 0.0 (TID 30)

18/04/21 06:23:23 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:671088640+134217728

18/04/21 06:23:23 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:23 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:23 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:24 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000028_28' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000028

18/04/21 06:23:24 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000028_28: Committed

18/04/21 06:23:24 INFO executor.Executor: Finished task 28.0 in stage 0.0 (TID 28). 2360 bytes result sent to driver

18/04/21 06:23:24 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 0.0 (TID 31, localhost, executor driver, partition 31, ANY, 2274 bytes)

18/04/21 06:23:24 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 0.0 (TID 28) in 4140 ms on localhost (executor driver) (28/39)

18/04/21 06:23:24 INFO executor.Executor: Running task 31.0 in stage 0.0 (TID 31)

18/04/21 06:23:24 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-11.csv:805306368+13877504

18/04/21 06:23:24 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:24 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:24 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:24 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000026_26' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000026

18/04/21 06:23:24 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000026_26: Committed

18/04/21 06:23:24 INFO executor.Executor: Finished task 26.0 in stage 0.0 (TID 26). 2360 bytes result sent to driver

18/04/21 06:23:24 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 0.0 (TID 32, localhost, executor driver, partition 32, ANY, 2274 bytes)

18/04/21 06:23:24 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 0.0 (TID 26) in 4648 ms on localhost (executor driver) (29/39)

18/04/21 06:23:24 INFO executor.Executor: Running task 32.0 in stage 0.0 (TID 32)

18/04/21 06:23:24 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-12.csv:0+134217728

18/04/21 06:23:24 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:24 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:24 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:25 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000031_31' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622_0000_m_000031

18/04/21 06:23:25 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000031_31: Committed

18/04/21 06:23:25 INFO executor.Executor: Finished task 31.0 in stage 0.0 (TID 31). 2360 bytes result sent to driver

18/04/21 06:23:25 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 0.0 (TID 33, localhost, executor driver, partition 33, ANY, 2274 bytes)

18/04/21 06:23:25 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 0.0 (TID 31) in 931 ms on localhost (executor driver) (30/39)

18/04/21 06:23:25 INFO executor.Executor: Running task 33.0 in stage 0.0 (TID 33)

18/04/21 06:23:25 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:134217728+134217728

18/04/21 06:23:25 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:25 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:25 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:26 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000029_29' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/ temporary/0/task_201804210622_0000_m_000029

18/04/21 06:23:26 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000029_29: Committed

18/04/21 06:23:26 INFO executor.Executor: Finished task 29.0 in stage 0.0 (TID 29). 2360 bytes result sent to driver

18/04/21 06:23:26 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 0.0 (TID 34, localhost, executor driver, partition 34, ANY, 2274 bytes)

18/04/21 06:23:26 INFO executor.Executor: Running task 34.0 in stage 0.0 (TID 34)

18/04/21 06:23:26 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:268435456+134217728

18/04/21 06:23:26 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 0.0 (TID 29) in 3496 ms on localhost (executor driver) (31/39)

18/04/21 06:23:26 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:26 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:26 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:26 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210622_0000_m_000030_30' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/_temporary/0/task_201804210622_0000_m_000030

18/04/21 06:23:26 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622_0000_m_000030_30: Committed

18/04/21 06:23:26 INFO executor.Executor: Finished task 30.0 in stage 0.0 (TID 30). 2360 bytes result sent to driver

18/04/21 06:23:26 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 0.0 (TID 35, localhost, executor driver, partition 35, ANY, 2274 bytes)

18/04/21 06:23:26 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 0.0 (TID 30) in 3132 ms on localhost (executor driver) (32/39)

18/04/21 06:23:26 INFO executor.Executor: Running task 35.0 in stage 0.0 (TID 35)

18/04/21 06:23:26 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:402653184+134217728

18/04/21 06:23:26 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:26 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:26 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:28 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210622_0000_m_000032_32' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter_By_RateCodeId_SparkRDD/_temporary/0/task_201804210622_0000_m_000032

18/04/21 06:23:28 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622_0000_m_000032_32: Committed

18/04/21 06:23:28 INFO executor.Executor: Finished task 32.0 in stage 0.0 (TID 32). 2360 bytes result sent to driver

18/04/21 06:23:28 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 0.0 (TID 36, localhost, executor driver, partition 36, ANY, 2274 bytes)

18/04/21 06:23:28 INFO executor.Executor: Running task 36.0 in stage 0.0 (TID 36)

18/04/21 06:23:28 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:536870912+134217728

18/04/21 06:23:28 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 0.0 (TID 32) in 3561 ms on localhost (executor driver) (33/39)

18/04/21 06:23:28 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:28 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:28 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:28 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000033_33' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000033

18/04/21 06:23:28 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000033_33: Committed

18/04/21 06:23:28 INFO executor.Executor: Finished task 33.0 in stage 0.0 (TID 33). 2360 bytes result sent to driver

18/04/21 06:23:28 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 0.0 (TID 37, localhost, executor driver, partition 37, ANY, 2274 bytes)

18/04/21 06:23:28 INFO executor.Executor: Running task 37.0 in stage 0.0 (TID 37)

18/04/21 06:23:28 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 0.0 (TID 33) in 3665 ms on localhost (executor driver) (34/39)

18/04/21 06:23:28 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:671088640+134217728

18/04/21 06:23:28 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:28 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:29 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:30 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000034_34' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000034

18/04/21 06:23:30 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000034_34: Committed

18/04/21 06:23:30 INFO executor.Executor: Finished task 34.0 in stage 0.0 (TID 34). 2360 bytes result sent to driver

18/04/21 06:23:30 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 0.0 (TID 38, localhost, executor driver, partition 38, ANY, 2274 bytes)

18/04/21 06:23:30 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 0.0 (TID 34) in 4148 ms on localhost (executor driver) (35/39)

18/04/21 06:23:30 INFO executor.Executor: Running task 38.0 in stage 0.0 (TID 38)

18/04/21 06:23:30 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow_tripdata_2017-12.csv:805306368+32783040

18/04/21 06:23:30 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/21 06:23:30 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/21 06:23:30 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:30 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000035_35' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000035

18/04/21 06:23:30 INFO mapred.SparkHadoopMapRedUtil: attempt_201804210622_0000_m_000035_35: Committed

18/04/21 06:23:30 INFO executor.Executor: Finished task 35.0 in stage 0.0 (TID 35). 2360 bytes result sent to driver

18/04/21 06:23:30 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 0.0 (TID 35) in 4183 ms on localhost (executor driver) (36/39)

18/04/21 06:23:31 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:31 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804210622_0000_m_000036_36' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task_201804210622_0000_m_000036

18/04/21 06:23:31 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622 0000 m 000036 36: Committed

18/04/21 06:23:31 INFO executor.Executor: Finished task 36.0 in stage 0.0 (TID 36). 2360 bytes result sent to driver

18/04/21 06:23:31 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 0.0 (TID 36) in 3268 ms on localhost (executor driver) (37/39)

18/04/21 06:23:31 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:31 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210622 0000 m 000038 38' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task 201804210622 0000 m 000038

18/04/21 06:23:31 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622 0000 m 000038 38: Committed

18/04/21 06:23:31 INFO executor.Executor: Finished task 38.0 in stage 0.0 (TID 38). 2360 bytes result sent to driver

18/04/21 06:23:31 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 0.0 (TID 38) in 687 ms on localhost (executor driver) (38/39)

18/04/21 06:23:31 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:31 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804210622 0000 m 000037 37' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/Filter By RateCodeId SparkRDD/ temporary/0/task 201804210622 0000 m 000037

18/04/21 06:23:31 INFO mapred.SparkHadoopMapRedUtil: attempt 201804210622 0000 m 000037 37: Committed

18/04/21 06:23:31 INFO executor.Executor: Finished task 37.0 in stage 0.0 (TID 37). 2360 bytes result sent to driver

18/04/21 06:23:31 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 0.0 (TID 37) in 2817 ms on localhost (executor driver) (39/39)

18/04/21 06:23:31 INFO scheduler.DAGScheduler: ResultStage 0 (saveAsTextFile at FilterByRateCodeIdRDD.java:42) finished in 35.378 s

18/04/21 06:23:31 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

18/04/21 06:23:31 INFO scheduler.DAGScheduler: Job 0 finished: saveAsTextFile at FilterByRateCodeIdRDD.java:42, took 35.501015 s

18/04/21 06:23:31 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/21 06:23:32 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.161:4040

18/04/21 06:23:32 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

18/04/21 06:23:32 INFO storage.MemoryStore: MemoryStore cleared

18/04/21 06:23:32 INFO storage.BlockManager: BlockManager stopped

18/04/21 06:23:32 INFO storage.BlockManagerMaster: BlockManagerMaster stopped

18/04/21 06:23:32 INFO scheduler.OutputCommitCoordinator\$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

18/04/21 06:23:32 INFO remote.RemoteActorRefProvider\$RemotingTerminator: Shutting down remote daemon.

18/04/21 06:23:32 INFO remote.RemoteActorRefProvider\$RemotingTerminator: Remote daemon shut down; proceeding with flushing remote transports.

18/04/21 06:23:32 INFO spark.SparkContext: Successfully stopped SparkContext

18/04/21 06:23:32 INFO util.ShutdownHookManager: Shutdown hook called

18/04/21 06:23:32 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-8c8ebda4-a11c-440d-841d-3edab7d3450a

18/04/21 06:23:32 INFO Remoting: Remoting shut down

18/04/21 06:23:32 INFO remote.RemoteActorRefProvider\$RemotingTerminator: Remoting shut down.

[root@ip-10-0-0-161 spark Assignment]#

3.groupedRecordsRDD

[root@ip-10-0-0-161 spark Assignment]# spark-submit --class spark.question3.groupedRecordsRDD --master yarn --deploy-mode client --name aws_spark --conf "spark.app.id=spark_aws_run" groupedRecordss.jar /user/root/spark_assignment/input_dataset/yellow_tripdata* /user/root/spark_assignment/output/grouped_Records_SparkRDD

18/04/23 18:39:55 INFO spark.SparkContext: Running Spark version 1.6.0

18/04/23 18:39:55 INFO spark.SecurityManager: Changing view acls to: root

18/04/23 18:39:55 INFO spark.SecurityManager: Changing modify acls to: root

18/04/23 18:39:55 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set(root)

18/04/23 18:39:56 INFO util.Utils: Successfully started service 'sparkDriver' on port 43340.

18/04/23 18:39:56 INFO slf4j.Slf4jLogger: Slf4jLogger started

18/04/23 18:39:56 INFO Remoting: Starting remoting

18/04/23 18:39:56 INFO Remoting: Remoting started; listening on addresses
:[akka.tcp://sparkDriverActorSystem@10.0.0.161:36927]

18/04/23 18:39:56 INFO Remoting: Remoting now listens on addresses:
[akka.tcp://sparkDriverActorSystem@10.0.0.161:36927]

18/04/23 18:39:56 INFO util.Utils: Successfully started service 'sparkDriverActorSystem' on port 36927.

18/04/23 18:39:56 INFO spark.SparkEnv: Registering MapOutputTracker

18/04/23 18:39:56 INFO spark.SparkEnv: Registering BlockManagerMaster

18/04/23 18:39:56 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-c7014ad0-28bd-4512-bfc2-0d176f0d4d5b

18/04/23 18:39:56 INFO storage.MemoryStore: MemoryStore started with capacity 530.0 MB

18/04/23 18:39:56 INFO spark.SparkEnv: Registering OutputCommitCoordinator

18/04/23 18:39:56 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.

18/04/23 18:39:56 INFO ui.SparkUI: Started SparkUI at http://10.0.0.161:4040

18/04/23 18:39:56 INFO spark.SparkContext: Added JAR file:/root/spark_Assignment/groupedRecordss.jar at
spark://10.0.0.161:43340/jars/groupedRecordss.jar with timestamp 1524508796979

18/04/23 18:39:57 INFO executor.Executor: Starting executor ID driver on host localhost

18/04/23 18:39:57 INFO util.Utils: Successfully started service
'org.apache.spark.network.netty.NettyBlockTransferService' on port 40937.

18/04/23 18:39:57 INFO netty.NettyBlockTransferService: Server created on 40937

18/04/23 18:39:57 INFO storage.BlockManager: external shuffle service port = 7337

18/04/23 18:39:57 INFO storage.BlockManagerMaster: Trying to register BlockManager

18/04/23 18:39:57 INFO storage.BlockManagerMasterEndpoint: Registering block manager localhost:40937 with
530.0 MB RAM, BlockManagerId(driver, localhost, 40937)

18/04/23 18:39:57 INFO storage.BlockManagerMaster: Registered BlockManager

18/04/23 18:39:58 INFO scheduler.EventLoggingListener: Logging events to hdfs://ip-10-0-0-
161.ec2.internal:8020/user/spark/applicationHistory/local-1524508797020

18/04/23 18:39:58 INFO spark.SparkContext: Registered listener
com.cloudera.spark.lineage.ClouderaNavigatorListener

18/04/23 18:39:58 INFO storage.MemoryStore: Block broadcast 0 stored as values in memory (estimated size
302.5 KB, free 529.7 MB)

18/04/23 18:39:58 INFO storage.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 26.0 KB, free 529.7 MB)

18/04/23 18:39:58 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on localhost:40937 (size: 26.0 KB, free: 530.0 MB)

18/04/23 18:39:58 INFO spark.SparkContext: Created broadcast 0 from textFile at groupedRecordsRDD.java:36

18/04/23 18:39:58 INFO mapred.FileInputFormat: Total input paths to process : 6

18/04/23 18:39:59 INFO spark.SparkContext: Starting job: sortByKey at groupedRecordsRDD.java:55

18/04/23 18:39:59 INFO scheduler.DAGScheduler: Registering RDD 3 (mapToPair at groupedRecordsRDD.java:47)

18/04/23 18:39:59 INFO scheduler.DAGScheduler: Got job 0 (sortByKey at groupedRecordsRDD.java:55) with 39 output partitions

18/04/23 18:39:59 INFO scheduler.DAGScheduler: Final stage: ResultStage 1 (sortByKey at groupedRecordsRDD.java:55)

18/04/23 18:39:59 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 0)

18/04/23 18:39:59 INFO scheduler.DAGScheduler: Missing parents: List(ShuffleMapStage 0)

18/04/23 18:39:59 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 0 (MapPartitionsRDD[3] at mapToPair at groupedRecordsRDD.java:47), which has no missing parents

18/04/23 18:39:59 INFO storage.MemoryStore: Block broadcast_1 stored as values in memory (estimated size 5.4 KB, free 529.7 MB)

18/04/23 18:39:59 INFO storage.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 2.9 KB, free 529.7 MB)

18/04/23 18:39:59 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on localhost:40937 (size: 2.9 KB, free: 530.0 MB)

18/04/23 18:39:59 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1004

18/04/23 18:39:59 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ShuffleMapStage 0 (MapPartitionsRDD[3] at mapToPair at groupedRecordsRDD.java:47) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/04/23 18:39:59 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 39 tasks

18/04/23 18:39:59 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, executor driver, partition 0, ANY, 2257 bytes)

18/04/23 18:39:59 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, localhost, executor driver, partition 1, ANY, 2257 bytes)

18/04/23 18:39:59 INFO executor.Executor: Running task 0.0 in stage 0.0 (TID 0)

18/04/23 18:39:59 INFO executor.Executor: Running task 1.0 in stage 0.0 (TID 1)

18/04/23 18:39:59 INFO executor.Executor: Fetching spark://10.0.0.161:43340/jars/groupedRecordss.jar with timestamp 1524508796979

18/04/23 18:39:59 INFO spark.ExecutorAllocationManager: New executor driver has registered (new total is 1)

18/04/23 18:39:59 INFO util.Utils: Fetching spark://10.0.0.161:43340/jars/groupedRecordss.jar to /tmp/spark-5af2f31d-30bf-46b3-b768-8b0160dadd71/userFiles-f99e87e8-4a30-435c-82da-0c9736ed111f/fetchFileTemp8122011153433450846.tmp

18/04/23 18:39:59 INFO executor.Executor: Adding file:/tmp/spark-5af2f31d-30bf-46b3-b768-8b0160dadd71/userFiles-f99e87e8-4a30-435c-82da-0c9736ed111f/groupedRecordss.jar to class loader

18/04/23 18:39:59 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow tripdata 2017-07.csv:134217728+134217728

18/04/23 18:39:59 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow tripdata 2017-07.csv:0+134217728

18/04/23 18:39:59 INFO Configuration.deprecation: mapred.tip.id is deprecated. Instead, use mapreduce.task.id

18/04/23 18:39:59 INFO Configuration.deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id

18/04/23 18:39:59 INFO Configuration.deprecation: mapred.task.is.map is deprecated. Instead, use mapreduce.task.ismap

18/04/23 18:39:59 INFO Configuration.deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.partition

18/04/23 18:39:59 INFO Configuration.deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id

18/04/23 18:40:00 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:00 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 2 total executors!

18/04/23 18:40:01 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:01 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 3 total executors!

18/04/23 18:40:02 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:02 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 4 total executors!

18/04/23 18:40:02 INFO executor.Executor: Finished task 1.0 in stage 0.0 (TID 1). 2219 bytes result sent to driver

18/04/23 18:40:02 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, localhost, executor driver, partition 2, ANY, 2257 bytes)

18/04/23 18:40:02 INFO executor.Executor: Running task 2.0 in stage 0.0 (TID 2)

18/04/23 18:40:02 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:268435456+134217728

18/04/23 18:40:02 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 0.0 (TID 1) in 3359 ms on localhost (executor driver) (1/39)

18/04/23 18:40:02 INFO executor.Executor: Finished task 0.0 in stage 0.0 (TID 0). 2274 bytes result sent to driver

18/04/23 18:40:02 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, localhost, executor driver, partition 3, ANY, 2257 bytes)

18/04/23 18:40:02 INFO executor.Executor: Running task 3.0 in stage 0.0 (TID 3)

18/04/23 18:40:02 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 3490 ms on localhost (executor driver) (2/39)

18/04/23 18:40:02 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:402653184+134217728

18/04/23 18:40:03 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:03 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 5 total executors!

18/04/23 18:40:04 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:04 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 6 total executors!

18/04/23 18:40:04 INFO executor.Executor: Finished task 2.0 in stage 0.0 (TID 2). 2274 bytes result sent to driver

18/04/23 18:40:04 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 0.0 (TID 4, localhost, executor driver, partition 4, ANY, 2257 bytes)

18/04/23 18:40:04 INFO executor.Executor: Running task 4.0 in stage 0.0 (TID 4)

18/04/23 18:40:04 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:536870912+134217728

18/04/23 18:40:04 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 2377 ms on localhost (executor driver) (3/39)

18/04/23 18:40:05 INFO executor.Executor: Finished task 3.0 in stage 0.0 (TID 3). 2274 bytes result sent to driver

18/04/23 18:40:05 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 0.0 (TID 5, localhost, executor driver, partition 5, ANY, 2257 bytes)

18/04/23 18:40:05 INFO executor.Executor: Running task 5.0 in stage 0.0 (TID 5)

18/04/23 18:40:05 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-07.csv:671088640+86047889

18/04/23 18:40:05 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 0.0 (TID 3) in 2360 ms on localhost (executor driver) (4/39)

18/04/23 18:40:05 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:05 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 7 total executors!

18/04/23 18:40:06 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:06 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 8 total executors!

18/04/23 18:40:06 INFO executor.Executor: Finished task 5.0 in stage 0.0 (TID 5). 2274 bytes result sent to driver

18/04/23 18:40:06 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 0.0 (TID 6, localhost, executor driver, partition 6, ANY, 2257 bytes)

18/04/23 18:40:06 INFO executor.Executor: Running task 6.0 in stage 0.0 (TID 6)

18/04/23 18:40:06 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:0+134217728

18/04/23 18:40:06 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 0.0 (TID 5) in 1326 ms on localhost (executor driver) (5/39)

18/04/23 18:40:06 INFO executor.Executor: Finished task 4.0 in stage 0.0 (TID 4). 2274 bytes result sent to driver

18/04/23 18:40:06 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, localhost, executor driver, partition 7, ANY, 2257 bytes)

18/04/23 18:40:06 INFO executor.Executor: Running task 7.0 in stage 0.0 (TID 7)

18/04/23 18:40:06 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:134217728+134217728

18/04/23 18:40:06 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 2092 ms on localhost (executor driver) (6/39)

18/04/23 18:40:07 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:07 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 9 total executors!

18/04/23 18:40:08 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:08 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 10 total executors!

18/04/23 18:40:08 INFO executor.Executor: Finished task 6.0 in stage 0.0 (TID 6). 2274 bytes result sent to driver

18/04/23 18:40:08 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, localhost, executor driver, partition 8, ANY, 2257 bytes)

18/04/23 18:40:08 INFO executor.Executor: Running task 8.0 in stage 0.0 (TID 8)

18/04/23 18:40:08 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 2174 ms on localhost (executor driver) (7/39)

18/04/23 18:40:08 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow tripdata 2017-08.csv:268435456+134217728

18/04/23 18:40:09 INFO executor.Executor: Finished task 7.0 in stage 0.0 (TID 7). 2274 bytes result sent to driver

18/04/23 18:40:09 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, localhost, executor driver, partition 9, ANY, 2257 bytes)

18/04/23 18:40:09 INFO executor.Executor: Running task 9.0 in stage 0.0 (TID 9)

18/04/23 18:40:09 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark assignment/input dataset/yellow tripdata 2017-08.csv:402653184+134217728

18/04/23 18:40:09 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 2154 ms on localhost (executor driver) (8/39)

18/04/23 18:40:09 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:09 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 11 total executors!

18/04/23 18:40:10 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:10 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 12 total executors!

18/04/23 18:40:10 INFO executor.Executor: Finished task 8.0 in stage 0.0 (TID 8). 2274 bytes result sent to driver

18/04/23 18:40:10 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, localhost, executor driver, partition 10, ANY, 2257 bytes)

18/04/23 18:40:10 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 2103 ms on localhost (executor driver) (9/39)

18/04/23 18:40:10 INFO executor.Executor: Running task 10.0 in stage 0.0 (TID 10)

18/04/23 18:40:10 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:536870912+134217728

18/04/23 18:40:11 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:11 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 13 total executors!

18/04/23 18:40:11 INFO executor.Executor: Finished task 9.0 in stage 0.0 (TID 9). 2274 bytes result sent to driver

18/04/23 18:40:11 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, localhost, executor driver, partition 11, ANY, 2257 bytes)

18/04/23 18:40:11 INFO executor.Executor: Running task 11.0 in stage 0.0 (TID 11)

18/04/23 18:40:11 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 2324 ms on localhost (executor driver) (10/39)

18/04/23 18:40:11 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-08.csv:671088640+71329760

18/04/23 18:40:12 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:12 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 14 total executors!

18/04/23 18:40:12 INFO executor.Executor: Finished task 11.0 in stage 0.0 (TID 11). 2274 bytes result sent to driver

18/04/23 18:40:12 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 0.0 (TID 12, localhost, executor driver, partition 12, ANY, 2257 bytes)

18/04/23 18:40:12 INFO executor.Executor: Running task 12.0 in stage 0.0 (TID 12)

18/04/23 18:40:12 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:0+134217728

18/04/23 18:40:12 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 0.0 (TID 11) in 1114 ms on localhost (executor driver) (11/39)

18/04/23 18:40:12 INFO executor.Executor: Finished task 10.0 in stage 0.0 (TID 10). 2274 bytes result sent to driver

18/04/23 18:40:12 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13, localhost, executor driver, partition 13, ANY, 2257 bytes)

18/04/23 18:40:12 INFO executor.Executor: Running task 13.0 in stage 0.0 (TID 13)

18/04/23 18:40:12 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 2290 ms on localhost (executor driver) (12/39)

18/04/23 18:40:12 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:134217728+134217728

18/04/23 18:40:13 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:13 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 15 total executors!

18/04/23 18:40:14 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:14 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 16 total executors!

18/04/23 18:40:14 INFO executor.Executor: Finished task 12.0 in stage 0.0 (TID 12). 2274 bytes result sent to driver

18/04/23 18:40:14 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, localhost, executor driver, partition 14, ANY, 2257 bytes)

18/04/23 18:40:14 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 2394 ms on localhost (executor driver) (13/39)

18/04/23 18:40:14 INFO executor.Executor: Running task 14.0 in stage 0.0 (TID 14)

18/04/23 18:40:14 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:268435456+134217728

18/04/23 18:40:15 INFO executor.Executor: Finished task 13.0 in stage 0.0 (TID 13). 2274 bytes result sent to driver

18/04/23 18:40:15 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, localhost, executor driver, partition 15, ANY, 2257 bytes)

18/04/23 18:40:15 INFO executor.Executor: Running task 15.0 in stage 0.0 (TID 15)

18/04/23 18:40:15 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 0.0 (TID 13) in 2278 ms on localhost (executor driver) (14/39)

18/04/23 18:40:15 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:402653184+134217728

18/04/23 18:40:15 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:15 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 17 total executors!

18/04/23 18:40:16 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:16 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 18 total executors!

18/04/23 18:40:17 INFO executor.Executor: Finished task 14.0 in stage 0.0 (TID 14). 2274 bytes result sent to driver

18/04/23 18:40:17 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, localhost, executor driver, partition 16, ANY, 2257 bytes)

18/04/23 18:40:17 INFO executor.Executor: Running task 16.0 in stage 0.0 (TID 16)

18/04/23 18:40:17 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 0.0 (TID 14) in 2121 ms on localhost (executor driver) (15/39)

18/04/23 18:40:17 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:536870912+134217728

18/04/23 18:40:17 INFO executor.Executor: Finished task 15.0 in stage 0.0 (TID 15). 2274 bytes result sent to driver

18/04/23 18:40:17 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 0.0 (TID 17, localhost, executor driver, partition 17, ANY, 2257 bytes)

18/04/23 18:40:17 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 2062 ms on localhost (executor driver) (16/39)

18/04/23 18:40:17 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:17 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 19 total executors!

18/04/23 18:40:17 INFO executor.Executor: Running task 17.0 in stage 0.0 (TID 17)

18/04/23 18:40:17 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-09.csv:671088640+117983715

18/04/23 18:40:18 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:18 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 20 total executors!

18/04/23 18:40:19 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:19 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to request 21 total executors!

18/04/23 18:40:19 INFO executor.Executor: Finished task 16.0 in stage 0.0 (TID 16). 2274 bytes result sent to driver

18/04/23 18:40:19 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 0.0 (TID 18, localhost, executor driver, partition 18, ANY, 2257 bytes)

18/04/23 18:40:19 INFO executor.Executor: Running task 18.0 in stage 0.0 (TID 18)

18/04/23 18:40:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:0+134217728

18/04/23 18:40:19 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 2372 ms on localhost (executor driver) (17/39)

18/04/23 18:40:19 INFO executor.Executor: Finished task 17.0 in stage 0.0 (TID 17). 2274 bytes result sent to driver

18/04/23 18:40:19 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 0.0 (TID 19, localhost, executor driver, partition 19, ANY, 2257 bytes)

18/04/23 18:40:19 INFO executor.Executor: Running task 19.0 in stage 0.0 (TID 19)

18/04/23 18:40:19 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:134217728+134217728

18/04/23 18:40:19 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 0.0 (TID 17) in 2230 ms on localhost (executor driver) (18/39)

18/04/23 18:40:21 INFO executor.Executor: Finished task 18.0 in stage 0.0 (TID 18). 2274 bytes result sent to driver

18/04/23 18:40:21 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 0.0 (TID 20, localhost, executor driver, partition 20, ANY, 2257 bytes)

18/04/23 18:40:21 INFO executor.Executor: Running task 20.0 in stage 0.0 (TID 20)

18/04/23 18:40:21 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:268435456+134217728

18/04/23 18:40:21 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 0.0 (TID 18) in 2190 ms on localhost (executor driver) (19/39)

18/04/23 18:40:21 INFO executor.Executor: Finished task 19.0 in stage 0.0 (TID 19). 2274 bytes result sent to driver

18/04/23 18:40:21 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 0.0 (TID 21, localhost, executor driver, partition 21, ANY, 2257 bytes)

18/04/23 18:40:21 INFO executor.Executor: Running task 21.0 in stage 0.0 (TID 21)

18/04/23 18:40:21 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:402653184+134217728

18/04/23 18:40:21 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 0.0 (TID 19) in 2197 ms on localhost (executor driver) (20/39)

18/04/23 18:40:21 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:23 INFO executor.Executor: Finished task 21.0 in stage 0.0 (TID 21). 2274 bytes result sent to driver

18/04/23 18:40:23 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 0.0 (TID 22, localhost, executor driver, partition 22, ANY, 2257 bytes)

18/04/23 18:40:23 INFO executor.Executor: Running task 22.0 in stage 0.0 (TID 22)

18/04/23 18:40:23 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 0.0 (TID 21) in 2055 ms on localhost (executor driver) (21/39)

18/04/23 18:40:23 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:536870912+134217728

18/04/23 18:40:23 INFO executor.Executor: Finished task 20.0 in stage 0.0 (TID 20). 2274 bytes result sent to driver

18/04/23 18:40:23 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 0.0 (TID 23, localhost, executor driver, partition 23, ANY, 2257 bytes)

18/04/23 18:40:23 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 0.0 (TID 20) in 2106 ms on localhost (executor driver) (22/39)

18/04/23 18:40:23 INFO executor.Executor: Running task 23.0 in stage 0.0 (TID 23)

18/04/23 18:40:23 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:671088640+134217728

18/04/23 18:40:23 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:25 INFO executor.Executor: Finished task 22.0 in stage 0.0 (TID 22). 2274 bytes result sent to driver

18/04/23 18:40:25 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 0.0 (TID 24, localhost, executor driver, partition 24, ANY, 2257 bytes)

18/04/23 18:40:25 INFO executor.Executor: Running task 24.0 in stage 0.0 (TID 24)

18/04/23 18:40:25 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 0.0 (TID 22) in 2062 ms on localhost (executor driver) (23/39)

18/04/23 18:40:25 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-10.csv:805306368+56688482

18/04/23 18:40:25 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:25 INFO executor.Executor: Finished task 23.0 in stage 0.0 (TID 23). 2274 bytes result sent to driver

18/04/23 18:40:25 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 0.0 (TID 25, localhost, executor driver, partition 25, ANY, 2257 bytes)

18/04/23 18:40:25 INFO executor.Executor: Running task 25.0 in stage 0.0 (TID 25)

18/04/23 18:40:25 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 0.0 (TID 23) in 2055 ms on localhost (executor driver) (24/39)

18/04/23 18:40:25 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:0+134217728

18/04/23 18:40:25 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:26 INFO executor.Executor: Finished task 24.0 in stage 0.0 (TID 24). 2274 bytes result sent to driver

18/04/23 18:40:26 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 0.0 (TID 26, localhost, executor driver, partition 26, ANY, 2257 bytes)

18/04/23 18:40:26 INFO executor.Executor: Running task 26.0 in stage 0.0 (TID 26)

18/04/23 18:40:26 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 0.0 (TID 24) in 949 ms on localhost (executor driver) (25/39)

18/04/23 18:40:26 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:134217728+134217728

18/04/23 18:40:26 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:27 INFO executor.Executor: Finished task 25.0 in stage 0.0 (TID 25). 2274 bytes result sent to driver

18/04/23 18:40:27 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 0.0 (TID 27, localhost, executor driver, partition 27, ANY, 2257 bytes)

18/04/23 18:40:27 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 0.0 (TID 25) in 2203 ms on localhost (executor driver) (26/39)

18/04/23 18:40:27 INFO executor.Executor: Running task 27.0 in stage 0.0 (TID 27)

18/04/23 18:40:27 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:268435456+134217728

18/04/23 18:40:28 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:28 INFO executor.Executor: Finished task 26.0 in stage 0.0 (TID 26). 2274 bytes result sent to driver

18/04/23 18:40:28 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 0.0 (TID 28, localhost, executor driver, partition 28, ANY, 2257 bytes)

18/04/23 18:40:28 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 0.0 (TID 26) in 1990 ms on localhost (executor driver) (27/39)

18/04/23 18:40:28 INFO executor.Executor: Running task 28.0 in stage 0.0 (TID 28)

18/04/23 18:40:28 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:402653184+134217728

18/04/23 18:40:28 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:30 INFO executor.Executor: Finished task 27.0 in stage 0.0 (TID 27). 2274 bytes result sent to driver

18/04/23 18:40:30 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 0.0 (TID 29, localhost, executor driver, partition 29, ANY, 2257 bytes)

18/04/23 18:40:30 INFO executor.Executor: Running task 29.0 in stage 0.0 (TID 29)

18/04/23 18:40:30 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:536870912+134217728

18/04/23 18:40:30 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 0.0 (TID 27) in 2052 ms on localhost (executor driver) (28/39)

18/04/23 18:40:30 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:30 INFO executor.Executor: Finished task 28.0 in stage 0.0 (TID 28). 2274 bytes result sent to driver

18/04/23 18:40:30 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 0.0 (TID 28) in 2043 ms on localhost (executor driver) (29/39)

18/04/23 18:40:30 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 0.0 (TID 30, localhost, executor driver, partition 30, ANY, 2257 bytes)

18/04/23 18:40:30 INFO executor.Executor: Running task 30.0 in stage 0.0 (TID 30)

18/04/23 18:40:30 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:671088640+134217728

18/04/23 18:40:30 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:32 INFO executor.Executor: Finished task 29.0 in stage 0.0 (TID 29). 2274 bytes result sent to driver

18/04/23 18:40:32 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 0.0 (TID 31, localhost, executor driver, partition 31, ANY, 2257 bytes)

18/04/23 18:40:32 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 0.0 (TID 29) in 2283 ms on localhost (executor driver) (30/39)

18/04/23 18:40:32 INFO executor.Executor: Running task 31.0 in stage 0.0 (TID 31)

18/04/23 18:40:32 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-11.csv:805306368+13877504

18/04/23 18:40:32 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:32 INFO executor.Executor: Finished task 31.0 in stage 0.0 (TID 31). 2274 bytes result sent to driver

18/04/23 18:40:32 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 0.0 (TID 32, localhost, executor driver, partition 32, ANY, 2257 bytes)

18/04/23 18:40:32 INFO executor.Executor: Running task 32.0 in stage 0.0 (TID 32)

18/04/23 18:40:32 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 0.0 (TID 31) in 255 ms on localhost (executor driver) (31/39)

18/04/23 18:40:32 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:0+134217728

18/04/23 18:40:32 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:33 INFO executor.Executor: Finished task 30.0 in stage 0.0 (TID 30). 2274 bytes result sent to driver

18/04/23 18:40:33 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 0.0 (TID 33, localhost, executor driver, partition 33, ANY, 2257 bytes)

18/04/23 18:40:33 INFO executor.Executor: Running task 33.0 in stage 0.0 (TID 33)

18/04/23 18:40:33 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 0.0 (TID 30) in 2304 ms on localhost (executor driver) (32/39)

18/04/23 18:40:33 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:134217728+134217728

18/04/23 18:40:33 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:34 INFO executor.Executor: Finished task 32.0 in stage 0.0 (TID 32). 2274 bytes result sent to driver

18/04/23 18:40:34 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 0.0 (TID 34, localhost, executor driver, partition 34, ANY, 2257 bytes)

18/04/23 18:40:34 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 0.0 (TID 32) in 2448 ms on localhost (executor driver) (33/39)

18/04/23 18:40:34 INFO executor.Executor: Running task 34.0 in stage 0.0 (TID 34)

18/04/23 18:40:34 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:268435456+134217728

18/04/23 18:40:35 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:35 INFO executor.Executor: Finished task 33.0 in stage 0.0 (TID 33). 2274 bytes result sent to driver

18/04/23 18:40:35 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 0.0 (TID 35, localhost, executor driver, partition 35, ANY, 2257 bytes)

18/04/23 18:40:35 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 0.0 (TID 33) in 2812 ms on localhost (executor driver) (34/39)

18/04/23 18:40:35 INFO executor.Executor: Running task 35.0 in stage 0.0 (TID 35)

18/04/23 18:40:35 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:402653184+134217728

18/04/23 18:40:35 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:37 INFO executor.Executor: Finished task 34.0 in stage 0.0 (TID 34). 2274 bytes result sent to driver

18/04/23 18:40:37 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 0.0 (TID 36, localhost, executor driver, partition 36, ANY, 2257 bytes)

18/04/23 18:40:37 INFO executor.Executor: Running task 36.0 in stage 0.0 (TID 36)

18/04/23 18:40:37 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:536870912+134217728

18/04/23 18:40:37 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 0.0 (TID 34) in 2477 ms on localhost (executor driver) (35/39)

18/04/23 18:40:37 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:38 INFO executor.Executor: Finished task 35.0 in stage 0.0 (TID 35). 2274 bytes result sent to driver

18/04/23 18:40:38 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 0.0 (TID 37, localhost, executor driver, partition 37, ANY, 2257 bytes)

18/04/23 18:40:38 INFO executor.Executor: Running task 37.0 in stage 0.0 (TID 37)

18/04/23 18:40:38 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 0.0 (TID 35) in 2182 ms on localhost (executor driver) (36/39)

18/04/23 18:40:38 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:671088640+134217728

18/04/23 18:40:38 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:39 INFO executor.Executor: Finished task 36.0 in stage 0.0 (TID 36). 2274 bytes result sent to driver

18/04/23 18:40:39 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 0.0 (TID 38, localhost, executor driver, partition 38, ANY, 2257 bytes)

18/04/23 18:40:39 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 0.0 (TID 36) in 2100 ms on localhost (executor driver) (37/39)

18/04/23 18:40:39 INFO executor.Executor: Running task 38.0 in stage 0.0 (TID 38)

18/04/23 18:40:39 INFO rdd.HadoopRDD: Input split: hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/input_dataset/yellow_tripdata_2017-12.csv:805306368+32783040

18/04/23 18:40:39 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:40 INFO executor.Executor: Finished task 37.0 in stage 0.0 (TID 37). 2274 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 0.0 (TID 37) in 2029 ms on localhost (executor driver) (38/39)

18/04/23 18:40:40 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:40 INFO executor.Executor: Finished task 38.0 in stage 0.0 (TID 38). 2274 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 0.0 (TID 38) in 625 ms on localhost (executor driver) (39/39)

18/04/23 18:40:40 INFO scheduler.DAGScheduler: ShuffleMapStage 0 (mapToPair at groupedRecordsRDD.java:47) finished in 41.013 s

18/04/23 18:40:40 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

18/04/23 18:40:40 INFO scheduler.DAGScheduler: looking for newly runnable stages

18/04/23 18:40:40 INFO scheduler.DAGScheduler: running: Set()

18/04/23 18:40:40 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 1)

18/04/23 18:40:40 INFO scheduler.DAGScheduler: failed: Set()

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[7] at sortByKey at groupedRecordsRDD.java:55), which has no missing parents

18/04/23 18:40:40 INFO storage.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 4.6 KB, free 529.7 MB)

18/04/23 18:40:40 INFO storage.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 2.5 KB, free 529.7 MB)

18/04/23 18:40:40 INFO storage.BlockManagerInfo: Added broadcast 2 piece0 in memory on localhost:40937 (size: 2.5 KB, free: 530.0 MB)

18/04/23 18:40:40 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1004

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ResultStage 1 (MapPartitionsRDD[7] at sortByKey at groupedRecordsRDD.java:55) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/04/23 18:40:40 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 39 tasks

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 39, localhost, executor driver, partition 1, NODE_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 40, localhost, executor driver, partition 2, NODE_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 1.0 in stage 1.0 (TID 39)

18/04/23 18:40:40 INFO executor.Executor: Running task 2.0 in stage 1.0 (TID 40)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 4 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 4 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 2.0 in stage 1.0 (TID 40). 1158 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 41, localhost, executor driver, partition 3, NODE_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Finished task 1.0 in stage 1.0 (TID 39). 1158 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Running task 3.0 in stage 1.0 (TID 41)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 1.0 (TID 40) in 55 ms on localhost (executor driver) (1/39)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 1.0 (TID 42, localhost, executor driver, partition 4, NODE_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 4.0 in stage 1.0 (TID 42)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 1.0 (TID 39) in 65 ms on localhost (executor driver) (2/39)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 3.0 in stage 1.0 (TID 41). 1157 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 4.0 in stage 1.0 (TID 42). 1157 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 43, localhost, executor driver, partition 0, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 1.0 (TID 44, localhost, executor driver, partition 5, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 5.0 in stage 1.0 (TID 44)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 1.0 (TID 41) in 27 ms on localhost (executor driver) (3/39)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 0.0 in stage 1.0 (TID 43)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 5.0 in stage 1.0 (TID 44). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 0.0 in stage 1.0 (TID 43). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 1.0 (TID 42) in 21 ms on localhost (executor driver) (4/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 1.0 (TID 45, localhost, executor driver, partition 6, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 1.0 (TID 46, localhost, executor driver, partition 7, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 6.0 in stage 1.0 (TID 45)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 1.0 (TID 44) in 14 ms on localhost (executor driver) (5/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 43) in 15 ms on localhost (executor driver) (6/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 6.0 in stage 1.0 (TID 45). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 1.0 (TID 47, localhost, executor driver, partition 8, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 1.0 (TID 45) in 10 ms on localhost (executor driver) (7/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 8.0 in stage 1.0 (TID 47)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 7.0 in stage 1.0 (TID 46)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 8.0 in stage 1.0 (TID 47). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 1.0 (TID 48, localhost, executor driver, partition 9, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 1.0 (TID 47) in 11 ms on localhost (executor driver) (8/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 7.0 in stage 1.0 (TID 46). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Running task 9.0 in stage 1.0 (TID 48)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 1.0 (TID 49, localhost, executor driver, partition 10, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 1.0 (TID 46) in 24 ms on localhost (executor driver) (9/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 10.0 in stage 1.0 (TID 49)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 9.0 in stage 1.0 (TID 48). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 10.0 in stage 1.0 (TID 49). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 1.0 (TID 50, localhost, executor driver, partition 11, PROCESS LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 1.0 (TID 51, localhost, executor driver, partition 12, PROCESS LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 11.0 in stage 1.0 (TID 50)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 12.0 in stage 1.0 (TID 51)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 11.0 in stage 1.0 (TID 50). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 1.0 (TID 48) in 22 ms on localhost (executor driver) (10/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 12.0 in stage 1.0 (TID 51). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 1.0 (TID 52, localhost, executor driver, partition 13, PROCESS LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 1.0 (TID 53, localhost, executor driver, partition 14, PROCESS LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 13.0 in stage 1.0 (TID 52)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 1.0 (TID 49) in 20 ms on localhost (executor driver) (11/39)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 1.0 (TID 50) in 14 ms on localhost (executor driver) (12/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 1.0 (TID 51) in 15 ms on localhost (executor driver) (13/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 13.0 in stage 1.0 (TID 52). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 1.0 (TID 54, localhost, executor driver, partition 15, PROCESS LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 15.0 in stage 1.0 (TID 54)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 14.0 in stage 1.0 (TID 53)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 1.0 (TID 52) in 11 ms on localhost (executor driver) (14/39)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 15.0 in stage 1.0 (TID 54). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 1.0 (TID 55, localhost, executor driver, partition 16, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 16.0 in stage 1.0 (TID 55)

18/04/23 18:40:40 INFO executor.Executor: Finished task 14.0 in stage 1.0 (TID 53). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 1.0 (TID 54) in 11 ms on localhost (executor driver) (15/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 1.0 (TID 56, localhost, executor driver, partition 17, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Finished task 16.0 in stage 1.0 (TID 55). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 1.0 (TID 57, localhost, executor driver, partition 18, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 18.0 in stage 1.0 (TID 57)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 1.0 (TID 53) in 23 ms on localhost (executor driver) (16/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 1.0 (TID 55) in 11 ms on localhost (executor driver) (17/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 17.0 in stage 1.0 (TID 56)

18/04/23 18:40:40 INFO executor.Executor: Finished task 18.0 in stage 1.0 (TID 57). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 1.0 (TID 58, localhost, executor driver, partition 19, PROCESS LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 19.0 in stage 1.0 (TID 58)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 1.0 (TID 57) in 10 ms on localhost (executor driver) (18/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 17.0 in stage 1.0 (TID 56). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 19.0 in stage 1.0 (TID 58). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 1.0 (TID 59, localhost, executor driver, partition 20, PROCESS LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 1.0 (TID 60, localhost, executor driver, partition 21, PROCESS LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 20.0 in stage 1.0 (TID 59)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 1.0 (TID 56) in 20 ms on localhost (executor driver) (19/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 20.0 in stage 1.0 (TID 59). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Running task 21.0 in stage 1.0 (TID 60)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 1.0 (TID 61, localhost, executor driver, partition 22, PROCESS LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 1.0 (TID 58) in 17 ms on localhost (executor driver) (20/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 1.0 (TID 59) in 11 ms on localhost (executor driver) (21/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 22.0 in stage 1.0 (TID 61)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 21.0 in stage 1.0 (TID 60). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 1.0 (TID 62, localhost, executor driver, partition 23, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 23.0 in stage 1.0 (TID 62)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 1.0 (TID 60) in 16 ms on localhost (executor driver) (22/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 22.0 in stage 1.0 (TID 61). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 1.0 (TID 63, localhost, executor driver, partition 24, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Finished task 23.0 in stage 1.0 (TID 62). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 1.0 (TID 61) in 14 ms on localhost (executor driver) (23/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 24.0 in stage 1.0 (TID 63)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 24.0 in stage 1.0 (TID 63). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 1.0 (TID 64, localhost, executor driver, partition 25, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 1.0 (TID 65, localhost, executor driver, partition 26, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 1.0 (TID 63) in 14 ms on localhost (executor driver) (24/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 1.0 (TID 62) in 18 ms on localhost (executor driver) (25/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 25.0 in stage 1.0 (TID 64)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 26.0 in stage 1.0 (TID 65)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 25.0 in stage 1.0 (TID 64). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 1.0 (TID 66, localhost, executor driver, partition 27, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 27.0 in stage 1.0 (TID 66)

18/04/23 18:40:40 INFO executor.Executor: Finished task 26.0 in stage 1.0 (TID 65). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 1.0 (TID 67, localhost, executor driver, partition 28, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 28.0 in stage 1.0 (TID 67)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 1.0 (TID 64) in 22 ms on localhost (executor driver) (26/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 27.0 in stage 1.0 (TID 66). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 1.0 (TID 68, localhost, executor driver, partition 29, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 29.0 in stage 1.0 (TID 68)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 1.0 (TID 65) in 17 ms on localhost (executor driver) (27/39)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 1.0 (TID 66) in 9 ms on localhost (executor driver) (28/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 28.0 in stage 1.0 (TID 67). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 29.0 in stage 1.0 (TID 68). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 1.0 (TID 69, localhost, executor driver, partition 30, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 1.0 (TID 70, localhost, executor driver, partition 31, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 1.0 (TID 67) in 14 ms on localhost (executor driver) (29/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 1.0 (TID 68) in 10 ms on localhost (executor driver) (30/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 30.0 in stage 1.0 (TID 69)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 31.0 in stage 1.0 (TID 70)

18/04/23 18:40:40 INFO executor.Executor: Finished task 30.0 in stage 1.0 (TID 69). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 1.0 (TID 71, localhost, executor driver, partition 32, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 32.0 in stage 1.0 (TID 71)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 31.0 in stage 1.0 (TID 70). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 32.0 in stage 1.0 (TID 71). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 1.0 (TID 72, localhost, executor driver, partition 33, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 1.0 (TID 73, localhost, executor driver, partition 34, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 1.0 (TID 70) in 19 ms on localhost (executor driver) (31/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 1.0 (TID 71) in 11 ms on localhost (executor driver) (32/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 33.0 in stage 1.0 (TID 72)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 1.0 (TID 69) in 26 ms on localhost (executor driver) (33/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 33.0 in stage 1.0 (TID 72). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 1.0 (TID 74, localhost, executor driver, partition 35, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 35.0 in stage 1.0 (TID 74)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 35.0 in stage 1.0 (TID 74). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 1.0 (TID 75, localhost, executor driver, partition 36, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 36.0 in stage 1.0 (TID 75)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 1.0 (TID 72) in 24 ms on localhost (executor driver) (34/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 1.0 (TID 74) in 12 ms on localhost (executor driver) (35/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 36.0 in stage 1.0 (TID 75). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 1.0 (TID 76, localhost, executor driver, partition 37, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 37.0 in stage 1.0 (TID 76)

18/04/23 18:40:40 INFO executor.Executor: Running task 34.0 in stage 1.0 (TID 73)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 1.0 (TID 75) in 8 ms on localhost (executor driver) (36/39)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 37.0 in stage 1.0 (TID 76). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 34.0 in stage 1.0 (TID 73). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 1.0 (TID 77, localhost, executor driver, partition 38, PROCESS_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 38.0 in stage 1.0 (TID 77)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 1.0 (TID 76) in 9 ms on localhost (executor driver) (37/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 1.0 (TID 73) in 34 ms on localhost (executor driver) (38/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 38.0 in stage 1.0 (TID 77). 1153 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 1.0 (TID 77) in 8 ms on localhost (executor driver) (39/39)

18/04/23 18:40:40 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool

18/04/23 18:40:40 INFO scheduler.DAGScheduler: ResultStage 1 (sortByKey at groupedRecordsRDD.java:55) finished in 0.277 s

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Job 0 finished: sortByKey at groupedRecordsRDD.java:55, took 41.417362 s

18/04/23 18:40:40 INFO spark.SparkContext: Starting job: collect at groupedRecordsRDD.java:61

18/04/23 18:40:40 INFO spark.MapOutputTrackerMaster: Size of output statuses for shuffle 0 is 196 bytes

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Registering RDD 5 (mapToPair at groupedRecordsRDD.java:55)

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Got job 1 (collect at groupedRecordsRDD.java:61) with 5 output partitions

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Final stage: ResultStage 4 (collect at groupedRecordsRDD.java:61)

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 3)

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Missing parents: List(ShuffleMapStage 3)

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 3 (MapPartitionsRDD[5] at mapToPair at groupedRecordsRDD.java:55), which has no missing parents

18/04/23 18:40:40 INFO storage.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 4.8 KB, free 529.7 MB)

18/04/23 18:40:40 INFO storage.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 2.6 KB, free 529.7 MB)

18/04/23 18:40:40 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on localhost:40937 (size: 2.6 KB, free: 530.0 MB)

18/04/23 18:40:40 INFO spark.SparkContext: Created broadcast 3 from broadcast at DAGScheduler.scala:1004

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Submitting 39 missing tasks from ShuffleMapStage 3 (MapPartitionsRDD[5] at mapToPair at groupedRecordsRDD.java:55) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

18/04/23 18:40:40 INFO scheduler.TaskSchedulerImpl: Adding task set 3.0 with 39 tasks

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 3.0 (TID 78, localhost, executor driver, partition 1, NODE_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 3.0 (TID 79, localhost, executor driver, partition 2, NODE_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 2.0 in stage 3.0 (TID 79)

18/04/23 18:40:40 INFO executor.Executor: Running task 1.0 in stage 3.0 (TID 78)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 1.0 in stage 3.0 (TID 78). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 3.0 (TID 80, localhost, executor driver, partition 3, NODE_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 3.0 (TID 78) in 49 ms on localhost (executor driver) (1/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 3.0 in stage 3.0 (TID 80)

18/04/23 18:40:40 INFO executor.Executor: Finished task 2.0 in stage 3.0 (TID 79). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 3.0 (TID 81, localhost, executor driver, partition 4, NODE LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 3.0 (TID 79) in 68 ms on localhost (executor driver) (2/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 4.0 in stage 3.0 (TID 81)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 39 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 3.0 in stage 3.0 (TID 80). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 3.0 (TID 82, localhost, executor driver, partition 0, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 3.0 (TID 80) in 40 ms on localhost (executor driver) (3/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 0.0 in stage 3.0 (TID 82)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 0.0 in stage 3.0 (TID 82). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 3.0 (TID 83, localhost, executor driver, partition 5, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 3.0 (TID 82) in 19 ms on localhost (executor driver) (4/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 5.0 in stage 3.0 (TID 83)

18/04/23 18:40:40 INFO executor.Executor: Finished task 4.0 in stage 3.0 (TID 81). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 3.0 (TID 84, localhost, executor driver, partition 6, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 3.0 (TID 81) in 51 ms on localhost (executor driver) (5/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 6.0 in stage 3.0 (TID 84)

18/04/23 18:40:40 INFO executor.Executor: Finished task 5.0 in stage 3.0 (TID 83). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 3.0 (TID 85, localhost, executor driver, partition 7, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 3.0 (TID 83) in 23 ms on localhost (executor driver) (6/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 7.0 in stage 3.0 (TID 85)

18/04/23 18:40:40 INFO executor.Executor: Finished task 6.0 in stage 3.0 (TID 84). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 3.0 (TID 86, localhost, executor driver, partition 8, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 8.0 in stage 3.0 (TID 86)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 7.0 in stage 3.0 (TID 85). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 3.0 (TID 84) in 23 ms on localhost (executor driver) (7/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 3.0 (TID 87, localhost, executor driver, partition 9, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO executor.Executor: Finished task 8.0 in stage 3.0 (TID 86). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Running task 9.0 in stage 3.0 (TID 87)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 3.0 (TID 88, localhost, executor driver, partition 10, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 3.0 (TID 85) in 18 ms on localhost (executor driver) (8/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 3.0 (TID 86) in 14 ms on localhost (executor driver) (9/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 10.0 in stage 3.0 (TID 88)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 9.0 in stage 3.0 (TID 87). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 3.0 (TID 89, localhost, executor driver, partition 11, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO executor.Executor: Finished task 10.0 in stage 3.0 (TID 88). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 3.0 (TID 90, localhost, executor driver, partition 12, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 3.0 (TID 87) in 21 ms on localhost (executor driver) (10/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 3.0 (TID 88) in 20 ms on localhost (executor driver) (11/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 11.0 in stage 3.0 (TID 89)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 12.0 in stage 3.0 (TID 90)

18/04/23 18:40:40 INFO executor.Executor: Finished task 11.0 in stage 3.0 (TID 89). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 3.0 (TID 91, localhost, executor driver, partition 13, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 3.0 (TID 89) in 15 ms on localhost (executor driver) (12/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 13.0 in stage 3.0 (TID 91)

18/04/23 18:40:40 INFO executor.Executor: Finished task 12.0 in stage 3.0 (TID 90). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 13.0 in stage 3.0 (TID 91). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 3.0 (TID 92, localhost, executor driver, partition 14, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 3.0 (TID 93, localhost, executor driver, partition 15, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 3.0 (TID 90) in 26 ms on localhost (executor driver) (13/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 3.0 (TID 91) in 16 ms on localhost (executor driver) (14/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 15.0 in stage 3.0 (TID 93)

18/04/23 18:40:40 INFO executor.Executor: Running task 14.0 in stage 3.0 (TID 92)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 15.0 in stage 3.0 (TID 93). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 14.0 in stage 3.0 (TID 92). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 3.0 (TID 94, localhost, executor driver, partition 16, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 3.0 (TID 95, localhost, executor driver, partition 17, PROCESS LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 3.0 (TID 93) in 21 ms on localhost (executor driver) (15/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 3.0 (TID 92) in 22 ms on localhost (executor driver) (16/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 17.0 in stage 3.0 (TID 95)

18/04/23 18:40:40 INFO executor.Executor: Running task 16.0 in stage 3.0 (TID 94)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 17.0 in stage 3.0 (TID 95). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 16.0 in stage 3.0 (TID 94). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 3.0 (TID 96, localhost, executor driver, partition 18, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 3.0 (TID 97, localhost, executor driver, partition 19, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 3.0 (TID 95) in 14 ms on localhost (executor driver) (17/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 18.0 in stage 3.0 (TID 96)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 3.0 (TID 94) in 16 ms on localhost (executor driver) (18/39)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 18.0 in stage 3.0 (TID 96). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 3.0 (TID 98, localhost, executor driver, partition 20, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 3.0 (TID 96) in 13 ms on localhost (executor driver) (19/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 19.0 in stage 3.0 (TID 97)

18/04/23 18:40:40 INFO executor.Executor: Running task 20.0 in stage 3.0 (TID 98)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 19.0 in stage 3.0 (TID 97). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 20.0 in stage 3.0 (TID 98). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 3.0 (TID 99, localhost, executor driver, partition 21, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 3.0 (TID 100, localhost, executor driver, partition 22, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 3.0 (TID 97) in 26 ms on localhost (executor driver) (20/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 3.0 (TID 98) in 16 ms on localhost (executor driver) (21/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 22.0 in stage 3.0 (TID 100)

18/04/23 18:40:40 INFO executor.Executor: Running task 21.0 in stage 3.0 (TID 99)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 22.0 in stage 3.0 (TID 100). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 21.0 in stage 3.0 (TID 99). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 3.0 (TID 101, localhost, executor driver, partition 23, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 3.0 (TID 102, localhost, executor driver, partition 24, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 3.0 (TID 100) in 16 ms on localhost (executor driver) (22/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 3.0 (TID 99) in 17 ms on localhost (executor driver) (23/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 24.0 in stage 3.0 (TID 102)

18/04/23 18:40:40 INFO executor.Executor: Running task 23.0 in stage 3.0 (TID 101)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 24.0 in stage 3.0 (TID 102). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 3.0 (TID 103, localhost, executor driver, partition 25, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 3.0 (TID 102) in 14 ms on localhost (executor driver) (24/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 25.0 in stage 3.0 (TID 103)

18/04/23 18:40:40 INFO executor.Executor: Finished task 23.0 in stage 3.0 (TID 101). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 3.0 (TID 104, localhost, executor driver, partition 26, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 26.0 in stage 3.0 (TID 104)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 25.0 in stage 3.0 (TID 103). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 3.0 (TID 105, localhost, executor driver, partition 27, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 3.0 (TID 101) in 24 ms on localhost (executor driver) (25/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 3.0 (TID 103) in 12 ms on localhost (executor driver) (26/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 27.0 in stage 3.0 (TID 105)

18/04/23 18:40:40 INFO executor.Executor: Finished task 26.0 in stage 3.0 (TID 104). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 3.0 (TID 106, localhost, executor driver, partition 28, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 28.0 in stage 3.0 (TID 106)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 3.0 (TID 104) in 16 ms on localhost (executor driver) (27/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 28.0 in stage 3.0 (TID 106). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 3.0 (TID 107, localhost, executor driver, partition 29, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO executor.Executor: Finished task 27.0 in stage 3.0 (TID 105). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 3.0 (TID 108, localhost, executor driver, partition 30, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 30.0 in stage 3.0 (TID 108)

18/04/23 18:40:40 INFO executor.Executor: Running task 29.0 in stage 3.0 (TID 107)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 30.0 in stage 3.0 (TID 108). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 3.0 (TID 105) in 24 ms on localhost (executor driver) (28/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 3.0 (TID 106) in 22 ms on localhost (executor driver) (29/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 3.0 (TID 109, localhost, executor driver, partition 31, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 3.0 (TID 108) in 17 ms on localhost (executor driver) (30/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 29.0 in stage 3.0 (TID 107). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 3.0 (TID 110, localhost, executor driver, partition 32, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 32.0 in stage 3.0 (TID 110)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 3.0 (TID 107) in 20 ms on localhost (executor driver) (31/39)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Running task 31.0 in stage 3.0 (TID 109)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 32.0 in stage 3.0 (TID 110). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 3.0 (TID 111, localhost, executor driver, partition 33, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO executor.Executor: Finished task 31.0 in stage 3.0 (TID 109). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 3.0 (TID 110) in 15 ms on localhost (executor driver) (32/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 3.0 (TID 112, localhost, executor driver, partition 34, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 3.0 (TID 109) in 24 ms on localhost (executor driver) (33/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 34.0 in stage 3.0 (TID 112)

18/04/23 18:40:40 INFO executor.Executor: Running task 33.0 in stage 3.0 (TID 111)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 33.0 in stage 3.0 (TID 111). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 3.0 (TID 113, localhost, executor driver, partition 35, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 3.0 (TID 111) in 21 ms on localhost (executor driver) (34/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 35.0 in stage 3.0 (TID 113)

18/04/23 18:40:40 INFO executor.Executor: Finished task 34.0 in stage 3.0 (TID 112). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 3.0 (TID 114, localhost, executor driver, partition 36, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 3.0 (TID 112) in 21 ms on localhost (executor driver) (35/39)

18/04/23 18:40:40 INFO executor.Executor: Finished task 35.0 in stage 3.0 (TID 113). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 3.0 (TID 115, localhost, executor driver, partition 37, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 37.0 in stage 3.0 (TID 115)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 3.0 (TID 113) in 14 ms on localhost (executor driver) (36/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 36.0 in stage 3.0 (TID 114)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 36.0 in stage 3.0 (TID 114). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 3.0 (TID 116, localhost, executor driver, partition 38, PROCESS_LOCAL, 1942 bytes)

18/04/23 18:40:40 INFO executor.Executor: Finished task 37.0 in stage 3.0 (TID 115). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 3.0 (TID 114) in 18 ms on localhost (executor driver) (37/39)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 3.0 (TID 115) in 14 ms on localhost (executor driver) (38/39)

18/04/23 18:40:40 INFO executor.Executor: Running task 38.0 in stage 3.0 (TID 116)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 38.0 in stage 3.0 (TID 116). 1306 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 3.0 (TID 116) in 14 ms on localhost (executor driver) (39/39)

18/04/23 18:40:40 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool

18/04/23 18:40:40 INFO scheduler.DAGScheduler: ShuffleMapStage 3 (mapToPair at groupedRecordsRDD.java:55) finished in 0.368 s

18/04/23 18:40:40 INFO scheduler.DAGScheduler: looking for newly runnable stages

18/04/23 18:40:40 INFO scheduler.DAGScheduler: running: Set()

18/04/23 18:40:40 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 4)

18/04/23 18:40:40 INFO scheduler.DAGScheduler: failed: Set()

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Submitting ResultStage 4 (MapPartitionsRDD[9] at mapToPair at groupedRecordsRDD.java:58), which has no missing parents

18/04/23 18:40:40 INFO storage.MemoryStore: Block broadcast 4 stored as values in memory (estimated size 4.3 KB, free 529.7 MB)

18/04/23 18:40:40 INFO storage.MemoryStore: Block broadcast 4 piece0 stored as bytes in memory (estimated size 2.5 KB, free 529.7 MB)

18/04/23 18:40:40 INFO storage.BlockManagerInfo: Added broadcast 4 piece0 in memory on localhost:40937 (size: 2.5 KB, free: 530.0 MB)

18/04/23 18:40:40 INFO spark.SparkContext: Created broadcast 4 from broadcast at DAGScheduler.scala:1004

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Submitting 5 missing tasks from ResultStage 4 (MapPartitionsRDD[9] at mapToPair at groupedRecordsRDD.java:58) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4))

18/04/23 18:40:40 INFO scheduler.TaskSchedulerImpl: Adding task set 4.0 with 5 tasks

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 4.0 (TID 117, localhost, executor driver, partition 0, NODE LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 4.0 (TID 118, localhost, executor driver, partition 1, NODE_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO executor.Executor: Running task 1.0 in stage 4.0 (TID 118)

18/04/23 18:40:40 INFO executor.Executor: Running task 0.0 in stage 4.0 (TID 117)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 4 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 4 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 1.0 in stage 4.0 (TID 118). 1132 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 0.0 in stage 4.0 (TID 117). 1132 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 4.0 (TID 119, localhost, executor driver, partition 2, NODE_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 4.0 (TID 120, localhost, executor driver, partition 3, NODE_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 4.0 (TID 118) in 24 ms on localhost (executor driver) (1/5)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 4.0 (TID 117) in 24 ms on localhost (executor driver) (2/5)

18/04/23 18:40:40 INFO executor.Executor: Running task 3.0 in stage 4.0 (TID 120)

18/04/23 18:40:40 INFO executor.Executor: Running task 2.0 in stage 4.0 (TID 119)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 4 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 4 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 3.0 in stage 4.0 (TID 120). 1133 bytes result sent to driver

18/04/23 18:40:40 INFO executor.Executor: Finished task 2.0 in stage 4.0 (TID 119). 1133 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 4.0 (TID 121, localhost, executor driver, partition 4, NODE_LOCAL, 1953 bytes)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 4.0 (TID 120) in 14 ms on localhost (executor driver) (3/5)

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 4.0 (TID 119) in 17 ms on localhost (executor driver) (4/5)

18/04/23 18:40:40 INFO executor.Executor: Running task 4.0 in stage 4.0 (TID 121)

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Getting 4 non-empty blocks out of 39 blocks

18/04/23 18:40:40 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:40 INFO executor.Executor: Finished task 4.0 in stage 4.0 (TID 121). 1124 bytes result sent to driver

18/04/23 18:40:40 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 4.0 (TID 121) in 22 ms on localhost (executor driver) (5/5)

18/04/23 18:40:40 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool

18/04/23 18:40:40 INFO scheduler.DAGScheduler: ResultStage 4 (collect at groupedRecordsRDD.java:61) finished in 0.057 s

18/04/23 18:40:40 INFO scheduler.DAGScheduler: Job 1 finished: collect at groupedRecordsRDD.java:61, took 0.460112 s

(4, 85287)

(3, 292090)

(2, 17648209)

(1, 36492406)

18/04/23 18:40:41 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/23 18:40:41 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/23 18:40:41 INFO spark.SparkContext: Starting job: saveAsTextFile at groupedRecordsRDD.java:67

18/04/23 18:40:41 INFO spark.MapOutputTrackerMaster: Size of output statuses for shuffle 0 is 196 bytes

18/04/23 18:40:41 INFO spark.MapOutputTrackerMaster: Size of output statuses for shuffle 1 is 190 bytes

18/04/23 18:40:41 INFO scheduler.DAGScheduler: Got job 2 (saveAsTextFile at groupedRecordsRDD.java:67) with 5 output partitions

18/04/23 18:40:41 INFO scheduler.DAGScheduler: Final stage: ResultStage 7 (saveAsTextFile at groupedRecordsRDD.java:67)

18/04/23 18:40:41 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 6)

18/04/23 18:40:41 INFO scheduler.DAGScheduler: Missing parents: List()

18/04/23 18:40:41 INFO scheduler.DAGScheduler: Submitting ResultStage 7 (MapPartitionsRDD[10] at saveAsTextFile at groupedRecordsRDD.java:67), which has no missing parents

18/04/23 18:40:41 INFO storage.MemoryStore: Block broadcast_5 stored as values in memory (estimated size 81.9 KB, free 529.6 MB)

18/04/23 18:40:41 INFO storage.MemoryStore: Block broadcast_5_piece0 stored as bytes in memory (estimated size 29.7 KB, free 529.6 MB)

18/04/23 18:40:41 INFO storage.BlockManagerInfo: Added broadcast_5_piece0 in memory on localhost:40937 (size: 29.7 KB, free: 529.9 MB)

18/04/23 18:40:41 INFO spark.SparkContext: Created broadcast 5 from broadcast at DAGScheduler.scala:1004

18/04/23 18:40:41 INFO scheduler.DAGScheduler: Submitting 5 missing tasks from ResultStage 7 (MapPartitionsRDD[10] at saveAsTextFile at groupedRecordsRDD.java:67) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4))

18/04/23 18:40:41 INFO scheduler.TaskSchedulerImpl: Adding task set 7.0 with 5 tasks

18/04/23 18:40:41 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 7.0 (TID 122, localhost, executor driver, partition 0, NODE LOCAL, 1953 bytes)

18/04/23 18:40:41 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 7.0 (TID 123, localhost, executor driver, partition 1, NODE LOCAL, 1953 bytes)

18/04/23 18:40:41 INFO executor.Executor: Running task 1.0 in stage 7.0 (TID 123)

18/04/23 18:40:41 INFO executor.Executor: Running task 0.0 in stage 7.0 (TID 122)

18/04/23 18:40:41 INFO storage.ShuffleBlockFetcherIterator: Getting 4 non-empty blocks out of 39 blocks

18/04/23 18:40:41 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms

18/04/23 18:40:41 INFO storage.ShuffleBlockFetcherIterator: Getting 4 non-empty blocks out of 39 blocks

18/04/23 18:40:41 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:41 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/23 18:40:41 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/23 18:40:41 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/23 18:40:41 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/23 18:40:41 INFO output.FileOutputCommitter: Saved output of task 'attempt 201804231840 0007 m 000001 123' to hdfs://ip-10-0-0-

161.ec2.internal:8020/user/root/spark_assignment/output/grouped_Records_SparkRDD/_temporary/0/task_201804231840_0007_m_000001

18/04/23 18:40:41 INFO mapred.SparkHadoopMapRedUtil: attempt_201804231840_0007_m_000001_123: Committed

18/04/23 18:40:41 INFO executor.Executor: Finished task 1.0 in stage 7.0 (TID 123). 2038 bytes result sent to driver

18/04/23 18:40:41 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 7.0 (TID 124, localhost, executor driver, partition 2, NODE_LOCAL, 1953 bytes)

18/04/23 18:40:41 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 7.0 (TID 123) in 118 ms on localhost (executor driver) (1/5)

18/04/23 18:40:41 INFO executor.Executor: Running task 2.0 in stage 7.0 (TID 124)

18/04/23 18:40:41 INFO storage.ShuffleBlockFetcherIterator: Getting 4 non-empty blocks out of 39 blocks

18/04/23 18:40:41 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:41 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/23 18:40:41 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/23 18:40:41 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804231840_0007_m_000002_124' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/grouped_Records_SparkRDD/_temporary/0/task_201804231840_0007_m_000002

18/04/23 18:40:41 INFO mapred.SparkHadoopMapRedUtil: attempt_201804231840_0007_m_000002_124: Committed

18/04/23 18:40:41 INFO executor.Executor: Finished task 2.0 in stage 7.0 (TID 124). 2052 bytes result sent to driver

18/04/23 18:40:41 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 7.0 (TID 125, localhost, executor driver, partition 3, NODE_LOCAL, 1953 bytes)

18/04/23 18:40:41 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 7.0 (TID 124) in 59 ms on localhost (executor driver) (2/5)

18/04/23 18:40:41 INFO executor.Executor: Running task 3.0 in stage 7.0 (TID 125)

18/04/23 18:40:41 INFO storage.ShuffleBlockFetcherIterator: Getting 4 non-empty blocks out of 39 blocks

18/04/23 18:40:41 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:41 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/23 18:40:41 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/23 18:40:41 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804231840_0007_m_000000_122' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/grouped Records SparkRDD/ temporary/0/task 201804231840_0007_m_000000

18/04/23 18:40:41 INFO mapred.SparkHadoopMapRedUtil: attempt_201804231840_0007_m_000000_122: Committed

18/04/23 18:40:41 INFO executor.Executor: Finished task 0.0 in stage 7.0 (TID 122). 2052 bytes result sent to driver

18/04/23 18:40:41 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 7.0 (TID 126, localhost, executor driver, partition 4, NODE LOCAL, 1953 bytes)

18/04/23 18:40:41 INFO executor.Executor: Running task 4.0 in stage 7.0 (TID 126)

18/04/23 18:40:41 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 7.0 (TID 122) in 504 ms on localhost (executor driver) (3/5)

18/04/23 18:40:41 INFO storage.ShuffleBlockFetcherIterator: Getting 4 non-empty blocks out of 39 blocks

18/04/23 18:40:41 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

18/04/23 18:40:41 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

18/04/23 18:40:41 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup temporary folders under output directory:false, ignore cleanup failures: false

18/04/23 18:40:41 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804231840_0007_m_000004_126' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/grouped Records SparkRDD/ temporary/0/task 201804231840_0007_m_000004

18/04/23 18:40:41 INFO mapred.SparkHadoopMapRedUtil: attempt_201804231840_0007_m_000004_126: Committed

18/04/23 18:40:41 INFO executor.Executor: Finished task 4.0 in stage 7.0 (TID 126). 2052 bytes result sent to driver

18/04/23 18:40:41 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 7.0 (TID 126) in 44 ms on localhost (executor driver) (4/5)

18/04/23 18:40:41 INFO output.FileOutputCommitter: Saved output of task 'attempt_201804231840_0007_m_000003_125' to hdfs://ip-10-0-0-161.ec2.internal:8020/user/root/spark_assignment/output/grouped Records SparkRDD/ temporary/0/task 201804231840_0007_m_000003

18/04/23 18:40:41 INFO mapred.SparkHadoopMapRedUtil: attempt_201804231840_0007_m_000003_125: Committed

18/04/23 18:40:41 INFO executor.Executor: Finished task 3.0 in stage 7.0 (TID 125). 2052 bytes result sent to driver

18/04/23 18:40:41 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 7.0 (TID 125) in 445 ms on localhost (executor driver) (5/5)

18/04/23 18:40:41 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 7.0, whose tasks have all completed, from pool

18/04/23 18:40:41 INFO scheduler.DAGScheduler: ResultStage 7 (saveAsTextFile at groupedRecordsRDD.java:67) finished in 0.617 s

18/04/23 18:40:41 INFO scheduler.DAGScheduler: Job 2 finished: saveAsTextFile at groupedRecordsRDD.java:67, took 0.644055 s

18/04/23 18:40:41 WARN spark.SparkContext: Requesting executors is only supported in coarse-grained mode

18/04/23 18:40:41 INFO ui.SparkUI: Stopped Spark web UI at http://10.0.0.161:4040

18/04/23 18:40:42 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

18/04/23 18:40:42 INFO storage.MemoryStore: MemoryStore cleared

18/04/23 18:40:42 INFO storage.BlockManager: BlockManager stopped

18/04/23 18:40:42 INFO storage.BlockManagerMaster: BlockManagerMaster stopped

18/04/23 18:40:42 INFO scheduler.OutputCommitCoordinator\$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

18/04/23 18:40:42 INFO remote.RemoteActorRefProvider\$RemotingTerminator: Shutting down remote daemon.

18/04/23 18:40:42 INFO remote.RemoteActorRefProvider\$RemotingTerminator: Remote daemon shut down; proceeding with flushing remote transports.

18/04/23 18:40:42 INFO spark.SparkContext: Successfully stopped SparkContext

18/04/23 18:40:42 INFO util.ShutdownHookManager: Shutdown hook called

18/04/23 18:40:42 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-5af2f31d-30bf-46b3-b768-8b0160dadd71

18/04/23 18:40:42 INFO Remoting: Remoting shut down

18/04/23 18:40:42 INFO remote.RemoteActorRefProvider\$RemotingTerminator: Remoting shut down.

[root@ip-10-0-0-161 spark Assignment]#

Logs for Pig Assignment

- lookup:

```
[root@ip-10-0-0-161 pig_Assignment]# pig lookUp.pig
```

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

```
log4j:WARN Please initialize the log4j system properly.
```

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

```
2018-04-21 17:28:43,878 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.14.0 (rUnversioned directory) compiled Jan 06 2018, 13:37:28
```

```
2018-04-21 17:28:43,878 [main] INFO org.apache.pig.Main - Logging error messages to: /root/pig_Assignment/pig_1524331723859.log
```

```
2018-04-21 17:28:44,687 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /root/.pigbootup not found
```

```
2018-04-21 17:28:44,785 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
```

```
2018-04-21 17:28:44,785 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-21 17:28:44,785 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://ip-10-0-0-161.ec2.internal:8020
```

```
2018-04-21 17:28:45,248 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-21 17:28:45,283 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-21 17:28:45,316 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-21 17:28:45,362 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

2018-04-21 17:28:45,391 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-21 17:28:45,422 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-21 17:28:45,442 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-21 17:28:45,469 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-21 17:28:46,188 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 2 time(s).

2018-04-21 17:28:46,188 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_LONG 1 time(s).

2018-04-21 17:28:46,188 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_INT 2 time(s).

2018-04-21 17:28:46,188 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 1 time(s).

2018-04-21 17:28:46,207 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: RANK,FILTER

2018-04-21 17:28:46,240 [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]}

2018-04-21 17:28:46,262 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator

2018-04-21 17:28:46,334 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

2018-04-21 17:28:46,369 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 2

2018-04-21 17:28:46,369 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 2

2018-04-21 17:28:46,469 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-21 17:28:46,642 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job

2018-04-21 17:28:46,706 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent

2018-04-21 17:28:46,706 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-04-21 17:28:46,706 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress

2018-04-21 17:28:48,928 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job8817998570827096936.jar

2018-04-21 17:28:52,648 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job8817998570827096936.jar created

2018-04-21 17:28:52,648 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.jar is deprecated. Instead, use mapreduce.job.jar

2018-04-21 17:28:52,667 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2018-04-21 17:28:52,697 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

2018-04-21 17:28:52,699 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address

2018-04-21 17:28:52,705 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-21 17:28:52,722 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-21 17:28:53,109 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 6

2018-04-21 17:28:53,109 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 6

2018-04-21 17:28:53,146 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 36

2018-04-21 17:28:53,893 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:36

2018-04-21 17:28:54,030 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1524330558439_0003

2018-04-21 17:28:54,275 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1524330558439_0003

2018-04-21 17:28:54,308 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-0-0-161.ec2.internal:8088/proxy/application_1524330558439_0003/

2018-04-21 17:28:54,309 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1524330558439_0003

2018-04-21 17:28:54,309 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Processing aliases file

2018-04-21 17:28:54,309 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
detailed locations: M: file[5,7] C: R:

2018-04-21 17:28:54,341 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0%
complete

2018-04-21 17:29:41,178 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 4%
complete

2018-04-21 17:30:21,391 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 8%
complete

2018-04-21 17:31:08,172 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12%
complete

2018-04-21 17:31:50,029 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 16%
complete

2018-04-21 17:32:39,308 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 20%
complete

2018-04-21 17:33:18,712 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 25%
complete

2018-04-21 17:33:20,248 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50%
complete

2018-04-21 17:33:20,487 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -
mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces

2018-04-21 17:33:20,583 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job

2018-04-21 17:33:20,594 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-04-21 17:33:31,141 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
creating jar file Job991615374572054034.jar

2018-04-21 17:33:34,758 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file
Job991615374572054034.jar created

2018-04-21 17:33:34,771 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Setting up single store job

2018-04-21 17:33:34,793 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1
map-reduce job(s) waiting for submission.

2018-04-21 17:33:34,796 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy -
Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-21 17:33:34,803 [JobControl] INFO
org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead,
use fs.defaultFS

2018-04-21 17:33:34,875 [JobControl] INFO
org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 36

2018-04-21 17:33:34,875 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process
: 36

2018-04-21 17:33:34,876 [JobControl] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths
(combined) to process : 48

2018-04-21 17:33:34,900 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -
number of splits:48

2018-04-21 17:33:34,942 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1524330558439_0004

2018-04-21 17:33:35,189 [JobControl] INFO
org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application
application_1524330558439_0004

2018-04-21 17:33:35,191 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-0-0-161.ec2.internal:8088/proxy/application_1524330558439_0004/

2018-04-21 17:33:35,294 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1524330558439_0004

2018-04-21 17:33:35,294 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases NoHeader

2018-04-21 17:33:35,294 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: NoHeader[8,11] C: R:

2018-04-21 17:34:24,811 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 54% complete

2018-04-21 17:35:06,840 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 58% complete

2018-04-21 17:35:51,856 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 63% complete

2018-04-21 17:36:22,180 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 67% complete

2018-04-21 17:36:59,701 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 71% complete

2018-04-21 17:37:31,135 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
100% complete

2018-04-21 17:37:31,149 [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	2018-04-21 17:28:46	2018-04-21 17:37:31	RANK,FILTER

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature
job_1524330558439_0003	36	0	28	14	19	19	n/a	n/a	n/a	n/a	file	MAP_ONLY

job_1524330558439_0004	48	0	16	8	12	12	n/a	n/a	n/a	n/a		
NoHeader	MAP_ONLY		/user/root/pigAssignment2/Q1Output,									

Input(s):

Successfully read 54518004 records (4810073810 bytes) from:
"/user/root/spark_assignment/input_dataset/yellow_tripdata*"

Output(s):

Successfully stored 1 records (99 bytes) in: "/user/root/pigAssignment2/Q1Output"

Counters:

Total records written : 1

Total bytes written : 99

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job_1524330558439_0003 -> job_1524330558439_0004,

job_1524330558439_0004

2018-04-21 17:37:32,151 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-10-0-0-161.ec2.internal/10.0.0.161:45255. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-04-21 17:37:33,152 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-10-0-0-161.ec2.internal/10.0.0.161:45255. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-04-21 17:37:34,155 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-10-0-0-161.ec2.internal/10.0.0.161:45255. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-04-21 17:37:34,260 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-04-21 17:37:34,474 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-04-21 17:37:34,578 [main] WARN

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 5 time(s).

2018-04-21 17:37:34,578 [main] WARN

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning ACCESSING_NON_EXISTENT_FIELD 20 time(s).

2018-04-21 17:37:34,578 [main] INFO

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

[root@ip-10-0-0-161 pig_Assignment]#

FilterByRateCode:

```
[root@ip-10-0-0-161 pig_Assignment]# pig FilterByRateCodeId.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.
```

```
2018-04-21 17:17:31,451 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.14.0 (rUnversioned directory) compiled Jan 06 2018, 13:37:28
```

```
2018-04-21 17:17:31,452 [main] INFO org.apache.pig.Main - Logging error messages to: /root/pig_Assignment/pig_1524331051432.log
```

```
2018-04-21 17:17:32,211 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /root/.pigbootup not found
```

```
2018-04-21 17:17:32,304 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
```

```
2018-04-21 17:17:32,304 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-21 17:17:32,304 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://ip-10-0-0-161.ec2.internal:8020
```

```
2018-04-21 17:17:32,778 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-21 17:17:32,812 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-21 17:17:32,843 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-21 17:17:32,876 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-21 17:17:32,917 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

2018-04-21 17:17:32,946 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-21 17:17:32,970 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-21 17:17:32,996 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-21 17:17:33,752 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_INT 1 time(s).

2018-04-21 17:17:33,771 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: RANK,FILTER

2018-04-21 17:17:33,801 [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]}

2018-04-21 17:17:33,819 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator

2018-04-21 17:17:33,886 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

2018-04-21 17:17:33,919 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 2

2018-04-21 17:17:33,919 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 2

2018-04-21 17:17:34,006 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-21 17:17:34,169 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job

2018-04-21 17:17:34,210 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent

2018-04-21 17:17:34,210 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-04-21 17:17:34,210 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress

2018-04-21 17:17:35,079 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job526568983127167961.jar

2018-04-21 17:17:38,842 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job526568983127167961.jar created

2018-04-21 17:17:38,842 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.jar is deprecated. Instead, use mapreduce.job.jar

2018-04-21 17:17:38,861 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2018-04-21 17:17:38,891 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

2018-04-21 17:17:38,892 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address

2018-04-21 17:17:38,897 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-21 17:17:38,910 [JobControl] INFO

org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-21 17:17:39,704 [JobControl] INFO

org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 6

2018-04-21 17:17:39,704 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 6

2018-04-21 17:17:39,742 [JobControl] INFO

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

2018-04-21 17:17:40,234 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:36

2018-04-21 17:17:40,756 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1524330558439_0001

2018-04-21 17:17:41,347 [JobControl] INFO

org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1524330558439_0001

2018-04-21 17:17:41,379 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-0-0-161.ec2.internal:8088/proxy/application_1524330558439_0001/

2018-04-21 17:17:41,380 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1524330558439_0001

2018-04-21 17:17:41,380 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases file

2018-04-21 17:17:41,380 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: file[5,7] C: R:

2018-04-21 17:17:41,417 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete

2018-04-21 17:18:33,813 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 4% complete

2018-04-21 17:19:19,034 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 8% complete

2018-04-21 17:20:03,715 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12% complete

2018-04-21 17:20:46,161 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 16% complete

2018-04-21 17:21:33,935 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 20% complete

2018-04-21 17:22:34,252 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 25% complete

2018-04-21 17:22:42,011 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50% complete

2018-04-21 17:22:42,423 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-04-21 17:22:42,504 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces

2018-04-21 17:22:42,688 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-04-21 17:22:42,725 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job

2018-04-21 17:22:42,730 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-04-21 17:22:45,988 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
creating jar file Job2111064601527516703.jar

2018-04-21 17:22:49,793 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file
Job2111064601527516703.jar created

2018-04-21 17:22:49,807 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Setting up single store job

2018-04-21 17:22:49,827 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1
map-reduce job(s) waiting for submission.

2018-04-21 17:22:49,829 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy -
Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-21 17:22:49,837 [JobControl] INFO

org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead,
use fs.defaultFS

2018-04-21 17:22:49,905 [JobControl] INFO

org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 36

2018-04-21 17:22:49,905 [JobControl] INFO

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

2018-04-21 17:22:49,907 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths
(combined) to process : 48

2018-04-21 17:22:50,338 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -
number of splits:48

2018-04-21 17:22:53,911 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter -
Submitting tokens for job: job_1524330558439_0002

2018-04-21 17:22:53,938 [JobControl] INFO

org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application
application_1524330558439_0002

2018-04-21 17:22:53,941 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to
track the job: http://ip-10-0-0-161.ec2.internal:8088/proxy/application_1524330558439_0002/

2018-04-21 17:22:53,941 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
HadoopJobId: job_1524330558439_0002

2018-04-21 17:22:53,941 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Processing aliases record

2018-04-21 17:22:53,941 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
detailed locations: M: record[9,9] C: R:

2018-04-21 17:24:02,651 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 54%
complete

2018-04-21 17:24:36,517 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 58%
complete

2018-04-21 17:25:23,229 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 62%
complete

2018-04-21 17:25:53,295 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 66%
complete

2018-04-21 17:26:32,349 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 70%
complete

2018-04-21 17:27:04,529 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 75%
complete

2018-04-21 17:27:11,494 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
100% complete

2018-04-21 17:27:11,519 [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	2018-04-21 17:17:34	2018-04-21 17:27:11	RANK,FILTER

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature
job_1524330558439_0001	36	0	39	15	22	20	n/a	n/a	n/a	n/a	file	MAP_ONLY
job_1524330558439_0002	48	0	19	5	13	12	n/a	n/a	n/a	n/a	record	MAP_ONLY /user/root/pigAssignment2/Q2Output,

Input(s):

Successfully read 54518004 records (4810073810 bytes) from:
"/user/root/spark_assignment/input_dataset/yellow_tripdata*"

Output(s):

Successfully stored 31029 records (3083370 bytes) in: "/user/root/pigAssignment2/Q2Output"

Counters:

Total records written : 31029

Total bytes written : 3083370

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job_1524330558439_0001 -> job_1524330558439_0002,

job_1524330558439_0002

2018-04-21 17:27:11,865 [main] WARN

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 6 time(s).

2018-04-21 17:27:11,865 [main] WARN

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Encountered Warning ACCESSING_NON_EXISTENT_FIELD 6 time(s).

2018-04-21 17:27:11,865 [main] INFO

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

[root@ip-10-0-0-161 pig_Assignment]#

- **groupedRecords**

```
[root@ip-10-0-0-161 pig_Assignment]# pig groupedRecords.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.
```

```
2018-04-22 10:00:33,099 [main] INFO  org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.14.0 (rUnversioned directory) compiled Jan 06 2018, 13:37:28
```

```
2018-04-22 10:00:33,099 [main] INFO  org.apache.pig.Main - Logging error messages to: /root/pig_Assignment/pig_1524391233078.log
```

```
2018-04-22 10:00:33,866 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup file /root/.pigbootup not found
```

```
2018-04-22 10:00:33,969 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
```

```
2018-04-22 10:00:33,969 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-22 10:00:33,969 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://ip-10-0-0-161.ec2.internal:8020
```

```
2018-04-22 10:00:34,617 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-22 10:00:34,656 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-22 10:00:34,693 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-22 10:00:34,741 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```

```
2018-04-22 10:00:34,783 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
```


2018-04-22 10:00:34,839 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-22 10:00:34,879 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-22 10:00:34,919 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-22 10:00:35,697 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_LONG 1 time(s).

2018-04-22 10:00:35,718 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY,ORDER_BY,RANK,FILTER

2018-04-22 10:00:35,750 [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]}

2018-04-22 10:00:35,773 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator

2018-04-22 10:00:35,850 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

2018-04-22 10:00:35,884 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.CombinerOptimizer - Choosing to move algebraic foreach to combiner

2018-04-22 10:00:35,903 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 4

2018-04-22 10:00:35,903 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 4

2018-04-22 10:00:36,009 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-22 10:00:36,203 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job

2018-04-22 10:00:36,289 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent

2018-04-22 10:00:36,289 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-04-22 10:00:36,289 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress

2018-04-22 10:00:37,430 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job3343307042968435569.jar

2018-04-22 10:00:41,029 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job3343307042968435569.jar created

2018-04-22 10:00:41,029 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.jar is deprecated. Instead, use mapreduce.job.jar

2018-04-22 10:00:41,048 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

2018-04-22 10:00:41,055 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.

2018-04-22 10:00:41,055 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache

2018-04-22 10:00:41,055 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

2018-04-22 10:00:41,091 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

2018-04-22 10:00:41,092 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address

2018-04-22 10:00:41,097 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-22 10:00:41,110 [JobControl] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-22 10:00:41,875 [JobControl] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 6

2018-04-22 10:00:41,876 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 6

2018-04-22 10:00:41,916 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 36

2018-04-22 10:00:42,772 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:36

2018-04-22 10:00:43,293 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1524388868016_0005

2018-04-22 10:00:43,544 [JobControl] INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1524388868016_0005

2018-04-22 10:00:43,576 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-0-0-161.ec2.internal:8088/proxy/application_1524388868016_0005/

2018-04-22 10:00:43,576 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1524388868016_0005

2018-04-22 10:00:43,576 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
Processing aliases file

2018-04-22 10:00:43,576 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
detailed locations: M: file[5,7],file[-1,-1] C: R:

2018-04-22 10:00:43,609 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0%
complete

2018-04-22 10:02:31,819 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 4%
complete

2018-04-22 10:04:06,757 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 8%
complete

2018-04-22 10:05:46,152 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12%
complete

2018-04-22 10:05:54,452 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 25%
complete

2018-04-22 10:05:54,607 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -

Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history
server

2018-04-22 10:05:54,813 [main] INFO org.apache.hadoop.conf.Configuration.deprecation -

mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces

2018-04-22 10:05:54,947 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate -

Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history
server

2018-04-22 10:05:54,957 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script

settings are added to the job

2018-04-22 10:05:54,969 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-04-22 10:05:54,970 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Reduce phase detected, estimating # of required reducers.

2018-04-22 10:05:54,971 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using
reducer estimator:

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator

2018-04-22 10:05:54,976 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
- BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=6310651376

2018-04-22 10:05:54,976 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Setting Parallelism to 7

2018-04-22 10:05:57,161 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
creating jar file Job5190776427308521641.jar

2018-04-22 10:06:00,733 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file
Job5190776427308521641.jar created

2018-04-22 10:06:00,743 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Setting up single store job

2018-04-22 10:06:00,744 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key
[pig.schematuple] is false, will not generate code.

2018-04-22 10:06:00,744 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting
process to move generated code to distributed cache

2018-04-22 10:06:00,744 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []

2018-04-22 10:06:00,789 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

2018-04-22 10:06:00,791 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-22 10:06:00,796 [JobControl] INFO

org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

2018-04-22 10:06:00,868 [JobControl] INFO

org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 36

2018-04-22 10:06:00,868 [JobControl] INFO

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

2018-04-22 10:06:00,869 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 48

2018-04-22 10:06:00,891 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:48

2018-04-22 10:06:01,341 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1524388868016_0006

2018-04-22 10:06:01,374 [JobControl] INFO

org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1524388868016_0006

2018-04-22 10:06:01,376 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-0-0-161.ec2.internal:8088/proxy/application_1524388868016_0006/

2018-04-22 10:06:01,376 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1524388868016_0006

2018-04-22 10:06:01,376 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases filteredRecords,filteredRecordsByPayment_type,result

2018-04-22 10:06:01,376 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
detailed locations: M: filteredRecords[7,18],result[9,9],filteredRecordsByPayment_type[8,32] C:
result[9,9],filteredRecordsByPayment_type[8,32] R: result[9,9]

2018-04-22 10:08:01,120 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 29%
complete

2018-04-22 10:09:38,724 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 33%
complete

2018-04-22 10:11:00,663 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 37%
complete

2018-04-22 10:11:23,799 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 42%
complete

2018-04-22 10:11:30,220 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 50%
complete

2018-04-22 10:11:32,464 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script
settings are added to the job

2018-04-22 10:11:32,472 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-04-22 10:11:32,473 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Reduce phase detected, estimating # of required reducers.

2018-04-22 10:11:32,473 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using
reducer estimator:

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator

2018-04-22 10:11:32,476 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
- BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=71

2018-04-22 10:11:32,476 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Setting Parallelism to 1

2018-04-22 10:11:35,025 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
creating jar file Job5784828988729493107.jar

2018-04-22 10:11:38,617 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file
Job5784828988729493107.jar created

2018-04-22 10:11:38,627 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Setting up single store job

2018-04-22 10:11:38,627 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key
[pig.schematuple] is false, will not generate code.

2018-04-22 10:11:38,627 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting
process to move generated code to distributed cache

2018-04-22 10:11:38,627 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []

2018-04-22 10:11:38,649 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1
map-reduce job(s) waiting for submission.

2018-04-22 10:11:38,651 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy -
Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-22 10:11:38,655 [JobControl] INFO

org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead,
use fs.defaultFS

2018-04-22 10:11:38,718 [JobControl] INFO

org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 7

2018-04-22 10:11:38,719 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 7

2018-04-22 10:11:38,719 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-04-22 10:11:38,742 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1

2018-04-22 10:11:39,194 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1524388868016_0007

2018-04-22 10:11:39,421 [JobControl] INFO

org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1524388868016_0007

2018-04-22 10:11:39,423 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-0-0-161.ec2.internal:8088/proxy/application_1524388868016_0007/

2018-04-22 10:11:39,423 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1524388868016_0007

2018-04-22 10:11:39,423 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases order_by_data

2018-04-22 10:11:39,423 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: order_by_data[10,16] C: R:

2018-04-22 10:11:49,475 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 62% complete

2018-04-22 10:11:55,739 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 75% complete

2018-04-22 10:11:59,979 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job

2018-04-22 10:11:59,989 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

2018-04-22 10:11:59,989 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Reduce phase detected, estimating # of required reducers.

2018-04-22 10:11:59,989 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Setting Parallelism to 1

2018-04-22 10:12:02,547 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
creating jar file Job5778560540297047920.jar

2018-04-22 10:12:06,079 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file
Job5778560540297047920.jar created

2018-04-22 10:12:06,089 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler -
Setting up single store job

2018-04-22 10:12:06,089 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key
[pig.schematuple] is false, will not generate code.

2018-04-22 10:12:06,089 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting
process to move generated code to distributed cache

2018-04-22 10:12:06,089 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key
[pig.schematuple.classes] with classes to deserialize []

2018-04-22 10:12:06,110 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1
map-reduce job(s) waiting for submission.

2018-04-22 10:12:06,112 [JobControl] INFO org.apache.hadoop.yarn.client.RMProxy -
Connecting to ResourceManager at ip-10-0-0-161.ec2.internal/10.0.0.161:8032

2018-04-22 10:12:06,120 [JobControl] INFO

org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead,
use fs.defaultFS

2018-04-22 10:12:06,585 [JobControl] INFO

org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 7

2018-04-22 10:12:06,586 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 7

2018-04-22 10:12:06,586 [JobControl] INFO

org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1

2018-04-22 10:12:07,014 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1

2018-04-22 10:12:07,450 [JobControl] INFO org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_1524388868016_0008

2018-04-22 10:12:07,677 [JobControl] INFO

org.apache.hadoop.yarn.client.api.impl.YarnClientImpl - Submitted application application_1524388868016_0008

2018-04-22 10:12:07,680 [JobControl] INFO org.apache.hadoop.mapreduce.Job - The url to track the job: http://ip-10-0-0-161.ec2.internal:8088/proxy/application_1524388868016_0008/

2018-04-22 10:12:07,680 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1524388868016_0008

2018-04-22 10:12:07,680 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases order_by_data

2018-04-22 10:12:07,680 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: order_by_data[10,16] C: R:

2018-04-22 10:12:17,780 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 87% complete

2018-04-22 10:12:28,311 [main] INFO

org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete

2018-04-22 10:12:28,330 [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	2018-04-22 10:00:36	2018-04-22 10:12:28	GROUP_BY,ORDER_BY,RANK,FILTER

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReductime	Alias	Feature
job_1524388868016_0005	36	0	27	21	23	23	n/a	n/a	n/a	n/a	file	MAP_ONLY
job_1524388868016_0006	48	7	20	10	16	17	47	411	6			filteredRecords,filteredRecordsByPayment_type,result GROUP_BY,COMBINER
job_1524388868016_0007	1	1	2	2	2	2	3	33	3		order_by_data	SAMPLER
job_1524388868016_0008	1	1	2	2	2	2	3	33	3		order_by_data	ORDER_BY /user/root/pigAssignment2/Q3Output,

Input(s):

Successfully read 54518004 records (4810073810 bytes) from:
"/user/root/spark_assignment/input_dataset/yellow_tripdata*"

Output(s):

Successfully stored 4 records (39 bytes) in: "/user/root/pigAssignment2/Q3Output"

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_1524388868016_0005 -> job_1524388868016_0006,

job_1524388868016_0006 -> job_1524388868016_0007,

job_1524388868016_0007 -> job_1524388868016_0008,

job_1524388868016_0008

2018-04-22 10:12:29,440 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-10-0-0-161.ec2.internal/10.0.0.161:33557. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-04-22 10:12:30,441 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-10-0-0-161.ec2.internal/10.0.0.161:33557. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-04-22 10:12:31,441 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-10-0-0-161.ec2.internal/10.0.0.161:33557. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-04-22 10:12:31,546 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-04-22 10:12:32,713 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-10-0-0-161.ec2.internal/10.0.0.161:45805. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-04-22 10:12:33,713 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-10-0-0-161.ec2.internal/10.0.0.161:45805. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-04-22 10:12:34,714 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: ip-10-0-0-161.ec2.internal/10.0.0.161:45805. Already tried 2 time(s); retry policy is RetryUpToMaximumCountWithFixedSleep(maxRetries=3, sleepTime=1000 MILLISECONDS)

2018-04-22 10:12:34,817 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-04-22 10:12:34,984 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

2018-04-22 10:12:35,053 [main] WARN
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 6 time(s).

2018-04-22 10:12:35,053 [main] WARN
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning ACCESSING_NON_EXISTENT_FIELD 96 time(s).

2018-04-22 10:12:35,054 [main] INFO
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!