

**CS-5513 Advance Database Management**

**Homework 4**

**Submitted By**

**Shyam Sundar Murali Krishnan**

**shyamkrishnan@ou.edu**

**113475983**

## TASK 1:

### QUERY A:

```
grunt> exec -param areacode="542" query.a.pig
2022-04-25 14:52:09,945 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 14:52:09,957 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 14:52:09,999 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY,FILTER
2022-04-25 14:52:10,008 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 14:52:10,008 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-04-25 14:52:10,008 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - (RULES ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter])
2022-04-25 14:52:10,016 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 14:52:10,017 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for data: $0, $3, $4
2022-04-25 14:52:10,019 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-04-25 14:52:10,020 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-04-25 14:52:10,020 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-04-25 14:52:10,026 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 14:52:10,027 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 14:52:10,028 [main] INFO org.apache.pig.tools.pigstats.MRScriptState - Pig script settings are added to the job
2022-04-25 14:52:10,028 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-04-25 14:52:10,029 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.
2022-04-25 14:52:10,029 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2022-04-25 14:52:10,029 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=12213
2022-04-25 14:52:10,029 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
2022-04-25 14:52:10,030 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2022-04-25 14:52:10,031 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2022-04-25 14:52:10,031 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2022-04-25 14:52:10,031 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Distributed cache not supported or needed in local mode. Setting key [pig.schematuple.local.dir] with code temp directory: /tmp/1650916330031-0
2022-04-25 14:52:10,040 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-04-25 14:52:10,043 [JobControl] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 14:52:10,307 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.9.2 0.17.0 mura0007 2022-04-25 14:52:10 2022-04-25 14:52:10 GROUP_BY,FILTER

Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime A
alias Feature Outputs
job_local798214372_0004 1 1 n/a n/a n/a n/a n/a n/a data,grou,new,result,sub_data GROUP_BY file:
/tmp/temp2009277357/tmp1879983222,
Input(s):
Successfully read 213 records from: "file:///oracle/stuhomes/mura0007/HW4-twitter_account.csv"
Output(s):
Successfully stored 1 records in: "file:/tmp/temp2009277357/tmp1879983222"

Counters:
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local798214372_0004

2022-04-25 14:52:10,308 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 14:52:10,308 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 14:52:10,309 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 14:52:10,312 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-04-25 14:52:10,312 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 14:52:10,312 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-25 14:52:10,321 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-04-25 14:52:10,321 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(542, {(t2go9708gh@ou.edu), (t0bf8ueggf@ou.edu), (r452d0wbba@aol.com)})
grunt>
```

```

mura0007@oracle18:~ 
grunt> exec -param areacode="847" query_a.pig
2022-04-25 15:43:43,805 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 15:43:44,160 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 15:43:44,353 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 15:43:44,373 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2022-04-25 15:43:44,377 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 6000: <line 11, column 0> Output Location Validation Failed for: '/oracle/stuhomes/mura0007' More info to follow:
Output directory file:/oracle/stuhomes/mura0007 already exists
Details at logfile: /oracle/stuhomes/mura0007/pig_1650919417276.log
grunt> [mura0007@oracle18 ~]$ pig -x local
22/04/25 15:45:48 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
22/04/25 15:45:48 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2022-04-25 15:45:48,164 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2022-04-25 15:45:48,164 [main] INFO org.apache.pig.Main - Logging error messages to: /oracle/stuhomes/mura0007/pig_1650919548160.log
2022-04-25 15:45:48,181 [main] INFO org.apache.pig.impl.Utils - Default bootup file /oracle/stuhomes/mura0007/pigbootup not found
2022-04-25 15:45:48,268 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-04-25 15:45:48,269 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2022-04-25 15:45:48,486 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 15:45:48,501 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-1e0ae815-a2e3-47b3-b05f-7ddcb72c90b6
2022-04-25 15:45:48,501 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt> clear

grunt> exec -param areacode="847" query_a.pig
2022-04-25 15:46:32,114 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 15:46:32,468 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 15:46:32,677 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 15:46:32,697 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2022-04-25 15:46:32,712 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY,FILTER
2022-04-25 15:46:32,727 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 15:46:32,755 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

3:48 PM 4/25/2022 [12]

mura0007@oracle18:~ 
2022-04-25 15:46:34,657 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.9.2 0.17.0 mura0007 2022-04-25 15:46:34 2022-04-25 15:46:34 GROUP_BY,FILTER

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime A
alias Feature Outputs job_local281413229_0002 1 n/a n/a n/a n/a n/a n/a data,grou,new,result,sub_data GROUP_BY file:
/tmp/temp1402191447/tmp276191781,

Input(s):
Successfully read 213 records from: "file:///oracle/stuhomes/mura0007/HW4-twitter_account.csv"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp1402191447/tmp276191781"

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

Job DAG:
job_local281413229_0002

2022-04-25 15:46:34,658 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 15:46:34,659 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 15:46:34,660 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 15:46:34,663 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-04-25 15:46:34,664 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 15:46:34,664 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-25 15:46:34,680 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-04-25 15:46:34,680 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(847,{(fjxrkljnln@gmail.com)})
grunt> [1]

```

3:47 PM 4/25/2022 [12]

```
mura0007@oracle18:~  
2022-04-25 15:46:34,660 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 15:46:34,663 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2022-04-25 15:46:34,664 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 15:46:34,664 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2022-04-25 15:46:34,680 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1  
2022-04-25 15:46:34,680 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(847, {((fjxrkljsnl@gmail.com))})  
grunt> clear  
  
grunt> exec -param areacode="430" query_a.pig  
2022-04-25 15:48:32,622 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 15:48:32,636 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 15:48:32,693 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 15:48:32,706 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY,FILTER  
2022-04-25 15:48:32,716 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 15:48:32,716 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.  
2022-04-25 15:48:32,717 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED={AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredictedPushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter}}  
2022-04-25 15:48:32,728 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 15:48:32,730 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for data: $0, $3, $4  
2022-04-25 15:48:32,732 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false  
2022-04-25 15:48:32,733 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1  
2022-04-25 15:48:32,733 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1  
2022-04-25 15:48:32,741 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 15:48:32,742 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 15:48:32,744 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job  
2022-04-25 15:48:32,744 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
2022-04-25 15:48:32,745 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.  
2022-04-25 15:48:32,745 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator  
3:49 PM 4/25/2022
```

```
mura0007@oracle18:~  
2022-04-25 15:48:33,472 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:  
HadoopVersion PigVersion UserId StartedAt FinishedAt Features  
2.9.2 0.17.0 mura0007 2022-04-25 15:48:33 2022-04-25 15:48:33 GROUP_BY,FILTER  
  
Success!  
  
Job Stats (time in seconds):  
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime A  
alias Feature Outputs  
job_local288220558_0004 1 1 n/a n/a n/a n/a n/a n/a data,grou,new,result,sub_data GROUP_BY file:  
/tmp/temp1402191447/tmp717571281,  
  
Input(s):  
Successfully read 213 records from: "file:///oracle/stuhomes/mura0007/HW4-twitter_account.csv"  
  
Output(s):  
Successfully stored 1 records in: "file:/tmp/temp1402191447/tmp717571281"  
  
Counters:  
Total records written : 1  
Total bytes written : 0  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_local288220558_0004  
  
2022-04-25 15:48:33,473 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 15:48:33,474 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 15:48:33,474 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 15:48:33,477 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2022-04-25 15:48:33,477 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 15:48:33,477 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2022-04-25 15:48:33,486 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1  
2022-04-25 15:48:33,486 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(430, {((rww6pa3x0y@ou.edu))})  
grunt>
```

## QUERY B:

```
mura0007@oracle18:~  
2022-04-25 16:20:26,294 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58  
2022-04-25 16:20:26,294 [main] INFO org.apache.pig.Main - Logging error messages to: /oracle/stuhomes/mura0007/pig_1650921626290.log  
2022-04-25 16:20:26,311 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /oracle/stuhomes/mura0007/.pigbootup not found  
2022-04-25 16:20:26,399 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.a  
ddress  
2022-04-25 16:20:26,400 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///br  
2022-04-25 16:20:26,614 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check  
sum  
2022-04-25 16:20:26,629 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-2cb2d3a9-d529-4a70-844c-e989aeab9fff  
2022-04-25 16:20:26,629 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false  
grunt> clear  
  
grunt> exec query_b.pig  
2022-04-25 16:20:42,202 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check  
sum  
2022-04-25 16:20:42,543 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check  
sum  
2022-04-25 16:20:42,685 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check  
sum  
2022-04-25 16:20:42,705 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapredu  
ce.output.txtoutputformat.separator  
2022-04-25 16:20:42,717 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY  
2022-04-25 16:20:42,732 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check  
sum  
2022-04-25 16:20:42,761 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - (RULES ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantC  
alculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, Predi  
catePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter])  
2022-04-25 16:20:42,812 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699400192 to monitor. collectionUs  
ageThreshold = 409500128, usageThreshold = 409500128  
2022-04-25 16:20:42,857 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic?  
false  
2022-04-25 16:20:42,869 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combine  
r  
2022-04-25 16:20:42,904 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1  
2022-04-25 16:20:42,904 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1  
2022-04-25 16:20:42,922 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check  
sum  
2022-04-25 16:20:42,932 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - session.id is deprecated. Instead, use dfs.metrics.session-id  
2022-04-25 16:20:42,968 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Initializing JVM Metrics with processName=JobTracker, sessionId=  
2022-04-25 16:20:42,992 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job  
2022-04-25 16:20:42,996 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, us  
e mapreduce.reduce.markreset.buffer.percent  
2022-04-25 16:20:42,996 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.per
```

Windows Type here to search 422 PM 4/25/2022

```
mura0007@oracle18:~  
Output(s):  
Successfully stored 15 records in: "file:/tmp/temp-1762407979/tmp980271540"  
  
Counters:  
Total records written : 15  
Total bytes written : 0  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
  
Job DAG:  
job_local151389184_0002  
  
2022-04-25 16:20:44,805 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea  
dy initialized  
2022-04-25 16:20:44,806 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea  
dy initialized  
2022-04-25 16:20:44,806 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea  
dy initialized  
2022-04-25 16:20:44,810 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYP  
E_CONVERSION FAILED 11 time(s).  
2022-04-25 16:20:44,810 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2022-04-25 16:20:44,811 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check  
sum  
2022-04-25 16:20:44,811 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2022-04-25 16:20:44,823 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1  
2022-04-25 16:20:44,823 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(Miami FL,15)  
(Dallas TX,14)  
(Lawton OK,18)  
(Norman OK,17)  
(Houston TX,19)  
(Trenton NJ,14)  
(Wichita KS,16)  
(Bismarck ND,18)  
(Corsicana TX,17)  
(Rochester NY,17)  
(Pittsburgh PA,14)  
(Kansas City MO,13)  
(Tallahassee FL,9)  
(Fort Lauderdale,11)  
(Fort Lauderdale FL,1)  
grunt>
```

Windows Type here to search 421 PM 4/25/2022

## QUERY C:

```
mura0007@oracle18:~  
(Trenton NJ,14)  
(Wichita KS,16)  
(Bismarck ND,18)  
(Corsicana TX,17)  
(Rochester NY,17)  
(Pittsburgh PA,14)  
(Kansas City MO,13)  
(Tallahassee FL,9)  
(Fort Lauderdale,11)  
(Fort Lauderdale FL,1)  
grunt> clear  
  
grunt> exec -param numtweets=150 query_c.pig  
2022-04-25 17:47:16,532 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 17:47:16,547 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 17:47:16,567 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 17:47:16,679 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 17:47:16,701 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN, GROUP_BY, FILTER  
2022-04-25 17:47:16,712 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 17:47:16,712 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.  
2022-04-25 17:47:16,712 [main] INFO org.apache.hadoop.mapred.newplan.logical.optimizer.LogicalPlanOptimizer - (RULES ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter])  
2022-04-25 17:47:16,719 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false  
2022-04-25 17:47:16,721 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combine first  
2022-04-25 17:47:16,723 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler$LastInputStreamingOptimizer - Rewrite: POackage->POForEach to POPackage(JoinPacker)  
2022-04-25 17:47:16,723 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 2  
2022-04-25 17:47:16,723 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 2  
2022-04-25 17:47:16,730 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 17:47:16,731 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 17:47:16,733 [main] INFO org.apache.pig.tools.pigstats.MRScriptState - Pig script settings are added to the job  
2022-04-25 17:47:16,733 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3  
5:48 PM 4/25/2022 10  
  
mura0007@oracle18:~  
dy initialized  
2022-04-25 17:47:18,281 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 17:47:18,283 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 11 time(s).  
2022-04-25 17:47:18,283 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2022-04-25 17:47:18,284 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum  
2022-04-25 17:47:18,284 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2022-04-25 17:47:18,293 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1  
2022-04-25 17:47:18,293 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(156,63.0)  
(161,8.0)  
(175,86.0)  
(182,88.0)  
(184,8.0)  
(191,38.0)  
(193,61.0)  
(214,55.0)  
(218,36.0)  
(219,2.0)  
(223,56.0)  
(231,65.0)  
(235,40.5)  
(240,8.0)  
(243,64.0)  
(245,85.0)  
(251,5.0)  
(253,48.0)  
(257,59.0)  
(259,89.0)  
(260,2.0)  
(262,12.0)  
(270,78.0)  
(278,29.0)  
(281,97.0)  
(287,70.5)  
(291,84.0)  
(296,57.0)  
(297,80.0)  
(299,59.0)  
(301,34.0)  
(302,8.0)  
(305,90.0)  
5:48 PM 4/25/2022 10
```

```

mura0007@oracle18:~ 
(699, 97.0)
(702, 19.0)
(703, 21.0)
(709, 34.0)
(710, 11.0)
(711, 63.0)
(715, 18.0)
(716, 46.0)
(723, 71.0)
(729, 71.0)
(727, 69.0)
(760, 97.0)
(772, 11.5)
(774, 99.0)
(783, 38.0)
(794, 27.5)
(806, 51.0)
(810, 29.0)
(812, 62.0)
(823, 17.0)
(833, 16.0)
(835, 27.0)
(841, 91.0)
(851, 13.0)
(866, 14.0)
(883, 95.0)
(884, 42.0)
(903, 80.0)
(907, 22.0)
(916, 20.0)
(926, 83.0)
(929, 1.0)
(931, 19.0)
(936, 39.0)
(939, 2.0)
(942, 33.0)
(946, 16.0)
(968, 41.0)
(979, 43.0)
(991, 29.0)
(993, 80.0)
(994, 23.0)
(995, 62.5)
grunt> 

mura0007@oracle18:~ 
(916, 20.0)
(926, 83.0)
(929, 1.0)
(931, 19.0)
(936, 39.0)
(939, 2.0)
(942, 33.0)
(946, 16.0)
(968, 41.0)
(979, 43.0)
(991, 29.0)
(993, 80.0)
(994, 23.0)
(995, 62.5)
grunt> clear

grunt> exec -param numtweets=300 query_c.pig
2022-04-25 17:51:24,666 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:51:24,677 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:51:24,691 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:51:24,763 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:51:24,777 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN,GROUP_BY,FILTER
2022-04-25 17:51:24,785 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:51:24,786 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-04-25 17:51:24,786 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2022-04-25 17:51:24,789 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-04-25 17:51:24,790 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combine
r
2022-04-25 17:51:24,791 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage ->POForEach to POackage(JoinPackager)
2022-04-25 17:51:24,791 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 2
2022-04-25 17:51:24,791 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 2
2022-04-25 17:51:24,797 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:51:24,797 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already
5:48 PM 66°F 4/25/2022 10
mura0007@oracle18:~ 
(916, 20.0)
(926, 83.0)
(929, 1.0)
(931, 19.0)
(936, 39.0)
(939, 2.0)
(942, 33.0)
(946, 16.0)
(968, 41.0)
(979, 43.0)
(991, 29.0)
(993, 80.0)
(994, 23.0)
(995, 62.5)
grunt> clear

mura0007@oracle18:~ 
(916, 20.0)
(926, 83.0)
(929, 1.0)
(931, 19.0)
(936, 39.0)
(939, 2.0)
(942, 33.0)
(946, 16.0)
(968, 41.0)
(979, 43.0)
(991, 29.0)
(993, 80.0)
(994, 23.0)
(995, 62.5)
grunt> clear

mura0007@oracle18:~ 
(916, 20.0)
(926, 83.0)
(929, 1.0)
(931, 19.0)
(936, 39.0)
(939, 2.0)
(942, 33.0)
(946, 16.0)
(968, 41.0)
(979, 43.0)
(991, 29.0)
(993, 80.0)
(994, 23.0)
(995, 62.5)
5:52 PM 66°F 4/25/2022 10

```

```
mura0007@oracle18:~  
dy initialized  
2022-04-25 17:51:26,133 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 17:51:26,134 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 17:51:26,135 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 17:51:26,136 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 17:51:26,136 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized  
2022-04-25 17:51:26,137 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED in time(s).  
2022-04-25 17:51:26,137 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2022-04-25 17:51:26,138 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum.  
2022-04-25 17:51:26,138 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2022-04-25 17:51:26,146 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1  
2022-04-25 17:51:26,146 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(301,34.0)  
(302,8.0)  
(305,90.0)  
(317,11.0)  
(328,17.0)  
(334,21.0)  
(345,24.0)  
(346,57.0)  
(347,43.0)  
(355,27.0)  
(360,55.0)  
(361,33.0)  
(369,65.5)  
(370,60.0)  
(375,99.0)  
(377,28.0)  
(378,99.0)  
(385,38.0)  
(393,91.0)  
(401,46.0)  
(403,61.0)  
(404,90.0)  
(414,51.0)  
(420,35.0)  
(437,0.0)
```



```
mura0007@oracle18:~  
(699,97.0)  
(702,19.0)  
(703,21.0)  
(709,34.0)  
(710,11.0)  
(711,83.0)  
(719,18.0)  
(716,46.0)  
(723,71.0)  
(725,71.0)  
(727,69.0)  
(760,97.0)  
(772,11.5)  
(774,99.0)  
(783,38.0)  
(794,27.5)  
(806,51.0)  
(810,29.0)  
(812,82.0)  
(823,17.0)  
(833,16.0)  
(835,27.0)  
(841,91.0)  
(851,13.0)  
(866,14.0)  
(883,95.0)  
(884,42.0)  
(903,80.0)  
(907,22.0)  
(916,20.0)  
(926,83.0)  
(929,1.0)  
(931,19.0)  
(936,39.0)  
(939,2.0)  
(942,33.0)  
(946,16.0)  
(968,41.0)  
(979,43.0)  
(991,29.0)  
(993,80.0)  
(994,23.0)  
(995,62.5)  
grunt> █
```



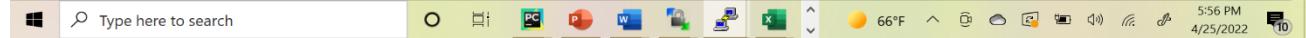
```

mura0007@oracle18:~ 
(931,19.0)
(936,59.0)
(939,2.0)
(942,33.0)
(946,16.0)
(960,41.0)
(979,43.0)
(991,29.0)
(993,80.0)
(994,23.0)
(995,62.5)
grunt> clear

grunt> exec -param numtweets=500 query_c.pig
2022-04-25 17:56:06,500 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:56:06,511 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:56:06,525 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:56:06,584 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:56:06,597 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH JOIN, GROUP BY, FILTER
2022-04-25 17:56:06,604 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:56:06,604 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-04-25 17:56:06,604 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2022-04-25 17:56:06,607 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic?
false
2022-04-25 17:56:06,608 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combine
r
2022-04-25 17:56:06,609 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler$LastInputStreamingOptimizer - Rewrite: POOPackage->POForEach to POOPackage(JoinPackager)
2022-04-25 17:56:06,609 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 2
2022-04-25 17:56:06,609 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 2
2022-04-25 17:56:06,614 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:56:06,615 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 17:56:06,616 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2022-04-25 17:56:06,616 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.per
5:56 PM
4/25/2022
mura0007@oracle18:~ 
(502,19.0)
(503,59.0)
(504,2.0)
(505,33.0)
(506,16.0)
(507,41.0)
(508,43.0)
(509,29.0)
(510,80.0)
(511,23.0)
(512,62.5)
(513,14.0)
(514,50.0)
(515,22.0)
(516,75.0)
(517,73.0)
(518,92.0)
(519,23.5)
(520,11.0)
(521,73.0)
(522,0.0)
(523,79.0)
(524,69.0)
(525,94.0)
(526,75.0)
(527,16.5)
(528,87.0)
(529,83.0)
(530,76.0)
(531,41.0)
(532,47.0)
(533,87.0)
(534,57.0)
(535,17.0)
(536,77.0)
(537,75.0)
(538,50.0)
(539,19.0)
2022-04-25 17:56:07,877 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 17:56:07,879 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 17:56:07,880 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 17:56:07,880 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2022-04-25 17:56:07,881 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 11 time(s).
2022-04-25 17:56:07,881 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-04-25 17:56:07,882 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-04-25 17:56:07,882 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-25 17:56:07,890 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-04-25 17:56:07,890 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(508,62.0)
(512,14.0)
(514,50.0)
(516,22.0)
(519,75.0)
(521,73.0)
(524,92.0)
(526,23.5)
(528,11.0)
(529,73.0)
(530,0.0)
(531,79.0)
(532,69.0)
(533,94.0)
(534,75.0)
(535,16.5)
(536,87.0)
(537,83.0)
(538,76.0)
(539,41.0)
(540,47.0)
(541,87.0)
(542,57.0)
(543,17.0)
(544,77.0)
(545,75.0)
(546,50.0)
(547,19.0)
5:56 PM
4/25/2022

```

```
mura0007@oracle18:~  
(699,97.0)  
(702,19.0)  
(703,21.0)  
(708,34.0)  
(710,11.0)  
(711,63.0)  
(719,18.0)  
(716,46.0)  
(723,71.0)  
(729,71.0)  
(727,69.0)  
(760,97.0)  
(772,11.5)  
(774,99.0)  
(783,38.0)  
(794,27.5)  
(806,51.0)  
(810,29.0)  
(812,62.0)  
(823,17.0)  
(833,16.0)  
(839,27.0)  
(841,91.0)  
(851,13.0)  
(866,14.0)  
(883,95.0)  
(884,42.0)  
(903,80.0)  
(907,22.0)  
(916,20.0)  
(926,63.0)  
(929,1.0)  
(931,19.0)  
(936,39.0)  
(939,2.0)  
(942,33.0)  
(946,16.0)  
(968,41.0)  
(979,43.0)  
(991,29.0)  
(993,80.0)  
(994,23.0)  
(995,62.5)  
grunt> █
```



## TASK 2:

```
mura0007@oracle18:~  
2022-04-27 02:36:21,044 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2022-04-27 02:36:21,044 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check  
sum  
2022-04-27 02:36:21,044 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized  
2022-04-27 02:36:21,052 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1  
2022-04-27 02:36:21,052 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(1239301, 0.0016242645744773976)  
(14936610, 0.004026547660689662)  
(19652836, 0.00820152754151653)  
(18848018, 0.008322103006705706)  
(20728561, 0.004955224985699292)  
(24182811, 0.004694836134169723)  
(35369214, 0.007546754461620854)  
(40302738, 0.0027825068671723)  
(44312605, 9.968701495586495e-4)  
(46948334, 0.006167422459677741)  
(50042330, 0.008064004996524578)  
(50570449, 0.0059965130439400445)  
(77000938, 0.00554143063820678)  
(90084099, 0.005046621735275526)  
(100322679, 0.00818270813135444)  
(110260678, 0.0012436497845760962)  
(131482972, 0.004547935181856279)  
(145910123, 0.005685139169620502)  
(165090253, 0.0028171913259234637)  
(167063179, 0.004694836134169723)  
(180463340, 0.015874687702328204)  
(180717062, 0.008115668775210484)  
(184426481, 0.007121455328021)  
(186353014, 0.003349278869448004)  
(195066320, 0.00576555295149095)  
(232414166, 0.003349700873170637)  
(243298366, 0.004413582111485571)  
(249950079, 0.007112043423624105)  
(269930499, 0.00920118413007234)  
(270231980, 0.007807887617693755)  
(270673659, 0.010440776825928456)  
(271136534, 0.005511153256348926)  
(271515500, 0.006487737885365764)  
(271658840, 0.00984943458190626)  
(271753649, 0.0067092535409961195)  
(272177272, 0.004189319920371904)  
(273022647, 0.00664145819148192)  
(273069036, 0.005143541722575296)
```

2:37 AM  
4/27/2022

```
mura0007@oracle18:~  
(40396946, 0.002881329462909466)  
(407030795, 0.0015285769326715448)  
(411000199, 0.001566912748149448)  
(411786274, 0.0024517071939249356)  
(4122443067, 0.00132313193973977244)  
(426907745, 0.003783431130543736)  
(434460832, 0.0022415792130437108)  
(435923452, 0.002241521973358008)  
(442334304, 0.0019039720588103002)  
(446783544, 0.004024043807496614)  
(453290398, 0.0016094849508795372)  
(455060666, 0.0030441682777768942)  
(456042335, 0.0010541915519047828)  
(456760820, 7.2664451401809448e-4)  
(458845777, 0.002946532793632773)  
(463999696, 0.002496739776436275)  
(464435768, 0.0035355948488563944)  
(465472520, 0.0010900794355825703)  
(466121896, 0.0035571257481229034)  
(466311355, 0.0023218857329289086)  
(467753757, 0.0025498618443839725)  
(478817325, 0.0022695692307229684)  
(488806995, 0.0022586721975335743)  
(497334912, 0.0019199170382512743)  
(506982155, 0.002892616734354793)  
(510896241, 0.001995557386565029)  
(512620911, 0.002488713504974417)  
(512638904, 0.0026789413934399376)  
(512896378, 0.0032108213297056274)  
(519281688, 0.002344025813292806)  
(521112781, 8.643169589254607B-4)  
(523832656, 0.0020209708515017476)  
(524620711, 0.0014002867295153333)  
(532562821, 0.00342784164842899)  
(533836053, 0.0016719287508044868)  
(536893070, 0.0022524809573247107)  
(540748208, 0.0015211433542595253)  
(543844138, 8.66316211978873B-4)  
(545106865, 0.0017200179535598503)  
(554003471, 0.001658816830442934)  
(554402185, 0.0018225322376770557)  
(555800132, 0.0024872411721139865)  
(563853564, 0.0016869647183944247)  
grunt>
```

2:38 AM  
4/27/2022