

Assignment Day 7

Task 1:

Write a program to implement wordcount using Pig.

Implementing Wordcount using pig in Map Reduce Mode (Default).

Input file is at HDFS:

```
[acadgild@localhost Assignment]$ hadoop fs -cat /hadoopdata/wordcount/test.txt
18/07/16 16:09:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Hi, This is a nano editor in Acadgild vm.
It is way simpler to use as compared to vim editor...!
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Assignment]$
```

Ans:

Note: Program files are properly documented for a detailed description of each instruction used within the program.

pig Task_1.pig

(Executes the command in Map Reduce mode on HDFS by default)

ScreenShot:

```
[acadgild@localhost Assignment]$ pig Task_1.pig
18/07/16 16:12:29 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/16 16:12:29 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/16 16:12:29 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-07-16 16:12:30,097 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-07-16 16:12:30,098 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/Assignment/pig_1531737759093.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2018-07-16 16:12:30,887 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using built-in java classes where applicable
2018-07-16 16:12:31,313 [main] INFO org.apache.pig.impl.util.Util - Default bootstrap file /home/acadgild/.pigbootstrap not found
2018-07-16 16:12:31,599 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2018-07-16 16:12:31,599 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:12:31,599 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2018-07-16 16:12:32,517 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task_1.pig-8c3051d8-8dad-4e0c-9556-61ce6b4b8f6d
2018-07-16 16:12:32,517 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-07-16 16:12:33,525 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:12:33,698 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-07-16 16:12:34,047 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY
2018-07-16 16:12:34,135 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:12:34,138 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-07-16 16:12:34,190 [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]}
2018-07-16 16:12:34,394 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Laer.MRCompiler - File concatenation threshold: 100 optimistic? false
2018-07-16 16:12:34,437 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreac h to combiner
```

Job DAG:
job_1531648281551_0013

```
2018-07-16 16:13:21,226 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-16 16:13:21,249 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESS
2018-07-16 16:13:21,359 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-16 16:13:21,372 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESS
2018-07-16 16:13:21,485 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-16 16:13:21,502 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCESS
2018-07-16 16:13:21,684 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-07-16 16:13:21,692 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:13:21,701 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-07-16 16:13:21,731 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-16 16:13:21,731 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(e,1)
(Hi,1)
(It,1)
(as,1)
(in,1)
(is,2)
(to,2)
(use,1)
(vim,1)
(vm.,1)
(way,1)
(This,1)
(nano,1)
(editor,1)
(simpler,1)
(Acadgild,1)
(compared,1)
(editor...,1)
2018-07-16 16:13:21,905 [main] INFO org.apache.pig.Main - Pig script completed in 52 seconds and 338 milliseconds (52338 ms)
```

Note: Program files are properly documented for a detailed description of each instruction used within the program.

Task 2:

Use **local mode** while running Pig and write Pig Latin script to get below results:

(Input Datasets)

```
[acadgild@localhost Assignment]$ ll
total 32
-rw-rw-r--. 1 acadgild acadgild 273 Jul 16 2018 employee_details.txt
-rw-rw-r--. 1 acadgild acadgild 79 Jul 11 12:03 employee_expenses.txt
-rw-rw-r--. 1 acadgild acadgild 755 Jul 16 16:12 Task_1.pig
-rw-rw-r--. 1 acadgild acadgild 802 Jul 17 2018 Task_2a.pig
-rw-rw-r--. 1 acadgild acadgild 882 Jul 17 2018 Task_2b.pig
-rw-rw-r--. 1 acadgild acadgild 1024 Jul 17 2018 Task_2c.pig
-rw-rw-r--. 1 acadgild acadgild 1156 Jul 17 2018 Task_2d.pig
-rw-rw-r--. 1 acadgild acadgild 1346 Jul 17 2018 Task_2e.pig
```

```

[acadgild@localhost Assignment]$ cat employee_details.txt
101,Amitabh,20000,1
102,Shahrukh,10000,2
103,Akshay,11000,3
104,Anubhav,5000,4
105,Pawan,2500,5
106,Aamir,25000,1
107,Salman,17500,2
108,Ranbir,14000,3
109,Katrina,1000,4
110,Priyanka,2000,5
111,Tushar,500,1
112,Ajay,5000,2
113,Jubeen,1000,1
114,Madhuri,2000,2
[acadgild@localhost Assignment]$ cat employee_expenses.txt
101      200
102      100
110      400
114      200
119      200
105      100
101      100
104      300
102      400
[acadgild@localhost Assignment]$

```

(a) Top 5 employees (employee id and employee name) with highest rating. (In case two employees have same rating, employee with name coming first in dictionary should get preference)

Ans:

pig -x local Task_2a.pig

(To Execute a pig program in local Mode)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
[acadgild@localhost Assignment]$ pig -x local Task_2a.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/07/16 16:24:42 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/16 16:24:42 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2018-07-16 16:24:42,177 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-07-16 16:24:42,177 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/Assignment/pig_153173
8482175.log
2018-07-16 16:24:42,204 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - user.name is deprecated. Instead, use mapreduce.job.user.name
2018-07-16 16:24:42,743 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/acadgild/.pigbootstrap not found
2018-07-16 16:24:43,085 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.job.tracker.address
2018-07-16 16:24:43,086 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:24:43,090 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2018-07-16 16:24:43,152 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task_2a.pig-7129282a-52f7-447a-8d75-6453ab39013d
2018-07-16 16:24:43,153 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-07-16 16:24:44,229 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:24:44,495 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using bu
iltin-java classes where applicable
2018-07-16 16:24:44,614 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: ORDER_BY,LIMIT
2018-07-16 16:24:44,710 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:24:44,840 [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]}
2018-07-16 16:24:44,882 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for A: $2
2018-07-16 16:24:44,983 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor.
collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-07-16 16:24:45,073 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100
optimistic? false
2018-07-16 16:24:45,154 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key O
2018-07-16 16:48:10,271 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,272 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,277 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,301 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,302 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,305 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,332 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,339 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,343 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,369 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,371 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,372 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessio
nId= - already initialized
2018-07-16 16:48:10,391 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-07-16 16:48:10,402 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:48:10,402 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-07-16 16:48:10,432 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-16 16:48:10,432 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(105,Pawan)
(110,Priyanka)
(104,Anubhav)
(109,Katrina)
(103,Akshay)
2018-07-16 16:48:10,547 [main] INFO org.apache.pig.Main - Pig script completed in 12 seconds and 274 milliseconds (12274 ms)
```

(b) Top 3 employees (employee id and employee name) with highest salary, whose employee id is an odd number. (In case two employees have same salary, employee with name coming first in dictionary should get preference)

Ans:

pig -x local Task_2b.pig

(To Execute a pig program in local Mode)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
[acadgild@localhost Assignment]$ pig -x local Task_2b.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2018-07-16 16:31:40 [main] INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2018-07-16 16:31:40 [main] INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2018-07-16 16:31:40,310 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-07-16 16:31:40,310 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/Assignment/pig_1531738900307.log
2018-07-16 16:31:40,337 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - user.name is deprecated. Instead, use mapreduce.job.user.name
2018-07-16 16:31:40,829 [main] INFO org.apache.pig.impl.util.Util - Default bootstrap file /home/acadgild/.pigbootstrap not found
2018-07-16 16:31:41,169 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.job.tracker.address
2018-07-16 16:31:41,171 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:31:41,177 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2018-07-16 16:31:41,239 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task_2b.pig-ed8f4024-1552-451d-9412-2b2cc4321998
2018-07-16 16:31:41,239 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-07-16 16:31:42,273 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:31:42,549 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using built-in java classes where applicable
2018-07-16 16:31:42,674 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: ORDER_BY,FILTER,LIMIT
2018-07-16 16:31:42,790 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:31:42,893 [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]}
2018-07-16 16:31:42,993 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor.
collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-07-16 16:31:43,097 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100
optimistic? false
2018-07-16 16:31:43,204 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapReduce node scope-29
2018-07-16 16:49:17,886 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,893 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,894 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,913 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,914 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,919 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,938 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,940 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,944 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,952 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,961 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,965 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= already initialized
2018-07-16 16:49:17,969 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-07-16 16:49:17,978 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:49:17,978 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-07-16 16:49:18,005 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-16 16:49:18,005 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(101,Amitabh)
(107,Salman)
(103,Akshay)
2018-07-16 16:49:18,120 [main] INFO org.apache.pig.Main - Pig script completed in 12 seconds and 551 milliseconds (12551 ms)
```

(c) Employee (employee id and employee name) with maximum expense (In case two employees have same expense, employee with name coming first in dictionary should get preference)

Ans:

pig -x local Task_2c.pig

(To Execute a pig program in local Mode)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
[acadgild@localhost Assignment]$ pig -x local Task_2c.pig
SLF4J: Class path contains multiple SLF4J bindings
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/07/16 16:50:58 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/16 16:50:58 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2018-07-16 16:50:58,263 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-07-16 16:50:58,264 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/Assignment/pig_1531740058260.log
2018-07-16 16:50:58,296 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - user.name is deprecated. Instead, use mapreduce.job.user.name
2018-07-16 16:50:58,841 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/acadgild/.pigbootstrap not found
2018-07-16 16:50:59,271 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.job.tracker.address
2018-07-16 16:50:59,272 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:50:59,279 [main] INFO org.apache.hadoop.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2018-07-16 16:50:59,356 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task_2c.pig-b5c17a56-2b35-4e4b-9a29-91f3adc40ae5
2018-07-16 16:50:59,356 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-07-16 16:51:00,610 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:51:00,790 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:51:00,927 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2018-07-16 16:51:01,058 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752
2018-07-16 16:51:01,093 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN,ORDER_BY,LIMIT
2018-07-16 16:51:01,198 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 16:51:01,382 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSet, LimitOptimizer, LoadTypeCastInserter, MergeFilter, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2018-07-16 16:51:01,679 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100
optimistic? false
2018-07-16 16:51:01,785 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key 0

nId= - already initialized
2018-07-16 21:25:38,830 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,843 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,844 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,846 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,855 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,856 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,857 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,869 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,870 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,871 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,897 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, session=
nId= - already initialized
2018-07-16 21:25:38,897 [main] WARN org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-16 21:25:38,928 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2018-07-16 21:25:38,928 [main] INFO org.apache.pig.Main - Pig script completed in 18 seconds and 737 milliseconds (18737 ms)
```

(d) List of employees (employee id and employee name) having entries in employee_expenses

file.

Ans:

pig -x local Task_2d.pig

(To Execute a pig program in local Mode)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
[acadgild@localhost Assignment]$ pig -x local Task_2d.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/07/16 21:28:16 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/16 21:28:16 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2018-07-16 21:28:16,836 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-07-16 21:28:16,837 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/Assignment/pig_1531756696833.log
2018-07-16 21:28:16,878 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - user.name is deprecated. Instead, use mapreduce.job.user.name
2018-07-16 21:28:17,519 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/acadgild/.pigbootstrap not found
2018-07-16 21:28:17,921 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.job.tracker.address
2018-07-16 21:28:17,922 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 21:28:17,927 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2018-07-16 21:28:18,019 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task_2d.pig-2991aa23-cf76-44a0-84e5-1e393d7a49b4
2018-07-16 21:28:18,019 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-07-16 21:28:19,359 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 21:28:19,605 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 21:28:19,957 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2018-07-16 21:28:20,174 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-07-16 21:28:20,207 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN, GROUP_BY
2018-07-16 21:28:20,328 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 21:28:20,501 [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]}
2018-07-16 21:28:20,608 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for A1: $1
2018-07-16 21:28:20,825 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false

Total records written : 6
Total bytes written : 6
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1579678763_0001    ->    job_local1351061141_0002,
job_local1351061141_0002

2018-07-16 21:28:27,412 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:28:27,420 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:28:27,425 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:28:27,456 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:28:27,461 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:28:27,466 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:28:27,484 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-07-16 21:28:27,506 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 21:28:27,506 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-07-16 21:28:27,538 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-16 21:28:27,538 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
101,Amitabh)
102,Shahrukh)
104,Anubhav)
105,Pawan)
110,Priyanka)
114,Madhuri)
2018-07-16 21:28:27,680 [main] INFO org.apache.pig.Main - Pig script completed in 12 seconds and 682 milliseconds (12682 ms)
```

(e) List of employees (employee id and employee name) having no entry in employee_expenses file.

Ans:

pig -x local Task_2e.pig

(To Execute a pig program in local Mode)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
[acadgild@localhost Assignment]$ pig -x local Task_2e.pig
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/07/16 21:44:18 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/16 21:44:18 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2018-07-16 21:44:18,923 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-07-16 21:44:18,924 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/Assignment/pig_1531757658920.log
2018-07-16 21:44:18,964 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - user.name is deprecated. Instead, use mapreduce.job.user.name
2018-07-16 21:44:19,583 [main] INFO org.apache.pig.impl.util.Util - Default bootstrap file /home/acadgild/.pigbootstrap not found
2018-07-16 21:44:20,040 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.job.tracker.address
2018-07-16 21:44:20,041 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 21:44:20,045 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2018-07-16 21:44:20,122 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task_2e.pig-9480b050-9ca3-4da4-bcc2-55663d4a4234
2018-07-16 21:44:20,123 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-07-16 21:44:21,441 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 21:44:21,649 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 21:44:21,910 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2018-07-16 21:44:22,085 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor.
collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-07-16 21:44:22,127 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN, GROUP_BY, FILTER
2018-07-16 21:44:22,237 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 21:44:22,348 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInsert, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInsert]}
2018-07-16 21:44:22,425 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for A1: $1
2018-07-16 21:44:22,611 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100
optimistic? false

Total records written : 8
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1932443027_0001 -> job_local505532985_0002,
job_local505532985_0002

2018-07-16 21:44:29,118 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:44:29,132 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:44:29,134 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:44:29,164 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:44:29,169 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:44:29,170 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2018-07-16 21:44:29,189 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-07-16 21:44:29,203 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-16 21:44:29,205 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2018-07-16 21:44:29,234 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-16 21:44:29,234 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(103,Akshay)
(106,Aamir)
(107,Salman)
(108,Ranbir)
(109,Katrina)
(111,Tushar)
(112,Ajay)
(113,Jubeen)
2018-07-16 21:44:29,369 [main] INFO org.apache.pig.Main - Pig script completed in 11 seconds and 937 milliseconds (11937 ms)
```

Task 3:

Implement the use case present in below blog link and share the complete steps along with screenshot(s) from your end.

<https://acadgild.com/blog/aviation-data-analysis-using-apache-pig/>

(a) Find out the top 5 most visited destinations.

Ans:

pig Task_3a.pig

(To Execute a pig program in MR Mode)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
[acadgild@localhost Assignment]$ pig Task_3a.pig
18/07/17 02:58:52 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/17 02:58:52 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/17 02:58:52 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-07-17 02:58:53,088 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-07-17 02:58:53,088 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/Assignment/pig_153177
6533083.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/Static
icLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2018-07-17 02:58:53,930 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using bu
iltin-java classes where applicable
2018-07-17 02:58:54,471 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.
jobtracker.address
2018-07-17 02:58:54,866 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 02:58:54,866 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs:/
/localhost:8020
2018-07-17 02:58:55,897 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task_3a.pig-78bedbda-4c2e-4b37-8efe-ecaada6847
81
2018-07-17 02:58:55,898 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-07-17 02:58:56,069 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 02:58:57,238 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 02:58:57,632 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 02:58:58,001 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN, GROUP_BY, ORDER_BY, FIL
TER, LIMIT
2018-07-17 02:58:58,107 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 02:58:58,116 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-07-17 02:58:58,221 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrun
e, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, Pre
dicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2018-07-17 02:58:58,413 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor.
collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-07-17 02:58:58,584 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCCompiler - File concatenation threshold: 100

2018-07-17 03:03:11,954 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EDED. Redirecting to job history server
2018-07-17 03:03:12,071 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:03:12,087 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EDED. Redirecting to job history server
2018-07-17 03:03:12,202 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:03:12,212 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EDED. Redirecting to job history server
2018-07-17 03:03:12,304 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:03:12,314 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EDED. Redirecting to job history server
2018-07-17 03:03:12,423 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:03:12,437 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EDED. Redirecting to job history server
2018-07-17 03:03:12,528 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:03:12,537 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EDED. Redirecting to job history server
2018-07-17 03:03:12,623 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:03:12,631 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EDED. Redirecting to job history server
2018-07-17 03:03:12,715 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:03:12,722 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EDED. Redirecting to job history server
2018-07-17 03:03:12,771 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:03:12,776 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EDED. Redirecting to job history server
2018-07-17 03:03:12,833 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-07-17 03:03:12,843 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 03:03:12,848 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-07-17 03:03:12,856 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-17 03:03:12,856 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(ATL,106898,ATL,Atlanta,USA)
(DEN,63003,DEN,Denver,USA)
(DFW,70657,DFW,Dallas-Fort Worth,USA)
(LAX,59969,LAX,Los Angeles,USA)
(ORD,108984,ORD,Chicago,USA)
2018-07-17 03:03:13,058 [main] INFO org.apache.pig.Main - Pig script completed in 4 minutes, 20 seconds and 574 milliseconds (260574 ms)
```

(b) Which month has seen the most number of cancellations due to bad weather?

Ans:

pig Task_3b.pig

(To Execute a pig program in MR Mode)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

ScreenShot:

```
[acadgild@localhost Assignment]$ pig Task_3b.pig
18/07/17 03:08:42 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/17 03:08:42 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/17 03:08:42 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2018-07-17 03:08:42,559 [main] INFO org.apache.pig.Main - Apache Pig Version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:49
2018-07-17 03:08:42,559 [main] INFO org.apache.pig.Main - Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/Assignment/pig_153177
7122534.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2018-07-17 03:08:43,575 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using built-in java classes where applicable
2018-07-17 03:08:45,215 [main] INFO org.apache.pig.impl.util.Util - Default bootstrap file /home/acadgild/.pigbootstrap not found
2018-07-17 03:08:46,066 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.job.tracker.address
2018-07-17 03:08:46,066 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 03:08:46,066 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2018-07-17 03:08:48,008 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-Task_3b.pig-9c43e864-4f2b-4298-a8df-693e4a07cd0a
2018-07-17 03:08:48,011 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
2018-07-17 03:08:48,287 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 03:08:50,218 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 03:08:51,000 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY,ORDER_BY,FILTER,LIMIT
2018-07-17 03:08:51,173 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 03:08:51,183 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-07-17 03:08:51,396 [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]}
2018-07-17 03:08:51,722 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor.
collectionUsageThreshold = 489350752, usageThreshold = 489350752
2018-07-17 03:08:52,045 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100
optimistic? false
2018-07-17 03:08:52,162 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.CombinerOptimizerUtil - Choosing to move algebraic foreac

2018-07-17 03:12:16,559 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:12:16,566 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EEDED. Redirecting to job history server
2018-07-17 03:12:16,684 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:12:16,711 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EEDED. Redirecting to job history server
2018-07-17 03:12:16,812 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:12:16,825 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EEDED. Redirecting to job history server
2018-07-17 03:12:16,899 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:12:16,907 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EEDED. Redirecting to job history server
2018-07-17 03:12:17,006 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:12:17,018 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EEDED. Redirecting to job history server
2018-07-17 03:12:17,094 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:12:17,100 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EEDED. Redirecting to job history server
2018-07-17 03:12:17,180 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:12:17,189 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EEDED. Redirecting to job history server
2018-07-17 03:12:17,278 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:12:17,289 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EEDED. Redirecting to job history server
2018-07-17 03:12:17,364 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:12:17,372 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EEDED. Redirecting to job history server
2018-07-17 03:12:17,459 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at localhost/127.0.0.1:8032
2018-07-17 03:12:17,465 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCC
EEDED. Redirecting to job history server
2018-07-17 03:12:17,574 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2018-07-17 03:12:17,581 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2018-07-17 03:12:17,582 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2018-07-17 03:12:17,591 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2018-07-17 03:12:17,591 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2018-07-17 03:12:17,591 [main] INFO org.apache.pig.Main - Pig script completed in 3 minutes, 35 seconds and 946 milliseconds (215946 ms)
```

(c) Top ten origins with the highest AVG departure delay.

Ans:

pig Task_3c.pig

(To Execute a pig program in MR Mode)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

```

[acadgild@localhost ~]$ pig -b Task_3c.pig
18/07/17 03:30:02 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/17 03:30:02 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/17 03:30:02 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:10:40
Logging error messages to: /home/acadgild/Desktop/TestHadoop/pig/Assignment/pig_1531778402526.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Default bootstrap file /home/acadgild/.pigbootstrap not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: hdfs://localhost:8020
Pig Script ID for the session: PIG-Task_3c.pig-432f23a7-d688-4141-8ad0-3128908d7927
ATS is disabled since yarn.timeline-service.enabled set to false
fs.default.name is deprecated. Instead, use fs.defaultFS
fs.default.name is deprecated. Instead, use fs.defaultFS
fs.default.name is deprecated. Instead, use fs.defaultFS
Pig features used in the script: HASH, JOIN, GROUP BY, ORDER BY, FILTER, LIMIT
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schematuple] was not set... will not generate code.
[RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]]
Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
File concatenation threshold: 100 optimistic? false
Choosing to move algebraic foreach to combiner
Using Secondary Key Optimization for MapReduce node scope-67
Using Secondary Key Optimization for MapReduce node scope-100
Rewrite: POPackage->POForEach to POPackage(JoinPackager)
MR plan size before optimization: 7
MR plan size after optimization: 7
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to ResourceManager at localhost/127.0.0.1:8032

Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Success!
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schematuple] was not set... will not generate code.
Total input paths to process : 1
Total input paths to process : 1
(CMX,Hancock,USA,116.1470588235294)
(PLN,Pellston,USA,93.76190476190476)
(SPI, Springfield, USA, 83.84873949579831)
(ALO, Waterloo, USA, 82.2258064516129)
(MQT, NA, USA, 79.55665024630542)
(ACY, Atlantic City, USA, 79.3103448275862)
(MOT, Minot, USA, 78.66165413533835)
(HHH, NA, USA, 76.53085464480874)
(EGE, Eagle, USA, 74.12891986062718)
(BGM, Binghamton, USA, 73.15533980582525)
Pig script completed in 5 minutes, 56 seconds and 380 milliseconds (356380 ms)

```

Ans:

(To Execute a pig program in MR Mode)

Note: Program files are properly documented for a detailed description of each instruction used within the program.

```
[acagild@localhost Assignment]$ pig -brief Task_3d.pig
18/07/17 03:37:09 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
18/07/17 03:37:09 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
18/07/17 03:37:09 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
Apache Pig version 0.16.0 (r17146530) compiled Jun 01 2016, 23:10:49
Logging error messages to: /home/acagild/.log/Desktop/TestHadoop/pig/Assignment/pig_1531778829911.log
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acagild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acagild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Default bootstrap file /home/acagild/.pigbootstrap not found
mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to hadoop file system at: hdfs://localhost:8020
Pig Script ID for the session: PIG-Task_3d.pig-85b7b580-0e71-423c-896b-a875de575b9a
ATS is disabled since yarn.timeline-service.enabled set to false
fs.default.name is deprecated. Instead, use fs.defaultFS
fs.default.name is deprecated. Instead, use fs.defaultFS
Pig features used in the script: GROUP_BY,ORDER_BY,FILTER,LIMIT
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.schematuple] was not set... will not generate code.
{RULES_ENABLED={AddForEach,ColumnMapKeyPrune,ConstantCalculator, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter}}
Selected heap (Tenured) of size 69072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
File concatenation threshold: 100 optimistic? false
Choosing to move algebraic foreach to combiner
Using Secondary Key Optimization for MapReduce node scope-41
MR plan size before optimization: 4
MR plan size after optimization: 4
fs.default.name is deprecated. Instead, use fs.defaultFS
Connecting to ResourceManager at localhost/127.0.0.1:8032
Pig script settings are added to the job
mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
```

```

Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Connecting to ResourceManager at localhost/127.0.0.1:8032
Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
Success!
fs.default.name is deprecated. Instead, use fs.defaultFS
Key [pig.scemathuple] was not set... will not generate code.
Total input paths to process : 1
Total input paths to process : 1
(ORD, LGA), 39)
(DAL, HOU), 35)
(DFW, LGA), 33)
(ATL, LGA), 32)
(ORD, SNA), 31)
(SLC, SUN), 31)
(MIA, LGA), 31)
(BUR, JFK), 29)
(HRL, HOU), 28)
(BUR, DFW), 25)
Pig script completed in 3 minutes, 35 seconds and 732 milliseconds (215732 ms)

```
