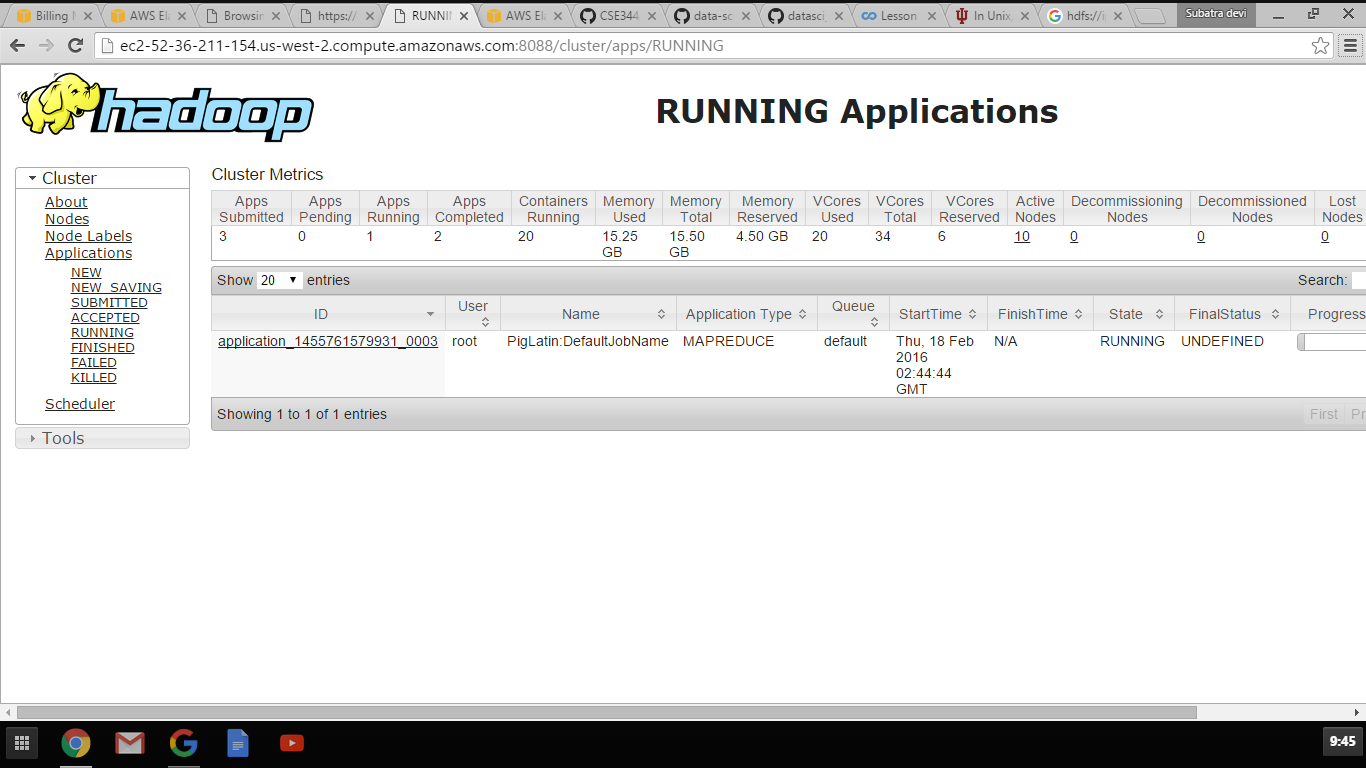
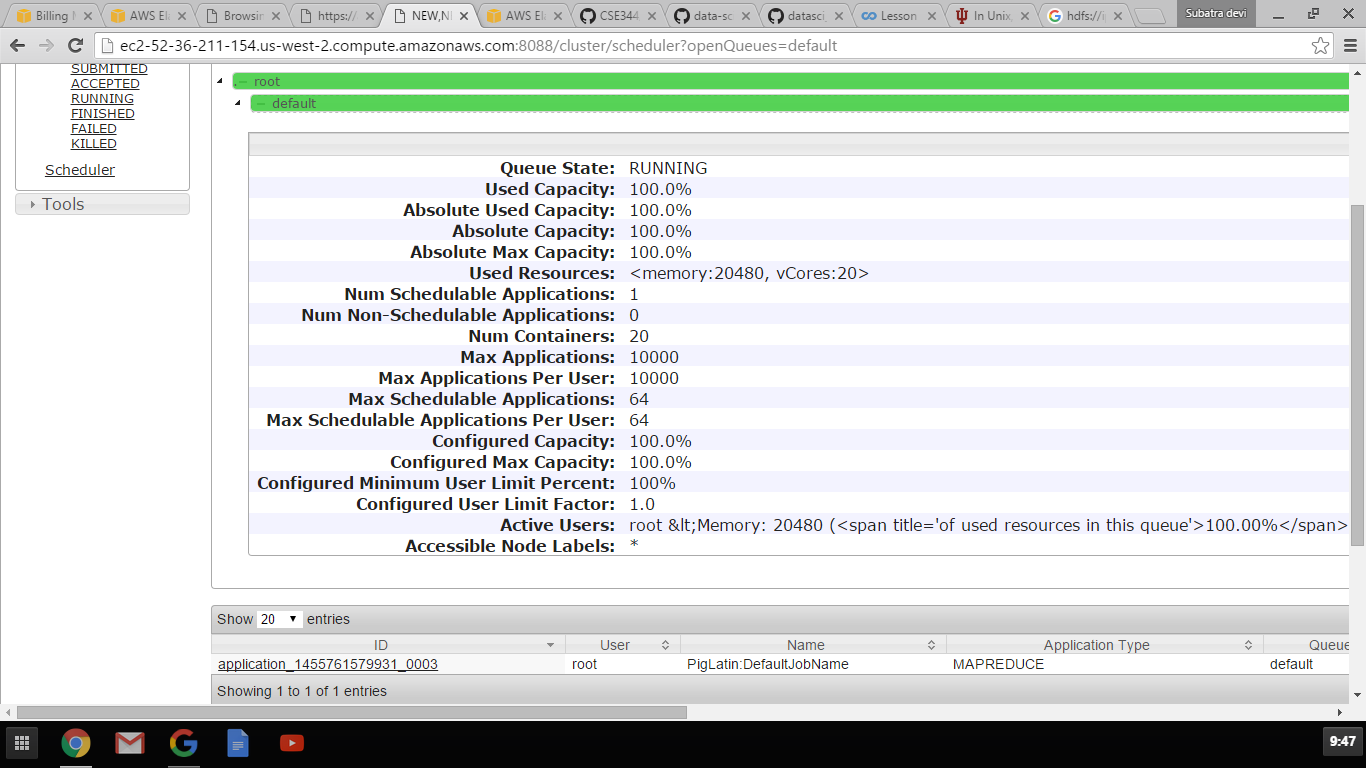
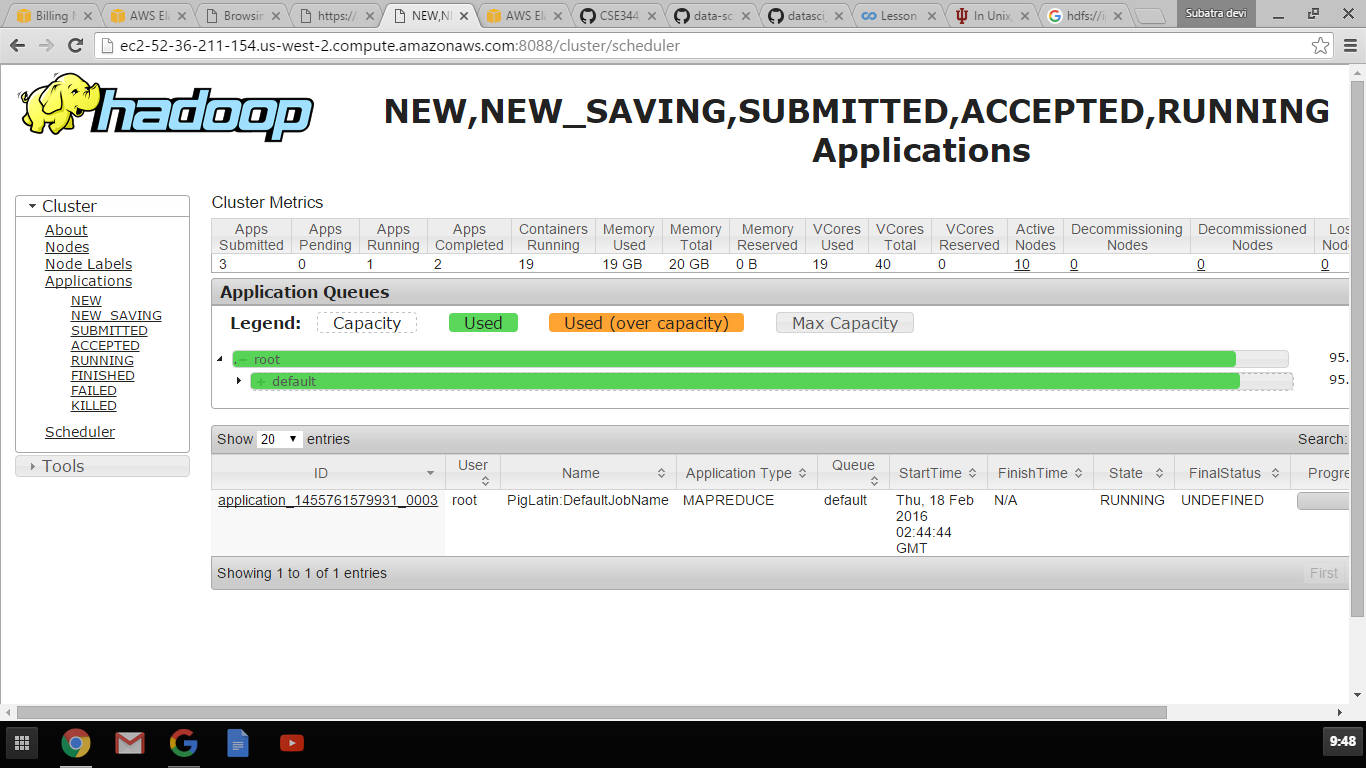


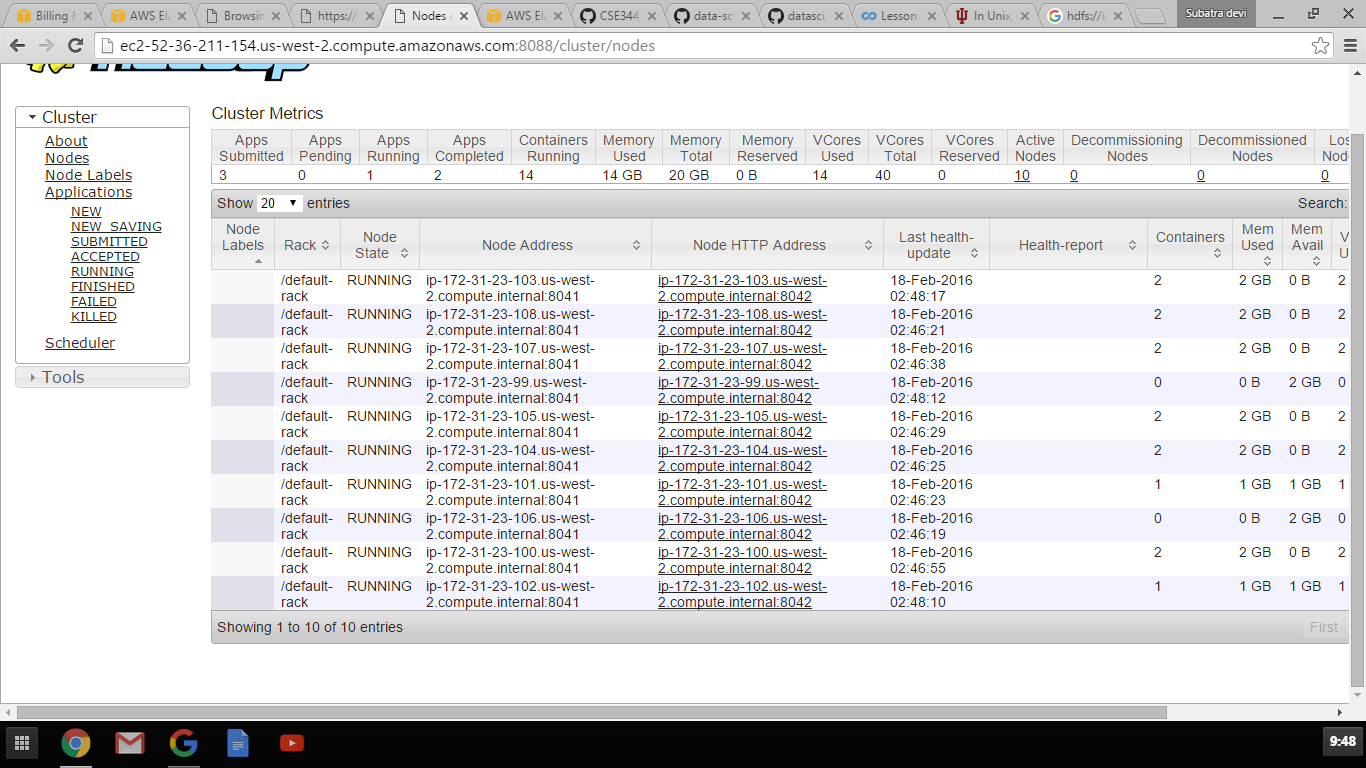
Problem 2B:

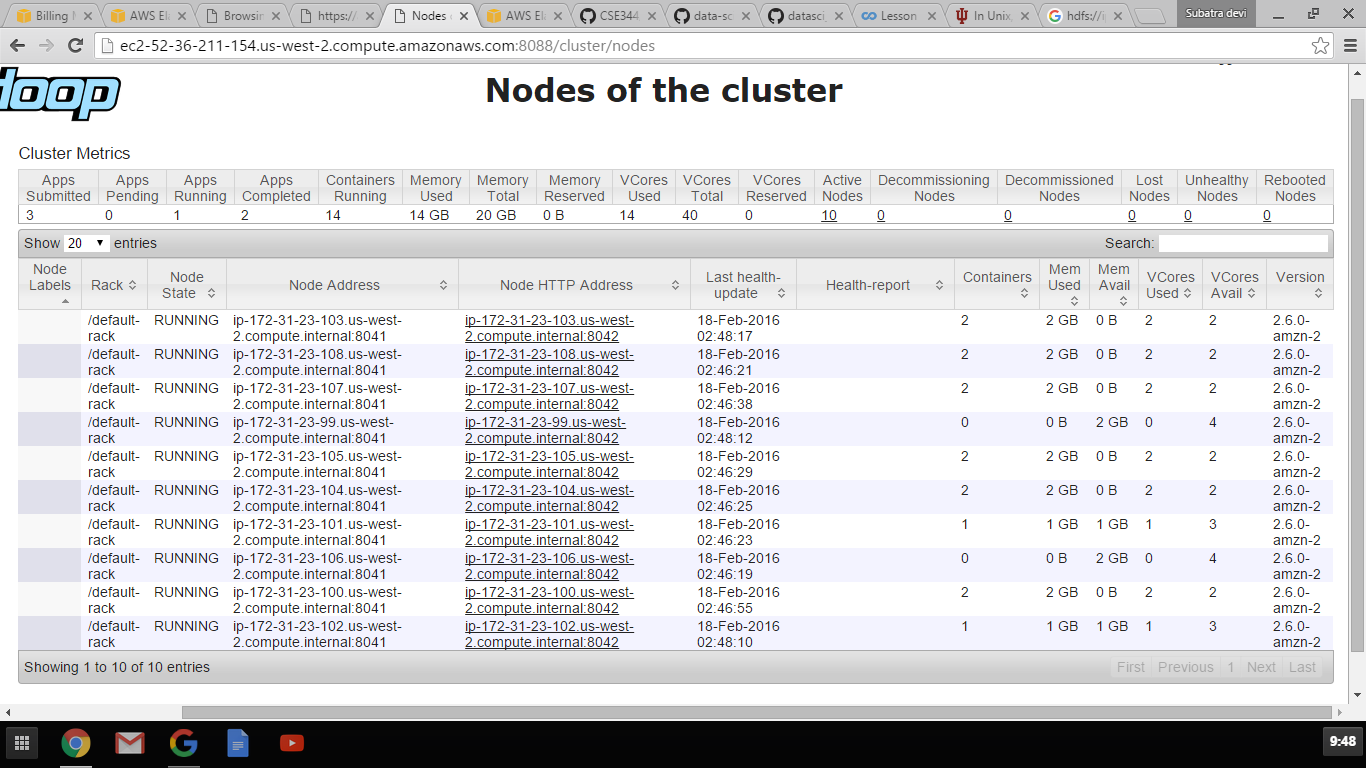


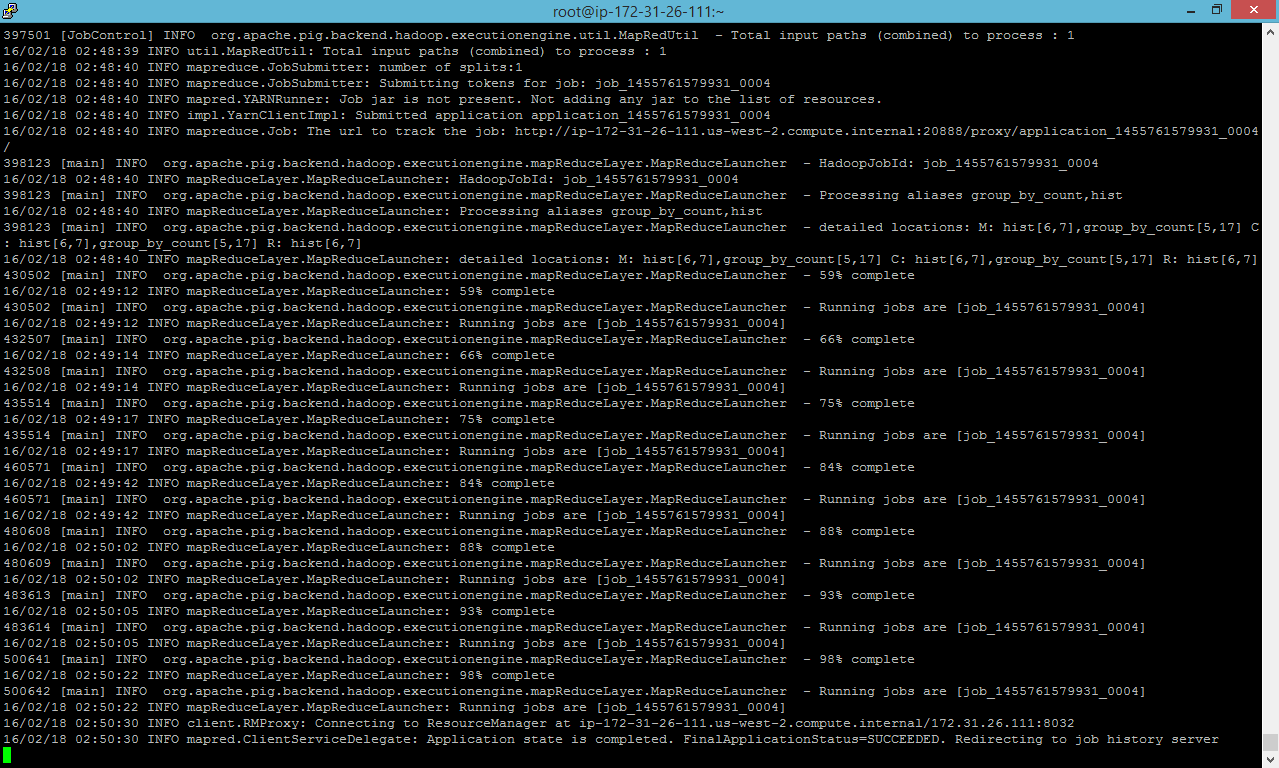


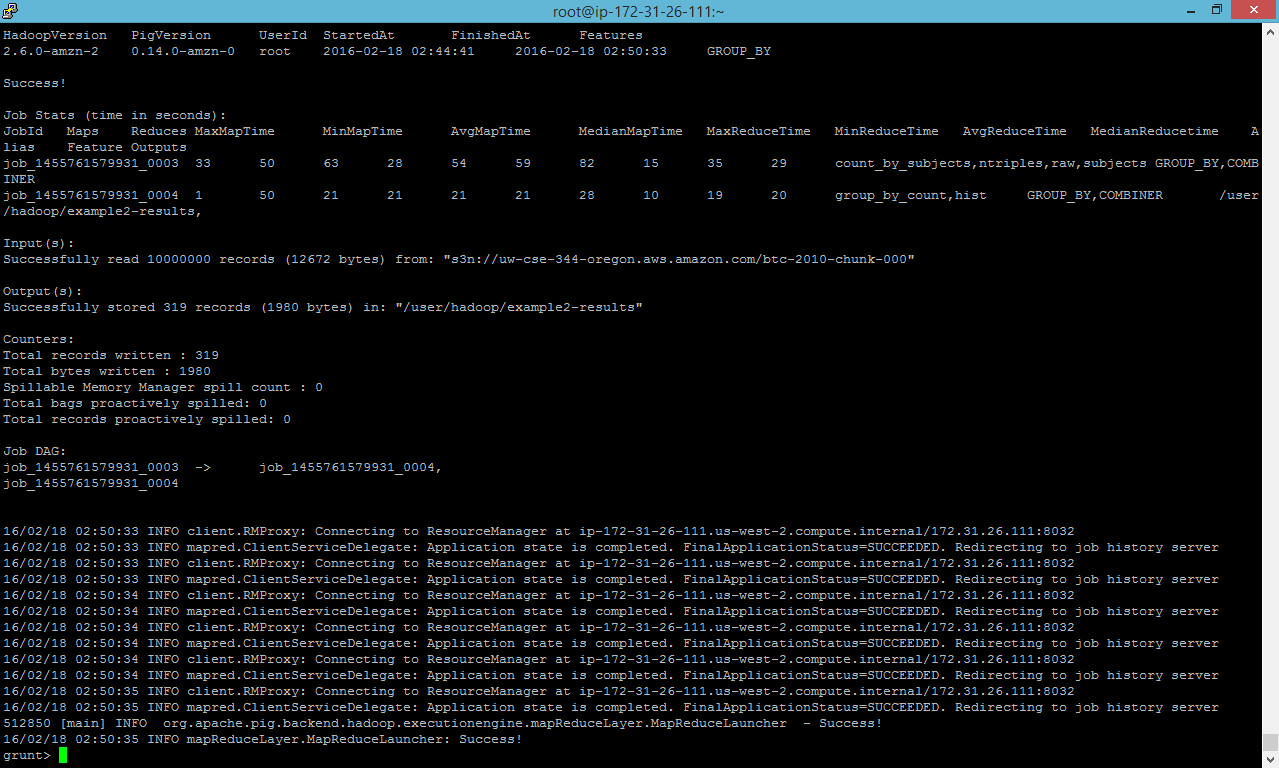




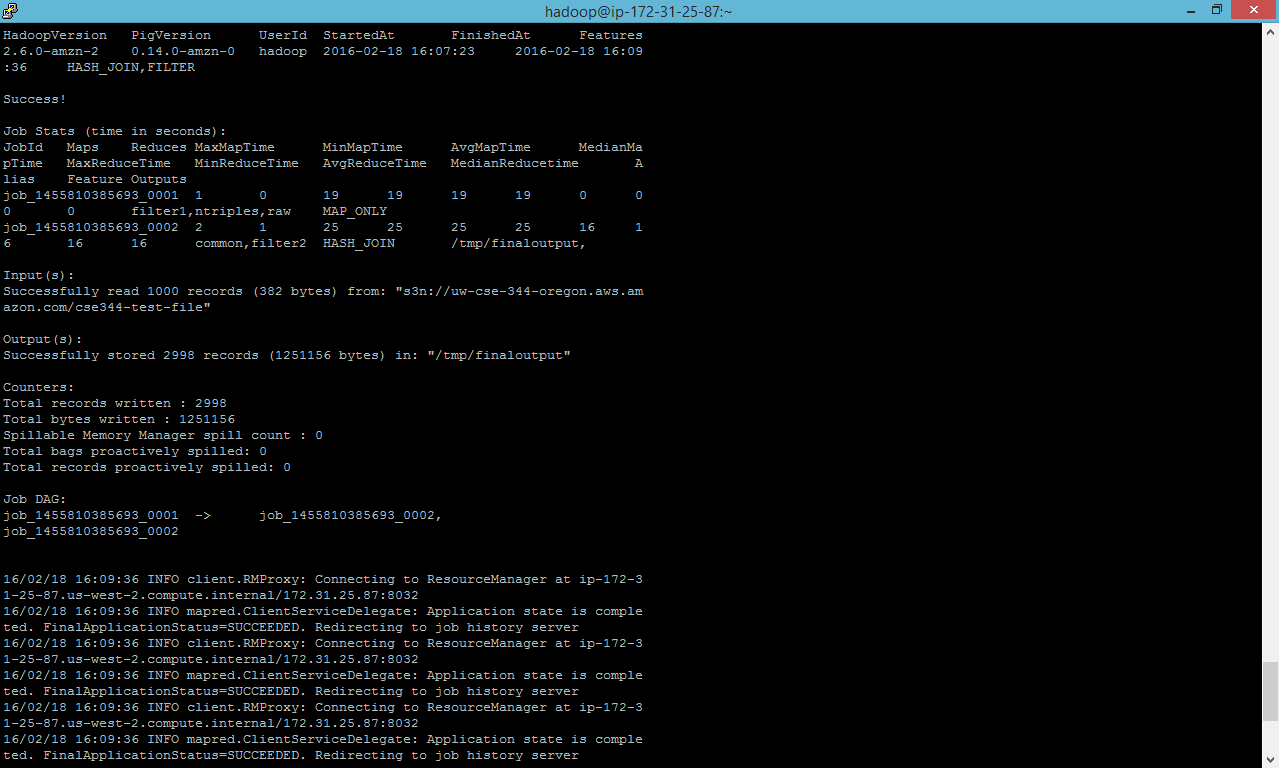








Problem 3A



Using username "hadoop".

Authenticating with public key "imported-openssh-key"

Last login: Thu Feb 18 15:50:36 2016

\_\_| \_\_|\_ )

\_| ( / Amazon Linux AMI

\_\_\_|\\_\_\_|\_\_\_|

https://aws.amazon.com/amazon-linux-ami/2015.09-release-notes/

41 package(s) needed for security, out of 69 available

Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRR

E::::::::::::::::::E M:::::::M M:::::::M R::::::::::::::R

EE:::::EEEEEEEEE:::E M::::::::M M::::::::M R:::::RRRRRR:::::R

E::::E EEEEE M:::::::::M M:::::::::M RR::::R R::::R

E::::E M::::::M:::M M:::M::::::M R:::R R::::R

E:::::EEEEEEEEEE M:::::M M:::M M:::M M:::::M R:::RRRRRR:::::R

E::::::::::::::E M:::::M M:::M:::M M:::::M R:::::::::::RR

E:::::EEEEEEEEEE M:::::M M:::::M M:::::M R:::RRRRRR::::R

E::::E M:::::M M:::M M:::::M R:::R R::::R

E::::E EEEEE M:::::M MMM M:::::M R:::R R::::R

EE:::::EEEEEEEE::::E M:::::M M:::::M R:::R R::::R

E::::::::::::::::::E M:::::M M:::::M RR::::R R::::R

EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-25-87 ~]$ pig

16/02/18 15:51:23 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL

16/02/18 15:51:23 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE

16/02/18 15:51:23 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType

92 [main] INFO org.apache.pig.Main - Apache Pig version 0.14.0-amzn-0 (r: un known) compiled Nov 16 2015, 21:51:33

16/02/18 15:51:23 INFO pig.Main: Apache Pig version 0.14.0-amzn-0 (r: unknown) c ompiled Nov 16 2015, 21:51:33

96 [main] INFO org.apache.pig.Main - Logging error messages to: /mnt/var/log /pig/pig\_1455810683627.log

16/02/18 15:51:23 INFO pig.Main: Logging error messages to: /mnt/var/log/pig/pig \_1455810683627.log

138 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/ha doop/.pigbootup not found

16/02/18 15:51:23 INFO util.Utils: Default bootup file /home/hadoop/.pigbootup n ot found

16/02/18 15:51:25 INFO Configuration.deprecation: mapred.job.tracker is deprecat ed. Instead, use mapreduce.jobtracker.address

16/02/18 15:51:25 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

1789 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://ip-172-31-25-87.us-west-2.comput e.internal:8020

16/02/18 15:51:25 INFO executionengine.HExecutionEngine: Connecting to hadoop fi le system at: hdfs://ip-172-31-25-87.us-west-2.compute.internal:8020

16/02/18 15:51:27 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> register s3n://uw-cse-344-oregon.aws.amazon.com/myudfs.jar

16/02/18 16:03:24 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:03:27 INFO fs.EmrFileSystem: Consistency disabled, using com.amazon. ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation

16/02/18 16:03:28 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInClust er: false instanceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEng ine: false maxMemoryMb: 3072 maxInstanceCount: 500 lastModified: 1455810399473

16/02/18 16:03:28 INFO metrics.MetricsSaver: Created MetricsSaver j-1LVOGTR4LHQB S:i-ff1c9638:RunJar:07979 period:60 /mnt/var/em/raw/i-ff1c9638\_20160218\_RunJar\_0 7979\_raw.bin

16/02/18 16:03:29 INFO s3n.S3NativeFileSystem: Opening 's3n://uw-cse-344-oregon. aws.amazon.com/myudfs.jar' for reading

16/02/18 16:03:29 INFO metrics.MetricsSaver: Thread 1 created MetricsLockFreeSav er 1

grunt> register s3n://uw-cse-344-oregon.aws.amazon.com/myudfs.jar

grunt> raw = LOAD 's3n://uw-cse-344-oregon.aws.amazon.com/cse344-test-file' USIN G TextLoader as (line:chararray);

16/02/18 16:06:30 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:06:30 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> ntriples = foreach raw generate FLATTEN(myudfs.RDFSplit3(line)) as (subje ct:chararray,predicate:chararray,object:chararray);

grunt> filter1 = FILTER ntriples BY subject matches '.\*business.\*';

grunt> filter2 = FOREACH filter1 GENERATE subject AS subject2, predicate AS pred icate2, object AS object2;

grunt> common = JOIN filter1 BY subject, filter2 BY subject2;

grunt> store common into '/tmp/finaloutput' using PigStorage();

16/02/18 16:07:20 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:07:21 INFO Configuration.deprecation: mapred.textoutputformat.separa tor is deprecated. Instead, use mapreduce.output.textoutputformat.separator

957732 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features us ed in the script: HASH\_JOIN,FILTER

16/02/18 16:07:21 INFO pigstats.ScriptState: Pig features used in the script: HA SH\_JOIN,FILTER

16/02/18 16:07:21 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

957916 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematup le] was not set... will not generate code.

16/02/18 16:07:21 INFO data.SchemaTupleBackend: Key [pig.schematuple] was not se t... will not generate code.

958109 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimize r - {RULES\_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByC onstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForE ach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatte n, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

16/02/18 16:07:21 INFO optimizer.LogicalPlanOptimizer: {RULES\_ENABLED=[AddForEac h, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptim izer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

958795 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MRCompiler - File concatenation threshold: 100 optimistic? false

16/02/18 16:07:22 INFO mapReduceLayer.MRCompiler: File concatenation threshold: 100 optimistic? false

958911 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage->POForEach to POPa ckage(JoinPackager)

16/02/18 16:07:22 INFO mapReduceLayer.MRCompiler$LastInputStreamingOptimizer: Re write: POPackage->POForEach to POPackage(JoinPackager)

958935 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MultiQueryOptimizer - MR plan size before optimization: 2

16/02/18 16:07:22 INFO mapReduceLayer.MultiQueryOptimizer: MR plan size before o ptimization: 2

958936 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MultiQueryOptimizer - Merged 0 out of total 2 MR operators.

16/02/18 16:07:22 INFO mapReduceLayer.MultiQueryOptimizer: Merged 0 out of total 2 MR operators.

958936 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MultiQueryOptimizer - MR plan size after optimization: 2

16/02/18 16:07:22 INFO mapReduceLayer.MultiQueryOptimizer: MR plan size after op timization: 2

16/02/18 16:07:22 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:07:22 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

960376 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

16/02/18 16:07:23 INFO mapreduce.MRScriptState: Pig script settings are added to the job

16/02/18 16:07:23 INFO Configuration.deprecation: mapred.job.reduce.markreset.bu ffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.perce nt

960407 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, se t to default 0.3

16/02/18 16:07:23 INFO mapReduceLayer.JobControlCompiler: mapred.job.reduce.mark reset.buffer.percent is not set, set to default 0.3

16/02/18 16:07:23 INFO Configuration.deprecation: mapred.output.compress is depr ecated. Instead, use mapreduce.output.fileoutputformat.compress

960427 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - This job cannot be converted run in-process

16/02/18 16:07:23 INFO mapReduceLayer.JobControlCompiler: This job cannot be con verted run in-process

961421 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/tmp/pig7958780584195752722tmp/myudfs.jar to DistributedCache through /tmp/temp749736356/tmp-58790875/myudfs.jar

16/02/18 16:07:24 INFO mapReduceLayer.JobControlCompiler: Added jar file:/tmp/pi g7958780584195752722tmp/myudfs.jar to DistributedCache through /tmp/temp74973635 6/tmp-58790875/myudfs.jar

961687 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp749736356/tmp1013510181/pig-0.14.0-amzn-0- core-h2.jar

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp7497363 56/tmp1013510181/pig-0.14.0-amzn-0-core-h2.jar

961759 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to D istributedCache through /tmp/temp749736356/tmp-813607174/automaton-1.11-8.jar

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp749736356/tm p-813607174/automaton-1.11-8.jar

961841 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp749736356/tmp-1844073509/antlr-runtime-3.4.jar

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp749736356/t mp-1844073509/antlr-runtime-3.4.jar

961954 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/hadoop/lib/guava-11.0.2.jar to Di stributedCache through /tmp/temp749736356/tmp1104406152/guava-11.0.2.jar

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop/lib/guava-11.0.2.jar to DistributedCache through /tmp/temp749736356/tmp 1104406152/guava-11.0.2.jar

962067 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/hadoop-mapreduce/joda-time-2.8.1. jar to DistributedCache through /tmp/temp749736356/tmp1008008783/joda-time-2.8.1 .jar

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop-mapreduce/joda-time-2.8.1.jar to DistributedCache through /tmp/temp7497 36356/tmp1008008783/joda-time-2.8.1.jar

962387 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Setting up single store job

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Setting up single stor e job

962463 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematu ple] is false, will not generate code.

16/02/18 16:07:26 INFO data.SchemaTupleFrontend: Key [pig.schematuple] is false, will not generate code.

962464 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

16/02/18 16:07:26 INFO data.SchemaTupleFrontend: Starting process to move genera ted code to distributed cacche

962467 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig. schematuple.classes] with classes to deserialize []

16/02/18 16:07:26 INFO data.SchemaTupleFrontend: Setting key [pig.schematuple.cl asses] with classes to deserialize []

962756 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

16/02/18 16:07:26 INFO mapReduceLayer.MapReduceLauncher: 1 map-reduce job(s) wai ting for submission.

16/02/18 16:07:26 INFO Configuration.deprecation: mapred.job.tracker.http.addres s is deprecated. Instead, use mapreduce.jobtracker.http.address

16/02/18 16:07:26 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:07:26 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:07:26 WARN mapreduce.JobSubmitter: No job jar file set. User classe s may not be found. See Job or Job#setJar(String).

16/02/18 16:07:27 INFO input.FileInputFormat: Total input paths to process : 1

963904 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths to process : 1

16/02/18 16:07:27 INFO util.MapRedUtil: Total input paths to process : 1

16/02/18 16:07:27 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library

16/02/18 16:07:27 INFO lzo.LzoCodec: Successfully loaded & initialized native-lz o library [hadoop-lzo rev 41f4e6be3ac5d6676a3464f77de79a33e8fdd9f3]

964094 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths (combined) to process : 1

16/02/18 16:07:27 INFO util.MapRedUtil: Total input paths (combined) to process : 1

16/02/18 16:07:27 INFO metrics.MetricsSaver: 1 aggregated HDFSWriteDelay 109 raw values into 2 aggregated values, total 2

16/02/18 16:07:27 INFO mapreduce.JobSubmitter: number of splits:1

16/02/18 16:07:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_14 55810385693\_0001

16/02/18 16:07:29 INFO mapred.YARNRunner: Job jar is not present. Not adding any jar to the list of resources.

16/02/18 16:07:30 INFO impl.YarnClientImpl: Submitted application application\_14 55810385693\_0001

16/02/18 16:07:30 INFO mapreduce.Job: The url to track the job: http://ip-172-31 -25-87.us-west-2.compute.internal:20888/proxy/application\_1455810385693\_0001/

967325 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - HadoopJobId: job\_1455810385693\_0001

16/02/18 16:07:30 INFO mapReduceLayer.MapReduceLauncher: HadoopJobId: job\_145581 0385693\_0001

967325 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Processing aliases filter1,ntriples,raw

16/02/18 16:07:30 INFO mapReduceLayer.MapReduceLauncher: Processing aliases filt er1,ntriples,raw

967325 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - detailed locations: M: raw[1,6],ntriples[-1,-1],filter1[3, 10] C: R:

16/02/18 16:07:30 INFO mapReduceLayer.MapReduceLauncher: detailed locations: M: raw[1,6],ntriples[-1,-1],filter1[3,10] C: R:

967344 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 0% complete

16/02/18 16:07:30 INFO mapReduceLayer.MapReduceLauncher: 0% complete

967344 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Running jobs are [job\_1455810385693\_0001]

16/02/18 16:07:30 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0001]

1007144 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - 25% complete

16/02/18 16:08:10 INFO mapReduceLayer.MapReduceLauncher: 25% complete

1007144 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - Running jobs are [job\_1455810385693\_0001]

16/02/18 16:08:10 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0001]

1012655 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - 50% complete

16/02/18 16:08:16 INFO mapReduceLayer.MapReduceLauncher: 50% complete

16/02/18 16:08:16 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:08:16 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:08:17 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:08:17 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:08:18 INFO Configuration.deprecation: mapred.reduce.tasks is depreca ted. Instead, use mapreduce.job.reduces

16/02/18 16:08:18 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:08:18 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

1015181 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pi g script settings are added to the job

16/02/18 16:08:18 INFO mapreduce.MRScriptState: Pig script settings are added to the job

1015183 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, s et to default 0.3

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: mapred.job.reduce.mark reset.buffer.percent is not set, set to default 0.3

1015189 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Reduce phase detected, estimating # of required reducers .

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: Reduce phase detected, estimating # of required reducers.

1015193 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.e xecutionengine.mapReduceLayer.InputSizeReducerEstimator

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: Using reducer estimato r: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducer Estimator

1015250 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalI nputFileSize=509324

16/02/18 16:08:18 INFO mapReduceLayer.InputSizeReducerEstimator: BytesPerReducer =1000000000 maxReducers=999 totalInputFileSize=509324

1015251 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Setting Parallelism to 1

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: Setting Parallelism to 1

1015251 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - This job cannot be converted run in-process

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: This job cannot be con verted run in-process

1015336 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/tmp/pig7958780584195752722tmp/myudfs.jar to DistributedCache through /tmp/temp749736356/tmp-1440031377/myudfs.jar

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: Added jar file:/tmp/pi g7958780584195752722tmp/myudfs.jar to DistributedCache through /tmp/temp74973635 6/tmp-1440031377/myudfs.jar

1015492 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/usr/lib/pig/pig-0.14.0-amzn-0-core-h2.ja r to DistributedCache through /tmp/temp749736356/tmp-1356663748/pig-0.14.0-amzn- 0-core-h2.jar

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp7497363 56/tmp-1356663748/pig-0.14.0-amzn-0-core-h2.jar

1015558 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp749736356/tmp-1444631415/automaton-1.11-8.jar

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp749736356/tm p-1444631415/automaton-1.11-8.jar

1015614 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp749736356/tmp-1658175706/antlr-runtime-3.4.ja r

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp749736356/t mp-1658175706/antlr-runtime-3.4.jar

1015701 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/usr/lib/hadoop/lib/guava-11.0.2.jar to D istributedCache through /tmp/temp749736356/tmp1071424277/guava-11.0.2.jar

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop/lib/guava-11.0.2.jar to DistributedCache through /tmp/temp749736356/tmp 1071424277/guava-11.0.2.jar

1015767 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/usr/lib/hadoop-mapreduce/joda-time-2.8.1 .jar to DistributedCache through /tmp/temp749736356/tmp1722450719/joda-time-2.8. 1.jar

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop-mapreduce/joda-time-2.8.1.jar to DistributedCache through /tmp/temp7497 36356/tmp1722450719/joda-time-2.8.1.jar

1015821 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Setting up single store job

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Setting up single stor e job

1015822 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schemat uple] is false, will not generate code.

16/02/18 16:08:19 INFO data.SchemaTupleFrontend: Key [pig.schematuple] is false, will not generate code.

1015823 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

16/02/18 16:08:19 INFO data.SchemaTupleFrontend: Starting process to move genera ted code to distributed cacche

1015823 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig .schematuple.classes] with classes to deserialize []

16/02/18 16:08:19 INFO data.SchemaTupleFrontend: Setting key [pig.schematuple.cl asses] with classes to deserialize []

1016005 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

16/02/18 16:08:19 INFO mapReduceLayer.MapReduceLauncher: 1 map-reduce job(s) wai ting for submission.

16/02/18 16:08:19 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:08:19 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:08:19 WARN mapreduce.JobSubmitter: No job jar file set. User classe s may not be found. See Job or Job#setJar(String).

16/02/18 16:08:19 INFO input.FileInputFormat: Total input paths to process : 1

1016149 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Ma pRedUtil - Total input paths to process : 1

16/02/18 16:08:19 INFO util.MapRedUtil: Total input paths to process : 1

1016149 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Ma pRedUtil - Total input paths (combined) to process : 1

16/02/18 16:08:19 INFO util.MapRedUtil: Total input paths (combined) to process : 1

16/02/18 16:08:19 INFO input.FileInputFormat: Total input paths to process : 1

1016159 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Ma pRedUtil - Total input paths to process : 1

16/02/18 16:08:19 INFO util.MapRedUtil: Total input paths to process : 1

1016159 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Ma pRedUtil - Total input paths (combined) to process : 1

16/02/18 16:08:19 INFO util.MapRedUtil: Total input paths (combined) to process : 1

16/02/18 16:08:19 INFO mapreduce.JobSubmitter: number of splits:2

16/02/18 16:08:20 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_14 55810385693\_0002

16/02/18 16:08:20 INFO mapred.YARNRunner: Job jar is not present. Not adding any jar to the list of resources.

16/02/18 16:08:20 INFO impl.YarnClientImpl: Submitted application application\_14 55810385693\_0002

16/02/18 16:08:20 INFO mapreduce.Job: The url to track the job: http://ip-172-31 -25-87.us-west-2.compute.internal:20888/proxy/application\_1455810385693\_0002/

1016614 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - HadoopJobId: job\_1455810385693\_0002

16/02/18 16:08:20 INFO mapReduceLayer.MapReduceLauncher: HadoopJobId: job\_145581 0385693\_0002

1016614 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - Processing aliases common,filter2

16/02/18 16:08:20 INFO mapReduceLayer.MapReduceLauncher: Processing aliases comm on,filter2

1016614 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - detailed locations: M: common[5,9],filter2[4,10],common[5 ,9] C: R:

16/02/18 16:08:20 INFO mapReduceLayer.MapReduceLauncher: detailed locations: M: common[5,9],filter2[4,10],common[5,9] C: R:

1063874 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - 75% complete

16/02/18 16:09:07 INFO mapReduceLayer.MapReduceLauncher: 75% complete

1063874 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - Running jobs are [job\_1455810385693\_0002]

16/02/18 16:09:07 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0002]

1083906 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - Running jobs are [job\_1455810385693\_0002]

16/02/18 16:09:27 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0002]

16/02/18 16:09:35 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:35 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

1092964 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - 100% complete

16/02/18 16:09:36 INFO mapReduceLayer.MapReduceLauncher: 100% complete

1092993 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - S cript Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features

2.6.0-amzn-2 0.14.0-amzn-0 hadoop 2016-02-18 16:07:23 2016-02-18 16:09 :36 HASH\_JOIN,FILTER

Success!

Job Stats (time in seconds):

JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMa pTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A lias Feature Outputs

job\_1455810385693\_0001 1 0 19 19 19 19 0 0 0 0 filter1,ntriples,raw MAP\_ONLY

job\_1455810385693\_0002 2 1 25 25 25 25 16 1 6 16 16 common,filter2 HASH\_JOIN /tmp/finaloutput,

Input(s):

Successfully read 1000 records (382 bytes) from: "s3n://uw-cse-344-oregon.aws.am azon.com/cse344-test-file"

Output(s):

Successfully stored 2998 records (1251156 bytes) in: "/tmp/finaloutput"

Counters:

Total records written : 2998

Total bytes written : 1251156

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1455810385693\_0001 -> job\_1455810385693\_0002,

job\_1455810385693\_0002

16/02/18 16:09:36 INFO mapreduce.SimplePigStats: Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features

2.6.0-amzn-2 0.14.0-amzn-0 hadoop 2016-02-18 16:07:23 2016-02-18 16:09 :36 HASH\_JOIN,FILTER

Success!

Job Stats (time in seconds):

JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMa pTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A lias Feature Outputs

job\_1455810385693\_0001 1 0 19 19 19 19 0 0 0 0 filter1,ntriples,raw MAP\_ONLY

job\_1455810385693\_0002 2 1 25 25 25 25 16 1 6 16 16 common,filter2 HASH\_JOIN /tmp/finaloutput,

Input(s):

Successfully read 1000 records (382 bytes) from: "s3n://uw-cse-344-oregon.aws.am azon.com/cse344-test-file"

Output(s):

Successfully stored 2998 records (1251156 bytes) in: "/tmp/finaloutput"

Counters:

Total records written : 2998

Total bytes written : 1251156

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1455810385693\_0001 -> job\_1455810385693\_0002,

job\_1455810385693\_0002

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:37 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:37 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:37 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:37 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

1093974 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - Success!

16/02/18 16:09:37 INFO mapReduceLayer.MapReduceLauncher: Success!

grunt> quit

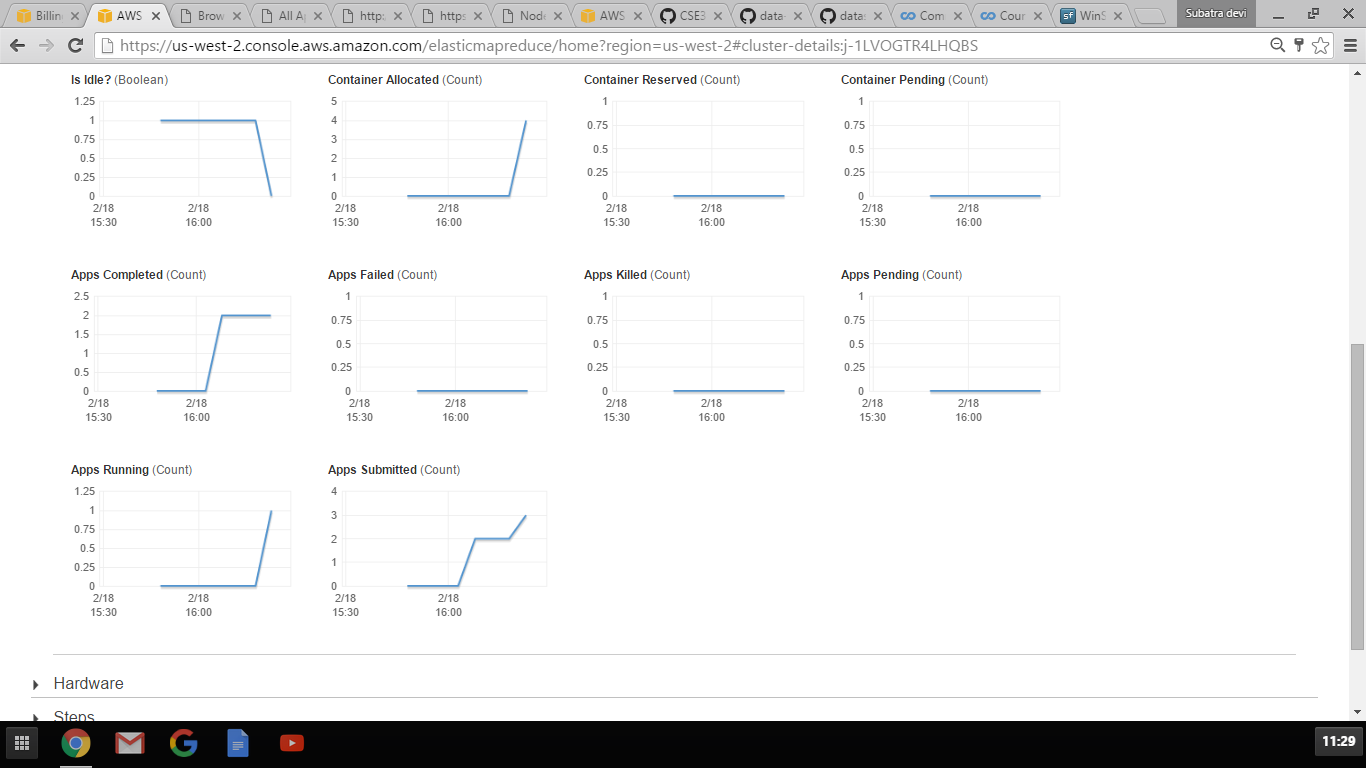
1225772 [main] INFO org.apache.pig.Main - Pig script completed in 20 minutes, 25 seconds and 967 milliseconds (1225967 ms)

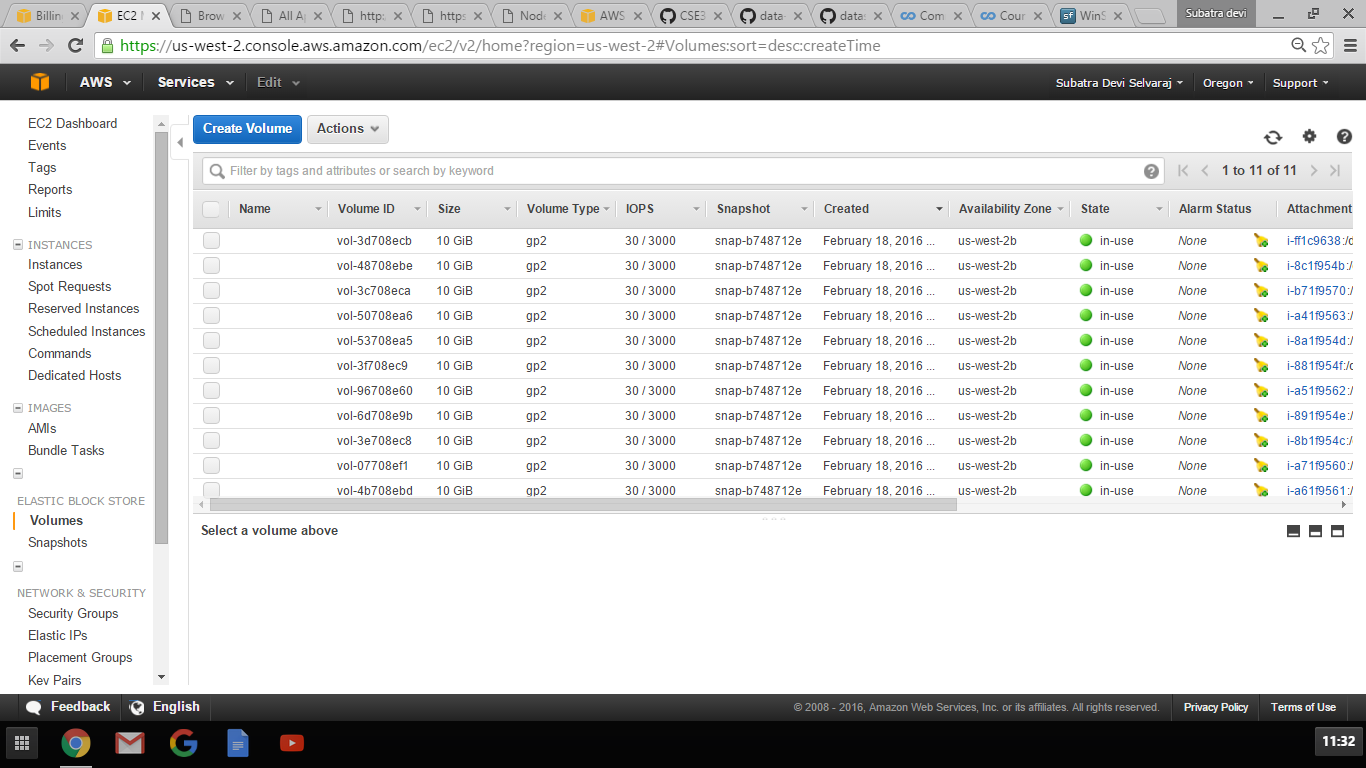
16/02/18 16:11:49 INFO pig.Main: Pig script completed in 20 minutes, 25 seconds and 967 milliseconds (1225967 ms)

[hadoop@ip-172-31-25-87 ~]$ hdfs dfs -getmerge /tmp/finaloutput /tmp/problem3A-r esults

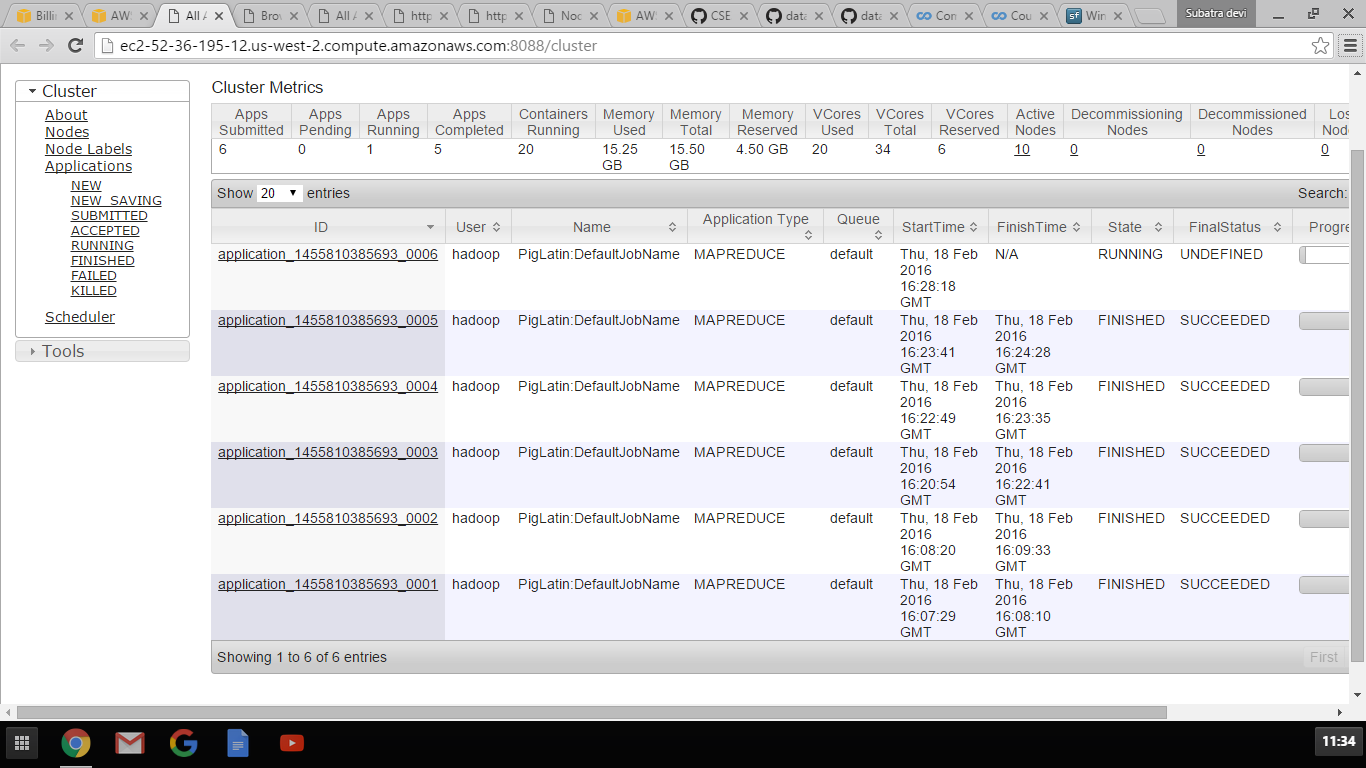
[hadoop@ip-172-31-25-87 ~]$

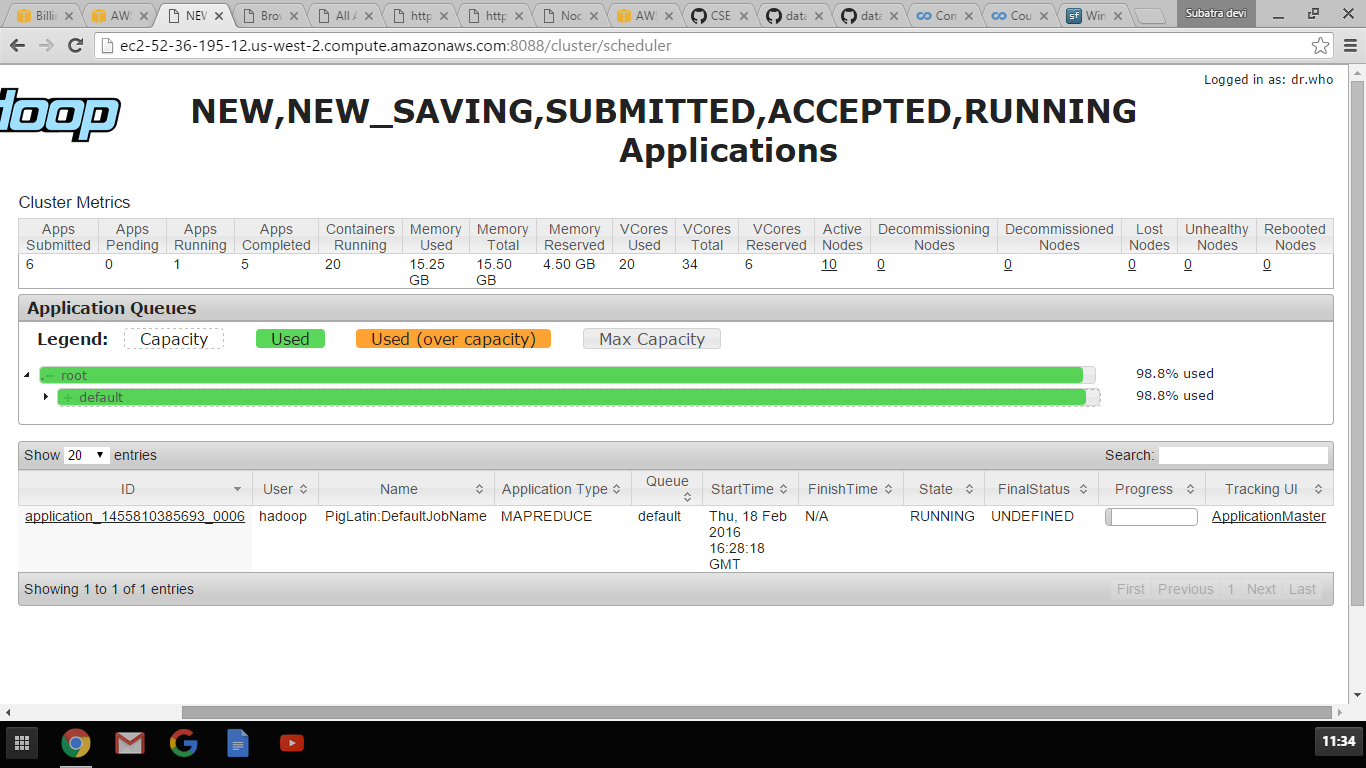
Problem 4 :

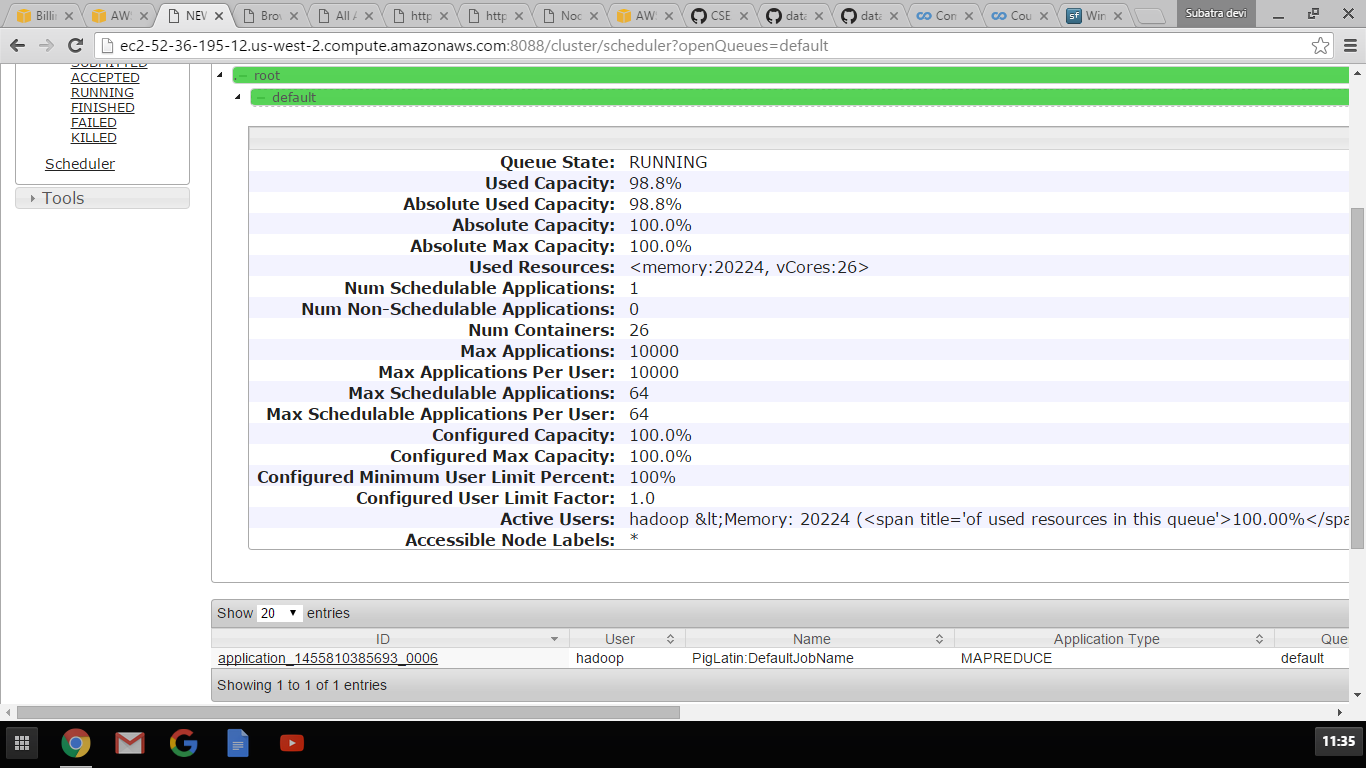




Resource Manager:







Using username "hadoop".

Authenticating with public key "imported-openssh-key"

Last login: Thu Feb 18 15:50:36 2016

\_\_| \_\_|\_ )

\_| ( / Amazon Linux AMI

\_\_\_|\\_\_\_|\_\_\_|

https://aws.amazon.com/amazon-linux-ami/2015.09-release-notes/

41 package(s) needed for security, out of 69 available

Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRR

E::::::::::::::::::E M:::::::M M:::::::M R::::::::::::::R

EE:::::EEEEEEEEE:::E M::::::::M M::::::::M R:::::RRRRRR:::::R

E::::E EEEEE M:::::::::M M:::::::::M RR::::R R::::R

E::::E M::::::M:::M M:::M::::::M R:::R R::::R

E:::::EEEEEEEEEE M:::::M M:::M M:::M M:::::M R:::RRRRRR:::::R

E::::::::::::::E M:::::M M:::M:::M M:::::M R:::::::::::RR

E:::::EEEEEEEEEE M:::::M M:::::M M:::::M R:::RRRRRR::::R

E::::E M:::::M M:::M M:::::M R:::R R::::R

E::::E EEEEE M:::::M MMM M:::::M R:::R R::::R

EE:::::EEEEEEEE::::E M:::::M M:::::M R:::R R::::R

E::::::::::::::::::E M:::::M M:::::M RR::::R R::::R

EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-25-87 ~]$ pig

16/02/18 15:51:23 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL

16/02/18 15:51:23 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE

16/02/18 15:51:23 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType

92 [main] INFO org.apache.pig.Main - Apache Pig version 0.14.0-amzn-0 (r: un known) compiled Nov 16 2015, 21:51:33

16/02/18 15:51:23 INFO pig.Main: Apache Pig version 0.14.0-amzn-0 (r: unknown) c ompiled Nov 16 2015, 21:51:33

96 [main] INFO org.apache.pig.Main - Logging error messages to: /mnt/var/log /pig/pig\_1455810683627.log

16/02/18 15:51:23 INFO pig.Main: Logging error messages to: /mnt/var/log/pig/pig \_1455810683627.log

138 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/ha doop/.pigbootup not found

16/02/18 15:51:23 INFO util.Utils: Default bootup file /home/hadoop/.pigbootup n ot found

16/02/18 15:51:25 INFO Configuration.deprecation: mapred.job.tracker is deprecat ed. Instead, use mapreduce.jobtracker.address

16/02/18 15:51:25 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

1789 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://ip-172-31-25-87.us-west-2.comput e.internal:8020

16/02/18 15:51:25 INFO executionengine.HExecutionEngine: Connecting to hadoop fi le system at: hdfs://ip-172-31-25-87.us-west-2.compute.internal:8020

16/02/18 15:51:27 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> register s3n://uw-cse-344-oregon.aws.amazon.com/myudfs.jar

16/02/18 16:03:24 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:03:27 INFO fs.EmrFileSystem: Consistency disabled, using com.amazon. ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation

16/02/18 16:03:28 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInClust er: false instanceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEng ine: false maxMemoryMb: 3072 maxInstanceCount: 500 lastModified: 1455810399473

16/02/18 16:03:28 INFO metrics.MetricsSaver: Created MetricsSaver j-1LVOGTR4LHQB S:i-ff1c9638:RunJar:07979 period:60 /mnt/var/em/raw/i-ff1c9638\_20160218\_RunJar\_0 7979\_raw.bin

16/02/18 16:03:29 INFO s3n.S3NativeFileSystem: Opening 's3n://uw-cse-344-oregon. aws.amazon.com/myudfs.jar' for reading

16/02/18 16:03:29 INFO metrics.MetricsSaver: Thread 1 created MetricsLockFreeSav er 1

grunt> register s3n://uw-cse-344-oregon.aws.amazon.com/myudfs.jar

grunt> raw = LOAD 's3n://uw-cse-344-oregon.aws.amazon.com/cse344-test-file' USIN G TextLoader as (line:chararray);

16/02/18 16:06:30 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:06:30 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> ntriples = foreach raw generate FLATTEN(myudfs.RDFSplit3(line)) as (subje ct:chararray,predicate:chararray,object:chararray);

grunt> filter1 = FILTER ntriples BY subject matches '.\*business.\*';

grunt> filter2 = FOREACH filter1 GENERATE subject AS subject2, predicate AS pred icate2, object AS object2;

grunt> common = JOIN filter1 BY subject, filter2 BY subject2;

grunt> store common into '/tmp/finaloutput' using PigStorage();

16/02/18 16:07:20 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:07:21 INFO Configuration.deprecation: mapred.textoutputformat.separa tor is deprecated. Instead, use mapreduce.output.textoutputformat.separator

957732 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features us ed in the script: HASH\_JOIN,FILTER

16/02/18 16:07:21 INFO pigstats.ScriptState: Pig features used in the script: HA SH\_JOIN,FILTER

16/02/18 16:07:21 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

957916 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematup le] was not set... will not generate code.

16/02/18 16:07:21 INFO data.SchemaTupleBackend: Key [pig.schematuple] was not se t... will not generate code.

958109 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimize r - {RULES\_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByC onstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForE ach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatte n, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

16/02/18 16:07:21 INFO optimizer.LogicalPlanOptimizer: {RULES\_ENABLED=[AddForEac h, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptim izer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

958795 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MRCompiler - File concatenation threshold: 100 optimistic? false

16/02/18 16:07:22 INFO mapReduceLayer.MRCompiler: File concatenation threshold: 100 optimistic? false

958911 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage->POForEach to POPa ckage(JoinPackager)

16/02/18 16:07:22 INFO mapReduceLayer.MRCompiler$LastInputStreamingOptimizer: Re write: POPackage->POForEach to POPackage(JoinPackager)

958935 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MultiQueryOptimizer - MR plan size before optimization: 2

16/02/18 16:07:22 INFO mapReduceLayer.MultiQueryOptimizer: MR plan size before o ptimization: 2

958936 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MultiQueryOptimizer - Merged 0 out of total 2 MR operators.

16/02/18 16:07:22 INFO mapReduceLayer.MultiQueryOptimizer: Merged 0 out of total 2 MR operators.

958936 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MultiQueryOptimizer - MR plan size after optimization: 2

16/02/18 16:07:22 INFO mapReduceLayer.MultiQueryOptimizer: MR plan size after op timization: 2

16/02/18 16:07:22 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:07:22 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

960376 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

16/02/18 16:07:23 INFO mapreduce.MRScriptState: Pig script settings are added to the job

16/02/18 16:07:23 INFO Configuration.deprecation: mapred.job.reduce.markreset.bu ffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.perce nt

960407 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, se t to default 0.3

16/02/18 16:07:23 INFO mapReduceLayer.JobControlCompiler: mapred.job.reduce.mark reset.buffer.percent is not set, set to default 0.3

16/02/18 16:07:23 INFO Configuration.deprecation: mapred.output.compress is depr ecated. Instead, use mapreduce.output.fileoutputformat.compress

960427 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - This job cannot be converted run in-process

16/02/18 16:07:23 INFO mapReduceLayer.JobControlCompiler: This job cannot be con verted run in-process

961421 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/tmp/pig7958780584195752722tmp/myudfs.jar to DistributedCache through /tmp/temp749736356/tmp-58790875/myudfs.jar

16/02/18 16:07:24 INFO mapReduceLayer.JobControlCompiler: Added jar file:/tmp/pi g7958780584195752722tmp/myudfs.jar to DistributedCache through /tmp/temp74973635 6/tmp-58790875/myudfs.jar

961687 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp749736356/tmp1013510181/pig-0.14.0-amzn-0- core-h2.jar

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp7497363 56/tmp1013510181/pig-0.14.0-amzn-0-core-h2.jar

961759 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to D istributedCache through /tmp/temp749736356/tmp-813607174/automaton-1.11-8.jar

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp749736356/tm p-813607174/automaton-1.11-8.jar

961841 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp749736356/tmp-1844073509/antlr-runtime-3.4.jar

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp749736356/t mp-1844073509/antlr-runtime-3.4.jar

961954 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/hadoop/lib/guava-11.0.2.jar to Di stributedCache through /tmp/temp749736356/tmp1104406152/guava-11.0.2.jar

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop/lib/guava-11.0.2.jar to DistributedCache through /tmp/temp749736356/tmp 1104406152/guava-11.0.2.jar

962067 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/hadoop-mapreduce/joda-time-2.8.1. jar to DistributedCache through /tmp/temp749736356/tmp1008008783/joda-time-2.8.1 .jar

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop-mapreduce/joda-time-2.8.1.jar to DistributedCache through /tmp/temp7497 36356/tmp1008008783/joda-time-2.8.1.jar

962387 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Setting up single store job

16/02/18 16:07:25 INFO mapReduceLayer.JobControlCompiler: Setting up single stor e job

962463 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematu ple] is false, will not generate code.

16/02/18 16:07:26 INFO data.SchemaTupleFrontend: Key [pig.schematuple] is false, will not generate code.

962464 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

16/02/18 16:07:26 INFO data.SchemaTupleFrontend: Starting process to move genera ted code to distributed cacche

962467 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig. schematuple.classes] with classes to deserialize []

16/02/18 16:07:26 INFO data.SchemaTupleFrontend: Setting key [pig.schematuple.cl asses] with classes to deserialize []

962756 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

16/02/18 16:07:26 INFO mapReduceLayer.MapReduceLauncher: 1 map-reduce job(s) wai ting for submission.

16/02/18 16:07:26 INFO Configuration.deprecation: mapred.job.tracker.http.addres s is deprecated. Instead, use mapreduce.jobtracker.http.address

16/02/18 16:07:26 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:07:26 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:07:26 WARN mapreduce.JobSubmitter: No job jar file set. User classe s may not be found. See Job or Job#setJar(String).

16/02/18 16:07:27 INFO input.FileInputFormat: Total input paths to process : 1

963904 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths to process : 1

16/02/18 16:07:27 INFO util.MapRedUtil: Total input paths to process : 1

16/02/18 16:07:27 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library

16/02/18 16:07:27 INFO lzo.LzoCodec: Successfully loaded & initialized native-lz o library [hadoop-lzo rev 41f4e6be3ac5d6676a3464f77de79a33e8fdd9f3]

964094 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths (combined) to process : 1

16/02/18 16:07:27 INFO util.MapRedUtil: Total input paths (combined) to process : 1

16/02/18 16:07:27 INFO metrics.MetricsSaver: 1 aggregated HDFSWriteDelay 109 raw values into 2 aggregated values, total 2

16/02/18 16:07:27 INFO mapreduce.JobSubmitter: number of splits:1

16/02/18 16:07:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_14 55810385693\_0001

16/02/18 16:07:29 INFO mapred.YARNRunner: Job jar is not present. Not adding any jar to the list of resources.

16/02/18 16:07:30 INFO impl.YarnClientImpl: Submitted application application\_14 55810385693\_0001

16/02/18 16:07:30 INFO mapreduce.Job: The url to track the job: http://ip-172-31 -25-87.us-west-2.compute.internal:20888/proxy/application\_1455810385693\_0001/

967325 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - HadoopJobId: job\_1455810385693\_0001

16/02/18 16:07:30 INFO mapReduceLayer.MapReduceLauncher: HadoopJobId: job\_145581 0385693\_0001

967325 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Processing aliases filter1,ntriples,raw

16/02/18 16:07:30 INFO mapReduceLayer.MapReduceLauncher: Processing aliases filt er1,ntriples,raw

967325 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - detailed locations: M: raw[1,6],ntriples[-1,-1],filter1[3, 10] C: R:

16/02/18 16:07:30 INFO mapReduceLayer.MapReduceLauncher: detailed locations: M: raw[1,6],ntriples[-1,-1],filter1[3,10] C: R:

967344 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 0% complete

16/02/18 16:07:30 INFO mapReduceLayer.MapReduceLauncher: 0% complete

967344 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Running jobs are [job\_1455810385693\_0001]

16/02/18 16:07:30 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0001]

1007144 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - 25% complete

16/02/18 16:08:10 INFO mapReduceLayer.MapReduceLauncher: 25% complete

1007144 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - Running jobs are [job\_1455810385693\_0001]

16/02/18 16:08:10 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0001]

1012655 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - 50% complete

16/02/18 16:08:16 INFO mapReduceLayer.MapReduceLauncher: 50% complete

16/02/18 16:08:16 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:08:16 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:08:17 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:08:17 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:08:18 INFO Configuration.deprecation: mapred.reduce.tasks is depreca ted. Instead, use mapreduce.job.reduces

16/02/18 16:08:18 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:08:18 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

1015181 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pi g script settings are added to the job

16/02/18 16:08:18 INFO mapreduce.MRScriptState: Pig script settings are added to the job

1015183 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, s et to default 0.3

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: mapred.job.reduce.mark reset.buffer.percent is not set, set to default 0.3

1015189 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Reduce phase detected, estimating # of required reducers .

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: Reduce phase detected, estimating # of required reducers.

1015193 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.e xecutionengine.mapReduceLayer.InputSizeReducerEstimator

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: Using reducer estimato r: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducer Estimator

1015250 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalI nputFileSize=509324

16/02/18 16:08:18 INFO mapReduceLayer.InputSizeReducerEstimator: BytesPerReducer =1000000000 maxReducers=999 totalInputFileSize=509324

1015251 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Setting Parallelism to 1

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: Setting Parallelism to 1

1015251 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - This job cannot be converted run in-process

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: This job cannot be con verted run in-process

1015336 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/tmp/pig7958780584195752722tmp/myudfs.jar to DistributedCache through /tmp/temp749736356/tmp-1440031377/myudfs.jar

16/02/18 16:08:18 INFO mapReduceLayer.JobControlCompiler: Added jar file:/tmp/pi g7958780584195752722tmp/myudfs.jar to DistributedCache through /tmp/temp74973635 6/tmp-1440031377/myudfs.jar

1015492 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/usr/lib/pig/pig-0.14.0-amzn-0-core-h2.ja r to DistributedCache through /tmp/temp749736356/tmp-1356663748/pig-0.14.0-amzn- 0-core-h2.jar

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp7497363 56/tmp-1356663748/pig-0.14.0-amzn-0-core-h2.jar

1015558 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp749736356/tmp-1444631415/automaton-1.11-8.jar

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp749736356/tm p-1444631415/automaton-1.11-8.jar

1015614 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp749736356/tmp-1658175706/antlr-runtime-3.4.ja r

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp749736356/t mp-1658175706/antlr-runtime-3.4.jar

1015701 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/usr/lib/hadoop/lib/guava-11.0.2.jar to D istributedCache through /tmp/temp749736356/tmp1071424277/guava-11.0.2.jar

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop/lib/guava-11.0.2.jar to DistributedCache through /tmp/temp749736356/tmp 1071424277/guava-11.0.2.jar

1015767 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Added jar file:/usr/lib/hadoop-mapreduce/joda-time-2.8.1 .jar to DistributedCache through /tmp/temp749736356/tmp1722450719/joda-time-2.8. 1.jar

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop-mapreduce/joda-time-2.8.1.jar to DistributedCache through /tmp/temp7497 36356/tmp1722450719/joda-time-2.8.1.jar

1015821 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.JobControlCompiler - Setting up single store job

16/02/18 16:08:19 INFO mapReduceLayer.JobControlCompiler: Setting up single stor e job

1015822 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schemat uple] is false, will not generate code.

16/02/18 16:08:19 INFO data.SchemaTupleFrontend: Key [pig.schematuple] is false, will not generate code.

1015823 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

16/02/18 16:08:19 INFO data.SchemaTupleFrontend: Starting process to move genera ted code to distributed cacche

1015823 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig .schematuple.classes] with classes to deserialize []

16/02/18 16:08:19 INFO data.SchemaTupleFrontend: Setting key [pig.schematuple.cl asses] with classes to deserialize []

1016005 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

16/02/18 16:08:19 INFO mapReduceLayer.MapReduceLauncher: 1 map-reduce job(s) wai ting for submission.

16/02/18 16:08:19 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:08:19 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:08:19 WARN mapreduce.JobSubmitter: No job jar file set. User classe s may not be found. See Job or Job#setJar(String).

16/02/18 16:08:19 INFO input.FileInputFormat: Total input paths to process : 1

1016149 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Ma pRedUtil - Total input paths to process : 1

16/02/18 16:08:19 INFO util.MapRedUtil: Total input paths to process : 1

1016149 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Ma pRedUtil - Total input paths (combined) to process : 1

16/02/18 16:08:19 INFO util.MapRedUtil: Total input paths (combined) to process : 1

16/02/18 16:08:19 INFO input.FileInputFormat: Total input paths to process : 1

1016159 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Ma pRedUtil - Total input paths to process : 1

16/02/18 16:08:19 INFO util.MapRedUtil: Total input paths to process : 1

1016159 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Ma pRedUtil - Total input paths (combined) to process : 1

16/02/18 16:08:19 INFO util.MapRedUtil: Total input paths (combined) to process : 1

16/02/18 16:08:19 INFO mapreduce.JobSubmitter: number of splits:2

16/02/18 16:08:20 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_14 55810385693\_0002

16/02/18 16:08:20 INFO mapred.YARNRunner: Job jar is not present. Not adding any jar to the list of resources.

16/02/18 16:08:20 INFO impl.YarnClientImpl: Submitted application application\_14 55810385693\_0002

16/02/18 16:08:20 INFO mapreduce.Job: The url to track the job: http://ip-172-31 -25-87.us-west-2.compute.internal:20888/proxy/application\_1455810385693\_0002/

1016614 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - HadoopJobId: job\_1455810385693\_0002

16/02/18 16:08:20 INFO mapReduceLayer.MapReduceLauncher: HadoopJobId: job\_145581 0385693\_0002

1016614 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - Processing aliases common,filter2

16/02/18 16:08:20 INFO mapReduceLayer.MapReduceLauncher: Processing aliases comm on,filter2

1016614 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - detailed locations: M: common[5,9],filter2[4,10],common[5 ,9] C: R:

16/02/18 16:08:20 INFO mapReduceLayer.MapReduceLauncher: detailed locations: M: common[5,9],filter2[4,10],common[5,9] C: R:

1063874 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - 75% complete

16/02/18 16:09:07 INFO mapReduceLayer.MapReduceLauncher: 75% complete

1063874 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - Running jobs are [job\_1455810385693\_0002]

16/02/18 16:09:07 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0002]

1083906 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - Running jobs are [job\_1455810385693\_0002]

16/02/18 16:09:27 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0002]

16/02/18 16:09:35 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:35 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

1092964 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - 100% complete

16/02/18 16:09:36 INFO mapReduceLayer.MapReduceLauncher: 100% complete

1092993 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - S cript Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features

2.6.0-amzn-2 0.14.0-amzn-0 hadoop 2016-02-18 16:07:23 2016-02-18 16:09 :36 HASH\_JOIN,FILTER

Success!

Job Stats (time in seconds):

JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMa pTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A lias Feature Outputs

job\_1455810385693\_0001 1 0 19 19 19 19 0 0 0 0 filter1,ntriples,raw MAP\_ONLY

job\_1455810385693\_0002 2 1 25 25 25 25 16 1 6 16 16 common,filter2 HASH\_JOIN /tmp/finaloutput,

Input(s):

Successfully read 1000 records (382 bytes) from: "s3n://uw-cse-344-oregon.aws.am azon.com/cse344-test-file"

Output(s):

Successfully stored 2998 records (1251156 bytes) in: "/tmp/finaloutput"

Counters:

Total records written : 2998

Total bytes written : 1251156

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1455810385693\_0001 -> job\_1455810385693\_0002,

job\_1455810385693\_0002

16/02/18 16:09:36 INFO mapreduce.SimplePigStats: Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features

2.6.0-amzn-2 0.14.0-amzn-0 hadoop 2016-02-18 16:07:23 2016-02-18 16:09 :36 HASH\_JOIN,FILTER

Success!

Job Stats (time in seconds):

JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMa pTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A lias Feature Outputs

job\_1455810385693\_0001 1 0 19 19 19 19 0 0 0 0 filter1,ntriples,raw MAP\_ONLY

job\_1455810385693\_0002 2 1 25 25 25 25 16 1 6 16 16 common,filter2 HASH\_JOIN /tmp/finaloutput,

Input(s):

Successfully read 1000 records (382 bytes) from: "s3n://uw-cse-344-oregon.aws.am azon.com/cse344-test-file"

Output(s):

Successfully stored 2998 records (1251156 bytes) in: "/tmp/finaloutput"

Counters:

Total records written : 2998

Total bytes written : 1251156

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1455810385693\_0001 -> job\_1455810385693\_0002,

job\_1455810385693\_0002

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:36 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:37 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:37 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:09:37 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:09:37 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

1093974 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLaye r.MapReduceLauncher - Success!

16/02/18 16:09:37 INFO mapReduceLayer.MapReduceLauncher: Success!

grunt> quit

1225772 [main] INFO org.apache.pig.Main - Pig script completed in 20 minutes, 25 seconds and 967 milliseconds (1225967 ms)

16/02/18 16:11:49 INFO pig.Main: Pig script completed in 20 minutes, 25 seconds and 967 milliseconds (1225967 ms)

[hadoop@ip-172-31-25-87 ~]$ hdfs dfs -getmerge /tmp/finaloutput /tmp/problem3A-r esults

[hadoop@ip-172-31-25-87 ~]$ pig

16/02/18 16:18:33 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL

16/02/18 16:18:33 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE

16/02/18 16:18:33 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType

93 [main] INFO org.apache.pig.Main - Apache Pig version 0.14.0-amzn-0 (r: un known) compiled Nov 16 2015, 21:51:33

16/02/18 16:18:34 INFO pig.Main: Apache Pig version 0.14.0-amzn-0 (r: unknown) c ompiled Nov 16 2015, 21:51:33

95 [main] INFO org.apache.pig.Main - Logging error messages to: /mnt/var/log /pig/pig\_1455812314029.log

16/02/18 16:18:34 INFO pig.Main: Logging error messages to: /mnt/var/log/pig/pig \_1455812314029.log

136 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/ha doop/.pigbootup not found

16/02/18 16:18:34 INFO util.Utils: Default bootup file /home/hadoop/.pigbootup n ot found

16/02/18 16:18:35 INFO Configuration.deprecation: mapred.job.tracker is deprecat ed. Instead, use mapreduce.jobtracker.address

16/02/18 16:18:35 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

1793 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://ip-172-31-25-87.us-west-2.comput e.internal:8020

16/02/18 16:18:35 INFO executionengine.HExecutionEngine: Connecting to hadoop fi le system at: hdfs://ip-172-31-25-87.us-west-2.compute.internal:8020

16/02/18 16:18:38 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> register s3n://uw-cse-344-oregon.aws.amazon.com/myudfs.jar

16/02/18 16:18:42 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:18:46 INFO fs.EmrFileSystem: Consistency disabled, using com.amazon. ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation

16/02/18 16:18:47 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInClust er: false instanceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEng ine: false maxMemoryMb: 3072 maxInstanceCount: 500 lastModified: 1455810399473

16/02/18 16:18:47 INFO metrics.MetricsSaver: Created MetricsSaver j-1LVOGTR4LHQB S:i-ff1c9638:RunJar:14722 period:60 /mnt/var/em/raw/i-ff1c9638\_20160218\_RunJar\_1 4722\_raw.bin

16/02/18 16:18:48 INFO s3n.S3NativeFileSystem: Opening 's3n://uw-cse-344-oregon. aws.amazon.com/myudfs.jar' for reading

16/02/18 16:18:48 INFO metrics.MetricsSaver: Thread 1 created MetricsLockFreeSav er 1

grunt> raw = LOAD 's3n://uw-cse-344-oregon.aws.amazon.com/btc-2010-chunk-000' US ING TextLoader as (line:chararray);

16/02/18 16:18:54 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:18:54 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> ntriples = foreach raw generate FLATTEN(myudfs.RDFSplit3(line)) as (subje ct:chararray,predicate:chararray,object:chararray);

grunt> filter1 = FILTER ntriples BY subject matches '.\*rdfabout\\.com.\*';

grunt> DESCRIBE filter1;

filter1: {subject: chararray,predicate: chararray,object: chararray}

grunt> filter2 = FOREACH filter1 GENERATE subject AS subject2, predicate AS pred icate2, object AS object2;

grunt> DESCRIBE filter2;

filter2: {subject2: chararray,predicate2: chararray,object2: chararray}

grunt> common = JOIN filter1 BY object, filter2 BY subject2;

grunt> DESCRIBE common;

common: {filter1::subject: chararray,filter1::predicate: chararray,filter1::obje ct: chararray,filter2::subject2: chararray,filter2::predicate2: chararray,filter 2::object2: chararray}

grunt> common\_distinct = DISTINCT common;

grunt> store common\_distinct into '/user/hadoop/example3-results' using PigStora ge();

16/02/18 16:20:50 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:20:50 INFO Configuration.deprecation: mapred.textoutputformat.separa tor is deprecated. Instead, use mapreduce.output.textoutputformat.separator

136592 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features us ed in the script: HASH\_JOIN,DISTINCT,FILTER

16/02/18 16:20:50 INFO pigstats.ScriptState: Pig features used in the script: HA SH\_JOIN,DISTINCT,FILTER

16/02/18 16:20:50 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

136653 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematup le] was not set... will not generate code.

16/02/18 16:20:50 INFO data.SchemaTupleBackend: Key [pig.schematuple] was not se t... will not generate code.

136727 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimize r - {RULES\_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByC onstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForE ach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatte n, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

16/02/18 16:20:50 INFO optimizer.LogicalPlanOptimizer: {RULES\_ENABLED=[AddForEac h, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptim izer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

137145 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MRCompiler - File concatenation threshold: 100 optimistic? false

16/02/18 16:20:51 INFO mapReduceLayer.MRCompiler: File concatenation threshold: 100 optimistic? false

137209 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage->POForEach to POPa ckage(JoinPackager)

16/02/18 16:20:51 INFO mapReduceLayer.MRCompiler$LastInputStreamingOptimizer: Re write: POPackage->POForEach to POPackage(JoinPackager)

137221 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MultiQueryOptimizer - MR plan size before optimization: 3

16/02/18 16:20:51 INFO mapReduceLayer.MultiQueryOptimizer: MR plan size before o ptimization: 3

137222 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MultiQueryOptimizer - Merged 0 out of total 2 MR operators.

16/02/18 16:20:51 INFO mapReduceLayer.MultiQueryOptimizer: Merged 0 out of total 2 MR operators.

137222 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MultiQueryOptimizer - MR plan size after optimization: 3

16/02/18 16:20:51 INFO mapReduceLayer.MultiQueryOptimizer: MR plan size after op timization: 3

16/02/18 16:20:51 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:20:51 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

137739 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

16/02/18 16:20:51 INFO mapreduce.MRScriptState: Pig script settings are added to the job

16/02/18 16:20:51 INFO Configuration.deprecation: mapred.job.reduce.markreset.bu ffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.perce nt

137914 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, se t to default 0.3

16/02/18 16:20:51 INFO mapReduceLayer.JobControlCompiler: mapred.job.reduce.mark reset.buffer.percent is not set, set to default 0.3

16/02/18 16:20:51 INFO Configuration.deprecation: mapred.output.compress is depr ecated. Instead, use mapreduce.output.fileoutputformat.compress

137929 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - This job cannot be converted run in-process

16/02/18 16:20:51 INFO mapReduceLayer.JobControlCompiler: This job cannot be con verted run in-process

138293 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/tmp/pig8339400829971654798tmp/myudfs.jar to DistributedCache through /tmp/temp-928277133/tmp-963971292/myudfs.jar

16/02/18 16:20:52 INFO mapReduceLayer.JobControlCompiler: Added jar file:/tmp/pi g8339400829971654798tmp/myudfs.jar to DistributedCache through /tmp/temp-9282771 33/tmp-963971292/myudfs.jar

138445 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp-928277133/tmp-174656221/pig-0.14.0-amzn-0 -core-h2.jar

16/02/18 16:20:52 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp-928277 133/tmp-174656221/pig-0.14.0-amzn-0-core-h2.jar

138501 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to D istributedCache through /tmp/temp-928277133/tmp461173956/automaton-1.11-8.jar

16/02/18 16:20:52 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-928277133/t mp461173956/automaton-1.11-8.jar

138545 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-928277133/tmp-1257977440/antlr-runtime-3.4.ja r

16/02/18 16:20:52 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-928277133/ tmp-1257977440/antlr-runtime-3.4.jar

138642 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/hadoop/lib/guava-11.0.2.jar to Di stributedCache through /tmp/temp-928277133/tmp-1005677653/guava-11.0.2.jar

16/02/18 16:20:52 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop/lib/guava-11.0.2.jar to DistributedCache through /tmp/temp-928277133/tm p-1005677653/guava-11.0.2.jar

138707 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/hadoop-mapreduce/joda-time-2.8.1. jar to DistributedCache through /tmp/temp-928277133/tmp861206267/joda-time-2.8.1 .jar

16/02/18 16:20:52 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop-mapreduce/joda-time-2.8.1.jar to DistributedCache through /tmp/temp-928 277133/tmp861206267/joda-time-2.8.1.jar

138847 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Setting up single store job

16/02/18 16:20:52 INFO mapReduceLayer.JobControlCompiler: Setting up single stor e job

138884 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematu ple] is false, will not generate code.

16/02/18 16:20:52 INFO data.SchemaTupleFrontend: Key [pig.schematuple] is false, will not generate code.

138885 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

16/02/18 16:20:52 INFO data.SchemaTupleFrontend: Starting process to move genera ted code to distributed cacche

138885 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig. schematuple.classes] with classes to deserialize []

16/02/18 16:20:52 INFO data.SchemaTupleFrontend: Setting key [pig.schematuple.cl asses] with classes to deserialize []

138977 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

16/02/18 16:20:52 INFO mapReduceLayer.MapReduceLauncher: 1 map-reduce job(s) wai ting for submission.

16/02/18 16:20:52 INFO Configuration.deprecation: mapred.job.tracker.http.addres s is deprecated. Instead, use mapreduce.jobtracker.http.address

16/02/18 16:20:52 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:20:53 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:20:53 WARN mapreduce.JobSubmitter: No job jar file set. User classe s may not be found. See Job or Job#setJar(String).

16/02/18 16:20:53 INFO input.FileInputFormat: Total input paths to process : 1

139592 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths to process : 1

16/02/18 16:20:53 INFO util.MapRedUtil: Total input paths to process : 1

16/02/18 16:20:53 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library

16/02/18 16:20:53 INFO lzo.LzoCodec: Successfully loaded & initialized native-lz o library [hadoop-lzo rev 41f4e6be3ac5d6676a3464f77de79a33e8fdd9f3]

139651 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths (combined) to process : 33

16/02/18 16:20:53 INFO util.MapRedUtil: Total input paths (combined) to process : 33

16/02/18 16:20:53 INFO metrics.MetricsSaver: 1 aggregated HDFSWriteDelay 109 raw values into 1 aggregated values, total 1

16/02/18 16:20:53 INFO mapreduce.JobSubmitter: number of splits:33

16/02/18 16:20:54 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_14 55810385693\_0003

16/02/18 16:20:54 INFO mapred.YARNRunner: Job jar is not present. Not adding any jar to the list of resources.

16/02/18 16:20:54 INFO impl.YarnClientImpl: Submitted application application\_14 55810385693\_0003

16/02/18 16:20:54 INFO mapreduce.Job: The url to track the job: http://ip-172-31 -25-87.us-west-2.compute.internal:20888/proxy/application\_1455810385693\_0003/

140909 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - HadoopJobId: job\_1455810385693\_0003

16/02/18 16:20:54 INFO mapReduceLayer.MapReduceLauncher: HadoopJobId: job\_145581 0385693\_0003

140909 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Processing aliases filter1,ntriples,raw

16/02/18 16:20:54 INFO mapReduceLayer.MapReduceLauncher: Processing aliases filt er1,ntriples,raw

140909 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - detailed locations: M: raw[1,6],ntriples[-1,-1],filter1[3, 10] C: R:

16/02/18 16:20:54 INFO mapReduceLayer.MapReduceLauncher: detailed locations: M: raw[1,6],ntriples[-1,-1],filter1[3,10] C: R:

140922 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 0% complete

16/02/18 16:20:54 INFO mapReduceLayer.MapReduceLauncher: 0% complete

140922 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Running jobs are [job\_1455810385693\_0003]

16/02/18 16:20:54 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0003]

198336 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 4% complete

16/02/18 16:21:52 INFO mapReduceLayer.MapReduceLauncher: 4% complete

198337 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Running jobs are [job\_1455810385693\_0003]

16/02/18 16:21:52 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0003]

203346 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 9% complete

16/02/18 16:21:57 INFO mapReduceLayer.MapReduceLauncher: 9% complete

203346 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Running jobs are [job\_1455810385693\_0003]

16/02/18 16:21:57 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0003]

238405 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 13% complete

16/02/18 16:22:32 INFO mapReduceLayer.MapReduceLauncher: 13% complete

238405 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Running jobs are [job\_1455810385693\_0003]

16/02/18 16:22:32 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0003]

251425 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 33% complete

16/02/18 16:22:45 INFO mapReduceLayer.MapReduceLauncher: 33% complete

16/02/18 16:22:45 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:22:45 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:22:47 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:22:47 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:22:47 INFO Configuration.deprecation: mapred.reduce.tasks is depreca ted. Instead, use mapreduce.job.reduces

16/02/18 16:22:47 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:22:47 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

253884 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

16/02/18 16:22:47 INFO mapreduce.MRScriptState: Pig script settings are added to the job

253889 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, se t to default 0.3

16/02/18 16:22:47 INFO mapReduceLayer.JobControlCompiler: mapred.job.reduce.mark reset.buffer.percent is not set, set to default 0.3

253896 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Reduce phase detected, estimating # of required reducers.

16/02/18 16:22:47 INFO mapReduceLayer.JobControlCompiler: Reduce phase detected, estimating # of required reducers.

253898 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.ex ecutionengine.mapReduceLayer.InputSizeReducerEstimator

16/02/18 16:22:47 INFO mapReduceLayer.JobControlCompiler: Using reducer estimato r: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducer Estimator

253969 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalIn putFileSize=13798728

16/02/18 16:22:47 INFO mapReduceLayer.InputSizeReducerEstimator: BytesPerReducer =1000000000 maxReducers=999 totalInputFileSize=13798728

253972 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Setting Parallelism to 1

16/02/18 16:22:47 INFO mapReduceLayer.JobControlCompiler: Setting Parallelism to 1

253973 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - This job cannot be converted run in-process

16/02/18 16:22:47 INFO mapReduceLayer.JobControlCompiler: This job cannot be con verted run in-process

254038 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/tmp/pig8339400829971654798tmp/myudfs.jar to DistributedCache through /tmp/temp-928277133/tmp56000334/myudfs.jar

16/02/18 16:22:47 INFO mapReduceLayer.JobControlCompiler: Added jar file:/tmp/pi g8339400829971654798tmp/myudfs.jar to DistributedCache through /tmp/temp-9282771 33/tmp56000334/myudfs.jar

254176 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp-928277133/tmp1606437908/pig-0.14.0-amzn-0 -core-h2.jar

16/02/18 16:22:48 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp-928277 133/tmp1606437908/pig-0.14.0-amzn-0-core-h2.jar

254216 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to D istributedCache through /tmp/temp-928277133/tmp-1918513333/automaton-1.11-8.jar

16/02/18 16:22:48 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-928277133/t mp-1918513333/automaton-1.11-8.jar

254258 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-928277133/tmp-1815687884/antlr-runtime-3.4.ja r

16/02/18 16:22:48 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-928277133/ tmp-1815687884/antlr-runtime-3.4.jar

254341 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/hadoop/lib/guava-11.0.2.jar to Di stributedCache through /tmp/temp-928277133/tmp1874919257/guava-11.0.2.jar

16/02/18 16:22:48 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop/lib/guava-11.0.2.jar to DistributedCache through /tmp/temp-928277133/tm p1874919257/guava-11.0.2.jar

254423 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/hadoop-mapreduce/joda-time-2.8.1. jar to DistributedCache through /tmp/temp-928277133/tmp-51314134/joda-time-2.8.1 .jar

16/02/18 16:22:48 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop-mapreduce/joda-time-2.8.1.jar to DistributedCache through /tmp/temp-928 277133/tmp-51314134/joda-time-2.8.1.jar

254467 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Setting up single store job

16/02/18 16:22:48 INFO mapReduceLayer.JobControlCompiler: Setting up single stor e job

254469 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematu ple] is false, will not generate code.

16/02/18 16:22:48 INFO data.SchemaTupleFrontend: Key [pig.schematuple] is false, will not generate code.

254469 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

16/02/18 16:22:48 INFO data.SchemaTupleFrontend: Starting process to move genera ted code to distributed cacche

254469 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig. schematuple.classes] with classes to deserialize []

16/02/18 16:22:48 INFO data.SchemaTupleFrontend: Setting key [pig.schematuple.cl asses] with classes to deserialize []

254650 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

16/02/18 16:22:48 INFO mapReduceLayer.MapReduceLauncher: 1 map-reduce job(s) wai ting for submission.

16/02/18 16:22:48 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:22:48 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:22:48 WARN mapreduce.JobSubmitter: No job jar file set. User classe s may not be found. See Job or Job#setJar(String).

16/02/18 16:22:48 INFO input.FileInputFormat: Total input paths to process : 33

254846 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths to process : 33

16/02/18 16:22:48 INFO util.MapRedUtil: Total input paths to process : 33

254848 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths (combined) to process : 1

16/02/18 16:22:48 INFO util.MapRedUtil: Total input paths (combined) to process : 1

16/02/18 16:22:48 INFO input.FileInputFormat: Total input paths to process : 33

254939 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths to process : 33

16/02/18 16:22:48 INFO util.MapRedUtil: Total input paths to process : 33

254941 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths (combined) to process : 1

16/02/18 16:22:48 INFO util.MapRedUtil: Total input paths (combined) to process : 1

16/02/18 16:22:49 INFO mapreduce.JobSubmitter: number of splits:2

16/02/18 16:22:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_14 55810385693\_0004

16/02/18 16:22:49 INFO mapred.YARNRunner: Job jar is not present. Not adding any jar to the list of resources.

16/02/18 16:22:49 INFO impl.YarnClientImpl: Submitted application application\_14 55810385693\_0004

16/02/18 16:22:49 INFO mapreduce.Job: The url to track the job: http://ip-172-31 -25-87.us-west-2.compute.internal:20888/proxy/application\_1455810385693\_0004/

255273 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - HadoopJobId: job\_1455810385693\_0004

16/02/18 16:22:49 INFO mapReduceLayer.MapReduceLauncher: HadoopJobId: job\_145581 0385693\_0004

255273 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Processing aliases common,filter2

16/02/18 16:22:49 INFO mapReduceLayer.MapReduceLauncher: Processing aliases comm on,filter2

255273 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - detailed locations: M: common[5,9],filter2[4,10],common[5, 9] C: R:

16/02/18 16:22:49 INFO mapReduceLayer.MapReduceLauncher: detailed locations: M: common[5,9],filter2[4,10],common[5,9] C: R:

285150 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 50% complete

16/02/18 16:23:19 INFO mapReduceLayer.MapReduceLauncher: 50% complete

285151 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Running jobs are [job\_1455810385693\_0004]

16/02/18 16:23:19 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0004]

302681 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 66% complete

16/02/18 16:23:36 INFO mapReduceLayer.MapReduceLauncher: 66% complete

302681 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Running jobs are [job\_1455810385693\_0004]

16/02/18 16:23:36 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0004]

16/02/18 16:23:39 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:23:39 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:23:39 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:23:40 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:23:40 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:23:40 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

306464 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

16/02/18 16:23:40 INFO mapreduce.MRScriptState: Pig script settings are added to the job

306465 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, se t to default 0.3

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: mapred.job.reduce.mark reset.buffer.percent is not set, set to default 0.3

306466 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Reduce phase detected, estimating # of required reducers.

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Reduce phase detected, estimating # of required reducers.

306471 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.ex ecutionengine.mapReduceLayer.InputSizeReducerEstimator

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Using reducer estimato r: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducer Estimator

306480 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalIn putFileSize=20746085

16/02/18 16:23:40 INFO mapReduceLayer.InputSizeReducerEstimator: BytesPerReducer =1000000000 maxReducers=999 totalInputFileSize=20746085

306481 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Setting Parallelism to 1

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Setting Parallelism to 1

306481 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - This job cannot be converted run in-process

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: This job cannot be con verted run in-process

306540 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/tmp/pig8339400829971654798tmp/myudfs.jar to DistributedCache through /tmp/temp-928277133/tmp1968780824/myudfs.jar

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Added jar file:/tmp/pi g8339400829971654798tmp/myudfs.jar to DistributedCache through /tmp/temp-9282771 33/tmp1968780824/myudfs.jar

306695 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp-928277133/tmp1579299614/pig-0.14.0-amzn-0 -core-h2.jar

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp-928277 133/tmp1579299614/pig-0.14.0-amzn-0-core-h2.jar

306753 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to D istributedCache through /tmp/temp-928277133/tmp-1747307723/automaton-1.11-8.jar

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-928277133/t mp-1747307723/automaton-1.11-8.jar

306814 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-928277133/tmp1527739897/antlr-runtime-3.4.jar

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-928277133/ tmp1527739897/antlr-runtime-3.4.jar

306898 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/hadoop/lib/guava-11.0.2.jar to Di stributedCache through /tmp/temp-928277133/tmp1297777846/guava-11.0.2.jar

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop/lib/guava-11.0.2.jar to DistributedCache through /tmp/temp-928277133/tm p1297777846/guava-11.0.2.jar

306961 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Added jar file:/usr/lib/hadoop-mapreduce/joda-time-2.8.1. jar to DistributedCache through /tmp/temp-928277133/tmp-1437186933/joda-time-2.8 .1.jar

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop-mapreduce/joda-time-2.8.1.jar to DistributedCache through /tmp/temp-928 277133/tmp-1437186933/joda-time-2.8.1.jar

307006 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Setting up single store job

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Setting up single stor e job

307007 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematu ple] is false, will not generate code.

16/02/18 16:23:40 INFO data.SchemaTupleFrontend: Key [pig.schematuple] is false, will not generate code.

307007 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

16/02/18 16:23:40 INFO data.SchemaTupleFrontend: Starting process to move genera ted code to distributed cacche

307008 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig. schematuple.classes] with classes to deserialize []

16/02/18 16:23:40 INFO data.SchemaTupleFrontend: Setting key [pig.schematuple.cl asses] with classes to deserialize []

307010 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .JobControlCompiler - Setting identity combiner class.

16/02/18 16:23:40 INFO mapReduceLayer.JobControlCompiler: Setting identity combi ner class.

307104 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

16/02/18 16:23:41 INFO mapReduceLayer.MapReduceLauncher: 1 map-reduce job(s) wai ting for submission.

16/02/18 16:23:41 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:23:41 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:23:41 WARN mapreduce.JobSubmitter: No job jar file set. User classe s may not be found. See Job or Job#setJar(String).

16/02/18 16:23:41 INFO input.FileInputFormat: Total input paths to process : 1

307349 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths to process : 1

16/02/18 16:23:41 INFO util.MapRedUtil: Total input paths to process : 1

307350 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.Map RedUtil - Total input paths (combined) to process : 1

16/02/18 16:23:41 INFO util.MapRedUtil: Total input paths (combined) to process : 1

16/02/18 16:23:41 INFO mapreduce.JobSubmitter: number of splits:1

16/02/18 16:23:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_14 55810385693\_0005

16/02/18 16:23:41 INFO mapred.YARNRunner: Job jar is not present. Not adding any jar to the list of resources.

16/02/18 16:23:41 INFO impl.YarnClientImpl: Submitted application application\_14 55810385693\_0005

16/02/18 16:23:41 INFO mapreduce.Job: The url to track the job: http://ip-172-31 -25-87.us-west-2.compute.internal:20888/proxy/application\_1455810385693\_0005/

307858 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - HadoopJobId: job\_1455810385693\_0005

16/02/18 16:23:41 INFO mapReduceLayer.MapReduceLauncher: HadoopJobId: job\_145581 0385693\_0005

307859 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Processing aliases

16/02/18 16:23:41 INFO mapReduceLayer.MapReduceLauncher: Processing aliases

307859 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - detailed locations: M: C: R:

16/02/18 16:23:41 INFO mapReduceLayer.MapReduceLauncher: detailed locations: M: C: R:

337997 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 83% complete

16/02/18 16:24:11 INFO mapReduceLayer.MapReduceLauncher: 83% complete

337997 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Running jobs are [job\_1455810385693\_0005]

16/02/18 16:24:11 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0005]

355028 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Running jobs are [job\_1455810385693\_0005]

16/02/18 16:24:28 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0005]

16/02/18 16:24:31 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:31 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:24:32 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:32 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:24:32 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:32 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

358638 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - 100% complete

16/02/18 16:24:32 INFO mapReduceLayer.MapReduceLauncher: 100% complete

358691 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Sc ript Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features

2.6.0-amzn-2 0.14.0-amzn-0 hadoop 2016-02-18 16:20:51 2016-02-18 16:24 :32 HASH\_JOIN,DISTINCT,FILTER

Success!

Job Stats (time in seconds):

JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMa pTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A lias Feature Outputs

job\_1455810385693\_0003 33 0 47 20 40 43 0 0 0 0 filter1,ntriples,raw MAP\_ONLY

job\_1455810385693\_0004 2 1 13 13 13 13 14 1 4 14 14 common,filter2 HASH\_JOIN

job\_1455810385693\_0005 1 1 14 14 14 14 13 1 3 13 13 DISTINCT /user/hadoop/example3-results,

Input(s):

Successfully read 10000000 records (12672 bytes) from: "s3n://uw-cse-344-oregon. aws.amazon.com/btc-2010-chunk-000"

Output(s):

Successfully stored 28461 records (13082183 bytes) in: "/user/hadoop/example3-re sults"

Counters:

Total records written : 28461

Total bytes written : 13082183

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1455810385693\_0003 -> job\_1455810385693\_0004,

job\_1455810385693\_0004 -> job\_1455810385693\_0005,

job\_1455810385693\_0005

16/02/18 16:24:32 INFO mapreduce.SimplePigStats: Script Statistics:

HadoopVersion PigVersion UserId StartedAt FinishedAt Features

2.6.0-amzn-2 0.14.0-amzn-0 hadoop 2016-02-18 16:20:51 2016-02-18 16:24 :32 HASH\_JOIN,DISTINCT,FILTER

Success!

Job Stats (time in seconds):

JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMa pTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime A lias Feature Outputs

job\_1455810385693\_0003 33 0 47 20 40 43 0 0 0 0 filter1,ntriples,raw MAP\_ONLY

job\_1455810385693\_0004 2 1 13 13 13 13 14 1 4 14 14 common,filter2 HASH\_JOIN

job\_1455810385693\_0005 1 1 14 14 14 14 13 1 3 13 13 DISTINCT /user/hadoop/example3-results,

Input(s):

Successfully read 10000000 records (12672 bytes) from: "s3n://uw-cse-344-oregon. aws.amazon.com/btc-2010-chunk-000"

Output(s):

Successfully stored 28461 records (13082183 bytes) in: "/user/hadoop/example3-re sults"

Counters:

Total records written : 28461

Total bytes written : 13082183

Spillable Memory Manager spill count : 0

Total bags proactively spilled: 0

Total records proactively spilled: 0

Job DAG:

job\_1455810385693\_0003 -> job\_1455810385693\_0004,

job\_1455810385693\_0004 -> job\_1455810385693\_0005,

job\_1455810385693\_0005

16/02/18 16:24:32 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:32 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:24:33 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:33 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:24:33 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:33 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:24:33 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:33 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:24:33 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:33 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:24:33 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:33 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:24:33 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:33 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:24:34 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:34 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

16/02/18 16:24:34 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:24:34 INFO mapred.ClientServiceDelegate: Application state is comple ted. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

360259 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer .MapReduceLauncher - Success!

16/02/18 16:24:34 INFO mapReduceLayer.MapReduceLauncher: Success!

grunt> quit

379819 [main] INFO org.apache.pig.Main - Pig script completed in 6 minutes, 20 seconds and 63 milliseconds (380063 ms)

16/02/18 16:24:53 INFO pig.Main: Pig script completed in 6 minutes, 20 seconds a nd 63 milliseconds (380063 ms)

^[[A[hadoop@ip-172-31-25-87

[hadoop@ip-172-31-25-87 ~]$ hdfs dfs -getmerge /tmp/finaloutput /tmp/problem3B-r esults

[hadoop@ip-172-31-25-87 ~]$ pig

16/02/18 16:26:41 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL

16/02/18 16:26:41 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE

16/02/18 16:26:41 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType

168 [main] INFO org.apache.pig.Main - Apache Pig version 0.14.0-amzn-0 (r: un known) compiled Nov 16 2015, 21:51:33

16/02/18 16:26:41 INFO pig.Main: Apache Pig version 0.14.0-amzn-0 (r: unknown) c ompiled Nov 16 2015, 21:51:33

176 [main] INFO org.apache.pig.Main - Logging error messages to: /mnt/var/log /pig/pig\_1455812801532.log

16/02/18 16:26:41 INFO pig.Main: Logging error messages to: /mnt/var/log/pig/pig \_1455812801532.log

275 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/ha doop/.pigbootup not found

16/02/18 16:26:41 INFO util.Utils: Default bootup file /home/hadoop/.pigbootup n ot found

16/02/18 16:26:43 INFO Configuration.deprecation: mapred.job.tracker is deprecat ed. Instead, use mapreduce.jobtracker.address

16/02/18 16:26:43 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

1928 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://ip-172-31-25-87.us-west-2.comput e.internal:8020

16/02/18 16:26:43 INFO executionengine.HExecutionEngine: Connecting to hadoop fi le system at: hdfs://ip-172-31-25-87.us-west-2.compute.internal:8020

16/02/18 16:26:46 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> register s3n://uw-cse-344-oregon.aws.amazon.com/myudfs.jar

16/02/18 16:26:53 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:26:56 INFO fs.EmrFileSystem: Consistency disabled, using com.amazon. ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation

16/02/18 16:26:56 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInClust er: false instanceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEng ine: false maxMemoryMb: 3072 maxInstanceCount: 500 lastModified: 1455810399473

16/02/18 16:26:56 INFO metrics.MetricsSaver: Created MetricsSaver j-1LVOGTR4LHQB S:i-ff1c9638:RunJar:17066 period:60 /mnt/var/em/raw/i-ff1c9638\_20160218\_RunJar\_1 7066\_raw.bin

16/02/18 16:26:58 INFO s3n.S3NativeFileSystem: Opening 's3n://uw-cse-344-oregon. aws.amazon.com/myudfs.jar' for reading

16/02/18 16:26:58 INFO metrics.MetricsSaver: Thread 1 created MetricsLockFreeSav er 1

grunt> raw = LOAD 's3n://uw-cse-344-oregon.aws.amazon.com/btc-2010-chunk-\*' USIN G TextLoader as (line:chararray);

16/02/18 16:27:02 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:27:02 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> ntriples = foreach raw generate FLATTEN(myudfs.RDFSplit3(line)) as (subje ct:chararray,predicate:chararray,object:chararray);

grunt> subjects = GROUP ntriples BY (subject) PARALLEL 50;

grunt> DESCRIBE subjects;

subjects: {group: chararray,ntriples: {(subject: chararray,predicate: chararray, object: chararray)}}

grunt> count\_by\_subjects = FOREACH subjects GENERATE FLATTEN($0), COUNT($1) AS c ount PARALLEL 50;

grunt> group\_by\_count = GROUP count\_by\_subjects BY count PARALLEL 50;

grunt> hist = FOREACH group\_by\_count GENERATE $0 AS x, COUNT($1) AS y;

grunt> store hist into '/user/hadoop/example4-results' using PigStorage();

16/02/18 16:28:10 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:28:10 INFO Configuration.deprecation: mapred.textoutputformat.separa tor is deprecated. Instead, use mapreduce.output.textoutputformat.separator

89137 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features use d in the script: GROUP\_BY

16/02/18 16:28:10 INFO pigstats.ScriptState: Pig features used in the script: GR OUP\_BY

16/02/18 16:28:10 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

89208 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematupl e] was not set... will not generate code.

16/02/18 16:28:10 INFO data.SchemaTupleBackend: Key [pig.schematuple] was not se t... will not generate code.

89265 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES\_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByCo nstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEa ch, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten , PushUpFilter, SplitFilter, StreamTypeCastInserter]}

16/02/18 16:28:10 INFO optimizer.LogicalPlanOptimizer: {RULES\_ENABLED=[AddForEac h, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptim izer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

89491 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. MRCompiler - File concatenation threshold: 100 optimistic? false

16/02/18 16:28:10 INFO mapReduceLayer.MRCompiler: File concatenation threshold: 100 optimistic? false

89568 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOp timizerUtil - Choosing to move algebraic foreach to combiner

16/02/18 16:28:10 INFO util.CombinerOptimizerUtil: Choosing to move algebraic fo reach to combiner

89575 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOp timizerUtil - Choosing to move algebraic foreach to combiner

16/02/18 16:28:10 INFO util.CombinerOptimizerUtil: Choosing to move algebraic fo reach to combiner

89611 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. MultiQueryOptimizer - MR plan size before optimization: 2

16/02/18 16:28:10 INFO mapReduceLayer.MultiQueryOptimizer: MR plan size before o ptimization: 2

89612 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. MultiQueryOptimizer - MR plan size after optimization: 2

16/02/18 16:28:10 INFO mapReduceLayer.MultiQueryOptimizer: MR plan size after op timization: 2

16/02/18 16:28:11 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:28:11 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

90157 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

16/02/18 16:28:11 INFO mapreduce.MRScriptState: Pig script settings are added to the job

16/02/18 16:28:11 INFO Configuration.deprecation: mapred.job.reduce.markreset.bu ffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.perce nt

90179 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

16/02/18 16:28:11 INFO mapReduceLayer.JobControlCompiler: mapred.job.reduce.mark reset.buffer.percent is not set, set to default 0.3

16/02/18 16:28:11 INFO Configuration.deprecation: mapred.output.compress is depr ecated. Instead, use mapreduce.output.fileoutputformat.compress

90185 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - Reduce phase detected, estimating # of required reducers.

16/02/18 16:28:11 INFO mapReduceLayer.JobControlCompiler: Reduce phase detected, estimating # of required reducers.

90189 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - Setting Parallelism to 50

16/02/18 16:28:11 INFO mapReduceLayer.JobControlCompiler: Setting Parallelism to 50

16/02/18 16:28:11 INFO Configuration.deprecation: mapred.reduce.tasks is depreca ted. Instead, use mapreduce.job.reduces

90190 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - This job cannot be converted run in-process

16/02/18 16:28:11 INFO mapReduceLayer.JobControlCompiler: This job cannot be con verted run in-process

90794 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - Added jar file:/tmp/pig8265574491528555244tmp/myudfs.jar t o DistributedCache through /tmp/temp616518894/tmp1751175778/myudfs.jar

16/02/18 16:28:12 INFO mapReduceLayer.JobControlCompiler: Added jar file:/tmp/pi g8265574491528555244tmp/myudfs.jar to DistributedCache through /tmp/temp61651889 4/tmp1751175778/myudfs.jar

90937 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - Added jar file:/usr/lib/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp616518894/tmp235914067/pig-0.14.0-amzn-0-co re-h2.jar

16/02/18 16:28:12 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp6165188 94/tmp235914067/pig-0.14.0-amzn-0-core-h2.jar

90983 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to Di stributedCache through /tmp/temp616518894/tmp1520141754/automaton-1.11-8.jar

16/02/18 16:28:12 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp616518894/tm p1520141754/automaton-1.11-8.jar

91041 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to D istributedCache through /tmp/temp616518894/tmp1868492832/antlr-runtime-3.4.jar

16/02/18 16:28:12 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp616518894/t mp1868492832/antlr-runtime-3.4.jar

91134 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - Added jar file:/usr/lib/hadoop/lib/guava-11.0.2.jar to Dis tributedCache through /tmp/temp616518894/tmp348943524/guava-11.0.2.jar

16/02/18 16:28:12 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop/lib/guava-11.0.2.jar to DistributedCache through /tmp/temp616518894/tmp 348943524/guava-11.0.2.jar

91208 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - Added jar file:/usr/lib/hadoop-mapreduce/joda-time-2.8.1.j ar to DistributedCache through /tmp/temp616518894/tmp1801700893/joda-time-2.8.1. jar

16/02/18 16:28:12 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/li b/hadoop-mapreduce/joda-time-2.8.1.jar to DistributedCache through /tmp/temp6165 18894/tmp1801700893/joda-time-2.8.1.jar

91328 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. JobControlCompiler - Setting up single store job

16/02/18 16:28:12 INFO mapReduceLayer.JobControlCompiler: Setting up single stor e job

91350 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematup le] is false, will not generate code.

16/02/18 16:28:12 INFO data.SchemaTupleFrontend: Key [pig.schematuple] is false, will not generate code.

91354 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process t o move generated code to distributed cacche

16/02/18 16:28:12 INFO data.SchemaTupleFrontend: Starting process to move genera ted code to distributed cacche

91355 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.s chematuple.classes] with classes to deserialize []

16/02/18 16:28:12 INFO data.SchemaTupleFrontend: Setting key [pig.schematuple.cl asses] with classes to deserialize []

91634 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

16/02/18 16:28:13 INFO mapReduceLayer.MapReduceLauncher: 1 map-reduce job(s) wai ting for submission.

16/02/18 16:28:13 INFO Configuration.deprecation: mapred.job.tracker.http.addres s is deprecated. Instead, use mapreduce.jobtracker.http.address

16/02/18 16:28:13 INFO client.RMProxy: Connecting to ResourceManager at ip-172-3 1-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 16:28:13 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 16:28:13 WARN mapreduce.JobSubmitter: No job jar file set. User classe s may not be found. See Job or Job#setJar(String).

16/02/18 16:28:13 INFO s3n.S3NativeFileSystem: listStatus s3n://uw-cse-344-orego n.aws.amazon.com/ with recursive false

16/02/18 16:28:14 INFO input.FileInputFormat: Total input paths to process : 251

92768 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapR edUtil - Total input paths to process : 251

16/02/18 16:28:14 INFO util.MapRedUtil: Total input paths to process : 251

16/02/18 16:28:14 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library

16/02/18 16:28:14 INFO lzo.LzoCodec: Successfully loaded & initialized native-lz o library [hadoop-lzo rev 41f4e6be3ac5d6676a3464f77de79a33e8fdd9f3]

93303 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapR edUtil - Total input paths (combined) to process : 7908

16/02/18 16:28:14 INFO util.MapRedUtil: Total input paths (combined) to process : 7908

16/02/18 16:28:14 INFO metrics.MetricsSaver: 1 aggregated HDFSWriteDelay 109 raw values into 1 aggregated values, total 1

16/02/18 16:28:17 INFO mapreduce.JobSubmitter: number of splits:7908

16/02/18 16:28:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_14 55810385693\_0006

16/02/18 16:28:18 INFO mapred.YARNRunner: Job jar is not present. Not adding any jar to the list of resources.

16/02/18 16:28:18 INFO impl.YarnClientImpl: Submitted application application\_14 55810385693\_0006

16/02/18 16:28:18 INFO mapreduce.Job: The url to track the job: http://ip-172-31 -25-87.us-west-2.compute.internal:20888/proxy/application\_1455810385693\_0006/

97096 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. MapReduceLauncher - HadoopJobId: job\_1455810385693\_0006

16/02/18 16:28:18 INFO mapReduceLayer.MapReduceLauncher: HadoopJobId: job\_145581 0385693\_0006

97096 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. MapReduceLauncher - Processing aliases count\_by\_subjects,ntriples,raw,subjects

16/02/18 16:28:18 INFO mapReduceLayer.MapReduceLauncher: Processing aliases coun t\_by\_subjects,ntriples,raw,subjects

97096 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. MapReduceLauncher - detailed locations: M: raw[1,6],ntriples[-1,-1],count\_by\_su bjects[4,20],subjects[3,11] C: count\_by\_subjects[4,20],subjects[3,11] R: count\_b y\_subjects[4,20]

16/02/18 16:28:18 INFO mapReduceLayer.MapReduceLauncher: detailed locations: M: raw[1,6],ntriples[-1,-1],count\_by\_subjects[4,20],subjects[3,11] C: count\_by\_subj ects[4,20],subjects[3,11] R: count\_by\_subjects[4,20]

97108 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. MapReduceLauncher - 0% complete

16/02/18 16:28:18 INFO mapReduceLayer.MapReduceLauncher: 0% complete

97108 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer. MapReduceLauncher - Running jobs are [job\_1455810385693\_0006]

16/02/18 16:28:18 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1 455810385693\_0006]

Using username "hadoop".

Authenticating with public key "imported-openssh-key"

Last login: Thu Feb 18 17:49:27 2016

\_\_| \_\_|\_ )

\_| ( / Amazon Linux AMI

\_\_\_|\\_\_\_|\_\_\_|

https://aws.amazon.com/amazon-linux-ami/2015.09-release-notes/

41 package(s) needed for security, out of 69 available

Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRR

E::::::::::::::::::E M:::::::M M:::::::M R::::::::::::::R

EE:::::EEEEEEEEE:::E M::::::::M M::::::::M R:::::RRRRRR:::::R

E::::E EEEEE M:::::::::M M:::::::::M RR::::R R::::R

E::::E M::::::M:::M M:::M::::::M R:::R R::::R

E:::::EEEEEEEEEE M:::::M M:::M M:::M M:::::M R:::RRRRRR:::::R

E::::::::::::::E M:::::M M:::M:::M M:::::M R:::::::::::RR

E:::::EEEEEEEEEE M:::::M M:::::M M:::::M R:::RRRRRR::::R

E::::E M:::::M M:::M M:::::M R:::R R::::R

E::::E EEEEE M:::::M MMM M:::::M R:::R R::::R

EE:::::EEEEEEEE::::E M:::::M M:::::M R:::R R::::R

E::::::::::::::::::E M:::::M M:::::M RR::::R R::::R

EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-25-87 ~]$ pig

16/02/18 18:01:12 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL

16/02/18 18:01:12 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE

16/02/18 18:01:12 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType

100 [main] INFO org.apache.pig.Main - Apache Pig version 0.14.0-amzn-0 (r: unknown) compiled Nov 16 2015, 21:51:33

16/02/18 18:01:12 INFO pig.Main: Apache Pig version 0.14.0-amzn-0 (r: unknown) compiled Nov 16 2015, 21:51:33

102 [main] INFO org.apache.pig.Main - Logging error messages to: /mnt/var/log/pig/pig\_1455818472285.log

16/02/18 18:01:12 INFO pig.Main: Logging error messages to: /mnt/var/log/pig/pig\_1455818472285.log

142 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/hadoop/.pigbootup not found

16/02/18 18:01:12 INFO util.Utils: Default bootup file /home/hadoop/.pigbootup not found

16/02/18 18:01:14 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address

16/02/18 18:01:14 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

1876 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://ip-172-31-25-87.us-west-2.compute.internal:8020

16/02/18 18:01:14 INFO executionengine.HExecutionEngine: Connecting to hadoop file system at: hdfs://ip-172-31-25-87.us-west-2.compute.internal:8020

16/02/18 18:01:18 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> register s3n://uw-cse-344-oregon.aws.amazon.com/myudfs.jar

16/02/18 18:01:41 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 18:01:44 INFO fs.EmrFileSystem: Consistency disabled, using com.amazon.ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation

16/02/18 18:01:45 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInCluster: false instanceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEngine: false maxMemoryMb: 3072 maxInstanceCount: 500 lastModified: 1455810399473

16/02/18 18:01:45 INFO metrics.MetricsSaver: Created MetricsSaver j-1LVOGTR4LHQBS:i-ff1c9638:RunJar:18683 period:60 /mnt/var/em/raw/i-ff1c9638\_20160218\_RunJar\_18683\_raw.bin

16/02/18 18:01:46 INFO s3n.S3NativeFileSystem: Opening 's3n://uw-cse-344-oregon.aws.amazon.com/myudfs.jar' for reading

16/02/18 18:01:46 INFO metrics.MetricsSaver: Thread 1 created MetricsLockFreeSaver 1

grunt> raw = LOAD 's3n://uw-cse-344-oregon.aws.amazon.com/btc-2010-chunk-\*' USING TextLoader as (line:chararray);

16/02/18 18:01:48 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 18:01:48 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> ntriples = foreach raw generate FLATTEN(myudfs.RDFSplit3(line)) as (subject:chararray,predicate:chararray,object:chararray);

grunt> subjects = GROUP ntriples BY (subject) PARALLEL 50;

grunt> count\_by\_subjects = FOREACH subjects GENERATE FLATTEN($0), COUNT($1) AS count PARALLEL 50;

grunt> group\_by\_count = GROUP count\_by\_subjects BY count PARALLEL 50;

grunt> hist = FOREACH group\_by\_count GENERATE $0 AS x, COUNT($1) AS y;

grunt> store hist into '/user/hadoop/example4-results' using PigStorage();

16/02/18 18:02:44 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 18:02:44 INFO Configuration.deprecation: mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator

92579 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP\_BY

16/02/18 18:02:44 INFO pigstats.ScriptState: Pig features used in the script: GROUP\_BY

16/02/18 18:02:44 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

92636 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.

16/02/18 18:02:44 INFO data.SchemaTupleBackend: Key [pig.schematuple] was not set... will not generate code.

92693 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES\_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

16/02/18 18:02:44 INFO optimizer.LogicalPlanOptimizer: {RULES\_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}

92920 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

16/02/18 18:02:45 INFO mapReduceLayer.MRCompiler: File concatenation threshold: 100 optimistic? false

92996 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combiner

16/02/18 18:02:45 INFO util.CombinerOptimizerUtil: Choosing to move algebraic foreach to combiner

93004 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreach to combiner

16/02/18 18:02:45 INFO util.CombinerOptimizerUtil: Choosing to move algebraic foreach to combiner

93056 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 2

16/02/18 18:02:45 INFO mapReduceLayer.MultiQueryOptimizer: MR plan size before optimization: 2

93056 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 2

16/02/18 18:02:45 INFO mapReduceLayer.MultiQueryOptimizer: MR plan size after optimization: 2

16/02/18 18:02:45 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 18:02:45 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-25-87.us-west-2.compute.internal/172.31.25.87:8032

93848 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job

16/02/18 18:02:46 INFO mapreduce.MRScriptState: Pig script settings are added to the job

16/02/18 18:02:46 INFO Configuration.deprecation: mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent

93873 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

16/02/18 18:02:46 INFO mapReduceLayer.JobControlCompiler: mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3

16/02/18 18:02:46 INFO Configuration.deprecation: mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress

93884 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.

16/02/18 18:02:46 INFO mapReduceLayer.JobControlCompiler: Reduce phase detected, estimating # of required reducers.

93884 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 50

16/02/18 18:02:46 INFO mapReduceLayer.JobControlCompiler: Setting Parallelism to 50

16/02/18 18:02:46 INFO Configuration.deprecation: mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces

93885 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process

16/02/18 18:02:46 INFO mapReduceLayer.JobControlCompiler: This job cannot be converted run in-process

94605 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/tmp/pig8182524164299324224tmp/myudfs.jar to DistributedCache through /tmp/temp-728273638/tmp1175041171/myudfs.jar

16/02/18 18:02:46 INFO mapReduceLayer.JobControlCompiler: Added jar file:/tmp/pig8182524164299324224tmp/myudfs.jar to DistributedCache through /tmp/temp-728273638/tmp1175041171/myudfs.jar

94743 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/lib/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp-728273638/tmp-1589097354/pig-0.14.0-amzn-0-core-h2.jar

16/02/18 18:02:46 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/lib/pig/pig-0.14.0-amzn-0-core-h2.jar to DistributedCache through /tmp/temp-728273638/tmp-1589097354/pig-0.14.0-amzn-0-core-h2.jar

94793 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-728273638/tmp-1876168206/automaton-1.11-8.jar

16/02/18 18:02:46 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-728273638/tmp-1876168206/automaton-1.11-8.jar

16/02/18 18:02:47 INFO metrics.MetricsSaver: 1 aggregated HDFSWriteDelay 70 raw values into 1 aggregated values, total 1

94847 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-728273638/tmp-1936914922/antlr-runtime-3.4.jar

16/02/18 18:02:47 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-728273638/tmp-1936914922/antlr-runtime-3.4.jar

94935 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/lib/hadoop/lib/guava-11.0.2.jar to DistributedCache through /tmp/temp-728273638/tmp-1129926257/guava-11.0.2.jar

16/02/18 18:02:47 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/lib/hadoop/lib/guava-11.0.2.jar to DistributedCache through /tmp/temp-728273638/tmp-1129926257/guava-11.0.2.jar

94993 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/lib/hadoop-mapreduce/joda-time-2.8.1.jar to DistributedCache through /tmp/temp-728273638/tmp-1732768022/joda-time-2.8.1.jar

16/02/18 18:02:47 INFO mapReduceLayer.JobControlCompiler: Added jar file:/usr/lib/hadoop-mapreduce/joda-time-2.8.1.jar to DistributedCache through /tmp/temp-728273638/tmp-1732768022/joda-time-2.8.1.jar

95116 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job

16/02/18 18:02:47 INFO mapReduceLayer.JobControlCompiler: Setting up single store job

95137 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.

16/02/18 18:02:47 INFO data.SchemaTupleFrontend: Key [pig.schematuple] is false, will not generate code.

95138 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche

16/02/18 18:02:47 INFO data.SchemaTupleFrontend: Starting process to move generated code to distributed cacche

95138 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

16/02/18 18:02:47 INFO data.SchemaTupleFrontend: Setting key [pig.schematuple.classes] with classes to deserialize []

95414 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.

16/02/18 18:02:47 INFO mapReduceLayer.MapReduceLauncher: 1 map-reduce job(s) waiting for submission.

16/02/18 18:02:47 INFO Configuration.deprecation: mapred.job.tracker.http.address is deprecated. Instead, use mapreduce.jobtracker.http.address

16/02/18 18:02:47 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-25-87.us-west-2.compute.internal/172.31.25.87:8032

16/02/18 18:02:47 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS

16/02/18 18:02:47 WARN mapreduce.JobSubmitter: No job jar file set. User classes may not be found. See Job or Job#setJar(String).

16/02/18 18:02:47 INFO s3n.S3NativeFileSystem: listStatus s3n://uw-cse-344-oregon.aws.amazon.com/ with recursive false

16/02/18 18:02:48 INFO input.FileInputFormat: Total input paths to process : 251

96479 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 251

16/02/18 18:02:48 INFO util.MapRedUtil: Total input paths to process : 251

16/02/18 18:02:48 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library

16/02/18 18:02:48 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 41f4e6be3ac5d6676a3464f77de79a33e8fdd9f3]

96956 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 7908

16/02/18 18:02:49 INFO util.MapRedUtil: Total input paths (combined) to process : 7908

16/02/18 18:02:52 INFO mapreduce.JobSubmitter: number of splits:7908

16/02/18 18:02:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1455810385693\_0007

16/02/18 18:02:53 INFO mapred.YARNRunner: Job jar is not present. Not adding any jar to the list of resources.

16/02/18 18:02:53 INFO impl.YarnClientImpl: Submitted application application\_1455810385693\_0007

16/02/18 18:02:53 INFO mapreduce.Job: The url to track the job: http://ip-172-31-25-87.us-west-2.compute.internal:20888/proxy/application\_1455810385693\_0007/

101459 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job\_1455810385693\_0007

16/02/18 18:02:53 INFO mapReduceLayer.MapReduceLauncher: HadoopJobId: job\_1455810385693\_0007

101459 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases count\_by\_subjects,ntriples,raw,subjects

16/02/18 18:02:53 INFO mapReduceLayer.MapReduceLauncher: Processing aliases count\_by\_subjects,ntriples,raw,subjects

101459 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: raw[1,6],ntriples[-1,-1],count\_by\_subjects[4,20],subjects[3,11] C: count\_by\_subjects[4,20],subjects[3,11] R: count\_by\_subjects[4,20]

16/02/18 18:02:53 INFO mapReduceLayer.MapReduceLauncher: detailed locations: M: raw[1,6],ntriples[-1,-1],count\_by\_subjects[4,20],subjects[3,11] C: count\_by\_subjects[4,20],subjects[3,11] R: count\_by\_subjects[4,20]

101470 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete

16/02/18 18:02:53 INFO mapReduceLayer.MapReduceLauncher: 0% complete

101471 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1455810385693\_0007]

16/02/18 18:02:53 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1455810385693\_0007]

3957493 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 4% complete

16/02/18 19:07:09 INFO mapReduceLayer.MapReduceLauncher: 4% complete

3957493 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job\_1455810385693\_0007]

16/02/18 19:07:09 INFO mapReduceLayer.MapReduceLauncher: Running jobs are [job\_1455810385693\_0007]